Patients recall worse preoperative pain after shoulder arthroplasty than originally reported: A study of recall accuracy using the American Shoulder and Elbow Surgeons score

Article in Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons ... [et al.] · October 2016
DOI: 10.1016/j.jse.2016.09.004

CITATION
1

READS
20

5 authors, including:

Xinning Li
Boston University
104 PUBLICATIONS 867 CITATIONS
SEE PROFILE

Edward J. Testa
Tufts University
5 PUBLICATIONS 3 CITATIONS
SEE PROFILE

Andrew Jawa
New England Baptist Hospital
51 PUBLICATIONS 640 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Commentary & Perspective Total Hip Arthroplasty View project
Patients recall worse preoperative pain after shoulder arthroplasty than originally reported: a study of recall accuracy using the American Shoulder and Elbow Surgeons score

Jeremiah T. Lowe, BA\textsuperscript{a,b}, Xinning Li, MD\textsuperscript{c}, Sydney M. Fasulo, BA\textsuperscript{a,b}, Edward J. Testa, BS\textsuperscript{d}, Andrew Jawa, MD\textsuperscript{a,b,*}

\textsuperscript{a}New England Baptist Hospital, Boston, MA, USA
\textsuperscript{b}Boston Sports and Shoulder Center, Waltham, MA, USA
\textsuperscript{c}Boston Medical Center, Boston, MA, USA
\textsuperscript{d}Tufts University School of Medicine, Boston, MA, USA

Background: Patient-reported outcome measures (PROMs) are valuable tools for quantifying outcomes of orthopedic surgery. However, when baseline scores are not obtained, there is considerable controversy about whether PROMs can be administered retrospectively for patients to recall their preoperative state. We investigated the accuracy of patient recall after total shoulder arthroplasty (TSA) using the American Shoulder and Elbow Surgeons (ASES) assessment score.

Methods: Recalled ASES scores were collected postoperatively at 6 weeks, 3 months, 6 months, and 12 months from 169 patients who previously completed baseline scores before TSA. The ASES total score was divided into its two subcomponents: functional ability and visual analog scale (VAS) for pain. We compared preoperative and recalled scores for each subcomponent and the total ASES score.

Results: Recalled ASES function scores were comparable to corresponding preoperative scores across all time points (analysis of variance, $P = .21$), but recalled VAS pain was significantly higher at all time points beyond 6 weeks after surgery ($P = .0001$ at 3 months; $P = .005$ at 6 months; and $P = .001$ at 12 months). As a result, the ASES total score was only comparable at 6 weeks after surgery ($P = .39$) and differed at all time points thereafter.

Conclusion: Patients are able to recall preoperative function with considerable accuracy for up to 12 months after TSA. However, beyond 6 weeks postoperatively, patients recall having worse pain than they originally reported, and recalled ASES total scores are unreliable as a result.

Level of evidence: Basic Science Study; Validation of Outcome Instruments

Keywords: American Shoulder and Elbow Surgeons score (ASES); patient recall; shoulder arthroplasty; outcomes; visual analog scale pain (VAS); patient-reported outcome measure (PROM)

For patients undergoing total shoulder arthroplasty (TSA), functional improvement is often measured using patient-reported outcome measures (PROMs) such as the American Shoulder and Elbow Surgeons (ASES) score. The ASES score,...
which consists of a functional ability section and the visual analog scale (VAS) for pain, has been validated as a responsive and reliable metric of shoulder function and disability.\textsuperscript{1,9,13} PROMs such as the ASES score are well-established measures for quantifying pain and function in outcomes research, which has become increasingly emphasized in orthopedics and medicine as a whole.\textsuperscript{15} To assess outcome trends among patients undergoing TSA, preoperative and postoperative ASES data must both be collected.\textsuperscript{17} However, when preoperative scores are not obtained, there is considerable controversy about whether PROMs, such as the ASES, can be administered retrospectively for patients to recall their preoperative pain and function.\textsuperscript{4,5,7,14,17}

As implementation of PROMs becomes a ubiquitous aspect of orthopedic surgery, likewise do situations arise where researchers lack sufficient preoperative data. This can be the case for surgeons transitioning to the use of the ASES questionnaire as a new measure for tracking outcomes or simply when researchers did not anticipate a need for preoperative scores. Stemming from this, recent studies have investigated whether PROMs can be retrospectively administered. Stepan et al\textsuperscript{14} reported that hand and elbow patients receiving a variety of treatments were able to accurately recall preoperative function for up to 2 years using the 11-item version of the Disabilities of the Arm, Shoulder and Hand (QuickDASH) questionnaire. In addition, Marsh et al,\textsuperscript{7} using their own self-validated questionnaire, found that total hip arthroplasty patients were able to recall preoperative health status 6 weeks after surgery, and Howell et al\textsuperscript{8} reported accurate recall of the Oxford Hip Score for up to 3 months. However, Wilson et al\textsuperscript{17} found unreliable recall accuracy using the Oxford Shoulder Score, and a large study of 770 total knee arthroplasty patients by Lingard et al\textsuperscript{5} concluded that recalled pain and function was poor at 3 months postoperatively, with patients tending to recall significantly greater pain than they originally reported.

We are unaware of any research regarding the recall accuracy of TSA patients using the ASES form, and studies that address recall longevity are sparse. Thus, the objective of the present study was to assess the accuracy and longevity of preoperative pain and function recall using the ASES form among TSA patients for up to 12 months to answer the question: Do patients remember their preoperative pain and function and if so, for how long? From the available research, we hypothesized that patients would be able to accurately recall their preoperative ASES score at initial follow-up intervals,\textsuperscript{7} but we expected accuracy to decline at subsequent visits to the point of unreliability as early as 3 months after surgery.\textsuperscript{5}

Materials and methods

All participants in the study underwent TSA by a single fellowship-trained, high-volume shoulder surgeon (A.J.). We obtained actual follow-up scores and recalled ASES scores, including the VAS pain score, which is a subcomponent of the questionnaire, from 193 TSA patients who came for postoperative appointments at 6 weeks, 3 months, 6 months, and 12 months from April to December 2015. Except for the 6-month follow-up, which is considered an optional appointment, all other follow-up intervals in the study are consistent with the surgeon’s protocol for postoperative appointments.

In accordance with the surgeon’s routine clinical practice, preoperative ASES scores were obtained at the appointment most immediately preceding surgery. At follow-up appointments, study patients were first asked to complete the ASES form about their postoperative function on the given day and then were asked to recall their preoperative state on a second form. To avoid potential bias, neither the surgeon nor study staff were present while patients completed the ASES questionnaires.

Patients in the study received anatomic, reverse, or revision TSA for treatment of degenerative joint disease, rotator cuff arthropathy, or failed previous arthroplasty, respectively (Table I). Among that population, individuals were included if they attended their follow-up appointments at the suggested intervals (6 weeks, 3 months, 6 months, 12 months) and had properly completed the preoperative ASES form.

The patient-reported section of the ASES standardized assessment form consists of the VAS pain score (rated from 0 to 10) and 10 functional questions that are specific to the upper extremity (rated from 0 to 3). As described by Richards et al,\textsuperscript{12} a standardized algorithm is applied using the selected numbers to calculate a score from 0 to 100, of which 50 points correspond to pain and 50 correspond to function. A low score indicates more limited function and higher pain. For the purposes of our comparative analysis, we separated the ASES total score into its subcomponents VAS pain (0 to 10 scale) and ASES function score (0 to 50 scale).

Statistical analysis

A descriptive analysis of continuous variables was performed and is reported in Table I. To evaluate differences between preoperative and recalled and ASES function scores over all time intervals, we performed a linear mixed-model analysis of variance with repeated measurements to reconcile absent data points. For post hoc analysis, we first evaluated the normality of the data at each time point using the Kolmogorov-Smirnov test. If the continuous variables of a given time-point satisfied a normal distribution, a paired $t$ test was used to assess the difference between preoperative ASES function and recalled ASES function scores at each time point, with $\alpha = 0.05$ as the level of significance. If data were not normally distributed, the Wilcoxon signed rank test was used. All analyses were repeated with preoperative and recalled VAS pain level.

For the purpose of regression analysis, we used the absolute difference between preoperative and recalled ASES total scores to represent recall accuracy, with a greater value indicating poorer recall accuracy. Univariate analysis (general linear model) was used to individually evaluate the relationship between relevant variables and recall accuracy. The variables considered were age, sex, type of shoulder arthroplasty; preoperative, recalled, and actual follow-up ASES total score; preoperative, recalled, and actual follow-up VAS pain score; number of days before the surgery date that preoperative questionnaires were completed, number of days after...
surgery when recalled and follow-up questionnaires were completed, and total time between the preoperative and follow-up appointment. Significance was assumed at $P < .05$. Any significant predictor variables were further assessed with multivariate analysis. An experienced biostatistician performed all data analysis using SAS 9.3 software (SAS Institute, Inc., Cary, NC, USA).

**Results**

Of the original 193 patients, 169 had sufficient ASES data to fulfill the 3 requisite data points (preoperative, follow-up, recall) and thereby qualify for inclusion. Age at surgery was 47 to 94 years, and mean ± standard deviation age of the pooled sample was 67.6 ± 8.2 years. Patients received anatomic TSA (39%), reverse TSA (56%), or revision TSA (5%).

Linear mixed-model analysis of variance with repeated measures for all time points indicated that preoperative and recalled VAS pain differed significantly ($P = .0001$), but the ASES function scores did not ($P = .21$). Post hoc analysis by Wilcoxon signed rank test or paired $t$ test indicated that preoperative and recalled VAS pain levels were comparable at 6 weeks but differed at all assessments thereafter (Table II).

Our univariate analysis of the pooled data to investigate any correlations with recall accuracy found no significant relationship with age, sex, type of TSA surgery, actual ASES at same-day follow-up, actual reported pain level at same-day follow-up, number of days between the preoperative appointment (when initial ASES was obtained) to follow-up, and number of days after surgery. However, a slight positive correlation for recall accuracy was found between the preoperative VAS pain level and the absolute difference in the preoperative and recalled ASES score ($r = 0.08$, $P = .0002$).

**Discussion**

Our findings suggest that although patients may be able to accurately recall their ASES function score, they are not able to accurately recall their preoperative pain from 3 months through 12 months after TSA. As a result, the ASES total score, which is the validated measure that combines pain and function, is unreliable beyond 6 weeks postoperatively. Our findings are generally consistent with the range of relevant orthopaedic studies indicating that recalled PROM scores might be reliable initially but become inaccurate as early as 3 months after surgery. At least 3 studies have found that patient recall of preoperative status after arthroplasty procedure is reliable in the short-term from 3 days to 3 months.$^{2,4,7}$ Other studies investigating recall accuracy over the long-term from 3 months to 2.5 years have concluded otherwise.$^{5,6,10}$ Lastly, Wilson et al$^{17}$ reported mixed results using the Oxford Shoulder Score among patients undergoing a range of shoulder operations at a mean of 50 days postoperatively.
Our mixed results suggest that patients are able to recall functional ability with considerable accuracy for up to 1 year, which is not the case for pain. On the one hand, this may reflect the concrete nature of the questions in the functional component of the ASES questionnaire (eg, “How difficult is it to reach a high shelf?”), which encourages patients to focus on distinctive actions and tasks that had been troublesome before the operation. On the other hand, pain level is a dynamic value that may be more difficult to pinpoint. Because the ASES total score (scale of 0 to 100 points) consists of up to 50 points derived from the VAS pain rating, poor recall of pain was enough to invalidate recalled total scores beyond 6 weeks after surgery.

Michener et al9 have criticized the combination of pain and function on the ASES questionnaire. They argue that pain is an impairment that should not be misconstrued with functional

### Table II Comparison of preoperative and recalled American Shoulder and Elbow Surgeons scores at each time point

<table>
<thead>
<tr>
<th>ASES components</th>
<th>Time after TSA</th>
<th>6 weeks</th>
<th>3 months</th>
<th>6 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-op†</td>
<td>VAS pain</td>
<td>6.2 ± 2.2</td>
<td>6.7 ± 1.8</td>
<td>6.5 ± 2.3</td>
<td>6.2 ± 2.4</td>
</tr>
<tr>
<td>AS</td>
<td>ASES function</td>
<td>14.1 ± 8.1</td>
<td>13.0 ± 6.8</td>
<td>13.7 ± 8.2</td>
<td>14.4 ± 9.2</td>
</tr>
<tr>
<td>AS</td>
<td>ASES total</td>
<td>32.8 ± 15.3</td>
<td>29.5 ± 12.1</td>
<td>31.3 ± 16.1</td>
<td>34.3 ± 14.4</td>
</tr>
<tr>
<td>Recalled‡</td>
<td>VAS pain</td>
<td>7.0 ± 1.9</td>
<td>7.9 ± 1.8</td>
<td>7.6 ± 2.3</td>
<td>7.5 ± 2.4</td>
</tr>
<tr>
<td>AS</td>
<td>ASES function</td>
<td>13.7 ± 9.6</td>
<td>12.2 ± 11.3</td>
<td>13.5 ± 9.5</td>
<td>12.8 ± 9.1</td>
</tr>
<tr>
<td>AS</td>
<td>ASES total</td>
<td>28.7 ± 16.1</td>
<td>22.4 ± 16.7</td>
<td>25.7 ± 17.2</td>
<td>25.3 ± 18.7</td>
</tr>
<tr>
<td>Mean difference (paired)§</td>
<td>VAS pain</td>
<td>−0.9 (−1.5 to −0.2)</td>
<td>−1.1 (−1.6 to −0.6)</td>
<td>−1.1 (−1.8 to −0.3)</td>
<td>−1.3 (−2.1 to −0.5)</td>
</tr>
<tr>
<td>AS</td>
<td>P value**</td>
<td>.21</td>
<td>.0001</td>
<td>.005</td>
<td>.001</td>
</tr>
<tr>
<td>AS</td>
<td>ASES function</td>
<td>0.4 (−2.4 to 3.2)</td>
<td>−0.2 (−2.9 to 2.5)</td>
<td>0.2 (−3.0 to 3.5)</td>
<td>1.6 (−1.4 to 4.6)</td>
</tr>
<tr>
<td>AS</td>
<td>P value</td>
<td>.62</td>
<td>.28</td>
<td>.45</td>
<td>.41</td>
</tr>
<tr>
<td>AS</td>
<td>ASES total</td>
<td>4.2 (−1.0 to 9.3)</td>
<td>5.4 (1.6 to 9.3)</td>
<td>5.6 (1.9 to 9.3)</td>
<td>9.0 (4.0 to 14.0)</td>
</tr>
<tr>
<td>AS</td>
<td>P value**</td>
<td>.39</td>
<td>.0001</td>
<td>.03</td>
<td>.005</td>
</tr>
</tbody>
</table>

ASES, American Shoulder and Elbow Surgeons score; TSA, total shoulder arthroplasty; VAS, visual analog scale for pain.

* Data are shown as mean ± standard deviation or as the mean difference (95% confidence interval).
† Patient-reported status obtained at preoperative appointment.
‡ Recalled preoperative status obtained at the indicated follow-up appointments.
§ Mean difference between preoperative and recalled values of matched pairs;
** Post hoc P values are from the Wilcoxon signed rank test or paired t test secondary to analysis of variance.

Our mixed results suggest that patients are able to recall functional ability with considerable accuracy for up to 1 year, which is not the case for pain. On the one hand, this may reflect the concrete nature of the questions in the functional component of the ASES questionnaire (eg, “How difficult is it to reach a high shelf?”), which encourages patients to focus on distinctive actions and tasks that had been troublesome before the operation. On the other hand, pain level is a dynamic value that may be more difficult to pinpoint. Because the ASES total score (scale of 0 to 100 points) consists of up to 50 points derived from the VAS pain rating, poor recall of pain was enough to invalidate recalled total scores beyond 6 weeks after surgery.

Michener et al9 have criticized the combination of pain and function on the ASES questionnaire. They argue that pain is an impairment that should not be misconstrued with functional
limitations within the same assessment score. In this case, we have found that the inclusion of pain does appear to negatively affect the accuracy of recalled ASES total scores. Contrary to our findings, Stepan et al. concluded that hand and elbow patients are able to accurately replicate baseline conditions on the QuickDASH for up to 2 years after treatment. The most probable explanation is that the QuickDASH is less influenced by patient-reported pain than the ASES score. As our results suggest, recall of preoperative pain is unreliable beyond 6 weeks, whereas recall of function appears to be accurate through 12 months.

Another possibility is that we have assessed a distinct population that is older and undergoing a more radical clinical transition than the patients sampled by Stepan et al. Their sample consisted of a younger population (mean age, 55 ± 12 years) with only a fraction receiving surgical treatment (8.6% to 37.1%) and some patients in each group receiving no treatment at all. By design, our sample only included patients undergoing TSA, and the mean age of our sample was 67.6 years. Patients receiving less drastic treatment measures, such as nonoperative treatment, as in the study by Stepan et al., would reasonably experience a subtler symptomatic change than those undergoing TSA.

Our observation that patients tend to recall more pain than originally reported is consistent with findings by Mancuso and Charlson, Lingard et al., Wilson et al., and Pellisé et al. This may reflect a bias of patients to overestimate their previous pain and disability level when experiencing improvement postoperatively. Similar phenomena have been described in clinical settings as a consequence of response shift, an effect that is particularly evident in research using PROMs. Our findings lend further support to the claim that as patients improve after surgery, memory of their preoperative condition becomes polarized relative to their improved postoperative state.

Our findings suggest that, in the long-term, patients may not have an entirely accurate reference on which to base their perceived improvement. This can be informative for surgeons in their approach to counseling both preoperatively and postoperatively. Given that preoperative expectations for improvement are significant determinants of patient satisfaction, it is important for patients to accurately perceive their improvement when following up after surgery. Surgeons and researchers should recognize that some of the improvement reported postoperatively by patients might be skewed by recall bias. Proper counseling can help patients manage expectations for realistic improvements and ultimately shape the perception of their outcome. Additional research could address specifically how skewed recall of the preoperative state affects patient satisfaction postoperatively.

The only factor demonstrating a significant correlation with recall accuracy was the preoperative VAS pain level (\( r = 0.08, P = .0002 \)). Notably, the correlation was incredibly weak. This suggests that higher levels of preoperative pain correspond weakly with poorer ASES score recall. Consistent with Stepan et al., no relationship was found between recall accuracy and patient age. We also did not observe any association between the type of surgery (TSA, RSA, and revision) and recall accuracy, which corroborates the finding of Stepan et al. that diagnosis or whether a patient underwent surgical intervention did not factor into recall.

This study is subject to certain limitations. Recalled scores were only obtained from patients who attended their follow-up appointments at the routine intervals, which introduces bias and limits the generalizability of our results relative to our target population. Patient inclusion was further limited by the availability of study staff to administer the ASES form. Patients who incorrectly or incompletely filled out the preoperative or recall ASES form were not included, which
may have resulted in an unrepresentative sample. Our results are only generalizable among patients undergoing TSA who self-report pain and function with the ASES form.

Lastly, the study lacks a prospective power analysis, and the sample size might therefore be inadequate to address our goals. Preceding studies with similar objectives have reported sample sizes that are comparable to the number of participants in our study. 7,14,17

Conclusions
Recalled ASES scores and VAS pain rating may be accurate for up to 6 weeks after TSA, which is consistent with existing studies using varying PROMs. The available evidence appears to legitimize the use of validated PROMs to retrospectively obtain preoperative levels of pain and function within 6 weeks postoperatively. Beyond that, however, patients recalled significantly higher VAS pain compared with their preoperative ratings, and the validated ASES cumulative score is therefore unreliable. Based on this finding and existing research, we do not recommend retrospective use of the ASES form with patients undergoing TSA beyond 6 weeks after surgery.

Disclaimer
Andrew Jawa has been a paid speaker for DJO Global. Xinning Li is a paid consultant for Mitek and Tornier. The other authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References
1. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). Arthritis Care Res (Hoboken) 2011;63(Suppl 11):S174-88. http://dx.doi.org/10.1002/acr.20630