Shoulder Arthroplasty Smart Score

Roche, C, Kumar, V, Overman, S, Simovitch, R, Flurin, P, Wright, T, Routman, H, Teredesai, A, & Zuckerman, J

Disclosures: C.P. Roche: 3A; Exactech, Inc. 4; Exactech Inc. V. Kumar: 3A; KenSci, Inc. S. Overman: 3B; KenSci, Inc. R. Simovitch: 3B; Exactech, Inc. 7A; Exactech, Inc. P. Flurin: 3B; Exactech, Inc. 7A; Exactech, Inc. T. Wright: 3B; Exactech, Inc. 7A; Exactech, Inc. H. Routman: 3B; Exactech, Inc. 7A; Exactech, Inc. A. Teredesai: 3A; KenSci, Inc. J . Zuckerman: 3B; Exactech, Inc. 7A; Exactech, Inc.

Introduction: We propose and evaluate a new clinical assessment tool to quantify outcomes following anatomic total shoulder arthroplasty (aTSA) and reverse TSA (rTSA). This TSA-specific outcome measure aims to provide greater insight into both the disease-progression and treatment by utilizing both subjective and objective measures previously demonstrated to be highly predictive of post-operative TSA outcomes. Doing so, may create a novel tool that better accounts for TSA outcomes variability. The goal of this study is to quantify and compare the psychometric properties of this machine learning derived assessment tool relative to 5 commonly used clinical measures to quantify TSA outcomes.

Methods: We analyzed a multi-center clinical outcomes database of shoulder arthroplasty patients who received a single platform shoulder prosthesis (Equinoxe, Exactech, Inc, Gainesville, FL) between November 2004 and December 2018. Every patient enrolled in this open-label clinical database provided consent. All data was collected using standardized forms at each of 30 different clinical sites according to an institutional review board-approved protocol. All primary aTSA and rTSA patients in the database with at least 2 years follow-up were included in this study. To ensure a homogenous dataset, patients with revisions, humeral fractures, endoprostheses, and hemiarthroplasty were excluded. The database contains demographic information, shoulder diagnoses, comorbidities, implant type, active and passive range of motion (ROM), radiographic findings, and 5 clinical outcome scores: ASES, Constant, University of California Los Angles (UCLA), Simple Shoulder Test (SST), and Shoulder Pain and Disability Index (SPADI), including the individual questions used to derive these 5 scores. The proposed TSA-specific clinical outcome measure, the Shoulder Arthroplasty Smart Score or SAS score (smartshoulderscore.com), is a multi-domain assessment consisting of 6 input questions, of which 3 are objective ROM measures and 3 are subjective measures of pain and function. The 6 input questions have an equal weight of 12.5 points each and one additional input, called the Composite ROM score, is calculated from the 3 objective ROM measures to transform those values into a functional score with an allocated weight of 25 points, yielding a score range of 0 to 100 points, with 100 the best score. These 6 questions were identified using a machine learning analysis of this database and were selected because they are among the most-predictive pre-operative inputs that influence post-operative TSA outcomes.[1-2] As all 6 inputs already existed in the database, the SAS score was able to be retrospectively calculated from each patient's pre-operative visit and all post-operative visits. Clinical data from 3,667 patients (1,594 aTSA, 2,073 rTSA) with 8,104 post-operative follow-up visits (3,878 aTSA, 4,226 rTSA) were used to quantify the psychometric properties (validity, responsiveness, and clinical interpretability) for the proposed SAS score and each of the SST, Constant, ASES, UCLA, and SPADI scores for aTSA and rTSA patients.

Results: Clinical outcomes for the SAS score and each of the SST, Constant, ASES, UCLA, and SPADI scores are presented in Table 1. Table 1 demonstrates convergent validity with all 6 outcome measures being moderately-to-highly correlated pre-operatively and highly-correlated post-operatively when quantifying TSA outcomes. The SAS score was most correlated with the UCLA score (pre-op R=0.78, post-op R=0.85) and least correlated with the SST (pre-op R=0.63, post-op R=0.75). The pre-op and post-op floor and ceiling (Table 2) effect analysis demonstrates that no clinical outcome score exhibited significant floor effects pre-operatively or post-operatively and no clinical outcome score exhibited significant ceiling effects pre-operatively; however, significant ceiling effects were present post-operatively for each of the SST (44.3%), UCLA (13.9%), ASES (18.7%), and SPADI (19.3%) measures. Ceiling effects were more pronounced for aTSA than rTSA, where each of the SST, UCLA, ASES, and SPADI measures had >20% of aTSA patients with ceiling effects. The SAS score had the least number of patients with floor and ceiling effects. A more-detailed analysis demonstrated that generally, males, younger patients, and Caucasians who received TSA were more likely to experience a ceiling effect, as compared to patients that were female, older, and of non-Caucasian race/ethnicity. The patient satisfaction anchor-based MCID and SCB thresholds for all 6 outcome measures is presented in Table 3. The anchor-based MCID and SCB thresholds for aTSA patients were larger than the MCID and SCB values for rTSA patients for all outcome measures, except the SST. The MCID and SCB thresholds for the SAS score were most similar in magnitude to the Constant score. The SRM and ES responsiveness for all 6 outcome measures is presented in Table 3. At 2 years minimum follow-up, all 6 clinical outcome measures detected a large effect with the UCLA exhibiting the most responsiveness and the SST exhibiting the least; each of the SAS, ASES, Constant, and SPADI scores had similarly large SRM and ES responsiveness.

Discussion: We present the first orthopaedic clinical outcome measure derived using machine learning and constructed of pre-operative inputs that are most-predictive of post-operative TSA outcomes. The results of this 3,667 TSA outcome study demonstrate that the SAS score has equivalent or better psychometric properties of validity, responsiveness, and clinical interpretability as the ASES, Constant, UCLA, SST, and SPADI scores. The SAS score has an appropriate response range with no floor/ceiling effects for aTSA or rTSA and no gender, age, or race/ethnicity response bias, as was observed with many of the other scores when quantifying TSA outcomes. These psychometric improvements were achieved despite the SAS score consisting of only 6 input questions; this efficient selection of 3 objective and 3 subjective measures is useful for both quality assurance and clinical research purposes, and represents a reduction of approximately half (or more) of the inputs required by ASES, SST, SPADI, and Constant scores. Such efficiency will likely reduce administrative burden and responder fatigue while improving patient compliance when performing TSA clinical research.

Significance: The 6 question SAS score is an efficient TSA-specific outcome measure with equivalent or better validity, responsiveness, and clinical interpretability as 5 other historical assessment tools. The SAS score has appropriate response range without floor/ceiling effects and without bias in any target patient characteristic, unlike the age, gender, or race/ethnicity bias observed in the ceiling scores with the other outcome measures. Due to these substantial benefits, we recommend the use of the novel SAS score for quantifying TSA outcomes.

References:

1. Kumar, V et al. What is the accuracy of Three Different Machine Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty. CORR. 2020.
2. Kumar**,** V et al. Using Machine Learning to Predict Clinical Outcomes After Shoulder Arthroplasty with a Minimal Feature Set. JSES. 2020.

Table 1. Top: Assessment of Validity: Comparison of Pre-operative, Post-operative, and Pre-to-Post-operative Improvement Outcomes Between the 6 Different Clinical Outcome Measures. Bottom: Assessment of Validity: Correlation of Pre-operative and Post-operative Clinical Outcome Measures Scores to each other and to Patient Satisfaction Assessment, where Pearson Coefficient 0-0.3 is considered poor correlation, 0.3-0.6 is moderately correlated, and 0.6-1.0 is highly correlated

Pre-op / Post-op / Pre-to-Post-op Improvement	aTSA + rTSA Cohort (Mean ± Std Dev)	aTSA Cohort (Mean ± Std Dev)	rTSA Cohort (Mean ± Std Dev)
SST	3.8 ± 2.9 / 10.1 ± 2.5 / 6.4 ± 3.3	4.1 ± 3.0 / 10.5 ± 2.3 / 6.5 ± 3.3	3.6 ± 2.8 / 9.8 ± 2.7 / 6.3 ± 3.3
Constant	36.8 ± 14.1 / 69.7 ± 14.3 / 33.5 ± 16.1	38.5 ± 13.9 / 71.6 ± 14.2 / 34.5 ± 15.8	35.5 ± 14.1 / 68.1 ± 14.1 / 32.8 ± 16.3
ASES	36.3 ± 16.1 / 83.4 ± 18.7 / 47.6 ± 21.7	36.3 ± 16.2 / 85.0 ± 18.5 / 49.7 ± 21.7	36.2 ± 16.0 / 82.0 ± 18.8 / 45.9 ± 21.6
UCLA	13.7 ± 4.1 / 30.4 ± 5.2 / 16.6 ± 5.9	14.3 ± 4.0 / 30.8 ± 5.3 / 16.7 ± 5.8	13.3 ± 4.2 / 30.0 ± 5.1 / 16.6 ± 6.0
SPADI	83.1 ± 23.2 / 20.6 ± 24.5 / -62.6 ± 29.0	82.4 ± 23.6 / 17.4 ± 22.7 / -65.7 ± 29.1	83.6 ± 22.9 / 23.5 ± 25.7 / -60.0 ± 28.6
SAS	46.3 ± 11.5 / 77.7 ± 12.2 / 31.7 ± 14.6	46.4 ± 10.8 / 80.4 ± 12.0 / 34.9 ± 13.7	46.2 ± 12.1 / 75.3 ± 11.9 / 29.2 ± 14.8

Combined aTSA + rTSA Cohort	Patient Satisfaction (Pre/Post)	SST (Pre/Post)	Constant (Pre/Post)	ASES (Pre/Post)	UCLA (Pre/Post)	SPADI (Pre/Post)
SST	0.096 / 0.435	1				
Constant	0.083 / 0.394	0.716 / 0.799	1			
ASES	0.075 / 0.460	0.717 / 0.841	0.653 / 0.810	1		
UCLA	0.064 / 0.521	0.630 / 0.756	0.773 / 0.806	0.771 / 0.873	1	
SPADI	-0.068 / -0.444	-0.825 / -0.890	-0.692 / -0.805	-0.810 / -0.924	-0.692 / -0.823	1
SAS	0.074 / 0.418	0.630 / 0.748	0.781 / 0.847	0.694 / 0.832	0.783 / 0.852	-0.694 / -0.825

Table 2. Assessment of Validity: Comparison of Floor and Ceiling Effects Across the 6 Different Clinical Outcome Measures

% Reports with a "Floor" Score	aTSA + rTSA Cohort (Pre-op/Post-op)	aTSA Cohort (Pre-op/Post-op)	rTSA Cohort (Pre-op/Post-op)
SST	10.2% / 0.4%	9.8% / 0.3%	10.5% / 0.4%
Constant	0.0% / 0.0%	0.0% / 0.0%	0.0% / 0.0%
Strength Component of Constant Score	67.0% / 15.9%	61.0% / 15.4%	72.0% / 16.3%
ASES	0.1% / 0.0%	0.1% / 0.0%	0.1% / 0.0%
UCLA	0.0% / 0.0%	0.0% / 0.0%	0.0% / 0.0%
SPADI	0.6% / 0.1%	0.6% / 0.0%	0.5% / 0.1%
SAS	0.0% / 0.0%	0.0% / 0.0%	0.0% / 0.0%

% Reports with a "Ceiling" Score	aTSA + rTSA Cohort (Pre-op/Post-op)	aTSA Cohort (Pre-op/Post-op)	rTSA Cohort (Pre-op/Post-op)
SST	1.0% / 44.3%	1.1% / 52.9%	1.0% / 36.7%
Constant	0.0% / 0.3%	0.0% / 0.5%	0.0% / 0.1%
Strength Component of Constant Score	0.2% / 0.7%	0.5% / 1.1%	0.0% / 0.4%
ASES	0.0% / 18.7%	0.0% / 25.1%	0.0% / 12.8%
UCLA	0.0% / 13.9%	0.0% / 20.8%	0.0% / 7.9%
SPADI	0.0% / 19.3%	0.0% / 25.6%	0.0% / 13.6%
SAS	0.0% / 0.1%	0.0% / 0.2%	0.0% / 0.0%

Table 3. Assessment of Responsiveness & Interpretability: Comparison of Patient Satisfaction Anchor-Based Minimal Clinically Important Difference (MCID) and Substantial Clinical Benefit (SCB) Thresholds, and Standardized Response Mean (SRM) & Effect Size (ES) Across the 6 Different Clinical Outcome Measures

Anchor MCID / Anchor SCB	aTSA + rTSA Cohort	aTSA Cohort	rTSA Cohort
SST	1.8 / 3.5	1.7 / 3.5	1.8 / 3.5
Constant	5.3 / 16.9	8.6 / 20.4	3.0 / 14.3
ASES	12.4 / 30.7	14.2 / 33.2	11.2 / 28.7
UCLA	7.9 / 11.8	8.1 / 12.6	7.7 / 11.3
SPADI	-20.4 / -44.0	-19.7 / -44.3	-21.3 / -43.7
SAS	6.1 / 16.6	8.5 / 19.2	4.9 / 14.4

SRM / ES	aTSA + rTSA Cohort	aTSA Cohort	rTSA Cohort
SST	1.91 / 2.20	1.96 / 2.19	1.88 / 2.23
Constant	2.09 / 2.38	2.18 / 2.49	2.02 / 2.32
ASES	2.19 / 2.95	2.29 / 3.06	2.12 / 2.87
UCLA	2.81 / 4.03	2.88 / 4.17	2.76 / 3.97
SPADI	-2.16 / -2.70	-2.26 / -2.79	-2.10 / -2.62
SAS	2.17 / 2.75	2.55 / 3.24	1.97 / 2.41