Using Machine Learning to Predict Clinical Outcomes After Shoulder Arthroplasty with a Minimal Feature Set

Vikas Kumar, PhD; Christopher Roche, MSE, MBA; Steve Overman, MD; Ryan Simovitch, MD; Pierre-Henri Flurin, MD; Thomas Wright, MD; Joseph Zuckerman, MD; Howard Routman, DO; & Ankur Teredesai, PhD

Affiliations

Vikas Kumar, PhD:
KenSci, 506 2nd Ave 21st Floor, Seattle, WA 98104
Vikas@kensci.com

Christopher Roche, MSE, MBA:
Exactech, 2320 NW 66th Court, Gainesville, FL 32653
Chris.Roche@exac.com

Steve Overman, MD:
KenSci, 506 2nd Ave 21st Floor, Seattle, WA 98104
drsteve@kensci.com

Ryan Simovitch, MD:
Hospital For Special Surgery – FL, West Palm Beach, FL 33401
simovitchr@hss.edu

Pierre-Henri Flurin, MD:
Bordeaux-Merignac Sport Clinic, 2 Rue Georges Negrevergne 33700 Merignac, France
phflurin@gmail.com

Thomas Wright, MD:
University of Florida Department of Orthopaedic Surgery, 3450 Hull Road Gainesville, FL 32611
wrightw@ortho.ufl.edu

Joseph Zuckerman, MD:
Department of Orthopedic Surgery at NYU Langone Orthopedic Hospital, 301 East 17th St. New York, NY 10003
joseph.zuckerman@nyumc.org

Howard Routman, DO:
Atlantis Orthopedics, 900 Village Square Crossing, #170 Palm Beach Gardens, FL 33410
HDR@floridashoulder.com

Ankur Teredesai, PhD:
KenSci, 506 2nd Ave 21st Floor, Seattle, WA 98104
ankur@kensci.com

Conflict of Interest Statement

Vikas Kumar, Steve Overman, and Ankur Teredesai are employed by Ken Sci, Inc.
Christopher Roche is employed by Exactech, Inc.
Ryan Simovitch and Howard Routman are consultants for Exactech, Inc.
Pierre-Henri Flurin, Thomas Wright, and Joseph Zuckerman are consultants for Exactech, Inc. and receive royalties on products related to this article.

Corresponding author:
Christopher Roche, MSE, MBA

Abstract

Background

A machine learning analysis was conducted on 5,774 shoulder arthroplasty patients to create predictive models for multiple clinical outcome measures after anatomic Total Shoulder Arthroplasty (aTSA) and reverse Total Shoulder Arthroplasty (rTSA). The goal of this study is to compare the accuracy associated with a full feature set predictive model (e.g. full model =291 parameters) and a minimal feature set model (e.g. abbreviated model =19 input parameters) to predict clinical outcomes in order to assess the efficacy of using a minimal feature set of inputs as a shoulder arthroplasty clinical decision-support tool.

Methods

Clinical data from 2,153 primary aTSA patients and 3,621 primary rTSA patients were analyzed using the XGBoost machine learning technique to create and test predictive models for multiple outcome measures at different post-operative timepoints using a full and abbreviated model. Mean absolute errors (MAE) quantified the difference between actual and predicted outcomes, and each model also predicted if a patient would experience clinical improvement greater than the minimal

clinically important difference (MCID) and substantial clinical benefit (SCB) patient satisfaction anchor-based thresholds for each outcome measure at 2-3 years after surgery.

Results

Across all post-operative timepoints analyzed, the full and abbreviated models had similar MAE for the American Shoulder and Elbow Surgeons (ASES) (full model = ±11.7 vs. abbreviated model = ±12.0), Constant (±8.9 vs. ±9.8), Global Shoulder Function (±1.4 vs. ±1.5), Visual Analog Scale (VAS) pain (±1.3 vs. ±1.4), active abduction (±20.4º vs. ±21.8º), forward elevation (±17.6º vs. ±19.2º), and external rotation (±12.2º vs. ±12.6º). Marginal improvements in MAE were observed for each outcome measure prediction when the abbreviated model was supplemented with implant size/type data and measurements of native glenoid anatomy. The full and abbreviated models each effectively risk-stratified patients using only pre-operative data by accurately identifying patients with improvement greater than the MCID and SCB thresholds.

Discussion

Our study demonstrated the full and abbreviated machine learning models achieved similar accuracy to predict clinical outcomes after aTSA and rTSA at multiple post-operative timepoints. These promising results demonstrate an efficient utilization of machine learning algorithms to predict clinical outcomes. The use of a minimal feature set of only 19 preoperative inputs suggests that this tool may be easily used during a surgical consultation to improve decision-making related to shoulder arthroplasty.

Keywords: Machine Learning, Predictive Outcomes Analytics, aTSA and rTSA Outcomes, Shoulder Arthroplasty

Level of Evidence: III (retrospective comparative study)

Introduction

Machine learning techniques can analyze clinical and patient-reported outcome data to create predictive models that can help physicians better understand their patients prior to treatment by quantifying patient-specific potential for improvement associated with different treatment options. The knowledge of these patient-specific outcome predictions is useful to better inform shared decision-making for both the patient and surgeon.[2,3,4,12] Machine learning models have recently been used to accurately predict clinical outcomes after anatomic (aTSA) and reverse (rTSA) total shoulder arthroplasty and risk-stratify patients based on predicted minimal clinically important difference (MCID) and substantial clinical benefit (SCB) improvement thresholds for different clinical outcome measures.[12] Deploying such a pre-operative prediction tool into clinical practice offers the potential to establish more accurate expectations of patient-specific improvement that can be achieved with shoulder arthroplasty, and also to align the patient and surgeon on what results to expect at different post-operative timepoints. The practical limitation of deploying such a tool into clinic is the large input burden often required by machine learning algorithms to generate patient-specific predictions; particularly since much of that data may not be routinely present in the patient's electronic medical record. As such, a prerequisite for a machine learning based clinical decision-support tool is the identification of a highly predictive minimal set of pre-operative inputs that can be readily-obtained as part of the normal standard of care.

To develop a clinical decision support tool utilizing a minimal feature set of pre-operative inputs, we first conducted a machine learning analysis on a multi-center clinical database of one platform shoulder prosthesis to create algorithms using the full set of pre-operative data to predict post-operative outcomes of various measures, at multiple post-operative timepoints after aTSA and rTSA. We developed these algorithms using a full feature set (e.g. the full model: n = 291 input parameters) and in the process of doing so, we identified a minimal feature set (e.g. the abbreviated model: n = 19 input parameters) consisting only of the most highly predictive features. Therefore, the goal of this study is to quantify and compare the accuracy associated with the full and abbreviated machine learning models to predict clinical outcomes after aTSA and rTSA.

Methods

We utilized the XGBoost[22] machine learning technique to analyze 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 the 30 clinical sites according to an institutional review board-approved protocol (NYU IRB Study #: i05-144_MOD41). Upon completion of each form, all forms are independently verified, and then are computer-scored on a secured IBM database. All primary aTSA and primary rTSA patients in the database with at least 3 months of follow-up were included. To ensure a homogenous dataset, patients with revisions, humeral fractures, endoprostheses, and hemiarthroplasty were excluded. Primary total shoulder arthroplasty patients who experienced complications and/or revisions were not excluded from this analysis as those experiences contribute to the outcomes variability of the aTSA and rTSA cohorts. These criteria resulted in pre-operative, intra-operative, and post-operative data from 5,774 patients with 17,427 post-operative follow-up visits available for analysis in this machine learning study. The full database contains 291 inputs, including: demographic, diagnoses, comorbidities, implant type, range of motion, radiographic findings, and clinical outcome metric scores [American Shoulder and Elbow Surgeons (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 each of these patient reported outcome scores. Range of motion assessment was performed by the implanting surgeon or their surrogate and was measured with a goniometer. This data was used to create predictive algorithms for the ASES, Constant, Global Shoulder Function score, Visual Analog Scale (VAS) pain score, active abduction, active forward elevation, and active external rotation with the arm at the side at multiple timepoints after aTSA or rTSA, including: 3-6 months, 6-9 months, 1 year [9-18 months], 2-3 years [18-36 months], 3-5 years [36-60 months], and 5+ years [60+ months].

XGBoost was used to create the predictive algorithms for both the full and abbreviated models. XGBoost is a supervised, ensemble machine learning technique of multiple-regression trees that are built by iteratively partitioning the training dataset into multiple small batches using a method called boosting.[22] The full model utilized all 291 inputs from the database; whereas, the abbreviated model utilized only a minimal feature set of 19 pre-operative inputs (Table 1) to predict the Global Shoulder Function score, the VAS pain score, active abduction, forward elevation, and external rotation. As described in Table 1, this minimal feature set is a selection of patient demographic, diagnoses, comorbidities, pre-operative range of motion, and patient responses to a few subjective questions. These specific 19 input parameters were identified using domain knowledge, prevalence of the feature, uniqueness of the feature values for patients, and finally the importance of features to the model. Of note, the uniqueness of a feature is computed using an information theory metric known as entropy, which measures if values in feature are highly uncertain and thus likely to be highly random across patients making it hard to outcomes based on that feature. Also, the importance of a feature to the model was computed based on F-scores determined from the XGBoost algorithm. The F-score quantifies the frequency that a particular feature is used as a candidate for the split by the decision-tree algorithm.[22] The prevalence, entropy, and F-score were used to determine individual ranking of each feature by combining into a single ranked list using Reciprocal Fusion Rank Score. When this abbreviated model is supplemented with the 10 additional questions needed to calculate the pre-operative

ASES score and the 20 additional questions needed to calculate the pre-operative Constant score, the abbreviated model can also be used to predict the ASES and Constant scores at each post-operative timepoint for both aTSA and rTSA. Finally, we conducted an additional analysis which supplemented the abbreviated model predictions with implant size/type data, and measurements of native glenoid version and inclination (i.e. beta angle), to simulate the additional predictive accuracy that could be acquired through utilization of data readily-available from CT-based pre-operative planning software.

Similar to the methodology in our previous work,[12] this data was split 2:1 into mutually exclusive datasets to build and test the predictive models using each of the full and abbreviated feature sets for each outcome metric at each post-operative timepoint. A random selection of 66.7% of the data defined the training cohort and the remaining 33.3% defined the validation test cohort. The performance of the full and abbreviated models to predict post-operative outcomes after aTSA and rTSA was quantified by the Mean Absolute Error (MAE) between the actual and predicted values for each outcome measure at each post-operative timepoint in the 33.3% validation test cohort. To evaluate the relative learning ability of the full and abbreviated XGBoost models, we also conducted a baseline average analysis (i.e. average error associated with selecting the cohort average) as the study control. Finally, a subgroup analysis was also performed using the XGBoost machine learning technique for the full and abbreviated models to predict if a patient would experience clinical improvement greater than the MCID[19] and SCB[20] patient satisfaction anchor-based thresholds for each outcome measure at 2-3 years follow-up. The performance of the full and abbreviated models to predict if a patient will achieve the MCID and SCB improvement thresholds was quantified using the classification metrics of precision (or positive predictive value, which quantifies the ability of a model to not identify a negative as positive), recall (or sensitivity, which quantifies the ability of a model to identify a positive as a positive), and the area under the receiver operating curve (AUROC).[2,9,11] As data science may be new to the reader, AUROC of 0.5 is considered random, >0.7 is considered acceptable, >0.8 is considered good, and >0.9 is considered excellent discrimination for a predictive model.[9,11]

Results

The clinical data from 2,153 primary aTSA patients (7,305 visits; average follow-up=46.4 ± 35.6 months) and 3,621 primary rTSA patients (10,122 visits; average follow-up=31.0 ± 25.8 months) was used to build and test predictive models at each post-operative timepoint: 3-6 months (aTSA=1282 and rTSA=2227 visits), 6-9 months (aTSA=658 and rTSA=1177 visits), 1 year (aTSA=1451 and rTSA=2445 visits), 2-3 years (aTSA=1347 and rTSA=1882 visits), 3-5 years (aTSA=1321 and rTSA=1482 visits), and 5+ years (aTSA=1246 and rTSA=909 visits). A summary of demographics, diagnoses, and comorbidities for the aTSA and rTSA patient cohorts are presented in Table 2. Pre-operative, post-operative, and pre-to-post-operative improvement in outcomes and also complication rates for the aTSA (Table 3) and rTSA (Table 4) patient cohorts at each follow-up duration are presented in Tables 3 and 4, respectively. aTSA and rTSA outcomes at each follow-up duration, stratified by age and gender, are presented in the online supplemental tables.

A comparison of the error between the actual and predicted outcomes in the validation dataset demonstrates that both the full and abbreviated XGBoost models had lower MAE relative to the baseline average study control MAE for all clinical outcome measure predictions, and for both aTSA and rTSA at all post-operative timepoints. (Table 5) MAE associated with each aTSA and rTSA outcome prediction at each post-operative timepoint are presented in the online supplemental tables. As described in these tables, the prediction accuracy observed for aTSA and rTSA were similar for both the full and abbreviated models; though, for both the full and abbreviated models, MAE was slightly higher at early post-operative timepoints as compared to later timepoints. A comparison of MAE between the full and abbreviated models demonstrates that each model had similar predictive accuracy for each of the ASES (full model = ±11.7 vs.

abbreviated model = ±12.0), Constant (±8.9 vs. ±9.8), Global Shoulder Function (±1.4 vs. ±1.5), VAS pain (±1.3 vs. ±1.4), active abduction (±20.4º vs. ±21.8º), forward elevation (±17.6º vs. ±19.2º), and external rotation (±12.2º vs. ±12.6º) outcome measures, despite the abbreviated model only utilizing a minimal feature set of pre-operative inputs to inform its predictions. Specifically, across all post-operative timepoints analyzed, the average difference in MAE between the full and abbreviated model predications was ±0.3 MAE for the ASES score (±0.3 aTSA, ±0.4 rTSA), ±0.9 for the Constant (±0.7 aTSA, ±0.8 rTSA), ±0.1 for the Global Shoulder Function(±0.1 aTSA, ±0.1 rTSA), ±0.1 for the VAS pain (±0.0 aTSA, ±0.2 rTSA), ±1.4º for abduction (±1.1 aTSA, ±1.2 rTSA), ±1.6º for forward elevation (±1.7 aTSA, ±1.4 rTSA), and ±0.4º for external rotation (±0.1 aTSA, ±0.4 rTSA). Of note, only marginal improvements in MAE were observed for each outcome measure prediction when the abbreviated model was supplemented with implant size/type data and measurements of native glenoid anatomy. (Table 5)

The full and abbreviated model predictions for MCID improvement for each outcome metric at 2-3 years follow-up is presented in Table 6. The full predictive models achieved 82-96% accuracy in MCID with an AUROC between 0.75-0.97 for aTSA patients and the abbreviated predictive models achieved 82-96% accuracy in MCID with an AUROC between 0.70-0.95 for aTSA patients. The full predictive models achieved 91-99% accuracy in MCID with an AUROC between 0.82-0.98 for rTSA patients and the abbreviated predictive models achieved 91-99% accuracy in MCID with an AUROC between 0.84-0.94 for rTSA patients. Similarly, the full and abbreviated model predictions for SCB improvement for each outcome metric at 2-3 years follow-up is presented in Table 7. The full predictive models achieved 79-90% accuracy in SCB with an AUROC between 0.74-0.90 for aTSA patients and the abbreviated predictive models achieved 76-90% accuracy in SCB with an AUROC between 0.70-0.89 for aTSA patients. Finally, the full predictive models achieved 83-92% accuracy in SCB with an AUROC between 0.78-0.88 for rTSA patients and the abbreviated predictive models achieved 81-90% accuracy in SCB with an AUROC between 0.70-0.87 for rTSA patients.

Discussion

The results of our 5,774 patient machine learning study demonstrate that an abbreviated model utilizing a minimal feature set of only 19 pre-operative inputs provides similar accuracy as the full model utilizing 291 inputs when predicting aTSA and rTSA outcomes at multiple post-operative timepoints. At each post-operative timepoint, the full and abbreviated models had similar MAE when predicting each outcome measure demonstrating the capability of the predictive algorithms to account for outcomes variability during both the recovery period and even into mid- and long-term followup as outcomes decline with age and deterioration. Additionally, no differences in accuracy were observed between aTSA and rTSA outcomes predictions at any timepoint for either the full or abbreviated models demonstrating the prediction algorithms were equally effective for each prosthesis type. Only minor improvements in the abbreviated model predictions were observed after incorporating implant size/type data and measurements of native glenoid version and inclination. Furthermore, the full and abbreviated models were equally effective at risk-stratifying patients using only pre-operative data, by accurately identifying patients at greatest risk for poor outcomes based upon MCID thresholds (full model accuracy >82%, AUROC >0.75 vs. abbreviated model accuracy >82%, AUROC >0.70), and identifying patients most likely to achieve excellent outcomes based upon SCB thresholds (full model accuracy >79%, AUROC >0.74 vs. abbreviated model accuracy >76%, AUROC >0.70) at 2-3 years follow-up for all outcome measurements.

Pre-operatively communicating the expected result from a proposed surgical treatment is an important component of informed consent. However, few surgeons can accurately predict the outcomes that a cohort of patients may achieve, nor do most know if a particular patient will do better or worse (and by how much) than the "average" patient. Machine learning based predictive outcomes algorithms are not a simple heuristic, rather these computational techniques analyze large quantities of clinical data and consider numerous parameters to inform its evidence-based outcomes predictions. As such, these predictions are clinically useful for shared decision-making and can be used to more effectively communicate expected outcomes and better inform the risks and benefits of a surgical procedure. For the shoulder surgeon, an evidenced-based tool that can accurately predict patient-specific outcomes after aTSA and rTSA from input of only 19 patient questions and active ROM measurements, has many practical applications. First, these predictions can better align the surgeon's objectives for the procedure with that of the patient and establish more accurate expectations of what can be achieved with shoulder arthroplasty, given each patient's unique demographics, diagnoses, and comorbidities. Better alignment between the patient and surgeon on what can be achieved with shoulder arthroplasty may translate to improved patient satisfaction with the procedure.[7,8,15,17] Furthermore, these predictions can aid the shoulder surgeon in selecting the best technique/treatment for a particular patient based upon a comparison of multiple different projected results, considering the patient's unique model inputs, as exemplified in this study by the comparative aTSA and rTSA outcome predictions. When these predictions are considered in a comparative manner, they function as a patient-specific tool to optimize clinical outcomes for various techniques and treatment options. Additionally, considering these predictions relative to the age and gender stratified outcomes for aTSA and rTSA (presented online) can be used as a quality assessment metric to assess performance at each post-operative timepoint. Future work should perform additional predictive analyses to compare and quantify the impact of complications on model predictions, particular as it relates to false-positive predictions for MCID and SCB, and also create patient-specific predictive models for complication risk associated with each of the various techniques and treatment options.

More controversial is the use of a predictive tool to identify if a specific patient is an appropriate candidate for an elective surgical treatment. The abbreviated model accurately identified >82% of patients that would achieve improvement greater than the MCID threshold and >76% of patients that would achieve improvement greater than the SCB threshold across all outcome metrics analyzed. Furthermore, the abbreviated model algorithms were associated with average MCID AUROC values of 0.82 for aTSA and 0.89 for rTSA and average SCB AUROC values of 0.85 for aTSA and 0.82 for rTSA. Thus, our AUROC results suggest these predictive algorithms created from a minimal feature set have on average, good (>0.8) to excellent (>0.9) discrimination of patients in the validation cohort to achieve MCID and SCB improvement. While these predictions can be helpful to pre-operatively identify patients who are good (or poor) candidates for these elective procedures, it must be acknowledged that each patient's needs are unique and different and that patient-specific requirements for pain relief and functional improvement may not align with the established MCID[19] or SCB[20] improvement thresholds. As such, this tool should not define who is eligible for surgical treatment; instead, it should be used to support treatment and never be misused to deny treatment.

Machine learning predictions using the full model, and its 291 feature inputs, have limited practical application for creating a decision-support tool that can be used in the typical clinical setting, due to the substantial input burden on the patient, office staff, and health care provider. Fortunately, our results demonstrate that similar levels of predictive accuracy can be attained using as few as 19 patient questions and active range of motion measurements. We observed minor improvements in predictive accuracy when the abbreviated model was supplemented with implant size/type data and measurements of native glenoid version and inclination; this additional data has the potentially to be seamlessly added to the model from CT-based pre-operative planning software without additional input required by the office staff. While future work is necessary to create and deploy the clinical software that utilizes these machine learning algorithms, the results of our study objectively demonstrate the efficacy of a minimal feature set algorithm to predict aTSA and rTSA outcomes at multiple post-operative timepoints. Furthermore, utilization of this minimal feature set comprised of the most predictive inputs represents an opportunity for more efficient data collection and resource utilization, as it is inferred from our results that the majority of the pre-operative data in the full model (including the majority of the questions from the 5 outcome metrics contained in the database: ASES, Constant, UCLA, SST, and SPADI) is superfluous, adding little additional predictive benefit to our model. Future work can apply these machine learning techniques to construct a new and more efficient shoulder arthroplasty specific patient reported outcome measure that eliminates inputs of little predictive value and only utilize those patient questions found to be highly predictive of post-operative outcomes and/or patient satisfaction.

Aside from our previous work,[12] the use of machine learning to predict outcomes after shoulder arthroplasty is novel, though a few studies have recently utilized machine learning to predict short-term complications after shoulder arthroplasty[10] and outcomes after hip[4] and knee[4,23] arthroplasty. Our machine learning analysis of aTSA and rTSA outcomes builds upon previous work[5,6,18-21] which used more traditional statistical techniques to compare aTSA and rTSA outcomes. Our results demonstrated similar MAE between aTSA and rTSA predictions for each clinical outcome measure at each post-operative timepoint; however, at earlier post-operative timepoints, we observed slightly higher MAE than later timepoints, despite having more data at those earlier timepoints. This finding is likely due to the greater variability in outcomes early due to patients having different healing rates, and perhaps also due to different methods and utilization of rehabilitation programs. As has been reported previously by Simovitch et al.[21] and Levy et al.[13], aTSA and rTSA patients can continue to experience improvement for up to 2 years after surgery, after which improvement plateaus; these findings are consistent with our own observations for both aTSA and rTSA cohorts in our dataset. (Tables 3 and 4) Since machine learning algorithms improve and reduce error by learning with new data, as additional clinical data is obtained, future work will refine these algorithms to further reduce model MAE and improve predictive accuracy.

Our study has several limitations. First, our clinical outcomes database is contributed to by 30 different sites/surgeons, and data from each site/surgeon inevitably contains some bias. As such, the derived models will also contain bias.[1,2,14,16] To reduce collection bias and input variability, all sites were trained to collect data using standardized data forms and all completed forms were independently verified. Second, each of the surgeons who contributed clinical data are experienced shoulder specialists who have multiple years of experience with the prosthesis utilized in this study; as such, these predictions may not translate to less-experienced surgeons or those who have not completed the learning curve with this platform shoulder prosthesis. Third, our clinical database consists only of patients who elected to undergo shoulder arthroplasty, and those patients are primarily elderly, non-Hispanic, Caucasians of European descent. For example, we do not collect data on individuals who were candidates but elected to forgo surgery due to comorbid illness, financial, or personal reasons. Therefore, model predictions may not be representative of the outcomes achieved by patients of different demographics, regions, or ethnicity/race and model predictions may be biased against patients too sick to safely undergo the procedure or patients whose condition was not sufficiently degenerative to have the procedure. Fourth, our models were developed from a dataset of primary aTSA and primary rTSA patients using one platform shoulder prosthesis, where patients with revisions, humeral fractures, or hemiarthroplasty were excluded; therefore, model predictions may not be appropriate for those excluded indications or other prosthesis types or designs. Fifth, our study utilized one tree-based machine learning technique to construct algorithms that quantify outcomes after shoulder arthroplasty, other techniques, such as deep-learning, could achieve better predictive accuracy than XGBoost, as has been shown previously[12] using the Wide and Deep[24] technique. Despite slight improvements in predictive accuracy using Wide and Deep, we utilized XGBoost in our study because its predictions are more interpretable, providing an F-score identification of the most meaningful parameters used by the model. Knowledge of the model input parameters driving up/down the patient-specific predictions can be clinically useful, particularly if those features are modifiable by patient. Sixth, our clinical database, while extensive, contains some missing data; fortunately, XGBoost manages missing values and data sparsity well and imputes missing values on its own by minimizing the error rate for each tree as it learns. Finally, while we utilized a minimal feature set of 19 of the most predictive features in our database, there may be other features that are more predictive and clinically meaningful that were not included in full model and are not currently collected in our clinical database. Future work should continue to refine the feature set to identify more clinically meaningful and highly predictive parameters that minimize model MAE while also minimizing the user input burden, thereby, ensuring the decision-support tool can be efficiently implemented in the clinical setting.

Conclusion

In conclusion, we utilized a commercially available supervised machine learning technique to analyze a clinical database of one platform shoulder prosthesis and constructed predictive algorithms using a full model (of 291 inputs) and an abbreviated model (of 19 inputs) and attained similar accuracy with each model to predict outcomes after shoulder arthroplasty at multiple post-operative timepoints in our study of 2,153 primary aTSA and 3,621 primary rTSA patients. The abbreviated prediction model was supplemented with implant size/type data and native glenoid version and inclination measurements, which demonstrated that marginal improvements can be achieved when incorporating pre-operative CT planning data. Finally, both the full and abbreviated model algorithms were able to pre-operatively risk-stratify patients based upon improvement predictions greater than the MCID and SCB patient-satisfaction thresholds for each outcome measure analyzed in our study. These promising results demonstrate an efficient utilization of machine learning algorithms to predict clinical outcomes. The use of a minimal feature set of only 19 preoperative inputs suggests that this tool may be easily used during a surgical consultation to improve decision-making related to shoulder arthroplasty.

Tables Descriptions

Table 1. Description of the Minimal Feature Set of Pre-operative Inputs Utilized by the Abbreviated Prediction Model.

Table 2. Comparison of Demographics, Diagnosis, and Comorbidities for the Primary aTSA and Primary rTSA Patients in this Study

Table 3. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration

Table 4. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration

Table 5. Comparison of the Mean Absolute Error (MAE) Associated with each Outcome Measure Prediction for the Full and Abbreviated Machine Learning Models. Note that these MAE Values are a Weighted Average over each Post-operative Timepoint (3-6 months, 6-9 months, 1 year, 2-3 years, 3-5 years, and 5+ years); Supplemental Tables are Included Online Reporting the MAE for each Outcome Measure at each Post-operative Timepoint.

Table 6. XGBoost predictions using the full and abbreviated models for aTSA & rTSA patients who experienced clinical improvement at 2-3 years follow-up greater than the MCID[19] threshold for multiple different outcome measures.

Table 7. XGBoost predictions using the full and abbreviated model for aTSA & rTSA patients who experienced clinical improvement at 2-3 years follow-up greater than the SCB[20] threshold for multiple different outcome measures.

Supplemental Tables (available online)

Supplemental Table 1. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age <60yrs at the time of surgery

Supplemental Table 2. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 60 to <70 yrs at the time of surgery

Supplemental Table 3. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 70 to <80 yrs at the time of surgery

Supplemental Table 4. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age ≥80yrs at the time of surgery

Supplemental Table 5. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age <60yrs at the time of surgery

Supplemental Table 6. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 60 to <70 yrs at the time of surgery

Supplemental Table 7. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 70 to <80 yrs at the time of surgery

Supplemental Table 8. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age ≥80yrs at the time of surgery

Supplemental Table 9. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age <60yrs at the time of surgery

Supplemental Table 10. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 60 to <70 yrs at the time of surgery

Supplemental Table 11. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 70 to <80 yrs at the time of surgery

Supplemental Table 12. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age ≥80yrs at the time of surgery

Supplemental Table 13. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age <60yrs at the time of surgery

Supplemental Table 14. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 60 to <70 yrs at the time of surgery

Supplemental Table 15. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 70 to <80 yrs at the time of surgery

Supplemental Table 16. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age ≥80yrs at the time of surgery

Supplemental Table 17. Comparison of Mean Absolute Error Associated with the ASES Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 18. Comparison of Mean Absolute Error Associated with the Constant Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 19. Comparison of Mean Absolute Error Associated with the Global Shoulder Function Score Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 20. Comparison of Mean Absolute Error Associated with the VAS Pain Score Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 21. Comparison of Mean Absolute Error Associated with the Active Abduction (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 22. Comparison of Mean Absolute Error Associated with the Active Forward Elevation (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

Supplemental Table 23. Comparison of Mean Absolute Error Associated with the Active External Rotation (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

Table 1. Description of the Minimal Feature Set of Pre-operative Inputs Utilized by the Abbreviated Prediction Model

Feature	Description and Unit	Range/Inputs
Age	Age of patient, years	18 to 115
Weight	Weight, lbs.	80 to 450
Height	Height, inches	48 to 80
Gender	Male or Female	Male or Female
Previous Shoulder Surgery?	Has the patient previously had a surgical operation on their shoulder?	Yes or No
Is surgery on the dominant shoulder?	Will the upcoming arthroplasty be on the patient's dominant shoulder?	Yes or No
Primary Diagnosis	What is the patient's diagnosis?	Osteoarthritis, Osteonecrosis, Rotator Cuff Tear, Rotator Cuff Tear Arthropathy, Rheumatoid Arthritis, and Post Traumatic Arthritis
Comorbidities	What are the patient's comorbidities?	No comorbidities, inflammatory arthritis, hypertension, heart disease, diabetes, chronic renal failure, and tobacco use
Preoperative Active Abduction	Active arm elevation in the frontal plane, degrees	0 to 180º
Preoperative Active Forward Elevation	Active arm elevation in the sagittal plane, degrees	0 to 180º
Preoperative Active External Rotation	Active lateral rotation of the arm, with the arm at the side, degrees	-90 to 90º
Preoperative Passive External Rotation	Passive lateral rotation of the arm, with the arm at the side, degrees	-90 to 90º
Preoperative Internal Rotation Score	Active medial rotation of the arm, with the arm at the side. Unitless, 8 point numeric scale with the following discreet assignments based on motion to vertebral segments: No motion = 0, hip = 1, buttocks = 2, sacrum = 3, L5 – L4 = 4, L3 – L1 = 5, T12 – T8 = 6, T7 or higher = 7.	0 to 7
Preoperative Global Shoulder Function Score	Global Shoulder Function score is patient assessment of their ability to use their shoulder prior to surgery. 11pt score (0-10), with 10 = full/ normal mobility	0 to 10
Preoperative Pain on a Daily Basis (ie Visual Analog Score (VAS))	VAS pain score is the patient assessment of the pain that they experience on a daily basis prior to surgery. 11pt score (0-10), with 10 = extreme pain	0 to 10

Preoperative Pain at Worst	Patient assessment of the worst pain that they experience on a daily basis prior to surgery. 11pt score (0-10), with 10 = extreme pain	0 to 10
Preoperative Pain when lying on your side	Patient assessment of the pain that they experience when lying on their effected side prior to surgery. 11pt score (0-10), with 10 = extreme pain	0 to 10
Preoperative Pain when touching back of neck	Patient assessment of the pain that they experience when touching back of neck prior to surgery. 11pt score (0-10), with 10 = extreme pain	0 to 10
Preoperative Pain when pushing with affected arm	Patient assessment of the pain that they experience when pushing with affected arm prior to surgery. 11pt score (0-10), with 10 = extreme pain	0 to 10

Table 2. Comparison of Demographics, Diagnosis, and Comorbidities for the Primary aTSA and Primary rTSA Patients in this Study

Demographics, Diagnosis, and Comorbidities	aTSA Patients	rTSA Patients
Age at Surgery	66.1 ± 9.2 yrs	72.5 ± 7.8 yrs
Gender	1111F/1027M/15Unk	2350F/1242M/29Unk
Height	66.6 ± 4.3 in	65.0 ± 4.0 in
Weight	188.9 ± 44.5 lbs	172.7 ± 41.0 lbs
Body Mass Index (BMI)	29.9 ± 6.3	28.7 ± 6.0
% Previous Shoulder Surgery	15.7%	24.7%
Surgery on Dominant Shoulder?	55.8%	62.2%
Osteoarthritis (OA) diagnosis	92.7%	53.2%
Osteonecrosis (ON) diagnosis	3.2%	2.4%
Rotator Cuff Tear (RCT) diagnosis	2.6%	39.3%
Cuff Tear Arthropathy (CTA) diagnosis	0.8%	38.4%
Rheumatoid Arthritis (RA) diagnosis	3.1%	3.5%
Post-traumatic Arthritis (PTA) diagnosis	2.1%	2.4%
No Comorbidities	35.9%	33.0%
Inflammatory Arthritis	11.7%	7.8%
Hypertension	47.4%	53.3%
Heart Disease	13.6%	16.2%

Diabetes	12.2%	13.7%
Chronic Renal Failure	1.2%	2.0%
Tobacco use?	9.8%	7.2%

Table 3. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	35.7 ± 16.4	38.3 ± 14.3	4.1 ± 2.0	6.4 ± 2.1	83.1 ± 30.8	97.1 ± 32.3	19.9 ± 19.6	NA
3 – 6 months	76.5 ± 18.9 / 40.4 ± 21.4	62.3 ± 15.5 / 24.4 ± 17.1	7.2 ± 2.2 / 3.0 ± 2.8	1.7 ± 2.1 / 4.7 ± 2.7	110.2 ± 32.0 / 26.3 ± 37.2	125.4 ± 33.5 / 29.3 ± 38.0	42.6 ± 18.7 / 22.4 ± 19.7	2.1% / 1.4%
6 – 9 months	82.4 ± 17.8 / 44.5 ± 21.5	68.1 ± 14.3 / 29.1 ± 15.9	7.7 ± 2.2 / 3.5 ± 2.7	1.3 ± 1.9 / 5.0 ± 2.7	119.8 ± 31.4 / 36.5 ± 37.2	136.0 ± 30.3 / 39.0 ± 36.2	49.5 ± 18.6 / 28.2 ± 20.8	1.5% / 0.8%
1 year	85.9 ± 17.1 / 49.2 ± 20.6	71.6 ± 13.6 / 33.1 ± 15.9	8.4 ± 1.9 / 4.3 ± 2.5	1.1 ± 1.9 / 5.3 ± 2.6	127.5 ± 32.0 / 44.8 ± 37.1	143.8 ± 29.0 / 47.3 ± 35.3	51.1 ± 18.7 / 30.7 ± 20.7	2.5% / 1.4%
2 – 3 years	87.4 ± 17.1 / 51.3 ± 20.9	73.4 ± 13.8 / 35.2 ± 16.0	8.6 ± 1.9 / 4.7 ± 2.5	1.0 ± 1.9 / 5.4 ± 2.6	129.1 ± 32.1 / 45.4 ± 39.1	146.5 ± 29.1 / 48.7 ± 37.2	52.4 ± 18.8 / 33.0 ± 22.0	3.4% / 2.0%
3 – 5 years	86.2 ± 17.4 / 50.5 ± 20.4	72.8 ± 13.6 / 35.5 ± 14.7	8.6 ± 2.0 / 4.5 ± 2.6	1.1 ± 1.9 / 5.3 ± 2.6	128.4 ± 31.6 / 44.3 ± 37.9	146.8 ± 29.2 / 49.4 ± 36.2	51.5 ± 18.9 / 31.8 ± 21.7	2.6% / 2.0%
5+ years (average follow-up = 90.0 months)	81.5 ± 20.3 / 47.1 ± 23.7	68.3 ± 14.7 / 32.7 ± 16.6	8.2 ± 2.2 / 4.4 ± 2.7	1.5 ± 2.3 / 5.0 ± 2.9	120.5 ± 31.9 / 37.3 ± 38.0	141.0 ± 31.6 / 44.8 ± 37.1	46.1 ± 20.0 / 31.1 ± 22.9	5.2% / 3.4%
							Full Primary aTSA Cohort	3.0%/1.9%

Table 4. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	34.7 ± 15.9	34.9 ± 14.4	3.7 ± 2.1	6.3 ± 2.2	72.7 ± 36.9	85.7 ± 39.3	18.5 ± 21.3	NA
3 – 6 months	73.4 ± 18.8 / 38.9 ± 20.5	58.3 ± 14.9 / 23.7 ± 16.8	6.9 ± 2.1 / 3.2 ± 2.8	1.7 ± 2.1 / 4.6 ± 2.7	104.9 ± 31.4 / 33.1 ± 39.6	120.9 ± 31.9 / 37.1 ± 42.1	28.5 ± 17.6 / 9.8 ± 21.4	3.1% / 0.7%
6 – 9 months	77.8 ± 18.1 / 41.2 ± 19.9	62.8 ± 13.7 / 28.2 ± 16.3	7.3 ± 2.0 / 3.7 ± 2.7	1.5 ± 2.0 / 4.5 ± 2.7	110.1 ± 29.7 / 39.4 ± 37.4	130.0 ± 29.1 / 46.6 ± 40.5	31.6 ± 17.7 / 13.6 ± 22.7	2.5% / 0.9%
1 year	81.2 ± 18.1 / 46.4 ± 20.7	67.0 ± 14.0 / 32.0 ± 16.2	7.9 ± 2.0 / 4.3 ± 2.6	1.3 ± 2.0 / 5.0 ± 2.7	120.6 ± 30.1 / 47.1 ± 39.7	137.8 ± 27.7 / 51.5 ± 41.2	35.7 ± 18.1 / 16.9 ± 22.2	2.3% / 1.0%
2 – 3 years	82.6 ± 18.1 / 46.7 ± 20.7	69.0 ± 13.8 / 34.1 ± 16.1	8.1 ± 1.9 / 4.4 ± 2.6	1.2 ± 2.0 / 5.0 ± 2.7	118.8 ± 30.6 / 46.5 ± 39.1	139.2 ± 26.9 / 54.2 ± 41.9	36.8 ± 17.6 / 18.8 ± 23.1	3.1% / 1.1%
3 – 5 years	82.2 ± 18.8 / 45.9 ± 21.6	68.0 ± 13.9 / 32.5 ± 15.7	8.2 ± 1.9 / 4.4 ± 2.7	1.2 ± 2.0 / 5.0 ± 2.7	117.9 ± 29.1 / 45.2 ± 39.1	137.3 ± 26.9 / 52.4 ± 41.7	36.3 ± 17.9 / 17.3 ± 23.4	2.7% / 1.3%
5+ years (average follow-up = 80.6 months)	79.9 ± 20.5 / 43.7 ± 23.5	65.7 ± 14.9 / 30.3 ± 17.3	7.9 ± 2.2 / 4.0 ± 2.9	1.4 ± 2.2 / 4.9 ± 2.9	112.3 ± 29.2 / 35.9 ± 39.6	130.7 ± 29.5 / 41.3 ± 42.9	32.3 ± 19.6 / 11.8 ± 25.3	2.5% / 1.1%
Full Primary rTSA Cohort								2.7%/1.0%

Table 5. Comparison of the Mean Absolute Error (MAE) Associated with each Outcome Measure Prediction for the Full and Abbreviated Machine Learning Models. Note that these MAE Values are a Weighted Average over each Post-operative Timepoint (3-6 months, 6-9 months, 1 year, 2-3 years, 3-5 years, and 5+ years); Supplemental Tables are Included Online Reporting the MAE for each Outcome Measure at each Post-operative Timepoint.

Prediction Error	ASES Weighted Average MAE (aTSA, rTSA)	Constant Weighted Average MAE (aTSA, rTSA)	Global Shoulder Function Weighted Average MAE (aTSA, rTSA)	VAS Pain Weighted Average MAE (aTSA, rTSA)	Active Abduction Weighted Average MAE (aTSA, rTSA)	Active Forward Elevation Weighted Average MAE (aTSA, rTSA)	Active External Rotation Weighted Average MAE (aTSA, rTSA)
Baseline Average	15.3 (15.6, 14.9)	12.5 (12.7, 12.5)	1.6 (1.6, 1.6)	1.6 (1.5, 1.6)	26.5º (26.4, 26.3)	23.0º (22.9, 23.2)	16.0º (16.2, 15.7)

XGBoost with Full Model	11.7 (11.6, 11.8)	8.9 (9.4, 9.1)	1.4 (1.4, 1.3)	1.3 (1.2, 1.3)	20.4º (20.9, 20.1)	17.6º (18.0, 17.8)	12.2º (13.1, 11.7)
XGBoost with Abbreviated Model	12.0 (11.9, 12.2)	9.8 (10.1, 9.9)	1.5 (1.5, 1.4)	1.4 (1.2, 1.5)	21.8º (22.0, 21.3)	19.2º (19.7, 19.2)	12.6º (13.2, 12.1)
MAE Difference (Full – Abbreviated)	0.3 (0.3, 0.4)	0.9 (0.7, 0.8)	0.1 (0.1, 0.1)	0.1 (0.0, 0.2)	1.4º (1.1, 1.2)	1.6º (1.7, 1.4)	0.4º (0.1, 0.4)
XGBoost with Abbreviated Model + Implant Data	12.0 (11.7, 12.0)	9.7 (9.8, 9.8)	1.4 (1.4, 1.4)	1.3 (1.2, 1.4)	21.7º (21.9, 21.1)	19.0º (19.2, 19.1)	12.4º (13.1, 12.0)

Table 6. XGBoost predictions using the full and abbreviated models for aTSA & rTSA patients who experienced clinical improvement at 2-3 years follow-up greater than the MCID[19] threshold for multiple different outcome measures.

MCID prediction	ASES (aTSA, rTSA)	Constant (aTSA, rTSA)	Global Shoulder Function (aTSA, rTSA)	VAS Pain (aTSA, rTSA)	Abduction (aTSA, rTSA)	Forward elevation (aTSA, rTSA)	External rotation (aTSA, rTSA)
MCID[19]	13.6 (17.0, 10.3)	5.7 (12.8, -0.3)	1.4 (1.7, 1.0)	1.6 (2.7, 1.4)	7.0 (13.9, -1.9)	12.0 (23.1, -2.9)	3.0 (14.5, -5.3)
Patient percent	77.9% (72.9%, 80.6%)	71.3% (63.0%, 77.6%)	75.7% (75.5%, 75.8%)	75.3% (66.5%, 77.9%)	83.5% (78.2%, 88.7%)	79.9% (73.4%, 92.7%)	84.7% (77.8%, 90.4%)
Precision (full model)	95% (94%, 95%)	96% (97%, 98%)	94% (96%, 93%)	92% (91%, 91%)	94% (91%, 98%)	92% (87%, 99%)	95% (90%, 99%)
Recall (full model)	99% (97%, 99%)	99% (99%, 100%)	96% (95%, 98%)	98% (97%, 99%)	94% (91%, 98%)	95% (88%, 99%)	96% (93%, 99%)
Accuracy (full model)	95% (94%, 95%)	97% (96%, 99%)	92% (93%, 92%)	93% (92%, 91%)	90% (86%, 98%)	89% (82%, 99%)	95% (92%, 99%)
AUC (full model)	0.90(0.90, 0.88)	0.95 (0.97, 0.96)	0.88 (0.91, 0.87)	0.87 (0.89, 0.82)	0.83 (0.80, 0.98)	0.79 (0.75, 0.95)	0.83 (0.78, 0.95)
Precision (abbreviated model)	91% (90%, 91%)	94% (93%, 94%)	93% (94%, 91%)	91% (89%, 91%)	90% (87%, 95%)	89% (90%, 97%)	89% (85%, 94%)

Recall (abbreviated model)	99% (97%, 99%)	99% (97%, 99%)	95% (94%, 97%)	96% (96%, 97%)	91% (89%, 95%)	95% (91%, 98%)	92% (88%, 96%)
Accuracy (abbreviated model)	93% (92%, 93%)	97% (96%, 99%)	90% (89%, 91%)	90% (88%, 92%)	84% (82%, 94%)	87% (87%, 98%)	87% (84%, 97%)
AUC (abbreviated model)	0.88 (0.87, 0.84)	0.94 (0.95, 0.93)	0.87 (0.89, 0.86)	0.87 (0.86, 0.84)	0.76 (0.72, 0.94)	0.72 (0.70, 0.89)	0.78 (0.73, 0.90)

Table 7. XGBoost predictions using the full and abbreviated model for aTSA & rTSA patients who experienced clinical improvement at 2-3 years follow-up greater than the SCB[20] threshold for multiple different outcome measures.

SCB prediction	ASES (aTSA, rTSA)	Constant (aTSA, rTSA)	Global Shoulder Function (aTSA, rTSA)	VAS Pain (aTSA, rTSA)	Abduction (aTSA, rTSA)	Forward elevation (aTSA, rTSA)	External rotation (aTSA, rTSA)
SCB[20]	31.5 (37.6, 25.9)	19.1 (25.4, 13.6)	3.1 (3.9, 2.4)	3.2 (3.8, 2.6)	28.5 (36.1, 19.6)	35.4 (45.5, 22.3)	11.7 (20.1, 3.6)
Patient percent	66.7% (57.3%, 73.1%)	62.3% (52.8%, 71.7%)	58.8% (57.4%, 70.7%)	62.8% (61.5%, 72.5%)	64.7% (55.2%, 73.9%)	61.3% (48.4%, 75.7%)	70.1% (64.6%, 82.0%)
Precision (full model)	88% (89%, 88%)	91% (93%, 94%)	83% (91%, 91%)	85% (92%, 88%)	86% (80%, 89%)	85% (85%, 91%)	85% (79%, 92%)
Recall (full model)	93% (93%, 94%)	95% (88%, 97%)	93% (86%, 90%)	99% (92%, 97%)	86% (87%, 87%)	91% (88%, 93%)	89% (92%, 93%)
Accuracy (full model)	87% (89%, 88%)	91% (90%, 92%)	85% (87%, 86%)	86% (89%, 88%)	82% (81%, 83%)	84% (86%, 88%)	81% (79%, 88%)
AUC (full model)	0.84 (0.89, 0.81)	0.90 (0.90, 0.88)	0.84 (0.87, 0.84)	0.82 (0.86, 0.81)	0.81 (0.80, 0.78)	0.83 (0.87, 0.83)	0.76 (0.74, 0.78)
Precision (abbreviated model)	87% (88%, 86%)	90% (91%, 92%)	82% (89%, 91%)	85% (92%, 88%)	82% (76%, 84%)	78% (74%, 86%)	82% (76%, 89%)
Recall (abbreviated model)	92% (92%, 93%)	95% (88%, 97%)	93% (86%, 90%)	99% (92%, 97%)	84% (83%, 84%)	88% (81%, 91%)	88% (89%, 91%)
Accuracy (abbreviated model)	87% (89%, 88%)	90% (88%, 90%)	84% (85%, 86%)	87% (90%, 88%)	80% (79%, 81%)	81% (82%, 81%)	79% (76%, 84%)
AUC (abbreviated model)	0.82 (0.86, 0.76)	0.89 (0.89, 0.87)	0.83 (0.86, 0.83)	0.85 (0.87, 0.82)	0.72 (0.74, 0.70)	0.74 (0.78, 0.70)	0.73 (0.70, 0.76)

Suppl Table 1. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age <60yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	31.1 ± 14.6	35.4 ± 14.0	4.0 ± 1.9	6.9 ± 2.0	82.1 ± 33.6	97.3 ± 37.6	20.3 ± 19.9	NA
3 – 6 months	65.7 ± 23.9 / 34.3 ± 24.6	55.0 ± 17.9 / 20.4 ± 18.5	6.5 ± 2.5 / 2.3 ± 3.1	2.9 ± 3.0 / 4.1 ± 3.3	102.7 ± 32.3 / 19.4 ± 36.4	118.3 ± 38.4 / 21.4 ± 41.6	39.8 ± 19.0 / 18.2 ± 20.2	2.9%/1.9%
6 – 9 months	68.9 ± 21.8 / 37.7 ± 23.2	58.9 ± 17.7 / 26.9 ± 20.8	6.7 ± 1.9 / 2.9 ± 2.6	2.7 ± 2.5 / 3.9 ± 2.9	105.0 ± 36.0 / 33.0 ± 43.8	125.1 ± 32.6 / 42.1 ± 43.9	43.0 ± 21.4 / 21.3 ± 23.4	5.0%/5.0%
1 year	75.0 ± 23.2 / 44.0 ± 24.1	65.2 ± 15.1 / 28.2 ± 15.8	8.1 ± 1.9 / 4.0 ± 2.3	2.1 ± 2.6 / 5.0 ± 3.2	122.4 ± 32.9 / 40.4 ± 37.7	138.7 ± 31.8 / 40.5 ± 36.6	48.7 ± 20.2 / 25.5 ± 20.7	4.9%/2.9%
2 – 3 years	82.7 ± 18.3 / 51.5 ± 20.9	68.1 ± 14.2 / 32.5 ± 16.3	8.2 ± 2.0 / 4.2 ± 2.3	1.3 ± 2.1 / 5.4 ± 2.8	124.2 ± 34.9 / 46.2 ± 43.6	146.3 ± 30.0 / 49.5 ± 38.8	51.7 ± 18.7 / 31.2 ± 22.5	5.9%/3.5%
3 – 5 years	80.3 ± 19.8 / 50.1 ± 21.4	68.7 ± 14.9 / 33.2 ± 12.5	8.0 ± 2.3 / 4.1 ± 2.4	1.8 ± 2.3 / 5.0 ± 2.6	117.2 ± 31.6 / 32.1 ± 41.0	145.6 ± 29.7 / 44.9 ± 38.8	46.6 ± 21.4 / 25.5 ± 26.4	7.3%/4.9%
5+ years	74.4 ± 20.8 / 47.5 ± 21.8	68.1 ± 18.0 / 34.9 ± 15.0	7.5 ± 2.4 / 4.1 ± 2.6	2.2 ± 2.6 / 5.1 ± 2.7	117.9 ± 33.3 / 38.9 ± 31.1	145.1 ± 38.7 / 41.5 ± 29.4	44.9 ± 19.7 / 28.0 ± 24.8	10.4%/10.4%
Primary aTSA Cohort, Female, age <60yrs at the time of surgery								5.8%/4.4%

Suppl Table 2. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 60 to <70 yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	33.6 ± 15.8	35.9 ± 13.1	3.9 ± 2.0	6.6 ± 2.1	81.4 ± 30.1	95.0 ± 33.4	21.2 ± 19.7	NA
3 – 6 months	75.5 ± 17.7 / 42.5 ± 20.5	60.4 ± 14.2 / 25.7 ± 15.9	7.2 ± 2.1 / 3.2 ± 2.7	1.7 ± 2.1 / 5.0 ± 2.6	104.8 ± 32.3 / 25.8 ± 38.9	121.6 ± 35.1 / 29.1 ± 42.2	42.2 ± 19.3 / 22.7 ± 19.7	4.3%/2.5%

6 – 9 months	80.5 ± 19.1 / 45.7 ± 23.1	66.0 ± 13.9 / 30.2 ± 16.0	7.8 ± 2.1 / 3.9 ± 2.8	1.6 ± 2.2 / 5.1 ± 2.7	120.3 ± 31.3 / 37.0 ± 39.5	137.1 ± 30.4 / 40.6 ± 38.3	52.2 ± 19.5 / 29.4 ± 21.6	3.0%/1.8%
1 year	85.9 ± 15.6 / 51.6 ± 19.2	70.1 ± 12.5 / 34.7 ± 14.1	8.5 ± 1.9 / 4.7 ± 2.4	1.0 ± 1.7 / 5.5 ± 2.4	125.9 ± 32.0 / 45.8 ± 36.9	145.2 ± 28.6 / 52.2 ± 35.8	51.5 ± 18.6 / 31.3 ± 21.6	3.9%/1.9%
2 – 3 years	85.8 ± 19.7 / 51.0 ± 23.1	70.9 ± 13.3 / 34.5 ± 15.0	8.6 ± 1.9 / 4.9 ± 2.6	1.2 ± 2.2 / 5.3 ± 2.8	128.2 ± 33.1 / 45.0 ± 40.0	147.4 ± 29.1 / 51.0 ± 38.0	55.2 ± 18.6 / 34.6 ± 22.3	4.8%/2.2%
3 – 5 years	84.5 ± 18.5 / 50.3 ± 20.7	69.9 ± 14.0 / 34.9 ± 15.8	8.4 ± 2.0 / 4.6 ± 2.6	1.2 ± 2.1 / 5.3 ± 2.6	124.8 ± 33.4 / 45.1 ± 38.3	142.5 ± 31.3 / 52.0 ± 37.4	51.5 ± 19.7 / 30.5 ± 22.4	4.0%/3.7%
5+ years	77.4 ± 23.1 / 47.0 ± 24.1	64.9 ± 15.6 / 33.0 ± 16.9	8.0 ± 2.2 / 4.8 ± 2.8	1.9 ± 2.7 / 5.1 ± 3.0	116.3 ± 32.7 / 36.4 ± 41.9	137.4 ± 35.3 / 48.2 ± 40.7	45.5 ± 21.6 / 31.2 ± 23.6	4.4%/2.8%
Primary aTSA Cohort, Female, age 60 to <70yrs at the time of surgery								4.1%/2.6%

Suppl Table 3. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 70 to <80 yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	32.3 ± 14.8	35.1 ± 12.8	3.9 ± 2.0	6.7 ± 2.0	78.7 ± 30.0	93.6 ± 32.3	20.0 ± 20.4	NA
3 – 6 months	76.2 ± 19.5 / 43.6 ± 23.2	59.5 ± 15.5 / 25.8 ± 17.7	7.3 ± 2.2 / 3.2 ± 2.7	1.7 ± 2.3 / 5.0 ± 2.9	107.0 ± 33.6 / 27.8 ± 37.8	120.3 ± 37.5 / 30.3 ± 38.8	43.2 ± 19.1 / 21.8 ± 21.0	2.8%/2.4%
6 – 9 months	81.9 ± 17.1 / 49.1 ± 20.9	65.7 ± 14.3 / 30.1 ± 17.3	7.6 ± 2.2 / 3.6 ± 2.6	1.0 ± 1.5 / 5.8 ± 2.5	115.1 ± 33.7 / 37.7 ± 41.7	132.6 ± 34.8 / 39.1 ± 38.5	49.4 ± 18.4 / 28.6 ± 21.7	0.9%/0.0%
1 year	86.7 ± 14.0 / 53.8 ± 19.3	69.5 ± 13.1 / 33.8 ± 16.6	8.5 ± 1.8 / 4.5 ± 2.4	0.8 ± 1.6 / 5.8 ± 2.4	125.6 ± 33.1 / 46.4 ± 38.0	141.3 ± 30.9 / 48.3 ± 35.5	51.6 ± 18.7 / 31.1 ± 20.7	1.7%/0.7%
2 – 3 years	87.1 ± 17.4 / 55.0 ± 20.7	72.0 ± 14.5 / 37.3 ± 17.7	8.7 ± 1.7 / 4.9 ± 2.3	0.9 ± 1.8 / 5.9 ± 2.5	127.9 ± 34.1 / 49.7 ± 42.2	143.1 ± 33.8 / 48.6 ± 40.2	52.1 ± 19.9 / 32.8 ± 23.5	3.6%/2.8%
3 – 5 years	86.3 ± 16.7 / 55.7 ± 18.7	71.6 ± 13.7 / 37.9 ± 15.3	8.6 ± 2.1 / 4.7 ± 2.7	1.0 ± 1.9 / 5.9 ± 2.4	130.9 ± 32.6 / 49.9 ± 39.7	144.2 ± 32.9 / 51.4 ± 39.4	55.5 ± 18.4 / 34.3 ± 21.2	2.9%/1.6%
5+ years	82.2 ± 19.2 / 49.8 ± 25.2	66.9 ± 13.2 / 33.1 ± 14.2	8.2 ± 2.1 / 4.8 ± 2.3	1.3 ± 2.2 / 5.1 ± 3.0	120.5 ± 32.3 / 41.3 ± 36.8	138.2 ± 32.9 / 44.3 ± 36.7	46.5 ± 20.8 / 33.1 ± 26.1	1.7%/1.3%
Primary aTSA Cohort, Female, age 70 to <80yrs at the time of surgery								2.4%/1.6%

Suppl Table 4. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age ≥80yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	30.4 ± 14.7	30.0 ± 12.1	3.6 ± 2.2	6.8 ± 2.1	72.4 ± 24.8	84.7 ± 28.0	16.9 ± 20.6	NA
3 – 6 months	77.4 ± 17.2 / 43.2 ± 18.0	60.8 ± 14.6 / 30.4 ± 13.9	7.1 ± 2.7 / 3.2 ± 3.2	1.6 ± 2.2 / 5.1 ± 2.6	110.6 ± 32.3 / 33.6 ± 32.7	119.0 ± 35.3 / 28.6 ± 37.6	38.7 ± 21.3 / 22.8 ± 16.9	0%/0%
6 – 9 months	86.4 ± 11.1 / 52.9 ± 18.4	61.8 ± 10.5 / 33.0 ± 16.4	8.5 ± 1.9 / 3.7 ± 2.3	0.9 ± 1.0 / 5.9 ± 2.6	110.9 ± 27.0 / 37.0 ± 22.8	117.6 ± 22.7 / 33.3 ± 28.9	43.4 ± 20.0 / 26.1 ± 17.6	0%/0%
1 year	85.3 ± 13.0 / 52.5 ± 15.3	69.0 ± 10.0 / 39.4 ± 13.8	8.6 ± 2.1 / 4.6 ± 2.1	0.7 ± 1.6 / 5.8 ± 2.6	118.3 ± 33.2 / 45.7 ± 34.4	128.6 ± 38.4 / 42.1 ± 42.4	50.3 ± 20.4 / 32.4 ± 18.6	0%/0%
2 – 3 years	84.7 ± 14.6 / 54.6 ± 20.9	68.5 ± 11.1 / 40.9 ± 16.3	8.1 ± 2.8 / 4.8 ± 2.5	0.8 ± 1.7 / 6.5 ± 2.5	123.2 ± 29.9 / 47.1 ± 32.5	136.8 ± 27.1 / 52.0 ± 29.0	54.5 ± 19.6 / 36.0 ± 23.6	0%/0%
3 – 5 years	85.1 ± 14.7 / 58.2 ± 19.0	70.4 ± 12.5 / 44.8 ± 18.4	8.3 ± 2.3 / 4.3 ± 2.5	0.6 ± 1.5 / 6.9 ± 2.1	120.2 ± 33.7 / 44.1 ± 36.4	138.3 ± 34.0 / 51.5 ± 34.5	55.1 ± 19.9 / 40.8 ± 24.7	0%/0%
5+ years	86.1 ± 16.0 / 51.5 ± 27.0	64.1 ± 11.9 / 35.4 ± 18.1	8.1 ± 1.9 / 4.4 ± 2.6	0.9 ± 1.8 / 5.3 ± 3.5	111.4 ± 28.9 / 40.7 ± 34.2	130.2 ± 32.0 / 42.4 ± 31.2	47.2 ± 17.4 / 37.6 ± 17.2	2.9%/2.9%
Primary aTSA Cohort, Female, age ≥80yrs at the time of surgery								0.4%/0.4%

Suppl Table 5. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age <60yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	36.4 ± 16.2	40.5 ± 13.9	4.1 ± 2.1	6.4 ± 2.1	85.5 ± 30.0	100.2 ± 30.5	16.3 ± 18.9	NA
3 – 6 months	71.9 ± 20.5 / 34.6 ± 22.4	62.0 ± 16.7 / 21.3 ± 17.9	6.8 ± 2.2 / 2.5 ± 2.5	2.3 ± 2.3 / 3.9 ± 2.7	110.4 ± 31.6 / 23.5 ± 37.2	127.9 ± 31.4 / 29.0 ± 37.3	38.9 ± 19.4 / 22.2 ± 19.8	0.5%/0.5%
6 – 9 months	79.4 ± 19.5 / 42.5 ± 21.2	69.6 ± 14.3 / 28.0 ± 14.5	7.5 ± 2.0 / 3.4 ± 2.6	1.7 ± 2.0 / 4.8 ± 2.5	121.2 ± 28.2 / 33.3 ± 34.1	140.4 ± 27.4 / 37.4 ± 32.3	45.0 ± 17.5 / 27.1 ± 18.3	0%/0%
1 year	80.4 ± 21.5 / 43.7 ± 23.2	71.2 ± 15.6 / 32.0 ± 17.0	7.9 ± 2.1 / 3.9 ± 2.8	1.7 ± 2.3 / 4.9 ± 2.7	124.7 ± 31.6 / 42.3 ± 37.3	143.3 ± 27.2 / 44.2 ± 33.6	46.9 ± 19.3 / 29.7 ± 20.9	2.9%/2.3%

2 – 3 years	83.9 ± 17.9 / 49.7 ± 19.6	73.8 ± 15.9 / 36.8 ± 15.5	8.3 ± 1.9 / 4.7 ± 2.3	1.5 ± 2.0 / 5.4 ± 2.4	124.7 ± 32.2 / 41.4 ± 36.6	147.1 ± 29.6 / 46.5 ± 33.4	48.8 ± 20.0 / 34.0 ± 21.8	2.0%/0.7%
3 – 5 years	83.3 ± 19.5 / 45.7 ± 22.2	75.5 ± 13.8 / 36.2 ± 14.0	8.2 ± 2.1 / 4.5 ± 2.8	1.5 ± 2.2 / 5.0 ± 2.8	129.8 ± 30.3 / 43.7 ± 35.9	148.0 ± 28.3 / 48.3 ± 28.7	48.1 ± 18.4 / 35.4 ± 21.2	1.1%/1.1%
5+ years	81.5 ± 19.0 / 46.1 ± 20.9	71.6 ± 14.5 / 31.3 ± 16.7	7.8 ± 2.3 / 4.1 ± 3.2	1.7 ± 2.2 / 5.0 ± 2.6	122.1 ± 32.3 / 34.8 ± 38.9	147.9 ± 28.9 / 43.6 ± 34.5	44.0 ± 20.7 / 35.4 ± 22.8	8.6%/5.9%
Primary aTSA Cohort, Male, age <60yrs at the time of surgery								2.6%/1.8%

Suppl Table 6. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 60 to <70 yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	40.4 ± 16.8	42.4 ± 15.0	4.6 ± 2.0	6.0 ± 2.2	86.2 ± 30.1	101.2 ± 29.1	19.8 ± 18.8	NA
3 – 6 months	81.9 ± 15.7 / 41.8 ± 19.5	68.1 ± 13.4 / 25.6 ± 16.0	7.6 ± 2.1 / 3.0 ± 2.7	1.3 ± 1.6 / 4.9 ± 2.6	117.5 ± 29.1 / 28.6 ± 35.5	132.7 ± 28.5 / 32.1 ± 32.4	44.7 ± 16.8 / 23.9 ± 18.2	1.4%/0.7%
6 – 9 months	87.9 ± 14.1 / 43.9 ± 19.9	72.4 ± 13.0 / 27.6 ± 14.7	8.1 ± 2.1 / 3.4 ± 2.6	1.0 ± 1.6 / 4.7 ± 2.8	124.7 ± 31.6 / 37.4 ± 34.7	140.8 ± 28.5 / 38.9 ± 34.7	51.2 ± 16.6 / 30.4 ± 19.7	0.7%/0%
1 year	89.8 ± 15.4 / 49.2 ± 19.5	75.9 ± 13.0 / 34.1 ± 15.2	8.6 ± 1.9 / 4.1 ± 2.3	0.9 ± 1.7 / 5.2 ± 2.4	132.1 ± 30.8 / 45.5 ± 35.9	147.9 ± 26.4 / 46.4 ± 32.4	51.7 ± 17.9 / 31.4 ± 19.8	1.4%/1.4%
2 – 3 years	91.3 ± 14.0 / 50.3 ± 18.5	77.7 ± 12.6 / 36.1 ± 13.4	8.8 ± 1.8 / 4.5 ± 2.3	0.8 ± 1.6 / 5.1 ± 2.4	132.1 ± 29.5 / 45.4 ± 35.6	149.3 ± 27.5 / 48.6 ± 34.0	51.9 ± 18.4 / 33.1 ± 20.1	1.8%/0.7%
3 – 5 years	89.6 ± 15.5 / 47.9 ± 20.0	76.8 ± 11.4 / 35.2 ± 11.3	8.9 ± 1.7 / 4.2 ± 2.5	1.0 ± 1.8 / 4.7 ± 2.6	133.8 ± 28.0 / 42.9 ± 32.4	152.3 ± 21.5 / 48.8 ± 30.5	51.3 ± 18.2 / 30.7 ± 19.3	0.7%/0.3%
5+ years	85.9 ± 17.6 / 46.7 ± 23.6	73.2 ± 13.6 / 33.8 ± 17.4	8.5 ± 1.9 / 4.2 ± 2.4	1.1 ± 1.9 / 4.9 ± 2.8	124.8 ± 31.5 / 37.7 ± 36.7	144.9 ± 18.4 / 46.1 ± 36.1	48.3 ± 18.4 / 29.3 ± 20.9	6.5%/4.0%
Primary aTSA Cohort, Male, age 60 to <70yrs at the time of surgery								2.2%/1.3%

Suppl Table 7. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 70 to <80 yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	41.2 ± 17.2	42.4 ± 15.7	4.4 ± 2.1	5.8 ± 2.3	89.6 ± 33.1	100.1 ± 33.6	22.1 ± 19.4	NA
3 – 6 months	82.0 ± 14.6 / 39.3 ± 18.5	66.1 ± 13.5 / 22.6 ± 18.2	7.3 ± 2.2 / 2.9 ± 3.0	1.2 ± 1.5 / 4.6 ± 2.3	115.6 ± 32.0 / 24.7 ± 39.0	128.5 ± 29.1 / 27.9 ± 38.7	45.7 ± 17.5 / 22.9 ± 20.8	0.6%/0.6%
6 – 9 months	86.6 ± 13.6 / 41.2 ± 19.6	71.9 ± 11.8 / 29.1 ± 14.3	7.8 ± 2.2 / 3.4 ± 3.0	1.1 ± 1.7 / 4.3 ± 2.8	123.2 ± 25.5 / 37.2 ± 29.6	135.6 ± 24.7 / 38.7 ± 29.5	49.2 ± 16.9 / 25.3 ± 21.4	1.3%/0%
1 year	88.9 ± 14.7 / 46.0 ± 20.3	74.4 ± 12.9 / 30.5 ± 17.1	8.6 ± 1.8 / 4.3 ± 2.5	1.0 ± 1.8 / 4.7 ± 2.9	132.8 ± 29.8 / 42.9 ± 38.4	146.4 ± 24.7 / 45.6 ± 36.6	53.4 ± 17.6 / 31.2 ± 20.9	2.4%/0.9%
2 – 3 years	89.7 ± 14.7 / 47.6 ± 21.4	75.7 ± 12.0 / 30.6 ± 18.8	8.6 ± 2.0 / 4.3 ± 2.8	0.9 ± 1.7 / 4.7 ± 2.9	133.5 ± 30.0 / 40.1 ± 38.3	146.7 ± 24.7 / 44.9 ± 39.6	51.7 ± 16.9 / 30.3 ± 21.9	4.1%/3.1%
3 – 5 years	89.9 ± 14.3 / 52.2 ± 17.9	74.3 ± 13.9 / 32.4 ± 17.4	9.0 ± 1.7 / 4.8 ± 2.3	0.7 ± 1.4 / 5.4 ± 2.3	130.7 ± 30.9 / 43.6 ± 41.4	145.5 ± 27.7 / 45.7 ± 41.9	50.1 ± 17.2 / 30.4 ± 20.3	2.5%/1.9%
5+ years	82.5 ± 20.1 / 43.2 ± 23.8	69.0 ± 14.0 / 28.6 ± 17.8	8.5 ± 2.1 / 4.0 ± 2.7	1.6 ± 2.3 / 4.6 ± 3.0	124.6 ± 28.7 / 33.5 ± 35.1	141.4 ± 27.3 / 38.7 ± 37.7	44.8 ± 18.1 / 25.9 ± 19.3	4.5%/1.9%
				Primary aTSA Cohort, Male, age 70 to <80yrs at the time of surgery				2.7%/1.6%

Suppl Table 8. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary aTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age ≥80yrs at the time of surgery

aTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	30.5 ± 20.0	37.6 ± 17.2	3.7 ± 2.3	7.0 ± 2.3	73.4 ± 28.1	88.0 ± 34.7	26.0 ± 16.7	NA
3 – 6 months	87.7 ± 8.9 / 56.3 ± 13.6	63.2 ± 13.4 / 34.2 ± 15.1	7.8 ± 1.9 / 4.1 ± 3.0	0.9 ± 1.2 / 6.1 ± 1.5	113.2 ± 27.0 / 42.7 ± 27.4	128.0 ± 26.6 / 37.3 ± 25.9	46.4 ± 22.5 / 17.7 ± 20.4	0%/0%
6 – 9 months	92.4 ± 9.7 / 47.3 ± 25.6	74.6 ± 14.8 / 34.1 ± 12.4	7.3 ± 3.9 / 3.1 ± 3.9	0.3 ± 0.7 / 6.1 ± 3.4	128.2 ± 32.6 / 51.6 ± 38.4	134.9 ± 34.2 / 36.9 ± 43.4	59.4 ± 22.6 / 34.2 ± 18.8	0%/0%
1 year	91.3 ± 7.2 / 55.5 ± 21.9	71.3 ± 12.1 / 35.4 ± 15.6	8.7 ± 1.2 / 5.0 ± 2.1	0.3 ± 0.6 / 7.0 ± 2.7	124.4 ± 31.4 / 54.9 ± 35.3	133.1 ± 30.4 / 53.5 ± 37.3	51.6 ± 21.7 / 29.7 ± 16.8	0%/0%
2 – 3 years	91.8 ± 11.4 / 65.4 ± 10.2	73.8 ± 11.0 / 39.1 ± 17.6	8.4 ± 2.6 / 5.6 ± 2.0	0.6 ± 1.5 / 6.5 ± 2.8	136.0 ± 32.8 / 71.3 ± 39.4	146.9 ± 25.2 / 54.7 ± 39.2	52.3 ± 21.8 / 30.6 ± 18.2	0%/0%

3 – 5 years	89.0 ± 13.3 / 51.6 ± 22.5	71.2 ± 11.2 / 24.0 ± 20.7	8.8 ± 1.6 / 5.9 ± 2.2	0.8 ± 1.7 / 6.3 ± 3.2	120.6 ± 23.6 / 54.0 ± 53.2	132.7 ± 26.8 / 43.0 ± 55.6	56.8 ± 16.6 / 34.2 ± 25.9	0%/0%
5+ years*	83.6 ± 18.0	61.0 ± 12.1	8.7 ± 2.3	0.7 ± 1.6	112.5 ± 21.0	120.0 ± 20.0	55.0 ± 23.5	0%/0%
Primary aTSA Cohort, Male, age ≥80yrs at the time of surgery								0%/0%

*Note: Insufficient follow-up data to calculate pre-to-post improvement beyond 5yrs follow-up for this primary aTSA patient cohort

Suppl Table 9. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age <60yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	27.4 ± 15.3	28.3 ± 12.6	2.8 ± 2.3	7.2 ± 2.1	70.1 ± 34.7	77.1 ± 33.9	20.7 ± 23.2	NA
3 – 6 months	56.7 ± 21.6 / 28.9 ± 21.9	45.2 ± 15.8 / 16.0 ± 17.5	5.6 ± 2.3 / 2.3 ± 3.0	3.5 ± 2.6 / 3.8 ± 2.7	100.7 ± 36.9 / 27.1 ± 49.4	111.0 ± 35.7 / 32.9 ± 46.4	28.1 ± 19.6 / 7.1 ± 25.0	2.8%/1.4%
6 – 9 months	62.5 ± 23.1 / 33.2 ± 16.3	52.7 ± 16.6 / 22.3 ± 15.0	6.6 ± 2.3 / 3.7 ± 2.8	3.0 ± 2.7 / 4.1 ± 2.7	108.4 ± 30.8 / 41.0 ± 40.8	124.2 ± 32.8 / 54.0 ± 37.2	27.0 ± 20.7 / 11.5 ± 24.9	4.7%/2.3%
1 year	68.3 ± 23.4 / 38.8 ± 21.0	57.1 ± 18.2 / 28.1 ± 21.0	7.0 ± 2.3 / 4.2 ± 2.9	2.8 ± 2.7 / 4.3 ± 2.8	122.5 ± 33.3 / 48.6 ± 50.1	128.5 ± 32.6 / 50.6 ± 44.4	34.6 ± 22.0 / 8.6 ± 24.7	0%/0%
2 – 3 years	68.4 ± 25.9 / 38.6 ± 23.3	58.7 ± 19.0 / 28.2 ± 18.3	7.0 ± 2.4 / 4.4 ± 3.3	2.5 ± 2.8 / 4.7 ± 2.9	121.5 ± 40.9 / 46.6 ± 55.9	132.8 ± 35.5 / 52.3 ± 48.6	35.7 ± 21.7 / 16.9 ± 24.5	5.5%/2.7%
3 – 5 years	67.8 ± 23.5 / 42.0 ± 25.7	58.1 ± 18.4 / 28.7 ± 15.8	7.0 ± 2.1 / 4.9 ± 3.0	2.6 ± 2.7 / 4.5 ± 3.1	118.6 ± 34.5 / 46.9 ± 42.1	137.6 ± 32.1 / 57.6 ± 35.3	34.3 ± 23.4 / 16.8 ± 23.2	2.8%/2.8%
5+ years	70.3 ± 21.7 / 40.7 ± 23.1	56.3 ± 17.8 / 30.2 ± 14.2	6.5 ± 2.3 / 3.8 ± 3.1	2.0 ± 2.5 / 5.3 ± 2.3	104.8 ± 36.1 / 39.1 ± 38.0	118.8 ± 42.1 / 50.0 ± 37.2	21.7 ± 31.0 / 17.2 ± 18.5	4.8%/2.4%
Primary rTSA Cohort, Female, age <60yrs at the time of surgery								3.2%/1.8%

Suppl Table 10. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 60 to <70 yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	32.5 ± 15.6	34.5 ± 14.0	3.9 ± 2.2	6.7 ± 2.2	75.0 ± 35.5	89.0 ± 38.5	20.1 ± 20.7	NA

3 – 6 months	70.8 ± 19.5 / 38.5 ± 21.1	56.8 ± 15.7 / 23.4 ± 17.5	6.7 ± 2.1 / 3.0 ± 2.9	2.0 ± 2.2 / 4.7 ± 2.8	105.5 ± 32.4 / 31.9 ± 40.1	120.8 ± 33.6 / 33.9 ± 42.2	29.2 ± 17.5 / 8.3 ± 20.6	4.6%/0.6%
6 – 9 months	75.5 ± 18.1 / 41.2 ± 18.2	60.7 ± 13.6 / 26.9 ± 15.9	7.3 ± 1.8 / 3.2 ± 2.7	1.7 ± 2.0 / 4.6 ± 2.3	111.3 ± 28.5 / 38.8 ± 37.3	130.1 ± 29.5 / 46.3 ± 38.6	35.5 ± 17.3 / 15.5 ± 20.9	1.9%/1.3%
1 year	80.1 ± 19.0 / 48.1 ± 20.6	67.0 ± 14.2 / 32.7 ± 15.7	7.8 ± 2.0 / 4.1 ± 2.7	1.4 ± 2.1 / 5.3 ± 2.8	125.0 ± 32.2 / 47.5 ± 39.6	141.1 ± 27.9 / 49.1 ± 39.6	38.9 ± 18.5 / 17.9 ± 21.5	1.6%/1.0%
2 – 3 years	81.7 ± 18.0 / 49.5 ± 19.6	67.9 ± 13.8 / 35.2 ± 15.4	8.1 ± 1.8 / 4.2 ± 2.5	1.3 ± 2.1 / 5.4 ± 2.6	120.4 ± 31.6 / 46.5 ± 37.4	140.8 ± 28.0 / 53.3 ± 39.8	37.9 ± 18.4 / 18.3 ± 22.4	3.4%/1.4%
3 – 5 years	81.3 ± 19.7 / 49.1 ± 20.5	65.9 ± 15.9 / 31.5 ± 16.4	8.1 ± 2.0 / 4.4 ± 2.3	1.3 ± 2.1 / 5.5 ± 2.7	118.7 ± 31.4 / 43.3 ± 38.1	137.8 ± 29.8 / 47.6 ± 41.8	35.5 ± 17.4 / 15.8 ± 22.9	2.2%/0.9%
5+ years	77.1 ± 22.1 / 46.7 ± 21.0	65.9 ± 16.0 / 34.2 ± 14.4	7.5 ± 2.2 / 4.2 ± 2.8	1.8 ± 2.5 / 5.3 ± 3.0	111.7 ± 31.6 / 36.6 ± 35.1	134.5 ± 31.8 / 42.8 ± 38.3	33.0 ± 18.8 / 11.1 ± 21.2	2.8%/0%
Primary rTSA Cohort, Female, age 60 to <70yrs at the time of surgery								2.9%/0.9%

Suppl Table 11. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age 70 to <80 yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	32.8 ± 15.2	33.9 ± 14.5	3.5 ± 2.1	6.5 ± 2.2	70.9 ± 37.5	85.5 ± 40.4	17.7 ± 20.9	NA
3 – 6 months	73.7 ± 18.0 / 41.5 ± 20.4	57.9 ± 14.8 / 24.4 ± 16.2	7.0 ± 2.1 / 3.4 ± 2.8	1.6 ± 2.0 / 5.0 ± 2.7	103.6 ± 32.6 / 35.3 ± 41.0	120.9 ± 33.0 / 39.2 ± 42.9	27.9 ± 17.1 / 10.5 ± 21.4	2.8%/0.7%
6 – 9 months	78.8 ± 16.7 / 44.7 ± 20.4	62.3 ± 13.1 / 28.8 ± 16.8	7.2 ± 2.0 / 3.9 ± 2.6	1.3 ± 1.9 / 4.9 ± 2.7	108.5 ± 31.2 / 40.0 ± 36.6	129.1 ± 31.2 / 46.2 ± 42.4	33.1 ± 16.8 / 15.2 ± 21.8	2.3%/0.3%
1 year	81.5 ± 16.9 / 48.6 ± 19.7	65.9 ± 13.2 / 31.4 ± 15.8	8.0 ± 1.9 / 4.4 ± 2.5	1.2 ± 1.9 / 5.3 ± 2.7	119.3 ± 29.5 / 47.6 ± 39.9	137.4 ± 28.0 / 51.7 ± 41.5	35.5 ± 17.4 / 17.4 ± 22.1	2.4%/0.7%
2 – 3 years	82.9 ± 17.2 / 48.8 ± 20.1	68.1 ± 12.7 / 34.7 ± 15.6	8.1 ± 1.9 / 4.6 ± 2.6	1.0 ± 1.9 / 5.2 ± 2.6	118.3 ± 29.3 / 49.3 ± 37.0	140.7 ± 25.5 / 58.5 ± 41.3	37.7 ± 17.3 / 20.6 ± 23.4	2.5%/0.7%
3 – 5 years	81.8 ± 17.4 / 47.7 ± 21.8	67.3 ± 12.1 / 32.2 ± 15.1	8.3 ± 1.7 / 4.6 ± 2.7	1.1 ± 1.9 / 5.3 ± 2.8	116.3 ± 27.3 / 48.3 ± 38.3	136.9 ± 26.0 / 54.9 ± 41.9	36.5 ± 17.8 / 18.0 ± 23.3	1.9%/0.4%
5+ years	78.4 ± 21.2 / 42.6 ± 25.4	63.5 ± 14.3 / 27.4 ± 18.0	7.9 ± 2.3 / 4.0 ± 3.1	1.4 ± 2.4 / 4.8 ± 3.1	110.6 ± 28.1 / 38.5 ± 40.3	129.3 ± 28.2 / 43.0 ± 43.3	31.4 ± 19.6 / 12.2 ± 26.6	0.9%/0%
Primary rTSA Cohort, Female, age 70 to <80yrs at the time of surgery								2.3%/0.5%

Suppl Table 12. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Female, Age ≥80yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	31.4 ± 14.2	30.1 ± 12.6	3.1 ± 2.1	6.5 ± 2.1	60.2 ± 30.8	71.6 ± 35.1	13.8 ± 19.8	NA
3 – 6 months	71.5 ± 18.7 / 40.8 ± 19.5	55.1 ± 13.4 / 26.4 ± 15.8	6.8 ± 2.2 / 3.6 ± 2.8	1.7 ± 2.2 / 4.8 ± 2.7	96.2 ± 27.7 / 38.5 ± 35.5	114.1 ± 31.9 / 45.5 ± 39.3	27.1 ± 18.8 / 12.6 ± 22.6	2.7%/0.8%
6 – 9 months	74.3 ± 18.2 / 40.0 ± 18.0	59.4 ± 12.1 / 29.8 ± 13.5	7.4 ± 2.0 / 4.3 ± 2.5	1.6 ± 2.3 / 4.4 ± 2.7	102.9 ± 29.2 / 43.5 ± 33.2	122.3 ± 28.9 / 51.0 ± 35.8	27.4 ± 17.1 / 13.3 ± 23.7	3.3%/0%
1 year	78.8 ± 16.3 / 48.0 ± 20.3	62.3 ± 12.4 / 33.8 ± 15.7	7.8 ± 2.0 / 4.5 ± 2.9	1.3 ± 1.9 / 5.4 ± 2.7	109.5 ± 29.2 / 51.6 ± 37.7	130.0 ± 29.0 / 60.9 ± 39.4	32.2 ± 18.5 / 18.2 ± 22.5	1.6%/0.3%
2 – 3 years	79.2 ± 17.9 / 47.3 ± 20.9	62.3 ± 14.2 / 34.2 ± 16.8	7.6 ± 2.3 / 4.5 ± 2.9	1.1 ± 2.0 / 5.4 ± 2.6	105.9 ± 30.7 / 49.3 ± 38.4	127.7 ± 30.9 / 59.7 ± 43.8	32.8 ± 17.1 / 19.7 ± 22.9	5.0%/0.9%
3 – 5 years	78.6 ± 18.9 / 48.0 ± 22.1	61.6 ± 13.1 / 35.1 ± 14.6	7.8 ± 2.2 / 4.4 ± 3.1	1.1 ± 1.9 / 5.6 ± 2.7	106.0 ± 27.4 / 49.6 ± 37.8	127.3 ± 28.6 / 62.7 ± 41.4	34.9 ± 18.2 / 17.0 ± 25.8	2.2%/0.5%
5+ years	80.7 ± 19.1 / 45.4 ± 26.6	61.5 ± 14.9 / 35.2 ± 15.9	7.8 ± 2.4 / 3.9 ± 3.4	1.1 ± 2.1 / 4.9 ± 3.3	105.2 ± 27.9 / 44.2 ± 35.0	120.1 ± 35.1 / 49.6 ± 44.6	34.2 ± 17.0 / 21.1 ± 25.8	4.4%/2.2%
Primary rTSA Cohort, Female, age ≥80yrs at the time of surgery								3.0%/0.7%

Suppl Table 13. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age <60yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	34.2 ± 15.5	35.7 ± 15.2	3.5 ± 2.2	6.1 ± 2.1	79.7 ± 40.6	86.8 ± 38.9	19.1 ± 23.8	NA
3 – 6 months	63.8 ± 22.0 / 29.3 ± 22.6	55.6 ± 17.7 / 20.3 ± 19.1	6.4 ± 2.2 / 2.8 ± 3.2	2.7 ± 2.5 / 3.2 ± 2.8	117.1 ± 31.2 / 37.1 ± 36.9	124.7 ± 32.2 / 36.5 ± 42.2	26.5 ± 16.5 / 7.5 ± 22.5	3.0%/0%
6 – 9 months	68.4 ± 20.8 / 36.3 ± 24.5	57.5 ± 17.9 / 27.7 ± 20.9	6.3 ± 2.3 / 2.7 ± 2.7	2.8 ± 2.7 / 3.7 ± 3.1	113.2 ± 32.2 / 43.2 ± 41.3	128.5 ± 33.8 / 50.5 ± 40.6	23.6 ± 16.3 / 6.7 ± 26.2	0%/0%
1 year	74.1 ± 25.4 / 39.7 ± 29.1	67.6 ± 19.4 / 33.9 ± 20.5	7.6 ± 2.1 / 4.2 ± 2.8	2.3 ± 2.8 / 3.7 ± 3.4	128.2 ± 32.2 / 46.3 ± 36.4	139.3 ± 29.7 / 50.3 ± 39.2	34.9 ± 18.6 / 17.0 ± 22.9	4.8%/1.6%

2 – 3 years	78.2 ± 20.0 / 40.2 ± 20.8	71.0 ± 16.4 / 37.3 ± 17.8	7.9 ± 1.7 / 4.3 ± 2.3	1.8 ± 2.1 / 4.0 ± 2.3	132.7 ± 33.6 / 47.5 ± 44.5	147.5 ± 25.2 / 52.6 ± 44.2	34.5 ± 20.3 / 18.1 ± 24.4	3.6%/3.6%
3 – 5 years	73.3 ± 23.4 / 29.6 ± 24.0	67.6 ± 20.5 / 33.3 ± 19.4	7.4 ± 2.3 / 4.3 ± 2.3	2.0 ± 2.5 / 3.4 ± 2.5	123.7 ± 36.0 / 45.5 ± 38.6	141.2 ± 33.0 / 50.9 ± 38.2	32.7 ± 20.4 / 27.8 ± 26.8	2.9%/2.9%
5+ years	71.0 ± 23.9 / 28.6 ± 29.1	67.2 ± 17.7 / 28.4 ± 20.1	7.1 ± 2.8 / 3.6 ± 2.7	2.3 ± 2.4 / 3.3 ± 3.1	119.1 ± 37.4 / 26.4 ± 32.0	135.4 ± 37.0 / 33.4 ± 37.4	29.4 ± 18.0 / 10.9 ± 26.5	3.9%/3.9%
Primary rTSA Cohort, Male, age <60yrs at the time of surgery								3.3%/1.8%

Suppl Table 14. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 60 to <70 yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	37.8 ± 15.7	38.0 ± 14.7	3.9 ± 1.9	6.0 ± 2.2	81.5 ± 38.6	92.0 ± 38.6	16.9 ± 21.1	NA
3 – 6 months	75.9 ± 17.2 / 37.4 ± 19.2	61.6 ± 13.9 / 23.7 ± 15.2	7.0 ± 2.0 / 2.9 ± 2.6	1.7 ± 1.9 / 4.3 ± 2.6	112.0 ± 31.2 / 30.3 ± 37.9	126.2 ± 29.6 / 35.1 ± 40.6	29.2 ± 17.5 / 13.0 ± 18.9	1.3%/0.4%
6 – 9 months	79.7 ± 16.7 / 41.1 ± 20.0	66.7 ± 12.4 / 29.8 ± 16.1	7.5 ± 1.7 / 3.6 ± 2.5	1.5 ± 1.8 / 4.3 ± 2.7	115.4 ± 27.0 / 35.1 ± 35.5	137.1 ± 23.4 / 45.1 ± 39.5	29.9 ± 17.7 / 16.2 ± 21.7	3.0%/3.0%
1 year	81.6 ± 18.4 / 43.3 ± 20.3	71.2 ± 14.0 / 32.9 ± 17.5	7.9 ± 2.0 / 4.1 ± 2.5	1.4 ± 1.9 / 4.5 ± 2.6	127.2 ± 29.4 / 44.9 ± 39.4	143.2 ± 25.6 / 50.5 ± 41.7	35.4 ± 17.0 / 19.0 ± 22.1	3.9%/2.8%
2 – 3 years	86.8 ± 15.9 / 48.0 ± 19.2	75.3 ± 11.9 / 36.0 ± 14.8	8.4 ± 1.6 / 4.6 ± 2.3	0.9 ± 1.7 / 5.0 ± 2.5	126.2 ± 28.1 / 44.8 ± 39.8	146.2 ± 19.5 / 54.7 ± 40.4	35.5 ± 16.2 / 20.0 ± 21.8	3.5%/1.3%
3 – 5 years	85.6 ± 18.0 / 45.5 ± 20.6	73.2 ± 14.0 / 32.5 ± 15.2	8.3 ± 1.9 / 3.9 ± 2.8	1.1 ± 2.1 / 4.8 ± 2.6	124.6 ± 28.8 / 38.7 ± 40.0	144.0 ± 23.7 / 45.9 ± 44.7	36.7 ± 17.5 / 18.4 ± 22.6	4.2%/2.4%
5+ years	83.9 ± 18.7 / 48.6 ± 20.4	69.5 ± 15.1 / 28.3 ± 19.6	8.2 ± 1.8 / 4.1 ± 2.4	1.2 ± 1.9 / 5.3 ± 2.7	113.8 ± 30.5 / 28.9 ± 43.2	133.6 ± 28.8 / 34.2 ± 49.2	28.9 ± 17.8 / 6.7 ± 25.4	3.5%/2.6%
Primary rTSA Cohort, Male, age 60 to <70yrs at the time of surgery								3.2%/1.9%

Suppl Table 15. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age 70 to <80 yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	41.0 ± 16.3	39.5 ± 14.4	4.0 ± 2.1	5.6 ± 2.3	79.2 ± 38.3	92.9 ± 40.9	21.3 ± 23.2	NA
3 – 6 months	79.2 ± 16.7 / 38.8 ± 20.0	62.8 ± 13.8 / 23.5 ± 17.8	7.3 ± 1.9 / 3.2 ± 2.7	1.3 ± 1.8 / 4.4 ± 2.6	107.4 ± 28.8 / 29.9 ± 39.8	123.9 ± 29.1 / 33.5 ± 43.5	29.4 ± 17.5 / 8.0 ± 22.2	3.8%/0.8%
6 – 9 months	81.7 ± 18.1 / 37.9 ± 20.6	66.9 ± 13.5 / 26.6 ± 17.7	7.6 ± 2.0 / 3.5 ± 2.9	1.2 ± 1.9 / 4.2 ± 2.6	114.7 ± 28.6 / 35.9 ± 42.2	134.3 ± 26.8 / 41.1 ± 43.9	31.3 ± 18.8 / 9.6 ± 24.9	3.4%/1.5%
1 year	85.6 ± 17.2 / 44.4 ± 21.0	71.8 ± 12.9 / 32.2 ± 16.8	8.2 ± 1.8 / 4.1 ± 2.6	1.0 ± 1.9 / 4.6 ± 2.7	122.8 ± 27.8 / 43.3 ± 40.1	140.5 ± 24.4 / 47.1 ± 41.9	36.3 ± 18.0 / 15.3 ± 22.4	2.7%/1.2%
2 – 3 years	86.2 ± 16.5 / 43.9 ± 22.0	73.2 ± 12.3 / 32.9 ± 17.2	8.3 ± 1.8 / 4.0 ± 2.6	1.0 ± 1.9 / 4.6 ± 2.7	121.8 ± 27.6 / 42.0 ± 39.7	139.3 ± 25.4 / 45.2 ± 42.2	38.0 ± 16.9 / 15.7 ± 24.4	1.5%/1.2%
3 – 5 years	87.4 ± 17.2 / 43.5 ± 20.7	72.5 ± 11.5 / 33.3 ± 16.4	8.5 ± 1.8 / 4.4 ± 2.6	0.9 ± 1.9 / 4.4 ± 2.6	121.5 ± 27.4 / 43.2 ± 40.8	139.1 ± 23.7 / 49.7 ± 41.3	37.4 ± 16.4 / 16.1 ± 22.6	3.7%/2.2%
5+ years	85.8 ± 15.4 / 43.8 ± 20.0	71.5 ± 10.2 / 31.4 ± 17.5	8.5 ± 1.6 / 4.2 ± 2.5	1.1 ± 1.8 / 4.7 ± 2.4	118.8 ± 24.6 / 31.3 ± 42.5	135.1 ± 19.8 / 36.6 ± 42.2	36.5 ± 18.6 / 9.7 ± 26.5	2.6%/1.9%
Primary rTSA Cohort, Male, age 70 to <80yrs at the time of surgery								2.9%/1.4%

Suppl Table 16. Comparison of Pre-operative, Post-operative, and Improvement in Outcomes for the Primary rTSA Patients Analyzed in this Study, Stratified by Follow-up Duration, Gender, and Age: Male, Age ≥80yrs at the time of surgery

rTSA Follow-up Duration	ASES (Post/Improve)	Constant (Post/Improve)	Global Shoulder Function (Post/Improve)	VAS Pain (Post/Improve)	Active Abduction (Post/Improve)	Active Forward Elevation (Post/Improve)	Active External Rotation (Post/Improve)	Adverse Event % / Revision %
Preoperative	39.9 ± 16.6	36.7 ± 13.6	4.1 ± 2.2	5.7 ± 2.3	68.5 ± 32.3	81.3 ± 35.2	21.8 ± 19.1	NA
3 – 6 months	77.1 ± 17.9 / 35.1 ± 19.9	59.3 ± 13.6 / 19.8 ± 16.8	7.1 ± 2.0 / 2.8 ± 2.6	1.7 ± 2.3 / 4.0 ± 2.8	103.9 ± 26.7 / 28.0 ± 33.8	120.4 ± 28.1 / 32.7 ± 37.1	30.2 ± 17.2 / 6.8 ± 18.8	3.0%/0.8%
6 – 9 months	85.2 ± 13.4 / 41.6 ± 17.4	65.1 ± 12.0 / 31.0 ± 13.5	8.0 ± 1.7 / 3.6 ± 2.7	0.9 ± 1.6 / 4.3 ± 2.5	109.0 ± 27.3 / 46.8 ± 33.0	129.1 ± 23.0 / 55.9 ± 32.5	33.3 ± 17.5 / 11.2 ± 19.6	0%/0%
1 year	83.9 ± 16.1 / 44.0 ± 19.9	66.4 ± 13.0 / 28.5 ± 13.4	8.0 ± 2.1 / 4.2 ± 2.5	1.1 ± 1.8 / 4.7 ± 2.6	117.0 ± 28.3 / 47.3 ± 37.6	134.6 ± 26.0 / 52.6 ± 41.6	34.5 ± 17.6 / 12.3 ± 21.8	1.6%/0.8%
2 – 3 years	83.2 ± 17.9 / 40.7 ± 20.0	68.4 ± 12.2 / 28.3 ± 14.3	8.1 ± 1.9 / 3.7 ± 2.5	1.2 ± 2.2 / 4.2 ± 2.9	115.5 ± 30.2 / 39.6 ± 36.1	135.0 ± 25.2 / 47.9 ± 37.7	37.2 ± 19.3 / 15.0 ± 19.6	3.7%/0%

3 – 5 years	87.9 ± 12.6 / 43.4 ± 16.5	72.2 ± 9.7 / 30.9 ± 14.2	8.8 ± 1.5 / 4.6 ± 2.3	0.7 ± 1.4 / 4.7 ± 2.3	121.8 ± 27.1 / 43.6 ± 37.5	136.5 ± 23.5 / 49.1 ± 34.9	41.1 ± 20.2 / 15.1 ± 23.6	3.6%/1.8%
5+ years	82.5 ± 26.2 / 27.6 ± 20.9	67.9 ± 19.2 / 27.8 ± 12.4	8.5 ± 2.3 / 3.8 ± 2.7	1.3 ± 2.6 / 3.1 ± 3.0	121.1 ± 32.1 / 42.3 ± 34.5	137.5 ± 28.8 / 53.2 ± 37.7	43.0 ± 23.3 / 21.4 ± 16.3	4.5%/0%
Primary rTSA Cohort, Male, age ≥80yrs at the time of surgery								2.5%/0.6%

Suppl Table 17. Comparison of Mean Absolute Error Associated with the ASES Predictions with the Full and Abbreviated Machine Learning Models

ASES Predictions	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	14.5 (14.9, 14.3)	14.4 (14.8, 14.1)	14.5 (14.8, 13.9)	15.8 (15.9, 15.0)	16.1 (16.2, 16.0)	16.9 (16.9, 16.9)	15.3 (15.6, 14.9)
XGBoost with Full Model	12.2 (12.9, 11.8)	12.2 (12.9, 12.0)	11.8 (11.1, 11.9)	11.1 (10.3, 11.7)	11.1 (10.8, 11.3)	12.2 (11.9, 12.5)	11.7 (11.6, 11.8)
XGBoost with Abbreviated Model	12.4 (13.0, 12.0)	12.4 (13.0, 12.1)	12.3 (11.5, 12.5)	11.5 (10.8, 12.1)	11.3 (11.0, 11.5)	12.4 (12.2, 12.7)	12.0 (11.9, 12.2)
MAE Difference (Full – Abbreviated)	0.2 (0.1, 0.2)	0.2 (0.1, 0.1)	0.5 (0.4, 0.6)	0.4 (0.1, 0.2)	0.2 (0.2, 0.2)	0.2 (0.3, 0.2)	0.3 (0.3, 0.4)
XGBoost with Abbreviated Model + Implant Data	12.4 (12.9, 12.09)	12.4 (12.9, 12.0)	12.2 (11.4, 12.4)	11.4 (10.6, 11.9)	11.2 (10.9, 11.4)	12.4 (12.0, 12.6)	12.0 (11.7, 12.0)

Suppl Table 18. Comparison of Mean Absolute Error Associated with the Constant Predictions with the Full and Abbreviated Machine Learning Models

Constant Predictions	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	13.1 (13.8, 13.0)	12.6 (13.6, 12.5)	12.4 (12.4, 12.4)	12.3 (12.3, 12.3)	12.3 (12.1, 12.5)	12.3 (12.1, 12.4)	12.5 (12.7, 12.5)

XGBoost with Full Model	10.1 (10.8, 10.0)	9.3 (10.1, 9.2)	9.2 (9.2, 9.2)	8.3 (8.2, 8.7)	7.9 (7.6, 8.2)	8.5 (8.2, 8.9)	8.9 (9.4, 9.1)
XGBoost with Abbreviated Model	10.3 (10.9, 10.1)	10.0 (10.9, 10.0)	10.3 (10.3, 10.2)	9.4 (9.3, 9.9)	9.0 (8.7, 9.7)	9.4 (9.1, 9.6)	9.8 (10.1, 9.9)
MAE Difference (Full – Abbreviated)	0.2 (0.1, 0.1)	0.7 (0.8, 0.8)	0.9 (1.1, 1.0)	1.1 (1.1, 1.2)	1.1 (1.1, 1.5)	0.9 (1.1, 0.7)	0.9 (0.7, 0.8)
XGBoost with Abbreviated Model + Implant Data	10.2 (10.9, 10.0)	10.0 (10.8, 9.9)	10.1 (10.2, 10.0)	9.2 (9.2, 9.8)	9.0 (8.7, 9.6)	9.3 (9.1, 9.5)	9.7 (9.8, 9.8)

Suppl Table 19. Comparison of Mean Absolute Error Associated with the Global Shoulder Function Score Predictions with the Full and Abbreviated Machine Learning Models

Shoulder Function Prediction	**MAE (aTSA, rTSA) 3-6 months**	**MAE (aTSA, rTSA) 6-9 months**	**MAE (aTSA, rTSA) 1yr**	**MAE (aTSA, rTSA) 2-3yrs**	**MAE (aTSA, rTSA) 3-5yrs**	**MAE (aTSA, rTSA) 5+yrs**	**Weighted Average MAE (aTSA, rTSA)**
Baseline Average	1.6 (1.5, 1.7)	1.6 (1.6, 1.7)	1.5 (1.5, 1.4)	1.5 (1.5, 1.5)	1.4 (1.4, 1.5)	1.7 (1.8, 1.7)	1.6 (1.6, 1.6)
XGBoost with Full Model	1.4 (1.3, 1.5)	1.5 (1.4, 1.5)	1.3 (1.4, 1.2)	1.4 (1.4, 1.3)	1.2 (1.2, 1.2)	1.5 (1.5, 1.4)	1.4 (1.4, 1.3)
XGBoost with Abbreviated Model	1.5 (1.4, 1.6)	1.5 (1.4, 1.6)	1.5 (1.6, 1.3)	1.4 (1.5, 1.4)	1.4 (1.4, 1.4)	1.5 (1.6, 1.5)	1.5 (1.5, 1.4)
MAE Difference (Full – Abbreviated)	0.1 (0.1, 0.1)	0.0 (0.0, 0.1)	0.2 (0.2, 0.1)	0.0 (0.1, 0.1)	0.2 (0.2, 0.2)	0.0 (0.1, 0.1)	0.1 (0.1, 0.1)
XGBoost with Abbreviated Model + Implant Data	1.4 (1.4, 1.5)	1.5 (1.4, 1.6)	1.4 (1.5, 1.2)	1.4 (1.5, 1.4)	1.3 (1.3, 1.3)	1.5 (1.5, 1.5)	1.4 (1.4, 1.4)

Suppl Table 20. Comparison of Mean Absolute Error Associated with the VAS Pain Score Predictions with the Full and Abbreviated Machine Learning Models

VAS Pain Prediction	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	1.6 (1.4, 1.6)	1.8 (1.7, 1.8)	1.4 (1.4, 1.5)	1.4 (1.4, 1.5)	1.5 (1.4, 1.5)	1.9 (1.9, 1.8)	1.6 (1.5, 1.6)
XGBoost with Full Model	1.3 (1.4, 1.5)	1.3 (1.2, 1.4)	1.2 (1.1, 1.3)	1.2 (1.1, 1.3)	1.2 (1.2, 1.2)	1.4 (1.4, 1.4)	1.3 (1.2, 1.3)
XGBoost with Abbreviated Model	1.4 (1.4, 1.6)	1.3 (1.2, 1.4)	1.5 (1.3, 1.6)	1.3 (1.2, 1.5)	1.4 (1.4, 1.4)	1.5 (1.6, 1.5)	1.4 (1.2, 1.5)
MAE Difference (Full – Abbreviated)	0.1 (0.0, 0.1)	0.0 (0.0, 0.0)	0.3 (0.2, 0.3)	0.1 (0.1, 0.2)	0.2 (0.2, 0.2)	0.1 (0.2, 0.1)	0.1 (0.0, 0.2)
XGBoost with Abbreviated Model + Implant Data	1.4 (1.4, 1.6)	1.3 (1.1, 1.3)	1.4 (1.3, 1.6)	1.2 (1.2, 1.4)	1.3 (1.3, 1.3)	1.5 (1.5, 1.5)	1.3 (1.2, 1.4)

Suppl Table 21. Comparison of Mean Absolute Error Associated with the Active Abduction (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

Abduction (º) Prediction	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	26.0 (26.5, 25.9)	26.3 (26.5, 26.1)	27.0 (26.3, 27.3)	28.3 (28.0, 28.1)	24.6 (24.4, 25.0)	26.3 (27.0, 24.0)	26.5 (26.4, 26.3)
XGBoost with Full Model	21.4 (22.4, 20.9)	22.0 (21.9, 22.9)	21.7 (21.2, 21.9)	19.6 (20.0, 18.2)	18.5 (18.0, 19.0)	18.1 (19.2, 16.6)	20.4 (20.9, 20.1)
XGBoost with Abbreviated Model	23.1 (23.9, 22.5)	23.4 (23.4, 23.2)	23.0 (23.0, 23.3)	20.2 (20.5, 19.7)	20.2 (20.1, 20.3)	19.8 (20.3, 18.8)	21.8 (22.0, 21.3)
MAE Difference (Full – Abbreviated)	1.5 (1.5, 1.6)	1.4 (1.5, 0.3)	1.3 (1.8, 1.4)	0.8 (0.5, 1.5)	1.7 (2.1, 1.3)	1.7 (1.1, 2.2)	1.4 (1.1, 1.2)
XGBoost with Abbreviated Model + Implant Data	22.9 (23.8, 22.2)	23.2 (23.2, 23.2)	23.0 (22.9, 23.1)	20.0 (20.3, 19.6)	20.0 (19.9, 20.2)	19.8 (20.1, 18.8)	21.7 (21.9, 21.1)

Suppl Table 22. Comparison of Mean Absolute Error Associated with the Active Forward Elevation (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

Forward Elevation (º) Prediction	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	24.9 (24.9, 24.9)	24.3 (24.5, 24.0)	22.6 (22.7, 22.4)	20.9 (20.0, 22.0)	21.4 (21.2, 22.7)	23.7 (24.0, 23.2)	23.0 (22.9, 23.2)
XGBoost with Full Model	20.5 (20.5, 20.4)	20.3 (20.4, 20.0)	17.4 (17.9, 17.0)	15.4 (15.0, 17.5)	15.2 (15.2, 15.3)	15.7 (16.1, 15.0)	17.6 (18.0, 17.8)
XGBoost with Abbreviated Model	21.9 (21.9, 21.9)	21.9 (22.0, 21.4)	19.0 (19.1, 18.7)	17.8 (17.5, 18.9)	16.9 (16.9, 17.0)	16.2 (16.9, 16.1)	19.2 (19.7, 19.2)
MAE Difference (Full – Abbreviated)	1.4 (1.4, 1.5)	1.6 (1.6, 1.4)	1.6 (1.2, 1.7)	2.4 (2.5, 1.4)	1.7 (1.7, 1.7)	0.5 (0.8, 1.1)	1.6 (1.7, 1.4)
XGBoost with Abbreviated Model + Implant Data	21.8 (21.8, 21.8)	21.8 (21.9, 21.4)	18.9 (19.1, 18.6)	17.0 (16.9, 17.8)	16.9 (16.9, 16.9)	16.2 (16.9, 16.1)	19.0 (19.2, 19.1)

Suppl Table 23. Comparison of Mean Absolute Error Associated with the Active External Rotation (in degrees) Predictions with the Full and Abbreviated Machine Learning Models

External Rotation (º) Prediction	MAE (aTSA, rTSA) 3-6 months	MAE (aTSA, rTSA) 6-9 months	MAE (aTSA, rTSA) 1yr	MAE (aTSA, rTSA) 2-3yrs	MAE (aTSA, rTSA) 3-5yrs	MAE (aTSA, rTSA) 5+yrs	Weighted Average MAE (aTSA, rTSA)
Baseline Average	15.6 (15.5, 15.7)	15.9 (15.9, 16.0)	15.6 (15.6, 15.2)	15.5 (16.0, 14.3)	17.0 (17.2, 16.7)	17.0 (17.9, 16.8)	16.0 (16.2, 15.7)
XGBoost with Full Model	12.3 (12.2, 12.3)	12.9 (13.1, 12.6)	12.7 (13.9, 12.0)	12.0 (13.0, 11.6)	12.2 (12.9, 11.7)	10.3 (11.1, 9.2)	12.2 (13.1, 11.7)
XGBoost with Abbreviated Model	12.8 (12.8, 12.8)	13.0 (13.2, 13.0)	12.9 (13.9, 12.4)	12.2 (13.3, 12.1)	12.9 (13.0, 12.3)	11.9 (12.5, 11.0)	12.6 (13.2, 12.1)

MAE Difference (Full – Abbreviated)	0.5 (0.6, 0.5)	0.1 (0.1, 0.4)	0.2 (0.0, 0.4)	0.2 (0.3, 0.5)	0.7 (0.1, 0.6)	1.6 (1.4, 1.8)	0.4 (0.1, 0.4)
XGBoost with Abbreviated Model + Implant Data	12.3 (12.3, 12.2)	12.9 (13.2, 13.0)	12.8 (13.0, 12.2)	11.9 (13.1, 11.7)	12.8 (13.0, 12.2)	11.9 (12.5, 11.0)	12.4 (13.1, 12.0)

References

1. Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. IEEE Intelligent Informatics Bulletin. 2018;1:1-6.
2. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med. 2018;56:516-524. doi: 10.1515/cclm-2017-0287.
3. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. Front Bioeng Biotechnol. 2018;6:75. doi: 10.3389/fbioe.2018.00075.
4. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? Clin Orthop Relat Res. 2019;477:1267-1279. doi: 10.1097/CORR.0000000000000687.
5. Friedman RJ, Cheung EV, Flurin PH, Wright T, Simovitch RW, Bolch C, Roche CP, Zuckerman JD. Are age and patient gender associated with different rates and magnitudes of clinical improvement after reverse shoulder arthroplasty? Clin Orthop Relat Res. 2018;476:1264-1273. doi: 10.1007/s11999.0000000000000270.
6. Friedman RJ, Eichinger J, Schoch B, Wright T, Zuckerman J, Flurin PH, Bolch C, Roche C. Preoperative parameters that predict postoperative patient reported outcome measures and range of motion with anatomic and reverse total shoulder arthroplasty. JSES Open Access. 2019;3:266-272. doi: 10.1016/j.jses.2019.09.010.

7. Gonzalez Saenz de Tejada M, Escobar A, Bilbao A, Herrera-Espiñeira C, García-Perez L, Aizpuru F, Sarasqueta C. A prospective study of the association of patient expectations with changes in health-related quality of life outcomes, following total joint replacement. BMC Musculoskelet Disord. 2014;15:248. doi: 10.1186/1471-2474-15-248.

8. Gonzalez Sáenz de Tejada M, Escobar A, Herrera C, García L, Aizpuru F, Sarasqueta C. Patient expectations and health-related quality of life outcomes following total joint replacement. Value Health. 2010;13:447-54. doi: 10.1111/j.1524-4733.2009.00685.x.

9. Gortmaker SL, Hosmer DW, Lemeshow S. Applied logistic regression. Contemp Sociol. 1994;23:159.

10. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, Liu JN. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg. 2019;28:e410-e421. doi: 10.1016/j.jse.2019.05.017.

11. Hosmer Jr, DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, NJ: John Wiley & Sons; 2013:177.

12. Kumar, V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. What is the accuracy of Three Different Machine Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty. CORR. In press. 2020.

13. Levy JC, Everding NG, Gil CC Jr, Stephens S, Giveans MR. Speed of recovery after shoulder arthroplasty: a comparison of reverse and anatomic total shoulder arthroplasty. J Shoulder Elbow Surg. 2014 Dec;23(12):1872-1881. doi: 10.1016/j.jse.2014.04.014.

14. Lipton ZC. The mythos of model interpretability. Queue, 2018;16(3):31-57. https://queue.acm.org/detail.cfm?id=3241340 Accessed: December 19, 2019.

15. Mahomed NN, Liang MH, Cook EF, Daltroy LH, Fortin PR, Fossel AH, Katz JN. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. J Rheumatol. 2002;29:1273-1279.

16. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216-9. doi:10.1056/NEJMp1606181.

17. Palazzo C, Jourdan C, Descamps S, Nizard R, Hamadouche M, Anract P, Boisgard S, Galvin M, Ravaud P, Poiraudeau S. Determinants of satisfaction 1 year after total hip arthroplasty: the role of expectations fulfilment. BMC Musculoskelet Disord. 2014;15:53. doi: 10.1186/1471-2474-15-53.

18. Parsons M, Routman H, Roche C, Friedman R. Patient-reported outcomes of reverse total shoulder arthroplasty: a comparative risk factor analysis of improved versus unimproved cases. JSES Open Access. 2019;3:174-178. doi: 10.1016/j.jses.2019.07.004.

19. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. J Shoulder Elbow Surg. 2018;27:298-305. doi: 10.1016/j.jse.2017.09.013.

20. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the substantial clinical benefit. J Shoulder Elbow Surg. 2018;27:903-911. doi: 10.1016/j.jse.2017.12.014.

21. 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;99:1801-1811. doi: 10.2106/JBJS.16.01387.

22. Torlay L, Perrone-Bertolotti M, Thomas E, Baciu M. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain Inform. 2017;4:159-169.

23. Twiggs JG, Wakelin EA, Fritsch BA, Liu DW, Solomon MI, Parker DA, Klasan A, Miles BP. Clinical and statistical validation of a probabilistic prediction tool of total knee arthroplasty outcome. J Arthroplasty. 2019;34:2624-2631. doi: 10.1016/j.arth.2019.06.007.

24. Zheng Z, Yang Y, Niu X, Dai HN, Zhou Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Transactions on Industrial Informatics*, 2017;14:1606-1615.