Unsupervised Machine Learning of the Combined Danish and Norwegian Knee Ligament Registers

Identification of 5 Distinct Patient Groups With Differing ACL Revision Rates

R. Kyle Martin,*†‡§ MD, Solvejg Wastvedt,‖ BA, Ayoosh Pareek,¶ MD, Andreas Persson,§#** MD, PhD, Håvard Visnes,§**†† MD, PhD, Anne Marie Fenstad,** MSc, Gilbert Moatshe,§# MD, PhD, Julian Wolfson,‖ PhD, Martin Lind,‡‡ MD, PhD, and Lars Engebretsen,§# MD, PhD
Investigation performed at the University of Minnesota, Minneapolis, Minnesota, USA

Background: Most clinical machine learning applications use a supervised learning approach using labeled variables. In contrast, unsupervised learning enables pattern detection without a prespecified outcome.

Purpose/Hypothesis: The purpose of this study was to apply unsupervised learning to the combined Danish and Norwegian knee ligament register (KLR) with the goal of detecting distinct subgroups. It was hypothesized that resulting groups would have differing rates of subsequent anterior cruciate ligament reconstruction (ACLR) revision.

Study Design: Cohort study; Level of evidence, 3.

Methods: K-prototypes clustering was performed on the complete case KLR data. After performing the unsupervised learning analysis, the authors defined clinically relevant characteristics of each cluster using variable summaries, surgeons' domain knowledge, and Shapley Additive exPlanations analysis.

Results: Five clusters were identified. Cluster 1 (revision rate, 9.9%) patients were young (mean age, 22 years; SD, 6 years), received hamstring tendon (HT) autograft (91%), and had lower baseline Knee injury and Osteoarthritis Outcome Score (KOOS) Sport and Recreation (Sports) scores (mean, 25.0; SD, 15.6). Cluster 2 (revision rate, 6.9%) patients received HT autograft (89%) and had higher baseline KOOS Sports scores (mean, 67.2; SD, 16.5). Cluster 3 (revision rate, 4.7%) patients received bone–patellar tendon–bone (BPTB) or quadriceps tendon (QT) autograft (94%) and had higher baseline KOOS Sports scores (mean, 65.8; SD, 16.4). Cluster 4 (revision rate, 4.1%) patients received BPTB or QT autograft (88%) and had low baseline KOOS Sports scores (mean, 20.5; SD, 14.0). Cluster 5 (revision rate, 3.1%) patients were older (mean age, 42 years; SD, 7 years), received HT autograft (89%), and had low baseline KOOS Sports scores (mean, 23.4; SD, 17.6).

Conclusion: Unsupervised learning identified 5 distinct KLR patient subgroups and each grouping was associated with a unique ACLR revision rate. Patients can be approximately classified into 1 of the 5 clusters based on only 3 variables: age, graft choice (HT, BPTB, or QT autograft), and preoperative KOOS Sports subscale score. If externally validated, the resulting groupings may enable quick risk stratification for future patients undergoing ACLR in the clinical setting. Patients in cluster 1 are considered high risk (9.9%), cluster 2 patients medium risk (6.9%), and patients in clusters 3 to 5 low risk (3.1%-4.7%) for revision ACLR.

Keywords: ACL revision; outcome prediction; machine learning; artificial intelligence; unsupervised learning

Machine learning represents an increasingly used approach within the orthopaedic literature due to the ability to process large volumes of complex data and develop clinically useful diagnostic, prognostic, or data collection models.[30,32] The 3 main categories of machine learning approaches are supervised learning, unsupervised learning, and reinforcement learning. Most of the orthopaedic studies to date have applied a supervised learning approach, referring to the analysis of labeled data. In the supervised learning approach, the computer algorithm is provided with variables that are labeled as either a "predictor" or an "outcome," and the model is tasked with predicting a specified outcome. In contrast, unsupervised learning involves the analysis of unlabeled data whereby the model

The American Journal of Sports Medicine
1–11
DOI: 10.1177/03635465231225215

is tasked with independently finding patterns in the data set. This process enables the interpretation and simplification of highly complex data through the identification of hidden structures and patterns.[7]

Within orthopaedic research, unsupervised learning approaches have recently been used to stratify groups of patients according to their risk of hip osteoarthritis progression[17] and to identify subphenotypes of osteoarthritis based on blood-based biochemical markers.[2] These examples highlight how a novel approach to a common problem can provide insight into the factors associated with complex clinical conditions. Outcome after anterior cruciate ligament (ACL) injury and subsequent ACL reconstruction (ACLR) is one such example of a clinical condition that evades complete understanding, despite troves of literature on the subject. Studies from the national knee ligament registers, Multicenter Orthopaedic Outcomes Network, and others have helped identify age, activity level, graft choice, fixation device, and posterior tibial slope as some factors that influence failure risk.[5,10,16,28,29,34,41] Despite recognition of these and other risk factors for a poor outcome,[11,19,25,35] along with recent advancements in surgical decision-making and techniques,[6,8,27,33,39] highly accurate clinical prediction models remain elusive. One constraint to accurate patient-specific outcome prediction is the sheer volume of risk factors that may contribute to a patient's outcome and, specifically, the limited ability to synthesize the complex and often unrecognized interactions between these factors.

The Norwegian Knee Ligament Register (NKLR) and Danish Knee Ligament Reconstruction Registry (DKRR) have been prospectively collecting data related to ACLR in their respective countries for nearly 20 years.[9,31] Since their inception, these national registers have produced several studies on ACL treatment and outcomes and have recently developed preliminary outcome prediction models using supervised machine learning methodology.[15,20-23] The present study sought to further investigate the factors associated with subsequent ACLR revision through the application of unsupervised learning techniques to the combined Norwegian and Danish knee ligament register (KLR). The primary goal of this analysis was to identify distinct subgroups of patients within the registers and determine if the rate of subsequent revision ACLR differs between the patient clusters. The hypothesis was that unsupervised learning would facilitate the grouping of patients based on common characteristics and that this would enable the identification of high- and low-risk groups of patients.

METHODS

Ethics

Informed consent was obtained prospectively from all patients enrolled in the NKLR and the Norwegian Data Inspectorate grants permission for the NKLR to collect, analyze, and publish on these health data. Data registration was performed according to European Union data protection rules, with all data deidentified before retrieval. The regional ethics committee stated that further ethics approval was not necessary.[9] Similarly, the DKRR prospectively obtained informed consent at the time of enrollment and patient data were deidentified before retrieval with no further ethics approval required.

Data Preparation

Patients with primary ACLR surgery dates from June 2004 through December 2020 were included. Patients with missing values for graft choice, those with graft choice recorded as "direct suture," and those with missing values for the indicator of revision surgery were excluded. Variables contained within the combined KLR and considered for analysis are shown in Table 1.

The activity that reportedly led to ACL injury was classified as a pivoting sport, nonpivoting sport, or other activity. Meniscal injuries were classified as present with repair, present without repair (no treatment or partial meniscectomy), or no meniscal injury. Cartilage injuries were grouped according to the International Cartilage Regeneration & Joint Preservation Society grading system and recorded as grade 1 or 2, grade 3 or 4, or no cartilage injury. Additionally, a predictor indicating if a patient

*Address correspondence to R. Kyle Martin, MD, Department of Orthopedic Surgery, University of Minnesota, 2512 South 7th St, Suite R200, Minneapolis, MN 55455, USA (email: rkylemmartin@gmail.com) (X: @UMNOrthoSurg; @RKMartin6; @nih_physicperf).
†Department of Orthopedic Surgery, University of Minnesota, Minneapolis, Minnesota, USA.
‡Department of Orthopedic Surgery, CentraCare, Saint Cloud, Minnesota, USA.
§Oslo Sport Trauma Research Center, Norwegian School of Sports Science, Oslo, Norway.
‖Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.
¶Department of Orthopedic Surgery, Hospital for Special Surgery, New York, New York, USA.
#Orthopaedic Clinic, Oslo University Hospital Ullevål, Oslo, Norway.
**Norwegian Knee Ligament Register, Haukeland University Hospital, Bergen, Norway.
††Department of Orthopedics, Sorlandet Hospital, Kristiansand, Norway.
‡‡Aarhus University Hospital, Aarhus, Denmark.
Submitted September 17, 2023; accepted November 29, 2023.
One or more of the authors has declared the following potential conflict of interest or source of funding: This study was funded by a Norwegian Centennial Chair seed grant. R.K.M. has received support for education from Gemini/Arthrex and Smith & Nephew and hospitality payments from Foundation Medical and Medical Device Business Services. G.M. has received consulting fees from Arthrex and IBSA. M.L. has received consulting fees from Smith & Nephew. L.E. has received research support from Biomet and Health Southeast Norway and royalties from Arthrex and Smith & Nephew. A.Pareek. has received support for education from Smith & Nephew and hospitality payments from Medical Device Business Services. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

TABLE 1
Patient Characteristics[a]

Variable	Combined Data, n = 62,955	Complete Case Data, N = 28,631
Revision	3205 (5.1)	1770 (6.2)
Mean follow-up time or time to revision, y	7.6 ± 4.5	8.2 ± 4.5
Mean age at surgery, y	28 ± 11	28 ± 10
Mean age at injury, y	27 ± 10	27 ± 10
Missing	1870	
Sex		
Male	36,509 (58)	15,671 (55)
Female	26,446 (42)	12,960 (45)
Mean presurgery KOOS QOL score	36.3 ± 18.0	36.5 ± 17.9
Missing	29,512	
Mean presurgery KOOS Sports score	41.2 ± 26.9	41.2 ± 26.8
Missing	29,708	
Below median on all presurgery KOOS Subscales	6372 (19)	5259 (18)
Missing	29,323	
Activity that led to injury		
Nonpivoting	20,391 (33)	8175 (29)
Pivoting	35,851 (57)	16,747 (58)
Other	6162 (9.9)	3709 (13)
Missing	551	
Meniscal injury		
Injury without repair	20,328 (32)	9568 (33)
Injury with repair	10,554 (17)	4640 (16)
None	32,061 (51)	14,423 (50)
Missing	12	
Cartilage injury (ICRS grade)		
1 or 2	8766 (14)	4195 (15)
3 or 4	3223 (5.1)	1627 (5.7)
None	50,878 (81)	22,809 (80)
Missing	88	
Graft choice		
BPTB	15,639 (25)	9000 (31)
Hamstring	43,518 (69)	18,356 (64)
QT/BQT	2520 (4.0)	888 (3.1)
Other	1278 (20)	387 (1.4)
Tibial fixation device		
Interference screw	55,792 (90)	25,759 (90)
Suspension/cortical device	3643 (5.9)	2031 (7.1)
Other	2356 (3.8)	841 (2.9)
Missing	1164	
Femoral fixation device		
Interference screw	16,434 (27)	8793 (31)
Suspension/cortical device	39,742 (65)	17,502 (61)
Other	4822 (7.9)	2336 (8.2)
Missing	1957	
Fixation device combination		
Interference screw × 2	15,865 (26)	8467 (30)
Interference/suspension	236 (0.4)	150 (0.5)
Suspension/cortical device × 2	2994 (4.9)	1540 (5.4)
Suspension/interference	34,895 (58)	15,493 (54)
Other	6529 (11)	2981 (10)
Missing	2436	
History of previous surgery on same knee[b]	10,312 (17)	4540 (16)
Missing	673	
History of previous cruciate ligament injury to opposite knee[b]	4839 (8.1)	1977 (6.9)
Missing	2946	

(continued)

TABLE 1
(continued)

Variable	Combined Data, n = 62,955	Complete Case Data, N = 28,631
Median time injury to surgery, y	1.63 [0.33-1.32]	0.61 [0.33-1.29]
Missing	2083	
Register		
DKRR	34,554 (55)	10,487 (37)
NKLR	28,401 (45)	18,144 (63)

[a]Data are presented as n, n (%), mean ± SD, or median [IQR]. BPTB, bone–patellar tendon–bone autograft; DKRR, Danish Knee Ligament Register; ICRS, International Cartilage Regeneration & Joint Preservation Society; KOOS, Knee injury and Osteoarthritis Outcome Score; NKLR, Norwegian Knee Ligament Register; QOL, Quality of Life subscale; QT/BQT, quadriceps tendon autograft, with or without bone; Sports, Sport and Recreation subscale.
[b]Surgery performed before primary anterior cruciate ligament reconstruction and enrollment in the register.

was below the median score in the respective register on all presurgery Knee injury and Osteoarthritis Outcome Score (KOOS) variables was also created.

Missing Data

A previous study applying supervised machine learning models to the combined KLR data guided the approach to missing data for this study.[20] Briefly, supervised learning models were trained and evaluated using complete data on all variables. This was then repeated using multiple imputation to assess the effect of restricting data to complete cases. This is a common technique for dealing with missing data that fills in incomplete values based on patterns in the data. Multiple imputation allowed the assessment of the reasonableness of restricting the analysis to complete cases and found that multiply imputed data were not notably different from the complete case analysis. This means that there was no meaningful advantage of data imputation for the predictive modeling. Therefore, for this study only patients with complete data on all predictors were included in the analysis.

Unsupervised Learning

The machine learning methods used in this analysis were all unsupervised, meaning the models were not trained to produce predictions for a specific outcome variable. Instead, unsupervised methods model how the data were organized with respect to a given set of predictor variables.[13] The applied unsupervised methods produced groups, or clusters, of observations with similar relationships among the predictor variables. Because unsupervised learning does not train and then test predictions, the sample was not split. The entire sample of patients was used to build the unsupervised models and characterize the resulting clusters. All analyses were conducted in R (Version 4.1.1; R Core Team).

Three unsupervised clustering methods were applied: k-means (function *kmeans*; package *stats*), agglomerative hierarchical clustering (function *hclust*; package *fastcluster*[26]), and k-prototypes (function *kproto*; package *clustMixType*[36]). K-means clustering required the user to prespecify the number of clusters. The algorithm then grouped the observations to minimize the sum of squares from points to the cluster centers.[12] To determine the number of clusters (k), a common technique called the elbow method was used. In this approach, clusters were computed for various possible values of k, and the within-cluster sums of squares were calculated and plotted against the value of k. The point at which this line bent sharply upward (the elbow) dictated the optimal number of clusters (Figure 1).

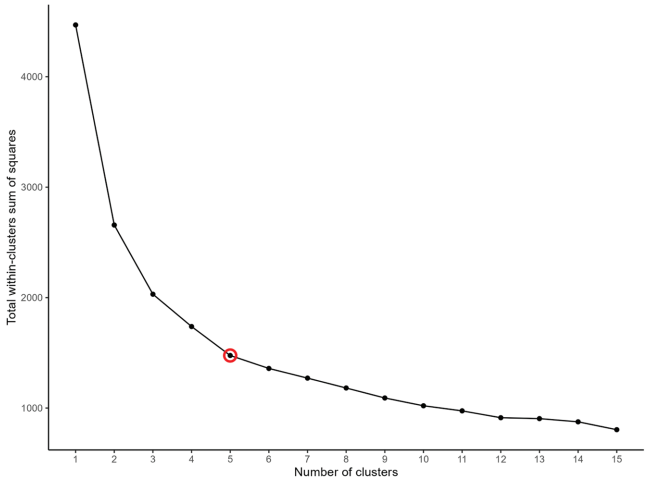

Figure 1. The elbow method to determine the number of clusters for unsupervised learning analysis. The location where the line bends sharply upward (circle) signifies the elbow, representing the optimal number of clusters.

This represented the fewest clusters that could be created without a sharp increase in within-cluster heterogeneity.[38] Agglomerative hierarchical clustering began with each observation in its own cluster and yielded many possible partitions of decreasing complexity, requiring the user to select a level of complexity (by specifying a desired number of clusters).

K-means and agglomerative hierarchical clustering only accommodates continuous predictor variables. To overcome this limitation, a third method, k-prototypes,

which accommodated mixed type predictors, was used. K-prototypes is similar to k-means in that it minimizes within-cluster distance from the cluster mean when assigning observations to a prespecified number of clusters. The distance metric was a weighted combination of Euclidean distance for continuous variables and the count of mismatched category labels for categorical variables. A data-driven technique was used to select the weighting parameter. The cluster "mean" was the mean for continuous variables and the mode for categorical variables. The elbow and silhouette methods were used to define the optimal number of k-prototypes clusters. The silhouette method identified the number of clusters that maximized between-cluster and minimized within-cluster dissimilarity.[36]

Measures of Cluster Quality

Unlike with supervised learning where models are trained on a training set and evaluated against observed labels on a test set, with unsupervised learning there are no labels for comparison. Assessing the quality of model results is therefore more challenging and typically relies on heuristic arguments and domain knowledge.[13] Therefore, a combination of 2 data-driven methods (elbow and silhouette) and domain knowledge was used to choose the number of clusters.

Model Interpretability and Clinical Relevance

To identify the defining characteristics of each cluster, 7 orthopaedic surgeons (R.K.M., A.Pareek., A.Persson., H.V., G.M., M.L., L.E.) with subspecialty training in sports medicine reviewed the patient groups and highlighted the clinically relevant features based on their domain knowledge and variable summaries. The goal was to define each cluster in terms that would enable the assignment of future patients to 1 of the 5 clusters. To aid in cluster interpretation, SHapley Additive exPlanations (SHAP) analysis was also performed.[18] This required a 2-step process: (1) build a classification model predicting clusters from input variables and (2) compute SHAP values for this classification model. First, a gradient boosting model was trained to predict the cluster number using all predictor variables originally used for clustering (R package *xgboost*). Gradient boosting is a tree-based machine learning method that can be used for classification with multiple classes, such as in this situation.[4] Next, SHAP values were computed for this model using built-in functions in the *xgboost* package. The SHAP values explained the contributions of input variables in each cluster by summarizing their influence on individual predictions. Cluster-specific Kaplan-Meier curves were created to describe each cluster's mean risk of revision surgery.

RESULTS

Participants

After data cleaning, a process whereby incorrect, duplicate, or incomplete data were removed or corrected, the combined register population consisted of 62,955 patients, 55% from the DKRR and 45% from the NKLR. The primary outcome, revision surgery, occurred in 5.1% of patients during a mean follow-up time of 7.6 years (SD, 4.5 years). The population was 55% male with median ages at primary injury and surgery of 24 years (IQR, 18-34 years) and 26 years (IQR, 20-36 years), respectively. After removing patients with missing predictor variables, the study population consisted of 28,631 patients. Characteristics of the study population at the time of surgery along with the complete case data set are presented in Table 1.

Clustering Results

The k-prototypes method was chosen because it accommodated both continuous and categorical predictors. The optimal number of clusters was set at 5 via a combination of the data-driven elbow and silhouette methods and domain knowledge (Figure 1). A description of the 5 clusters is presented in Table 2 and Figure 2. Figure 3 presents the SHAP values for all clusters. Cluster-specific Kaplan-Meier curves demonstrating the revision risk profiles for the 5 patient groups are presented in Figure 4.

Surgeon domain knowledge and SHAP values were used to interpret the variable summaries and simplify the distinguishing characteristics of each cluster for clinical relevance. Cluster 1 (revision rate, 9.9%) patients were young (mean age, 22 years; SD, 6 years) and more often female (60%), received hamstring tendon (HT) autograft (91%), and had lower baseline KOOS Sport and Recreation (Sports) scores (mean, 25.0; SD, 15.6). Cluster 2 (revision rate, 6.9%) patients received HT autograft (89%), were more often male (68%), and had higher baseline KOOS Sports scores (mean, 67.2; SD, 16.5). Cluster 3 (revision rate, 4.7%) patients received bone–patellar tendon–bone (BPTB) or quadriceps tendon (QT) autograft (94%) and had higher baseline KOOS Sports scores (mean, 65.8; SD, 16.4). Cluster 4 (revision rate, 4.1%) patients received BPTB or QT autograft (88%) and had low baseline KOOS Sports scores (mean, 20.5; SD, 14.0). Cluster 5 (revision rate, 3.1%) patients were older (mean, 42; SD, 7 years), underwent ACLR with HT autograft (89%), and had low baseline KOOS Sports scores (mean, 23.4; SD, 17.6).

DISCUSSION

The most important finding of this study was that unsupervised learning analysis of the combined KLR identified 5 distinct patient subgroups among patients undergoing primary ACLR, which are clinically distinguishable based on age, graft type, and baseline KOOS Sports score. Each grouping was associated with its own unique rate of subsequent ACLR revision. If externally validated, the results of this analysis could be applied in the clinical setting to classify patients into 1 of the 5 clusters. This would enable rapid estimation of the risk of subsequent revision ACLR and could be used to guide preoperative discussions and

TABLE 2
Characteristics of Clusters Using k-Prototypes Method[a]

Variable	Cluster 1, n = 7038	Cluster 2, n = 7693	Cluster 3, n = 4118	Cluster 4, n = 4852	Cluster 5, n = 4930
Revision	695 (9.9)	532 (6.9)	193 (4.7)	198 (4.1)	152 (3.1)
Mean follow-up time or time to revision, y	8.2 ± 4.3	8.5 ± 4.3	7.5 ± 4.8	8.0 ± 5.0	8.8 ± 4.2
Mean age at surgery	22 ± 6	25 ± 9	25 ± 9	30 ± 10	42 ± 7
Mean age at injury	21 ± 6	24 ± 8	23 ± 8	28 ± 9	40 ± 8
Sex					
Male	2808 (40)	5198 (68)	2473 (60)	3036 (63)	2156 (44)
Female	4230 (60)	2495 (32)	1645 (40)	1816 (37)	2774 (56)
Mean presurgery KOOS QOL score	29.7 ± 13.9	49.1 ± 16.1	47.6 ± 15.7	25.5 ± 13.4	28.4 ± 14.4
Mean presurgery KOOS Sports score	25.0 ± 15.6	67.2 ± 16.5	65.8 ± 16.4	20.5 ± 14.0	23.4 ± 17.6
Below median on all presurgery KOOS Subscales	1852 (26)	0 (0)	0 (0)	1738 (36)	1669 (34)
Activity that led to injury					
Nonpivoting	1524 (22)	1746 (23)	931 (23)	1069 (22)	2905 (59)
Pivoting	4863 (69)	5273 (69)	2730 (66)	2796 (58)	1085 (22)
Other	651 (9.2)	674 (8.8)	457 (11)	987 (20)	940 (19)
Meniscal injury					
Injury without repair	2182 (31)	2467 (32)	1277 (31)	1723 (36)	1919 (39)
Injury with repair	1305 (19)	1183 (15)	774 (19)	905 (19)	473 (9.6)
None	3551 (50)	4043 (53)	2067 (50)	2224 (46)	2538 (51)
Cartilage injury (ICRS grade)					
1 or 2	808 (11)	930 (12)	632 (15)	898 (19)	927 (19)
3 or 4	262 (3.7)	280 (3.6)	176 (4.3)	389 (8.0)	520 (11)
None	5968 (85)	6483 (84)	3310 (80)	3565 (73)	3483 (71)
Graft choice					
BPTB	424 (6.0)	579 (7.5)	3565 (87)	4035 (83)	397 (8.1)
Hamstring	6388 (91)	6884 (89)	224 (5.4)	478 (9.9)	4382 (89)
QT/BQT	152 (2.2)	142 (1.8)	270 (6.6)	252 (5.2)	72 (1.5)
Other	74 (1.1)	88 (1.1)	59 (1.4)	87 (1.8)	79 (1.6)
Tibial fixation device					
Interference screw	6101 (87)	6688 (87)	3980 (97)	4594 (95)	4396 (89)
Suspension/cortical device	700 (9.9)	771 (10)	59 (1.4)	123 (2.5)	378 (7.7)
Other	237 (3.4)	234 (3.0)	79 (1.9)	135 (2.8)	156 (3.2)
Femoral fixation device					
Interference screw	118 (1.7)	10 (0.1)	3955 (96)	4369 (90)	341 (6.9)
Suspension/cortical device	6284 (89)	6902 (90)	10 (0.2)	117 (2.4)	4189 (85)
Other	636 (9.0)	781 (10)	153 (3.7)	366 (7.5)	400 (8.1)
Fixation device combination					
Interference screw ×2	96 (1.4)	0 (0)	3845 (93)	4211 (87)	315 (6.4)
Interference screw femur/suspension tibia	15 (0.2)	8 (0.1)	50 (1.2)	63 (1.3)	14 (0.3)
Suspension/cortical device ×2	587 (8.3)	619 (8.0)	9 (0.2)	13 (0.3)	312 (6.3)
Suspension femur/interference screw tibia	5523 (78)	6102 (79)	0 (0)	97 (2.0)	3771 (76)
Other	817 (12)	964 (13)	214 (5.2)	468 (9.6)	518 (11)
History of previous surgery on opposite knee[b]	379 (5.4)	467 (6.1)	266 (6.5)	434 (8.9)	431 (8.7)
History of previous surgery on same knee[b]	1043 (15)	1002 (13)	472 (11)	878 (18)	1145 (23)
Median time injury to surgery, y	0.54 [0.30-1.13]	0.64 [0.37-1.36]	0.63 [0.34-1.22]	0.58 [0.30-1.29]	0.68 [0.36-1.56

[a]Data are presented as n (%), mean ± SD, or median [IQR]. BPTB, bone–patellar tendon–bone autograft; ICRS, International Cartilage Regeneration & Joint Preservation Society; KOOS, Knee injury and Osteoarthritis Outcome Score; QOL, Quality of Life subscale; QT/BQT, quadriceps tendon autograft, with or without bone.

[b]Surgery performed before primary anterior cruciate ligament reconstruction and enrollment in the register.

surgical decision-making with patients undergoing primary ACLR.

To our knowledge, this is the first unsupervised learning analysis of an ACLR database. Unsupervised learning is a useful adjunct to clinical risk prediction efforts, as it may find patterns in data sets like the KLR without manual specification, which can be used to guide decision-making and prognostication.[7] Unsupervised learning models consider all variables in the data set that are categorized as predictors and are blind to the outcome for each patient (in this case, revision surgery). The algorithm is then tasked with finding common groups of patients within the data set, breaking them into different clusters. These clusters are arrived upon through complex analysis that is not explicitly directed by human instruction. Once the clusters have been identified, the outcome can be assessed

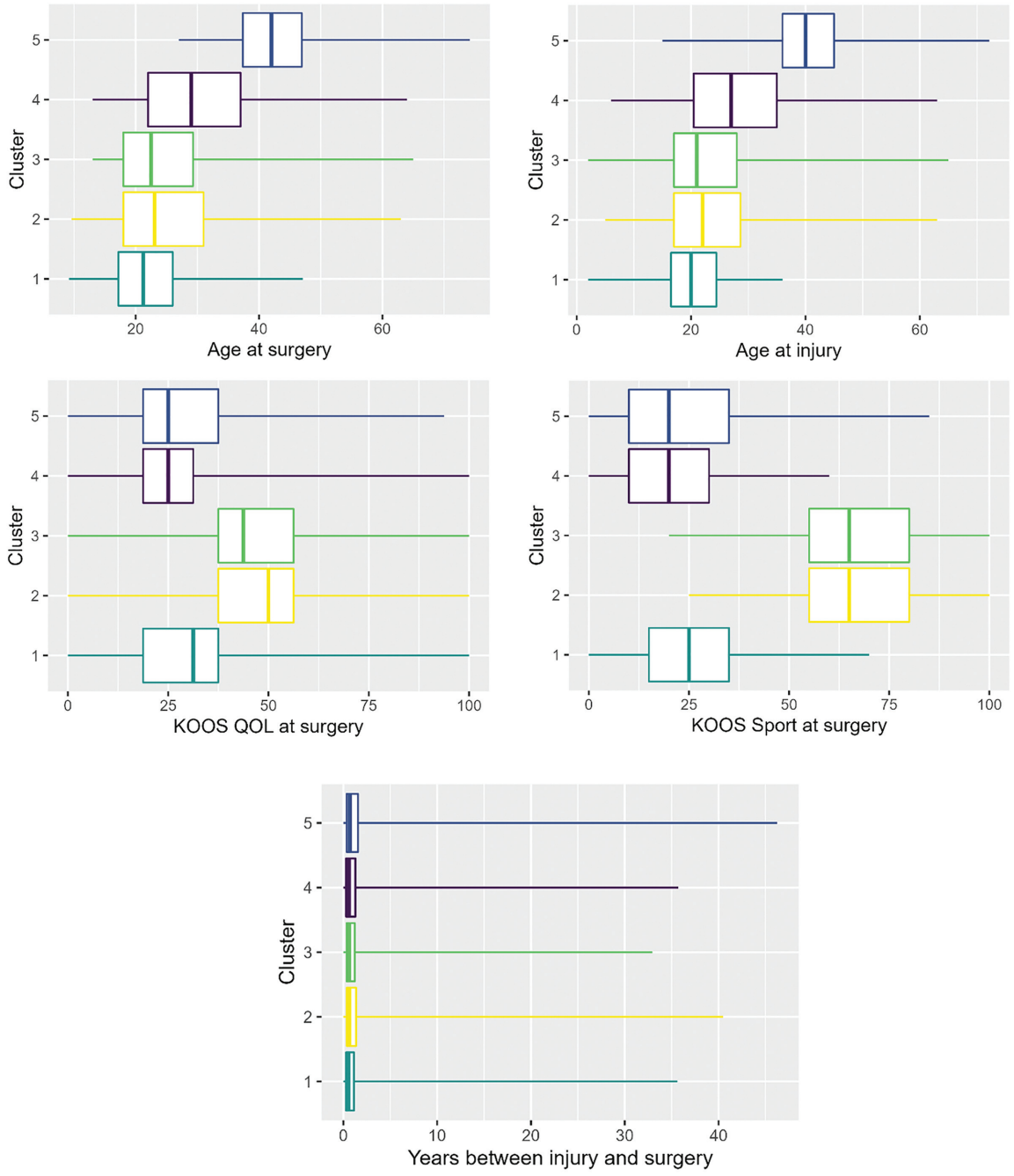

Figure 2. Continuous variable summaries by cluster. Box plots summarize the distributions of continuous predictor variables for each of the 5 patient subgroups identified with the unsupervised learning procedure. KOOS, Knee injury and Osteoarthritis Outcome Score; QOL, Quality of Life subscale; Sport, Sport and Recreation subscale.

in each group. In this study, revision rate was the primary outcome of interest, and this rate was different among each of the 5 patient groups. Similarly, the survival for each cluster was also distinct, allowing for a time-dependent cluster-based estimation of revision risk.

Accurately assigning a patient to 1 of the 5 clusters requires consideration of all variables included in the analysis. However, with so many predictor variables to consider, clinical interpretation and application of the patient subgroups can be challenging. To increase the clinical utility, the 5 patient clusters were reviewed by 7 subspecialty-trained orthopaedic sports medicine surgeons for defining characteristics. The recently developed SHAP analysis[18] was also applied to increase the explainability

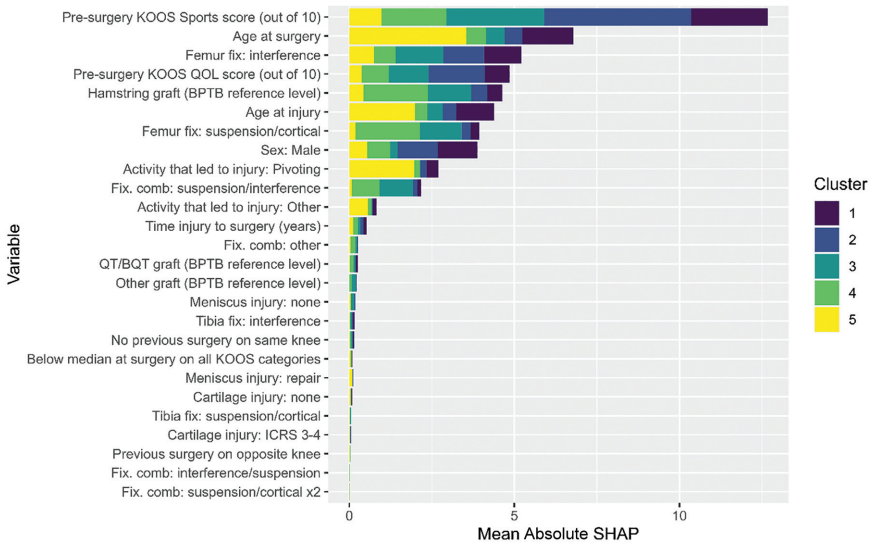

Figure 3. The plot shows mean absolute SHapley Additive exPlanations (SHAP) values by variable for all clusters. Colors in the plot show the contributions from observations assigned to each cluster. BPTB, bone–patellar tendon–bone autograft; comb, combined; fix., fixation; ICRS, International Cartilage Regeneration & Joint Preservation Society; KOOS, Knee injury and Osteoarthritis Outcome Score; QOL, Quality of Life subscale; QT/BQT, quadriceps tendon autograft, with or without bone; Sports, Sport and Recreation subscale.

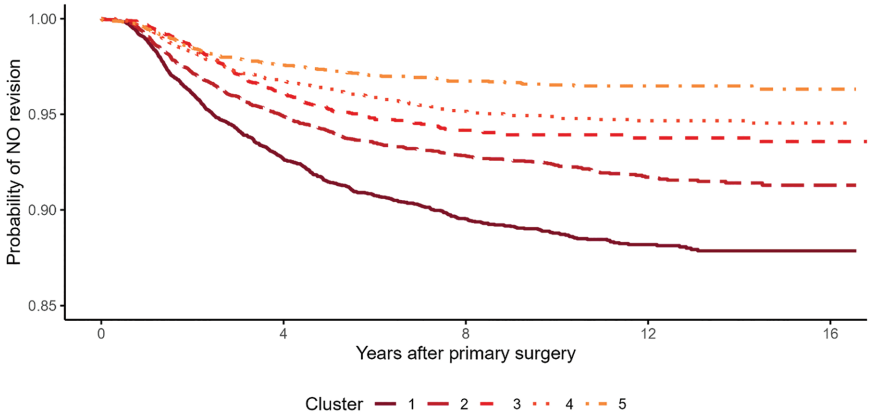

Figure 4. Kaplan-Meier survival curve for all 5 clusters.

of the model and decrease the black-box effect. The clusters were subsequently simplified into the following categories (Figure 5):

- Cluster 1: young patient with HT autograft and low baseline KOOS Sports score
- Cluster 2: patient with HT autograft and high baseline KOOS Sports score
- Cluster 3: patient with BPTB or QT autograft and high baseline KOOS Sports score
- Cluster 4: patient with BPTB or QT autograft and low baseline KOOS Sports score
- Cluster 5: older patient with HT autograft and low baseline KOOS Sports score

Based on the revision rates of each cluster, cluster 1 is considered high risk for revision surgery, cluster 2 is considered moderate risk for revision, and clusters 3 to 5 are considered low risk. While the overall revision rate in the KLR was 5.1%, nearly half (49%) of the patients fell into one of the low-risk categories (clusters 3-5) with a revision rate of 3.1% to 4.7%. On the other end of the spectrum, cluster 1 patients demonstrated a revision rate of nearly 10%.

Closer inspection of the highest risk cluster (cluster 1) reveals some interesting trends, including a higher proportion of patients with HT autograft, young age, female sex, and inferior baseline KOOS Sports scores. These factors become especially apparent when compared with clusters

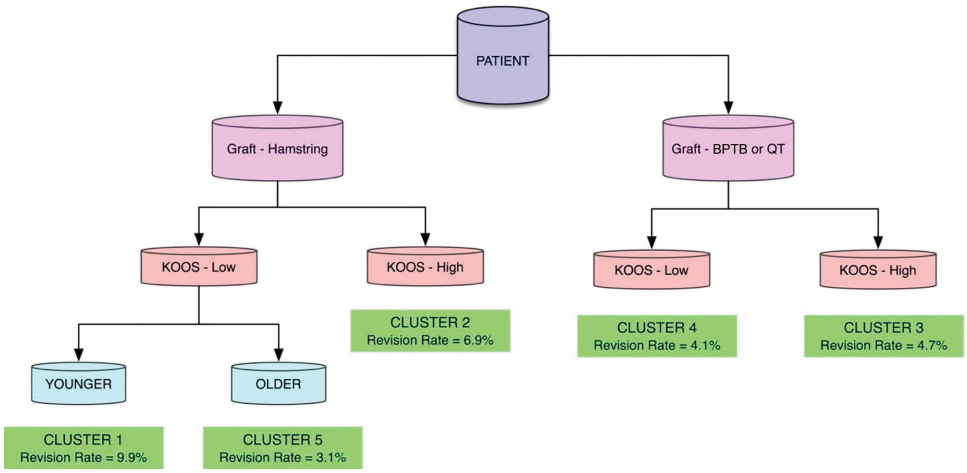

Figure 5. Tree diagram for approximate patient classification by cluster. BPTB, bone–patellar tendon–bone autograft; KOOS, Knee injury and Osteoarthritis Outcome Score (Sports subscale); QT, quadriceps tendon autograft, with or without bone.

2 and 5, which also consisted primarily of HT reconstruction but demonstrated revision rates closer to the mean. ACLR with HT autograft has previously been associated with higher revision surgery rates based on the NKLR.[28] Additionally, young age is a recognized risk factor for failure of ACLR.[14,40,41] Interestingly, the finding that young women receiving HT autograft (cluster 1) may be considered to have the highest risk for subsequent revision surgery is a novel finding. While it is generally accepted that female sex is associated with a higher risk of initial ACL injury,[37] it has not been found to be associated with higher ACLR revision rates.[1,3,14,24] Similarly, the authors are not aware of any literature associating preoperative KOOS Sports scores and subsequent revision risk. This unsupervised learning analysis suggests that because of the complex nature of the interactions between predictor variables, for some patients in certain circumstances, variables such as sex and preoperative patient-reported outcome measures may be important risk factors.

There are limitations to the present study. First, complete case data were available for less than half of the KLR, decreasing the number of patients available for analysis. Despite the missing data, however, >28,000 patients were included, which is sufficient for the purpose of unsupervised machine learning model development, and the inclusion of patients from 2 national databases increases generalizability. Another limitation is that the KLR is primarily composed of patients who received either HT, BPTB, or QT autograft. There were not enough patients receiving other graft choices such as allograft to have a meaningful effect on the clustering. These additional data would be useful in future studies to evaluate whether patients receiving allograft would form their own distinct clusters. The primary outcome measure of revision surgery represents another limitation, as some patients who experience graft failure or inferior clinical surgical outcome do not undergo subsequent revision surgery. Additionally, it is possible that an alternative unsupervised learning method may have yielded different results. There are several alternative approaches to unsupervised learning, such as principal component analysis, anomaly detection, and divisive hierarchical clustering, among others. However, the 3 unsupervised learning methods evaluated with this study represent the most common and appropriate for the data type and goals of this study. Finally, other factors potentially associated with failure of ACLR, such as pivot-shift grade, tibial slope, rehabilitation details, and surgical adjuncts such as lateral extra-articular tenodesis or anterolateral ligament reconstruction, were not captured in the KLR and were not considered in the analysis. The inclusion of these variables in future data collection may yield different clustering results.

There are also limitations to the clinical interpretability of this unsupervised analysis because of the complex determination of cluster characteristics. The simplified summary of each cluster may not consider certain relevant characteristics, which may lead to inaccurate risk estimation in the office setting. Considering, for example, that nearly 12% of the patients in cluster 4 received grafts other than BPTB or QT, suggests that there is more to the groupings than simply graft choice and KOOS Sports score. Similarly, continuous variables such as age and KOOS values can be challenging to interpret, for example, when defining what constitutes the cutoff point for high or low preoperative KOOS values. Finally, because of the nature of the study investigating revision rates of unsupervised learning–based clusters, the accuracy of the risk estimates was not externally validated. This represents the most important next step before prospective clinical application is recommended.

CONCLUSION

Unsupervised learning enabled the identification of 5 distinct KLR patient subgroups, and each grouping was associated with a unique ACLR revision rate. Patients can be approximately classified into 1 of the 5 clusters based on only 3 variables: age, graft choice (HT, BPTB, or QT

autograft), and preoperative KOOS Sports subscale score. If externally validated, the resulting groupings may enable quick risk stratification for future patients undergoing ACLR in the clinical setting. Patients in cluster 1 are considered high risk (9.9%) for subsequent revision ACLR, patients in cluster 2 medium risk (6.9%), and patients in clusters 3 to 5 low risk (3.1%-4.7%).

ORCID iDs

R. Kyle Martin https://orcid.org/0000-0001-9918-0264
Lars Engebretsen https://orcid.org/0000-0003-2294-921X

REFERENCES

1. Andernord D, Desai N, Björnsson H, Ylander M, Karlsson J, Samuelsson K. Patient predictors of early revision surgery after anterior cruciate ligament reconstruction: a cohort study of 16,930 patients with 2-year follow-up. *Am J Sports Med*. 2015;43(1):121-127.
2. Angelini F, Widera P, Mobasheri A, et al. Osteoarthritis endotype discovery via clustering of biochemical marker data. *Ann Rheum Dis*. 2022;81(5):666-675.
3. Capogna BM, Mahure SA, Mollon B, Duenes ML, Rokito AS. Young age, female gender, Caucasian race, and workers' compensation claim are risk factors for reoperation following arthroscopic ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2020;28(7):2213-2223.
4. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785-794.
5. Christensen JJ, Krych AJ, Engasser WM, Vanhees MK, Collins MS, Dahm DL. Lateral tibial posterior slope is increased in patients with early graft failure after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2015;43(10):2510-2514.
6. Devitt BM, Neri T, Fritsch BA. Combined anterolateral complex and anterior cruciate ligament injury: anatomy, biomechanics and management—state-of-the-art. *J ISAKOS*. 2023;8(1):37-46.
7. Eckhardt CM, Madjarova SJ, Williams RJ, et al. Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(2):376-381.
8. Getgood AMJ, Bryant DM, Litchfield R, et al. Lateral extra-articular tenodesis reduces failure of hamstring tendon autograft anterior cruciate ligament reconstruction: 2-year outcomes from the STABILITY study randomized clinical trial. *Am J Sports Med*. 2020;48(2):285-297.
9. Granan LP, Bahr R, Steindal K, Furnes O, Engebretsen L. Development of a national cruciate ligament surgery registry: the Norwegian National Knee Ligament Registry. *Am J Sports Med*. 2008;36(2):308-315.
10. Grassi A, Macchiarola L, Urrizola Barrientos F, et al. Steep posterior tibial slope, anterior tibial subluxation, deep posterior lateral femoral condyle, and meniscal deficiency are common findings in multiple anterior cruciate ligament failures: an MRI case-control study. *Am J Sports Med*. 2019;47(2):285-295.
11. Hamrin Senorski E, Svantesson E, Baldari A, et al. Factors that affect patient reported outcome after anterior cruciate ligament reconstruction—a systematic review of the Scandinavian knee ligament registers. *Br J Sports Med*. 2019;53(7):410-417.
12. Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. *Appl Stat*. 1979;28(1):100.
13. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. Springer New York; 2009.
14. Hettrich CM, Dunn WR, Reinke EK; MOON Group; Spindler KP. The rate of subsequent surgery and predictors after anterior cruciate ligament reconstruction: two- and 6-year follow-up results from a multicenter cohort. *Am J Sports Med*. 2013;41(7):1534-1540.
15. Kaarre J, Zsidai B, Narup E, et al. Scoping review on ACL surgery and registry data. *Curr Rev Musculoskelet Med*. 2022;15(5):385-393.
16. Kaeding CC, Pedroza AD, Reinke EK, Huston LJ; MOON Consortium; Spindler KP. Risk factors and predictors of subsequent ACL injury in either knee after ACL reconstruction: prospective analysis of 2488 primary ACL reconstructions from the MOON cohort. *Am J Sports Med*. 2015;43(7):1583-1590.
17. Ko S, Pareek A, Jo C, et al. Automated risk stratification of hip osteoarthritis development in patients with femoroacetabular impingement using an unsupervised clustering algorithm: a study from the Rochester Epidemiology Project. *Orthop J Sports Med*. 2021;9(11):232596712110506.
18. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Preprint published online November 24, 2017. *arxiv*. doi:10.48550/arXiv.1705.07874
19. Ma Y, Ao YF, Yu JK, Dai LH, Shao ZX. Failed anterior cruciate ligament reconstruction: analysis of factors leading to instability after primary surgery. *Chin Med J (Engl)*. 2013;126(2):280-285.
20. Martin RK, Wastvedt S, Pareek A, et al. Ceiling effect of the combined Norwegian and Danish Knee Ligament Registers limits anterior cruciate ligament reconstruction outcome prediction. *Am J Sports Med*. 2023;51(9):2324-2332.
21. Martin RK, Wastvedt S, Pareek A, et al. Machine learning algorithm to predict anterior cruciate ligament revision demonstrates external validity. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(2):368-375.
22. Martin RK, Wastvedt S, Pareek A, et al. Predicting anterior cruciate ligament reconstruction revision: a machine learning analysis utilizing the Norwegian Knee Ligament Register. *J Bone Joint Surg Am*. 2022;104(2):145-153.
23. Martin RK, Wastvedt S, Pareek A, et al. Predicting subjective failure of ACL reconstruction: a machine learning analysis of the Norwegian Knee Ligament Register and patient reported outcomes. *J ISAKOS*. 2022;7(3):1-9.
24. Mok AC, Fancher AJ, Vopat ML, et al. Sex-specific outcomes after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Orthop J Sports Med*. 2022;10(2):232596712210768.
25. MOON Knee Group; Spindler KP, Huston LJ, Chagin KM, et al. Ten-year outcomes and risk factors after anterior cruciate ligament reconstruction: a MOON longitudinal prospective cohort study. *Am J Sports Med*. 2018;46(4):815-825.
26. Müllner D. fastcluster: fast hierarchical, agglomerative clustering routines for *R* and *Python*. *J Stat Softw*. 2013;53(9):1-18.
27. Pang L, Li P, Li T, Li Y, Zhu J, Tang X. Arthroscopic anterior cruciate ligament repair versus autograft anterior cruciate ligament reconstruction: a meta-analysis of comparative studies. *Front Surg*. 2022;9:887522.
28. Persson A, Fjeldsgaard K, Gjertsen JE, et al. Increased risk of revision with hamstring tendon grafts compared with patellar tendon grafts after anterior cruciate ligament reconstruction: a study of 12,643 patients from the Norwegian Cruciate Ligament Registry, 2004-2012. *Am J Sports Med*. 2014;42(2):285-291.
29. Persson A, Kjellsen AB, Fjeldsgaard K, Engebretsen L, Espehaug B, Fevang JM. Registry data highlight increased revision rates for Endobutton/Biosure HA in ACL reconstruction with hamstring tendon autograft: a nationwide cohort study from the Norwegian Knee Ligament Registry, 2004-2013. *Am J Sports Med*. 2015;43(9):2182-2188.
30. Pruneski JA, Williams RJ, Nwachukwu BU, et al. The development and deployment of machine learning models. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(12):3917-3923.
31. Rahr-Wagner L, Lind M. The Danish Knee Ligament Reconstruction Registry. *Clin Epidemiol*. 2016;8:531-535.
32. Ramkumar PN, Pang M, Polisetty T, Helm JM, Karnuta JM. Meaningless applications and misguided methodologies in artificial intelligence–related orthopaedic research propagates hype over hope. *Arthroscopy*. 2022;38(9):2761-2766.
33. Riediger MD, Stride D, Coke SE, Kurz AZ, Duong A, Ayeni OR. ACL reconstruction with augmentation: a scoping review. *Curr Rev Musculoskelet Med*. 2019;12(2):166-172.

34. Salmon LJ, Heath E, Akrawi H, Roe JP, Linklater J, Pinczewski LA. 20-year outcomes of anterior cruciate ligament reconstruction with hamstring tendon autograft: the catastrophic effect of age and posterior tibial slope. *Am J Sports Med*. 2018;46(3):531-543.
35. Svantesson E, Hamrin Senorski E, Baldari A, et al. Factors associated with additional anterior cruciate ligament reconstruction and register comparison: a systematic review on the Scandinavian knee ligament registers. *Br J Sports Med*. 2019;53(7):418-425.
36. Szepannek G. clustMixType: user-friendly clustering of mixed-type data in R. *R J*. 2019;10(2):200.
37. The female ACL: why is it more prone to injury? *J Orthop*. 2016; 13(2):A1-A4.
38. University of Cincinnati. K-means cluster analysis. *UC Business Analytics R Programming Guide*. Accessed January 8, 2024. https://uc-r.github.io/kmeans_clustering
39. Vadhera AS, Knapik DM, Gursoy S, et al. Current concepts in anterior tibial closing wedge osteotomies for anterior cruciate ligament deficient knees. *Curr Rev Musculoskelet Med*. 2021;14(6):485-492.
40. Webster KE, Feller JA. Exploring the high reinjury rate in younger patients undergoing anterior cruciate ligament reconstruction. *Am J Sports Med*. 2016;44(11):2827-2832.
41. Webster KE, Feller JA, Leigh WB, Richmond AK. Younger patients are at increased risk for graft rupture and contralateral injury after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2014;42(3):641-647.