Artificial Intelligence Can Define and Predict the "Optimal Observed Outcome" After Anterior Shoulder Instability Surgery: An Analysis of 200 Patients With 11-Year Mean Follow-Up

Sara E. Till, M.S., Yining Lu, M.D., Anna K. Reinholz, B.S., Alexander M. Boos, B.A., Aaron J. Krych, M.D., Kelechi R. Okoroha, M.D., and Christopher L. Camp, M.D.

Purpose: The purpose of this study was to use unsupervised machine learning clustering to define the "optimal observed outcome" after surgery for anterior shoulder instability (ASI) and to identify predictors for achieving it. **Methods:** Medical records, images, and operative reports were reviewed for patients <40 years old undergoing surgery for ASI. Four unsupervised machine learning clustering algorithms partitioned subjects into "optimal observed outcome" or "suboptimal outcome" based on combinations of actually observed outcomes. Demographic, clinical, and treatment variables were compared between groups using descriptive statistics and Kaplan-Meier survival curves. Variables were assessed for prognostic value through multivariate stepwise logistic regression. **Results:** Two hundred patients with a mean follow-up of 11 years were included. Of these, 146 (64%) obtained the "optimal observed outcome," characterized by decreased: postoperative pain (23% vs 52%; $P < 0.001$), recurrent instability (12% vs 41%; $P < 0.001$), revision surgery (10% vs 24%; $P = 0.015$), osteoarthritis (OA) (5% vs 19%; $P = 0.005$), and restricted motion (161° vs 168°; $P = 0.001$). Forty-one percent of patients had a "perfect outcome," defined as ideal performance across all outcomes. Time from initial instability to presentation (odds ratio [OR] = 0.96; 95% confidence interval [CI], 0.92-0.98; $P = 0.006$) and habitual/voluntary instability (OR = 0.17; 95% CI, 0.04-0.77; $P = 0.020$) were negative predictors of achieving the "optimal observed outcome." A predilection toward subluxations rather than dislocations before surgery (OR = 1.30; 95% CI, 1.02-1.65; $P = 0.030$) was a positive predictor. Type of surgery performed was not a significant predictor. **Conclusion:** After surgery for ASI, 64% of patients achieved the "optimal observed outcome" defined as minimal postoperative pain, no recurrent instability or OA, low revision surgery rates, and increased range of motion, of whom only 41% achieved a "perfect outcome." Positive predictors were shorter time to presentation and predilection toward preoperative subluxations over dislocations. **Level of Evidence:** Retrospective cohort, level IV.

Anterior shoulder instability (ASI) remains one of the most common causes of shoulder dysfunction in athletic and active patients.[1-4] ASI is traditionally considered to be an injury primarily impacting the young athlete, and literature demonstrates a bimodal distribution with increased rates in fall-risk populations.[4-7] Treatment options depend on a multitude of factors including the precipitating event, the number of instability episodes, and the presence of concomitant injuries. Operative management is generally advised for patients with a history of multiple instability events, substantial bone loss, or other risk factors for recurrence (i.e., young age, male sex, contact sports, etc.).[8,9]

Because most patients undergoing surgery for ASI tend to be athletes and/or active individuals, there is a strong desire to return these patients to a high level of function after surgery. However, it is unclear if having them return to this high level of function adversely impacts long term outcomes. Significant improvements

From the Department of Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota, U.S.A.

The authors report the following potential conflict of interest or source of funding: This study was partially funded by the following: National Institute of Arthritis and Musculoskeletal and Skin Diseases for the Musculoskeletal Research Training Program (T32AR56950). This study used the resources of the Rochester Epidemiology Project (REP) medical records-linkage system, which is supported by the National Institute on Aging (NIA; AG 058738), by the Mayo Clinic Research Committee, and by fees paid annually by REP users. The content of this article is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health (NIH) or the Mayo Clinic. Full ICMJE author disclosure forms are available for this article online, as supplementary material.

Received March 20, 2023; accepted June 14, 2023.

Address correspondence to Christopher L. Camp, M.D., Mayo Clinic, Department of Orthopedic Surgery, 200 First St. SW, Rochester, MN 55905, U.S.A. E-mail: Camp.christopher@mayo.edu

2666-061X/23424

https://doi.org/10.1016/j.asmr.2023.100773

in mechanical function and daily pain with operative treatment for chronic ASI for mid- and short-term outcomes have been reported.[10,11] Data on the long-term outcomes are focused primarily on joint preservation and rates of osteoarthritis.[12-14] Although it is clearly desirable that patients achieve a "perfect outcome" by obtaining every one of these favorable outcomes (no pain, normal stability, full motion, no progression to OA, return to high demand activity/sports, etc.), it is not clear if some of these may be mutually exclusive from one another in actual observed outcomes. To date, the predictors of optimized functional outcome from a global perspective, as opposed to domain-specific perspectives, remain elusive.

This incomplete understanding may be attributable to an absence of robust statistical tools to fully explore these questions. Traditional statistical approaches are subject to an inherent degree of bias given the need for pre-specification of either an outcome of interest or a risk factor (i.e., recurrent instability) or a stratifying risk factor (i.e., smoking status). The determination of these variables are the results of investigator input, and although most are based on rigorous scientific reasoning or clinical experience, some can, at times, be arbitrary. Furthermore, traditional parametric regression models are limited in their validity and ability to adapt to population changes over time because of the specific assumptions about the distribution of data they require. Conversely, unsupervised machine learning clustering is data driven and may be able to detect intrinsic structure in data without human input, allowing for identification of factors that influence outcomes based on the aggregation and evaluation of numerous features rather than a single variable or researcher-selected variables, as is the case in traditional statistical methods. Additionally, although machine learning is frequently praised for its ability to analyze large quantities of data, it is well suited for work with smaller numbers of patients with a large number of features and high dimensional data. Clustering, specifically, can work well within cohort sizes of several hundred, and frequently algorithm performance can be assessed empirically based on comparisons of the clusters generated.[15]

Currently, there remains a gap in knowledge regarding reasonable expected functional outcomes for patients undergoing surgery for ASI. The purpose of this study was to use unsupervised machine learning clustering to define the "optimal observed outcome" after surgery for ASI and to identify predictors for achieving it. We hypothesized that a majority of patients would achieve the "optimal observed outcome" after surgery but that a relative minority would be able to achieve a "perfect outcome."

Methods

Data Source

Institutional review board approval was obtained from both the Mayo Clinic and Olmsted Medical Center (16-007084 and 042-OMC-16) for this retrospective cohort study. Patients who experienced anterior shoulder instability between January 1, 1994, and July 31, 2016, were identified using the Rochester Epidemiology Project, an established geographic database of more than 500,000 patients with complete medical records of all residents in Olmsted County, Minnesota, and neighboring counties in southeast Minnesota and western Wisconsin. Exact methods for abstracting information from this database and its generalizability have been previously described in detail.[16-18] Patients were identified with International Classification of Diseases, Revisions 9 and 10, diagnosis codes for shoulder instability. Patient charts were individually reviewed in detail to confirm the diagnosis of anterior shoulder instability, defined as a documented clinical diagnosis of either dislocation or subluxation by a consulting physician. Inclusion criteria consisted of patients (1) with 1 or more anterior shoulder instability events, (2) $<$40 years of age at the time of initial instability, because of potential confounding by pre-existing osteoarthritis or concurrent rotator cuff pathology, (3) treated surgically for ASI, (4) with a minimum of 2 years' follow-up, and (5) who gave consent for research. Patients with evidence of multi-directional instability or posterior-only shoulder instability based on chart reviews were excluded from analysis, as were those treated solely without surgery.

Variables and Outcomes

Variables documented by the outcomes collection platform were used for feature selection. These included patient demographic information (age, sex, body mass index, smoking status, occupation, activity level, sports involvement near date of injury, date of injury), surgical details (approach, type of repair, bone block augmentation, and concomitant rotator cuff/biceps treatment), and outcome information (recurrence of pain or instability, revision surgery, progression to radiographical osteoarthritis, and physical examination findings at final follow-up), and length of follow-up. Patients were deemed to have achieved either the "optimal observed outcome" or a "suboptimal outcome" based on clustering results. Patients were considered to have achieved a "perfect outcome" if they obtained all of the following ideal results for each outcome restoration of preoperative ROM to within 5° of normal, no recurrent instability, no revision surgery, no pain, full return to sports, no progression to OA, and no complications.

Unsupervised Clustering

Unsupervised clustering is a machine learning technique to produce optimized groupings of objects based on a predefined distance measure within a multidimensional feature space. This technique is distinct from supervised machine learning in that an outcome of interest is not explicitly supplied (hence, "unsupervised"), with the goal of elucidating intrinsic structural patterns within the data. For instance, rather than asking the model to identify differences based on patient who return to sport and those who do not, it naturally identifies outcome differences in patients across all outcome measures of interests, clustering them into groups based on the actually observed outcomes without knowing which of the outcomes are more desirable. This allows for a more unbiased assessment of actual outcomes compared to investigator-driven searches of the desired outcome(s) (i.e., this helps tell the story of what actually happens rather than what the investigators want to happen). Additionally, because this technique is exploratory and not predictive, a model is not generated, the classic train-test internal validation scheme does not apply, and the technique can be performed on smaller datasets than traditional supervised learning. Given the increased freedom in experimental design and ability to aggregate and analyze all the known variables for a patient, clustering is not subject to the limitations of single binomial/multinomial outcomes (i.e., repeat dislocation or not), which is useful when the goal is identification of general risk factors and outcomes, as opposed to modeling of individualized predictions. Furthermore, the fidelity of this technique can easily be empirically determined in the ability to yield useful differentiable clusters.

Unsupervised clustering was performed utilizing the following four candidate machine learning algorithms based on previously established clustering analyses and optimized based on Euclidean distance: unweighted pair group method with arithmetic mean, K-means clustering, agglomerative nesting of hierarchical clustering, and divisive analysis of hierarchical clustering.[19-22] Clustering performances were evaluated by using the 2 internal validation metrics: connectivity and silhouette coefficient, as well as 2 internal stability metrics: average distance and figure of merit. Internal validation metrics evaluates the quality of clustering based on the partitions produced and intrinsic characteristics of the objects in each partition, whereas internal stability measures the consistency of the results through repeated clustering following an iterative feature-elimination process. These internal validation values are calculated post-hoc following each cluster generation, and the final candidate clusters are selected based on optimization of all 4 machine learning algorithms (unweighted pair group method with arithmetic mean, K-means clustering, agglomerative nesting of hierarchical clustering, and divisive analysis of hierarchical clustering), as well as effective partitioning of distinct clusters after outlier elimination.

Briefly, connectivity is a measure of the degree to which nearest neighbors in the feature space are clustered together; it can take a value between 0 and infinity and should be minimized.[15] Conversely, the silhouette coefficient is calculated by taking the ratio of the mean intra-cluster distance between objects to the mean distance between a cluster and its nearest-neighbor cluster. The silhouette coefficient can take any value between -1 and 1, with the optimal clustering assignment maximizing its value. Average distance measures the changes in average distance between observations within the same cluster and figure of merit measures the variance within each cluster, respectively, after iterative elimination of features.[15] After clustering, the largest clusters were selected for pairwise comparisons while remaining clusters were denoted as outliers and excluded from analysis if no individual clusters accounted for more than 10% of the overall cohort size. A detailed explanation of the clustering algorithms and each optimization metric is provided in Appendix Table A1.

Statistical Analysis

Statistical analyses were conducted by use of RStudio software version 1.1.143 (R Foundation for Statistical Computing, Vienna, Austria). Univariate comparisons were performed using Welch's *t*-tests for continuous variables and χ^2 analyses for categorical variables between patients clustered on average CSO achievement and patients with high relative CSO achievement. After univariate analysis, a stepwise multivariable logistic regression controlling for patient demographic, intraoperative, and comorbid variables was used to identify independent risk factors that predicted average CSO achievement. Finally, a time-to-event analysis using Kaplan-Meier plots was performed to compare the natural postoperative clinical course between the optimal and suboptimal clusters. All statistical tests were 2-tailed, and the statistical difference was established with $\alpha \leq 0.05$.

Results

Population Demographics

A total of 654 patients had documented encounters for ASI during the study period. Of these, 228 underwent operative intervention for ASI and met inclusion for analysis. There was a median follow-up of 11.1 years. After clustering and outlier removal ($n = 28$), a total of 200 patients were included in the final 2 largest clusters for comparative analysis. Subsequent stratification for functional outcome improvement achievement resulted in 146 (64%) patients within the "optimal observed outcome" group, whereas 54 (36%) patients were clustered into the suboptimal functional outcomes group.

Demographics were largely similar between both cohorts, with the exception of 3 preoperative variables (Table 1). The mean age at instability diagnosis was younger among the "optimal observed outcome" cluster (21.4 ± 6.4) compared with the "suboptimal outcome" cluster (25.3 ± 7.8, $P < 0.001$). Those with "suboptimal outcome" experienced significantly longer time from initial instability to time of diagnosis (51.4 ± 64.2 vs 1.9 ± 4.4 months, $P < 0.001$) and were more likely to have a history of voluntary/habitual instability (6.2% vs 18.5%, $P = 0.018$). There were no differences between cohorts in age, history of traumatic instability, or the distribution of sports and athletic activities. Comparison of intraoperative findings between the 2 clusters found no significant differences based on labrum repair region, the use of bone block augmentation, or concurrent procedures (Table 1). Among patients with baseline magnetic resonance imaging scans (n = 138 [69%]), there were no significant differences in imaging findings between the groups (Appendix Table A2).

Clustering Results

Visualization of the 2 largest clusters in multidimensional feature space after outlier removal is provided in Figure 1.

Differences Between "Optimal Observed Outcomes" and "Suboptimal Outcomes" Clusters

After clustering, the suboptimal outcomes cluster demonstrated significantly increased rate of recurrent postoperative pain (51.9% vs 22.6%, $P < 0.001$), subjective reporting of postoperative pain as moderate (11.1% vs 1.4%) and severe (1.9% vs 0%, $P < 0.001$), rate of recurrent instability (40.7% vs 12.3%, $P < 0.001$), rate of revision surgery (24.1% vs 9.6%, $P < 0.015$), the onset of symptomatic OA (18.5% vs 4.8%, $P = 0.005$), and decreased forward elevation

Table 1. Baseline Characteristics of Cohorts After Clustering

Variables	Optimal Observed Outcome(N = 146)	Suboptimal Outcome (N = 54)	*P* Value
Demographics			
Male sex	121 (82.9%)	46 (85.2%)	0.860
Age at initial instability event	20.53 (6.55%)	20.46 (6.40%)	0.947
Age at instability diagnosis	21.37 (6.44%)	25.30 (7.84%)	*<0.001*
Months from initial instability to presentation	11.58 ± 15.21	55.86 ± 54.37	*<0.001*
Months from initial instability to surgery	24.80 (23.13%)	63.54 (52.70%)	*<0.001*
Age at initial surgery	22.22 (6.45%)	25.62 (7.64%)	*0.002*
Initial instability from acute trauma	138 (94.5%)	50 (92.6%)	0.862
Sport			0.331
None	37 (25.3%)	14 (25.9%)	
Contact/Weights	82 (56.2%)	23 (42.6%)	
Extreme	6 (4.1%)	3 (5.6%)	
Overhead	13 (8.9%)	9 (16.7%)	
Throwing	8 (5.5%)	5 (9.3%)	
Habitual voluntary dislocation	9 (6.2%)	10 (18.5%)	0.018
Preoperative radiograph			
Osteoarthritis present	1 (0.7%)	2 (3.7%)	0.366
Hills Sachs present	37 (25.3%)	16 (29.6%)	0.668
Bony Bankart present	12 (8.2%)	1 (1.9%)	0.194
Treatment			
Formal physical therapy	110 (75.3%)	41 (75.9%)	>0.999
Open surgical approach	36 (24.7%)	16 (29.6%)	0.596
Posterosuperior labral repair	3 (2.1%)	1 (1.9%)	>0.999
Posteroinferior labral repair	4 (2.7%)	2 (3.7%)	>0.999
Anteroinferior labral repair	122 (83.6%)	42 (77.8%)	0.461
Soft tissue Bankart repair	114 (78.1%)	38 (70.4%)	0.344
Anterosuperior labral repair	23 (15.8%)	6 (11.1%)	0.547
HAGL repair	0 (0.0%)	2 (3.7%)	0.124
Hill Sachs repair			0.387
No	141 (96.6%)	54 (100.0%)	
Remplissage	2 (1.4%)	0 (0.0%)	
Repair	3 (2.1%)	0 (0.0%)	
Biceps tenodesis	1 (0.7%)	0 (0.0%)	>0.999
Rotator cuff repair	5 (3.4%)	1 (1.9%)	0.911
Bone block augmentation	8 (5.5%)	2 (3.7%)	0.884

Italics indicates significance.

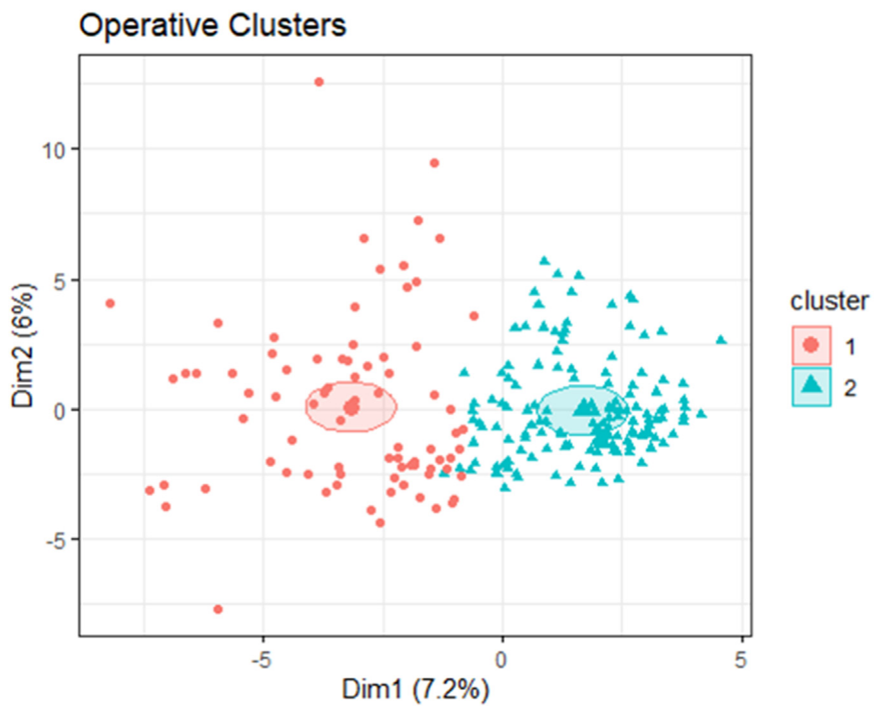

Fig 1. Final clusters after centering and scaling and exclusion of outlying centers. Patients are defined as "suboptimal outcomes" (cluster 1, n = 54) and "optimal outcomes" (cluster 2, n = 146). Dim1 and Dim 2 are the two principal components with the highest contribution to the clustering model based on principal component analysis.

(161 ± 18, 168 ± 11; $P < 0.001$). Full pairwise comparisons are provided in Table 2. Additionally, Kaplan-Meier analysis demonstrated that most adverse events among the "optimal observed outcome" cluster occurred within 24 months of the index surgery, whereas patients in the "suboptimal outcome" cluster continued to experience complications at both midterm and long-term follow-up (44.4 months to recurrence, $P < 0.001$, and 42.5 months to revision surgery, $P < 0.012$) (Fig 2).

Significant Predictors of "Optimal Observed Outcomes"

Significant predictors of "optimal outcomes" achievement after operative intervention were determined by a stepwise multivariable logistic regression model (Table 3). Months from initial instability to surgical consult (odds ratio [OR] = 0.95; 95% confidence interval [CI], 0.92-0.98; $P = 0.006$) and a diagnosis of habitual/voluntary instability (OR = 0.17; 95% CI, 0.04-0.77; $P = 0.020$) predicted significantly decreased likelihood for optimal functional achievement, whereas number of subluxations prior to surgery in a patient without a history of voluntary instability was a significant positive predictor of achievement (OR = 1.30; 95% CI, 1.02-1.65; $P = 0.030$).

Patients Achieving a "Perfect Outcome"

Of the 200 patients treated surgically for ASI, only 82 (41%) achieved the idealized "perfect outcome" that is the ultimate goal for all patients. The final outcomes for this select group of patients included full ROM, no recurrent instability, no complications, no pain, no arthritis, and the ability to return to their preinjury sports and level of activity.

Discussion

The most important finding of this study was that 64% of surgically treated ASI patients were able to achieve the "optimal observed outcome," of whom 41% of patients were able to achieve a "perfect outcome." The "optimal observed outcome" may represent a more realistic goal based on actual outcomes observed. This "optimal observed outcome" group demonstrated statistically significant decreases in recurrent pain (22.6% vs 51.9%), recurrent instability (12.3% vs 40.7%), progression to symptomatic arthritis (4.8% vs 18.5%), revision surgery (9.6% vs 24.1%), and restricted ROM (161° vs 168°) compared to the "suboptimal outcome" group. Further analysis demonstrated that predictors for belonging to the "suboptimal outcome" group included increased time from initial instability to consult, presence of habitual instability, and predilection for recurrent dislocations rather than subluxations before surgery. Achievement of "optimal observed outcome" was not influenced by labral repair location, concurrent procedures, or use of bone block augmentation.

Although there is complex interplay between domain-specific outcomes such as recurrent instability, pain, OA, and range of motion (ROM), the picture of

Table 2. Outcomes After Clustering

	Optimal Observed Outcome (N = 146)	Suboptimal Outcome (N = 54)	*P* Value
After initial consult			
Recurrent pain	122 (83.6%)	39 (72.2%)	0.111
Secondary to acute trauma	70 (47.9%)	20 (37.0%)	0.224
Recurrent Instability	115 (78.8%)	41 (75.9%)	0.812
After initial surgery			
Recurrent pain	33 (22.6%)	28 (51.9%)	*<0.001*
Months from surgery	20.9 (9.81)	51.1 (30.7)	*<0.001*
Secondary to acute trauma	21 (14.4%)	18 (33.3%)	*0.005*
Recurrent Instability	18 (12.3%)	22 (40.7%)	*<0.001*
Months from surgery	24.8 (6.18)	44.4 (23.26)	*<0.001*
Months from consult	24.2 (19.9)	55.3 (43.66)	*<0.001*
Months from initial instability	36.2 (23.0)	110.7 (31.62)	*<0.001*
Postoperative pain			*0.001*
None	127 (87.0%)	35 (64.8%)	
Mild	17 (11.6%)	12 (22.2%)	
Moderate	2 (1.4%)	6 (11.1%)	
Severe	0 (0.0%)	1 (1.9%)	
Adhesive capsulitis	1 (0.7%)	2 (3.7%)	0.366
Postoperative infection	2 (1.4%)	0 (0.0%)	0.949
Postoperative nerve injury	5 (3.4%)	2 (3.7%)	>0.999
Underwent revision surgery	14 (9.6%)	13 (24.1%)	*0.015*
Months from initial surgery	23.09 (6.88)	42.54 (19.45)	*<0.001*
Age at revision surgery	24.25 (3.77)	29.44 (5.54)	*<0.001*
Final follow-up			
Symptomatic osteoarthritis	7 (4.8%)	10 (18.5%)	*0.005*
Years from initial XR	12.29 (2.06)	12.20 (3.85)	
Forward elevation	168.10 (10.79)	160.95 (18.35)	*0.001*
External rotation	66.97 (17.55)	61.80 (19.42)	0.074
Internal rotation to T12 or higher	141 (96.6)	49 (90.7)	0.188

Italics indicates significance.

"optimal observed outcome" remains ill-defined in patients with ASI. Specifically, it is unclear whether improvements across all domains can be achieved or whether some are mutually exclusive from one another. Absence of recurrent instability events and avoidance of revision are 2 commonly used markers for better outcomes.[11,14,23,24] These 2 factors were also identified in this study as markers of improved outcome. Although return to play is often cited within athlete cohorts, return to baseline activity and ROM for the nonathlete is rarely discussed.[10,24] In this study, return to normal ROM, decreased postoperative pain, minimal recurrent pain, and no progression to osteoarthritis were concomitantly present in patients within the "optimal observed outcome" cohort. This suggests that these factors are not always mutually exclusive. These factors have been intermittently identified in the literature as individual markers of successful patient outcome but have yet to be recognized as components of a global functional outcome. Based on these findings, the definition of "optimal observed outcome" can be shifted from a singular outcome to achievement in multiple clinical domains to this more global perspective.

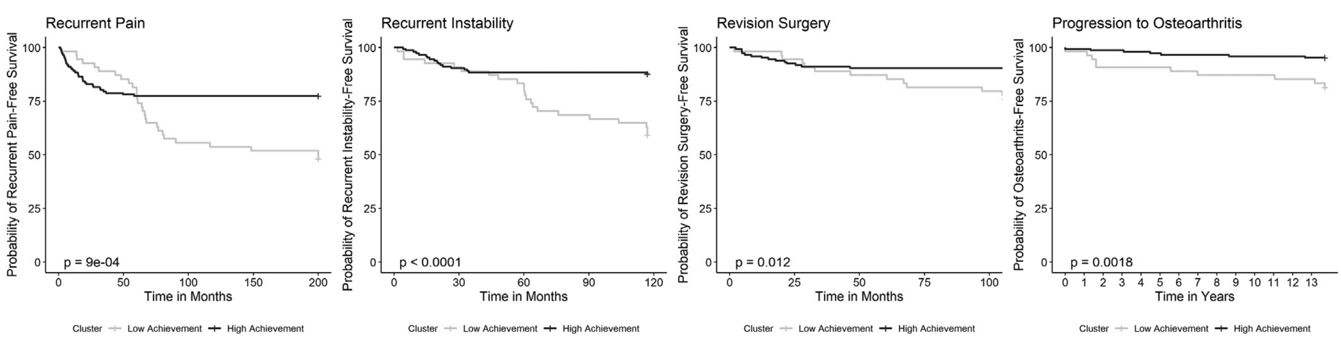

Fig 2. The natural time course of outcomes in operatively managed patients after their first presentation after anterior shoulder instability.

Table 3. Predictors of Optimal Observed Outcome After Operative Treatment of Anterior Shoulder Instability

	OR	95% CI	*P* Value
Months from initial instability to surgical consult	0.95	(0.92-0.98)	0.006
Number of subluxations prior to surgery	1.30	1.02-1.65	0.030
Habitual/voluntary instability	0.17	(0.04-0.77)	0.020

Although almost 2/3 of the cohort achieved optimal functional outcome, patients obviously may achieve varying levels of success within these 6 measurements. In this work, only 41% of patients were considered to have achieved a "perfect outcome" as evidenced by achieving the ideal outcome on all analyzed metrics. Although this has been and will continue to be the goal for our patients, this work suggests that this is not a realistic expectation for the majority of patients. Current literature reports instability recurrence ranging from 10% to 30% after operative treatment, with most estimates around 20% at 10-year follow-up.[25-27] This study demonstrated significantly lower rates of recurrent instability in patients with "optimal observed outcome" at 12.3%. These patients also showed increased gains in forward elevation and return to ROM, consistent with what is described in the literature.[25,28,29] Recurrent pain beyond the postoperative period was reported by just over one fifth of those with "optimal observed outcome." Osteoarthritis development has a broad range in the existing literature, with rates as high as 68% and as low as 11%.[13,14,29-31] Reoperation rates are most often reported in the context of specific technique, but overall rates have been reported at 14%.[31] Although the whole cohort revision rate of 13.5% is similar, revision rates dropped slightly (9.6%) for "optimal observed outcome" and drastically increased in patients with suboptimal outcomes (24.1%). Notably, achievement of "optimal observed outcome" was not influenced by the type of repair nor the need for larger, open procedures; this coincides with recent long-term outcome findings by Bernard et al.[32] Overall, patients should be aware that even with "optimal observed outcome," up to 22.6% may have recurrent pain, 9.8% may have recurrence or require revision surgery, and 4.8% will experience symptomatic OA.

Patients should also be counseled that almost half of those with "suboptimal outcome" will experience recurrent instability and postoperative pain. Approximately one quarter will require revision surgery and develop significant osteoarthritis, with many of these complications and sequelae occurring greater than 3 or 4 years after surgery. As a result of the delayed appearance of these outcomes in the "suboptimal outcome" group, these adverse events may be underreported in short terms studies with 2 to 4 years of mean follow-up.

Both patient factors and injury characteristics have been highly investigated for their roles in ASI and outcomes. Certain patient demographics, such as sex, have shown mixed effects, despite large discrepancies in instability rate.[26,28] Young age has been demonstrated to increase the risk for recurrent instability.[25,26,33,34] Delays in the time from injury to presentation and treatment has been demonstrated to worsen outcomes.[3,34] In the current study, younger age at diagnosis, younger age at first operation, decreased time from injury to treatment, and total number of subluxations (rather than frank dislocation) were all significantly associated with improved outcomes after surgical intervention. Although the "optimal observed outcome" cohort had a younger age at diagnosis, both groups had similar ages at initial instability. This raises the possibility that time from injury, rather than age, may be the more important factor in determining outcome.

A strength of this study comes from the ability to examine multiple outcomes, whereas multivariate regression and supervised machine learning have been shown to generate reliable and effective models for the prediction of singular outcomes. Clustering analysis reviews and groups outcomes based on the aggregation of all available information to enable the possible identification of clinically meaningful subgroups without the inherent bias present in traditional statistical models designed to find investigator-selected outcomes of interest. Additionally, this cohort closely reflects those in previously published literature, with the majority of patients being young males who commonly participated in contact or overhead/throwing sports and experienced an acute traumatic injury leading to the onset of symptoms.[7,35-38] These similarities increase the generalizability of the study and make the findings applicable to patients experiencing anterior shoulder instability.

Limitations

This study is not without limitations. The study was performed in a retrospective fashion and is subject to the common limitations of retrospective research. This includes variability in operative techniques and postoperative regimens; however, this variability is consistent with current standard of practice, where injury characteristics and nuances (such as bone loss) often determine the best surgical technique. Furthermore, we noted no significant differences in surgical techniques and intraoperative findings between the cohorts. The retrospective nature also led to a dearth of patient-reported outcomes. Although this limited our ability to define the "optimal observed outcome" from the patient's perspective, the majority of identified metrics are likely highly important to patients—such as recurrent instability or need for surgical revision. Although this study population was overwhelmingly male, this high prevalence of male patients is consistent with reported incidence of anterior shoulder injuries, and the 16.5% female population is in line with currently reported rates.[1,28,31] Finally, this study analyzed a relatively small cohort given the lengthy span of the study period and a minimum 2-year follow-up requirement.

Conclusion

After surgery for ASI, 64% of patients achieved the "optimal observed outcome" defined as minimal post-operative pain, no recurrent instability or OA, low revision surgery rates, and increased ROM, of whom only 41% achieved a "perfect outcome." Positive predictors were shorter time to presentation and predilection toward preoperative subluxations over dislocations.

Acknowledgments

The authors acknowledge the support from the Foderaro-Quattrone Musculoskeletal-Orthopaedic Surgery Research Innovation Fund.

References

1. Leland DP, Bernard CD, Keyt LK, et al. An age-based approach to anterior shoulder instability in patients under 40 years old: Analysis of a US population. *Am J Sports Med* 2020;48:56-62.
2. Leland DP, Parkes CW, Bernard CD, et al. Significant changes in the diagnosis, injury severity and treatment for anterior shoulder instability over time in a U.S. population. *Arthrosc Sports Med Rehabil* 2020;2(6):e761-e769.
3. Lu Y, Pareek A, Wilbur RR, Leland DP, Krych AJ, Camp CL. Understanding anterior shoulder instability through machine learning: New models that predict recurrence, progression to surgery, and development of arthritis. *Orthop J Sports Med* 2021;9(11):23259671211053326.
4. Wilbur RR, Shirley MB, Nauert RF, et al. Anterior shoulder instability in throwers and overhead athletes: Long-term outcomes in a geographic cohort. *Am J Sports Med* 2021;50:182-188.
5. Dugas JR, Crozier MW. Traumatic anterior instability: Treatment options for initial instability. *Sports Med Arthrosc Rev* 2018;26:95-101.
6. Khiami F, Gérometta A, Loriaut P. Management of recent first-time anterior shoulder dislocations. *Orthop Traumatol Surg Res* 2015;101(1, Supplement):S51-S57.
7. Zacchilli MA, Owens BD. Epidemiology of shoulder dislocations presenting to emergency departments in the United States. *J Bone Joint Surg Am* 2010;92:542-549.
8. Dekker TJ, Goldenberg B, Lacheta L, P Horan M, Millett PJ. Anterior shoulder instability in the professional athlete: Return to competition, time to return, and career length. *Orthop J Sports Med* 2020;8(11):2325967120959728.
9. Donohue MA, Owens BD, Dickens JF. Return to play following anterior shoulder dislocation and stabilization surgery. *Clin Sports Med* 2016;35:545-561.
10. Dickens JF, Rue J-P, Cameron KL, et al. Successful return to sport after arthroscopic shoulder stabilization versus nonoperative management in contact athletes with anterior shoulder instability: A prospective multicenter study. *Am J Sports Med* 2017;45:2540-2546.
11. Hurley ET, Manjunath AK, Bloom DA, et al. Arthroscopic Bankart repair versus conservative management for first-time traumatic anterior shoulder instability: A systematic review and meta-analysis. *Arthroscopy* 2020;36:2526-2532.
12. Buscayret F, Edwards TB, Szabo I, Adeleine P, Coudane H, Walch G. Glenohumeral arthrosis in anterior instability before and after surgical treatment: Incidence and contributing factors. *Am J Sports Med* 2004;32:1165-1172.
13. Cameron ML, Kocher MS, Briggs KK, Horan MP, Hawkins RJ. The prevalence of glenohumeral osteoarthrosis in unstable shoulders. *Am J Sports Med* 2003;31:53-55.
14. Hurley ET, Jamal MS, Ali ZS, Montgomery C, Pauzenberger L, Mullett H. Long-term outcomes of the Latarjet procedure for anterior shoulder instability: A systematic review of studies at 10-year follow-up. *J Shoulder Elbow Surg* 2019;28(2):e33-e39.
15. Brock G, Pihur V, Datta S, Datta S. clValid: An R package for cluster validation. *J Stat Software* 2008;25(4):1-22.
16. Rocca WA, Grossardt BR, Brue SM, et al. Data resource profile: Expansion of the Rochester Epidemiology Project medical records-linkage system (E-REP). *Int J Epidemiol* 2018;47:368-j.
17. St Sauver JL, Grossardt BR, Leibson CL, Yawn BP, Melton LJ 3rd, Rocca WA. Generalizability of epidemiological findings and public health decisions: an illustration from the Rochester Epidemiology Project. *Mayo Clin Proc* 2012;87:151-160.
18. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: The Rochester Epidemiology Project. *Am J Epidemiol* 2011;173:1059-1068.
19. Eckhardt CM, Madjarova SJ, Williams RJ, et al. Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surg Sports Traumatol Arthrosc* 2023;31:376-381.
20. Oyelade J, Isewon I, Oladipupo F, et al. Clustering algorithms: Their application to gene expression data. *Bioinform Biol Insights* 2016;10:237-253.
21. Datta S, Datta S. Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes. *BMC Bioinform* 2006;7:397.

22. Alashwal H, El Halaby M, Crouse JJ, Abdalla A, Moustafa AA. The application of unsupervised clustering methods to Alzheimer's disease. *Front Comput Neurosci* 2019;13:31.
23. Robinson CM, Howes J, Murdoch H, Will E, Graham C. Functional outcome and risk of recurrent instability after primary traumatic anterior shoulder dislocation in young patients. *J Bone Joint Surg Am* 2006;88:2326-2336.
24. Waterman B, Owens BD, Tokish JM. Anterior shoulder instability in the military athlete. *Sports Health* 2016;8: 514-519.
25. Aboalata M, Plath JE, Seppel G, Juretzko J, Vogt S, Imhoff AB. Results of arthroscopic Bankart repair for anterior-inferior shoulder instability at 13-year follow-up. *Am J Sports Med* 2017;45:782-787.
26. Panzram B, Kentar Y, Maier M, Bruckner T, Hetto P, Zeifang F. Mid-term to long-term results of primary arthroscopic Bankart repair for traumatic anterior shoulder instability: A retrospective study. *BMC Musculoskelet Disord* 2020;21:191.
27. Takatori N, Uchiyama Y, Imai T, Watanabe M. Long-term outcomes of open modified inferior capsular shift for traumatic anterior shoulder instability: Over 20 years of follow-up. *J Shoulder Elbow Surg* 2022;31:359-366.
28. Chen M, Feng S, Chen Y, et al. Sex-based differences in clinical outcomes after arthroscopic anterior shoulder stabilization: Results at 5-year follow-up. *Orthop J Sports Med* 2021;9(5):23259671211008841.
29. Wei J, Lu M, Zhao L, Zeng X, He L. Free bone grafting improves clinical outcomes in anterior shoulder instability with bone defect: A systematic review and meta-analysis of studies with a minimum of 1-year follow-up. *J Shoulder Elbow Surg* 2022;31(4):e190-e208.
30. Kruckeberg BM, Leland DP, Bernard CD, et al. Incidence of and risk factors for glenohumeral osteoarthritis after anterior shoulder instability: A US population-based study with average 15-year follow-up. *Orthop J Sports Med* 2020;8(11):2325967120962515.
31. Plath JE, Aboalata M, Seppel G, et al. Prevalence of and risk factors for dislocation arthropathy: Radiological long-term outcome of arthroscopic Bankart repair in 100 shoulders at an average 13-year follow-up. *Am J Sports Med* 2015;43:1084-1090.
32. Bernard CD, Leland DP, Keyt LK, et al. Although surgical techniques differ, similar outcomes can be obtained when operating after single versus multiple anterior shoulder dislocations. *Arthrosc Sports Med Rehabil* 2021;3(1): e163-e170.
33. Flinkkilä T, Knape R, Sirniö K, Ohtonen P, Leppilahti J. Long-term results of arthroscopic Bankart repair: Minimum 10 years of follow-up. *Knee Surg Sports Traumatol Arthrosc* 2018;26:94-99.
34. Porcellini G, Campi F, Pegreffi F, Castagna A, Paladini P. Predisposing factors for recurrent shoulder dislocation after arthroscopic treatment. *J Bone Joint Surg Am* 2009;91: 2537-2542.
35. Owens BD, Dawson L, Burks R, Cameron KL. Incidence of shoulder dislocation in the United States military: Demographic considerations from a high-risk population. *J Bone Joint Surg Am* 2009;91:791-796.
36. Owens BD, Duffey ML, Nelson BJ, DeBerardino TM, Taylor DC, Mountcastle SB. The incidence and characteristics of shoulder instability at the United States Military Academy. *Am J Sports Med* 2007;35:1168-1173.
37. Shah A, Judge A, Delmestri A, et al. Incidence of shoulder dislocations in the UK, 1995-2015: A population-based cohort study. *BMJ Open* 2017;7(11):e016112.
38. Kraeutler MJ, McCarty EC, Belk JW, et al. Descriptive epidemiology of the MOON Shoulder Instability Cohort. *Am J Sports Med* 2018;46:1064-1069.

Appendix Table A1. Detailed definition of terms frequently encountered in unsupervised machine learning and clustering

Term	Definition
Hierarchical clustering	An umbrella term for clustering algorithms that groups objects into clusters/subgroups based a calculated distance between each object that is described within a distance matrix. This can be performed either in a top-down manner (starting with all objects in the population of interest and dividing into smaller clusters) or a bottom-up manner (all individual cases in the population of interest as their own clusters and adding into form larger clusters), both of these are performed in an iterative fashion. The former method is known as agglomerative hierarchical clustering whereas the latter is known as divisive hierarchical clustering.
K-Means clustering	A method for reducing the variability within clusters using iterations based off an initial guess for the center of each cluster. After this initial guess each observation is placed into the cluster to which it is closest, and then the cluster centers are updated, and the entire process is repeated until the centers of each cluster no longer move. This initial guess can also be determined using another clustering algorithm, such as a hierarchical algorithm, to determine the starting points for the cluster centers.
Unweighted pair group method with arithmetic mean	An agglomerative hierarchical clustering method that constructs a dendrogram from the bottom-up as the 2 nearest clusters are combined to form a new higher-level cluster after each iteration until the desired number of clusters is achieved. The distance between any 2 clusters is calculated as the average distance between the elements in each cluster.
Divisive analysis of hierarchical clustering	A specific divisive hierarchical algorithm that begins with all observations in a single cluster and then subsequently divides the cluster until each cluster contains a single observation.
Agglomerative nesting of hierarchical clustering	An agglomerative hierarchical clustering method in which each observation starts as its own cluster and then the observations are collected until all similar points form a single cluster. The distance between 2 clusters is the mean of the dissimilarities between elements in 1 cluster and the elements in another cluster.
Connectivity	An internal validation measure of connectedness, or the extent to which observations are placed in the same cluster as their nearest neighbors in space, given as a value between 0 and infinity, with a smaller value reflecting a higher degree of connectivity.
Silhouette coefficient	A metric used to calculate the accuracy of a clustering technique based on the distinct separation of clusters. A value from -1 to 1 is given with values near 1 representing well-clustered observations, and values near -1 representing poorly clustered observations.
Average distance	A stability measurement computed from the average distance between observations that are placed in the same cluster by the clustering algorithm both based on the full data and then based on the data with a single variable removed. A value between 0 and infinity is assigned, with smaller values representing a better model.
Figure of merit	A stability measure based on the mean variance of observations in the deleted column within a cluster that is based on the undeleted variables, which estimates the mean error using predictions based on the cluster averages. The final score is averaged over all the deleted columns and is given as a value from 0 to infinity, with smaller values representing better performance.

Appendix Table A2. Baseline Findings in Those Undergoing Preoperative Magnetic Resonance Imaging

Variables	Optimal Observed Outcome (N = 100)	Suboptimal Outcome (N = 38)	*P*
Posterosuperior labral tear	11 (11.0%)	2 (5.3%)	0.481
SLAP anterior superior tear	23 (23.0%)	9 (23.7%)	>0.999
Posteroinferior labral tear	13 (13.0%)	5 (13.2%)	>0.999
Anterior inferior labral tear	90 (90.0%)	30 (78.9%)	0.150
Bony Bankart	28 (28.0%)	13 (34.2%)	0.614
Glenohumeral ligament tear			0.064
Complete	0 (0.0%)	2 (5.3%)	
Partial	18 (18.0%)	7 (18.4%)	
No	82 (82.0%)	29 (76.3%)	
Cartilage injury			0.657
Both	4 (4.0%)	0 (0.0%)	
Glenoid	54 (54.0%)	22 (57.9%)	
Humeral head	3 (3.0%)	1 (2.6%)	
No	39 (39.0%)	15 (39.5%)	
Hill Sachs (Yes)	80 (80.0%)	30 (78.9%)	>0.999
Biceps tendon pathology	3 (3.0%)	2 (5.3%)	0.900
Cuff tear	14 (14.0%)	6 (15.8%)	>0.999

Clustering was performed using cluster numbers ranging from a minimum of 2 and a maximum of 20 using the previously mentioned algorithms. Among these algorithm and cluster center combinations, AGNES clustering with 13 clusters was the best-performing candidate algorithm, with a connectivity of 20.9, a silhouette coefficient of 0.4, average distance of 99.5, and figure of merit of 2.9. Visualization of the 2 largest clusters in multidimensional feature space after outlier removal is provided in Figure 1, in which the 2 distinct clusters can be seen plotted after principal component analysis was performed to reduce the entire feature set of each patient to a 2-dimensional cartesian coordinate system.