Journal of ISAKOS 7 (2022) 1–9

Contents lists available at ScienceDirect

Journal of ISAKOS

journal homepage: www.elsevier.com/locate/jisakos

Original Research

Predicting subjective failure of ACL reconstruction: a machine learning analysis of the Norwegian Knee Ligament Register and patient reported outcomes

R. Kyle Martin, MD, FRCSC [a,b,*], Solvejg Wastvedt, BA [c], Ayoosh Pareek, MD [d], Andreas Persson, MD, PhD [e,f,g], Håvard Visnes, MD, PhD [f], Anne Marie Fenstad, MS [f], Gilbert Moatshe, MD, PhD [g,h], Julian Wolfson, PhD [c], Lars Engebretsen, MD, PhD [g,h]

[a] Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, MN, USA
[b] Department of Orthopaedic Surgery, CentraCare, Saint Cloud, MN, USA
[c] Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA
[d] Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, USA
[e] Department of Orthopedic Surgery, Martina Hansens Hospital, Bærum, Norway
[f] Norwegian Knee Ligament Register, Haukeland University Hospital, Norway
[g] Oslo Sport Trauma Research Center, Norwegian School of Sports Science, Oslo, Norway
[h] Orthopaedic Clinic, Oslo University Hospital Ullevål, Oslo, Norway

ARTICLE INFO

Keywords:
Machine learning
Artificial intelligence
ACL reconstruction
Subjective outcome

ABSTRACT

Objectives: Accurate prediction of outcome following anterior cruciate ligament (ACL) reconstruction is challenging, and machine learning has the potential to improve our predictive capability. The purpose of this study was to determine if machine learning analysis of the Norwegian Knee Ligament Register (NKLR) can (1) identify the most important risk factors associated with subjective failure of ACL reconstruction and (2) develop a clinically meaningful calculator for predicting the probability of subjective failure following ACL reconstruction.
Methods: Machine learning analysis was performed on the NKLR. All patients with 2-year follow-up data were included. The primary outcome was the probability of subjective failure 2 years following primary surgery, defined as a Knee Injury and Osteoarthritis Outcome Score (KOOS) Quality of Life (QoL) subscale score of <44. Data were split randomly into training (75%) and test (25%) sets. Four models intended for this type of data were tested: Lasso logistic regression, random forest, generalized additive model (GAM), and gradient boosted regression (GBM). These four models represent a range of approaches to statistical details like variable selection and model complexity. Model performance was assessed by calculating calibration and area under the curve (AUC).
Results: Of the 20,818 patients who met the inclusion criteria, 11,630 (56%) completed the 2-year follow-up KOOS QoL questionnaire. Of those with complete KOOS data, 22% reported subjective failure. The lasso logistic regression, GBM, and GAM all demonstrated AUC in the moderate range (0.67–0.68), with the GAM performing best (0.68; 95% CI 0.64–0.71). Lasso logistic regression, GBM, and the GAM were well-calibrated, while the random forest showed evidence of mis-calibration. The GAM was selected to create an in-clinic calculator to predict subjective failure risk at a patient-specific level (https://swastvedt.shinyapps.io/calculator_koosqol/).
Conclusion: Machine learning analysis of the NKLR can predict subjective failure risk following ACL reconstruction with fair accuracy. This algorithm supports the creation of an easy-to-use in-clinic calculator for point-of-care risk stratification. Clinicians can use this calculator to estimate subjective failure risk at a patient-specific level when discussing outcome expectations preoperatively.
Level of evidence: Level-III Retrospective review of a prospective national register.

* Corresponding author. Department of Orthopedic Surgery, University of Minnesota, 2512 South 7th Street, Suite R200 Minneapolis, MN 55455, USA. Tel.: +1 612 273 1177.
E-mail address: rkylemmartin@gmail.com (R.K. Martin).

https://doi.org/10.1016/j.jisako.2021.12.005
Received 23 September 2021; Accepted 30 December 2021
Available online 11 January 2022

What are the new findings?

- Machine learning analysis can be performed on a national knee ligament register to predict the risk of subjective failure following anterior cruciate ligament reconstruction
- An in-clinic calculator can guide clinical discussion and expectations at a patient-specific level
- Variables for predicting subjective failure following anterior cruciate ligament reconstruction are patient-related and non-modifiable by the surgeon

Introduction

Anterior cruciate ligament (ACL) reconstruction is a common orthopaedic procedure aimed at restoring function and stability following injury. Literature regarding the surgical outcome is often reported in relation to patient-reported outcome measures (PROM), and several risk factors for a poor outcome have been suggested [1–4]. Currently, however, the ability to use these predictors at the time of surgery to accurately predict which patients are at risk of experiencing a poor outcome is poor [1].

Recently, there has been an increased focus on the use of artificial intelligence and machine learning to improve predictive capability within several fields of medicine, including orthopaedic surgery [5–9]. These advanced statistical techniques utilise computer algorithms to model complex interactions between variables and may lead to improved capacity to predict the outcome. The "advanced" nature of these techniques is derived from the fact that the interactions can be more complex than with traditional statistics. Machine learning analyses can consider all possible interactions between variables in a database and determine the relationships to the desired outcome measure. The factors important for predicting outcomes can then be identified and used to develop the predictive algorithm. Often, minimal explicit and direct human computer programming is required, and the resulting algorithms can be used to prospectively predict the patient-specific outcome.

The Norwegian Knee Ligament Register (NKLR) has been prospectively collecting demographic, injury, surgical, and outcome data since 2004. It now includes over 25,000 patients who have undergone ACL reconstruction with high compliance across the country [10]. Several studies that have improved our understanding of ACL injuries have been based on the NKLR [11–14], and machine learning analysis allows deeper evaluation of factors associated with outcome [9]. There are currently no machine learning models to predict subjective outcomes following primary ACL reconstruction, and the development of such a tool could impact clinical practice by informing shared decision-making and outcome expectations.

The purpose of this study was to use machine learning analysis of the NKLR to (1) identify the most important risk factors associated with subjective failure of primary ACL reconstruction and (2) develop a clinically meaningful model for predicting subjective failure of primary ACL reconstruction. Subjective failure was defined as a Knee Injury and Osteoarthritis Outcome Score (KOOS) Quality of Life (QoL) subscale score of <44. This endpoint has been clinically validated as a marker of failure following ACL reconstruction [11]. The hypothesis was that machine learning analysis would facilitate accurate prediction of subjective failure for a patient undergoing primary ACL reconstruction.

Materials and methods

This manuscript was written in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement [15].

Data source

The NKLR is a nationwide register aiming to collect all reconstructive surgery on cruciate ligament injuries in Norway. Reporting has been mandatory since 2017, and the compliance of reporting to the register was 86% in 2017 to 2018 [10]. The patients are registered with their personal social security number, which allows them to be followed in case of later surgery independent of service provider. Patient-specific and intraoperative data are submitted to the NKLR by the surgeons (through an article or web-based form directly after surgery). The patients are to report KOOS preoperatively and at 2, 5, and 10 years of follow-up.

Ethics

Informed consent is obtained from all patients at time of enrolment in the NKLR. Based on this consent, the Norwegian Data Inspectorate provides permission for the NKLR to collect, analyse, and publish on health data. The registration of data was performed confidentially and according to Norwegian and European Union data protection rules, with all data de-identified prior to retrieval from the NKLR. The Regional Ethics Committee has previously determined that it is not necessary to obtain further ethical approval for Norwegian register-based studies [16].

Data preparation

This level-III retrospective review of a prospective national register included all patients contained within the NKLR with primary ACL reconstruction surgery dates from January 2004 through December 2018. Those with values for graft choice recorded as "direct suture," "other," or missing were excluded. Patients with other ligamentous injuries at the time of primary surgery or <2 years of follow-up were also excluded. Variables considered in the analysis are presented in Table 1. Variables were re-coded or newly defined for the following: years between injury and primary surgery; cartilage injury identified at surgery (none, ICRS 1–2, ICRS 3–4); meniscus injury identified at surgery (yes/no); graft choice (patellar tendon autograft, hamstring tendon autograft, other); fixation choice (interference screw, suspension/cortical device, other); and height and weight variables that combined data from the patient- and surgeon-reported variables. A predictor indicating if a patient was below the median score in all five KOOS categories at the time of primary surgery was also created, and predictors for KOOS QoL and Sports measures were scaled to a score out of 10.

Model creation

The primary outcome was the probability of subjective failure at 2 years following primary ACL reconstruction, as defined as a KOOS QoL score of <44. Cleaned data were randomly split into training (75%) and test (25%) sets that were used to fit and evaluate the models, respectively. The program R (version: 3.6.1, R Core Team 2019) was used to fit four machine learning models to the training data: lasso logistic regression, random forest, gradient boosted regression model (GBM), and generalized additive model (GAM) [17]. These four models are among the most commonly used for machine learning classification tasks and offer a range of approaches in terms of variable selection, optimisation technique, and complexity. Lasso logistic regression is a parametric, penalised regression model that selects a subset of variables for inclusion [18]. The random forest is a tree-based, nonparametric method [19]. GBMs are also nonparametric, meaning that they do not require pre-specification of a model structure and iteratively improve the model fit using all available variables [20,21]. GAM allow for machine-selected nonlinear relationships among a pre-specified group of variables [22]. Further description of each of the four machine learning models can be found in Appendix A.

An L1-regularised logistic regression model ("lasso logistic regression," package *glmnet*; lambda value selected via cross-validation) was

Table 1
Characteristics of patients.

Variable[a]	All N = 20,818	Complete 2-year Outcome Data N = 11,630
Follow-up time or time to revision	7.3 (3.9)	7.9 (3.6)
KOOS QOL <44 at 2 years	2,556 (22%)	2,556 (22%)
Missing	9,188	0
Age at surgery	28 (10)	29 (11)
Age at injury	26 (10)	27 (11)
Missing	1072	544
Sex		
Male	11,669 (56%)	5,836 (50%)
Female	9,149 (44%)	5,794 (50%)
Pre-surgery BMI	25.0 (3.7)	24.8 (3.7)
Missing	7,244	4,365
Pre-surgery KOOS QOL score (out of 10)	3.50 (1.83)	3.52 (1.83)
Missing	4,022	2,008
Pre-surgery KOOS Sports score (out of 10)	4.33 (2.71)	4.37 (2.69)
Missing	4,162	2,087
Below median on all pre-surgery KOOS	3,285 (19%)	1,806 (19%)
Missing	3,893	1,942
Activity that led to injury		
Non-pivoting	4,109 (25%)	2,392 (26%)
Pivoting	12,007 (75%)	6,716 (74%)
Other/Unknown	0 (0%)	0 (0%)
Missing	4,702	2,522
Meniscus injury	10,942 (53%)	5,927 (51%)
Cartilage injury		
ICRS 1-2	3,625 (17%)	2,016 (17%)
ICRS 3-4	993 (4.8%)	577 (5.0%)
None	16,200 (78%)	9,037 (78%)
Graft choice		
BPTB autograft	7,334 (35%)	3,782 (33%)
Hamstring autograft	13,197 (63%)	7,740 (67%)
Other	287 (1.4%)	108 (0.9%)
Tibia fixation device		
Interference screw	17,893 (89%)	9,905 (88%)
Suspension/cortical device	2,073 (10%)	1,245 (11%)
Other	152 (0.8%)	88 (0.8%)
Missing	700	392
Femur fixation device		
Interference screw	6,325 (31%)	3,314 (29%)
Suspension/cortical device	11,629 (57%)	6,613 (58%)
Other	2,484 (12%)	1,491 (13%)
Missing	380	212
Fixation device combination		
Interference screw x2	6,028 (30%)	3,163 (28%)
Interference/suspension	51 (0.3%)	17 (0.2%)
Suspension/cortical device x2	1,646 (8.2%)	1,011 (9.0%)
Suspension/interference	9,635 (48%)	5,410 (48%)
Other	2,634 (13%)	1,577 (14%)
Missing	824	452
Injured side		
Right	10,613 (51%)	5,871 (50%)
Left	10,205 (49%)	5,759 (50%)
Previous surgery on opposite knee	1,526 (7.3%)	786 (6.8%)
Previous surgery on same knee	3,784 (18%)	2,220 (19%)
Time injury to surgery (years)	1.71 (3.36)	1.81 (3.63)
Missing	1,076	546
Systemic Antibiotic Prophylaxis	20,669 (100%)	11,534 (99%)
Missing	51	34

[a] Statistics presented: Mean (SD); n (%).

applied to select variables for each outcome, and those with non-zero coefficients were retained (Fig. 1). Random forests (function *randomForest* from package *randomForest*) were trained for each outcome with minimum node size 5, 10 variables tried per split, 500 trees, and the full set of predictors (hyperparameters selected via cross-validation). GAMs (function *gam* from package *mgcv*) were trained with those variables selected in the lasso for the respective outcomes, using smooth terms for all continuous variables selected. Finally, GBMs (function *gbm* from package *gbm*) were trained using a shrinkage parameter of 0.01, minimum node size of 10, maximum tree depth of 3, 1000 trees, and the full set of predictors (hyperparameters selected via cross-validation). All four models were restricted to patients with complete data for the predictors used (Table 2a and Table 2b).

Model evaluation

Model performance was evaluated by calculating predicted probabilities of subjective failure at 2 years of follow-up for the hold-out test data using the trained models. Model calibration was assessed using the Hosmer–Lemeshow statistic (function *hoslem.test* in package *ResourceSelection*) [23]. Calibration refers to the accuracy of the predicted probabilities, comparing expected to actual observed outcomes. This statistic sums average misclassification in each predicted risk quintile and converts the sum into a chi-squared statistic. Larger calibration statistics correspond to smaller p values, and statistical significance means that the null hypothesis of perfect calibration is rejected. The area under the curve (AUC) was also calculated for each model along with confidence intervals for the AUC using bootstrap resampling (functions *auc* and *ci.auc* from package *pROC*).

Missing data

An inverse probability-weighted analysis was conducted to assess whether patients with complete follow-up KOOS QoL score data were fundamentally different from those with missing outcome data based on observed characteristics. Inverse probability weighting assigns each observation a weight based on the inverse of the probability of a patient with similar observed characteristics being present in the dataset. In this case, patients with combinations of predictor variables that are rare in the complete outcome dataset receive high weights. Conversely, patients with common predictor variables are down-weighted to adjust for their over-representation. The result of the weighting is a population that mimics what would have occurred if all patients were to have complete outcome data. The same models are then built on this weighted population and compared to the unweighted analysis. If the weighted models show substantively different results, this indicates that there may be fundamental differences between patients with complete and incomplete outcome data. If there is no substantive difference, this indicates that removing patients with incomplete outcome data does not jeopardise the results.

To assess the effect of excluding patients with missing predictor values from the models, the same four models were trained using multiple imputations to fill in missing values in the training data (function *mice* from package *mice*). As with the weighted models, if there is no substantive difference when using imputation, this indicates that removing patients with incomplete predictor data does not adversely affect the results.

Sources of funding

This study was funded by a Norwegian Centennial Chair seed grant. Funding supported the machine learning analysis and interpretation. The funding agencies had no direct role in the investigation.

Results

Data characteristics

Table 1 describes the characteristics of the registered population at the time of primary surgery and the variables included for analysis. After data cleaning, 20,818 patients met the inclusion criteria (Fig. 2). Of these patients, 11,630 (56%) had complete 2-year follow-up KOOS QoL data. Subjective failure (KOOS QoL score <44) occurred in 2,556 (22%) of the patients with complete outcome data. The population was approximately evenly split between male and female, with an average age (and standard deviation) of 29 ± 11 years at the time of primary surgery.

Fig. 1. Variable Importance. The four plots show relative feature importance in each of the machine learning models. The vertical axis is a variable importance score, which differs depending on the model. For the lasso logistic regression and GAM, the vertical axis is the absolute value of the variable coefficient (effect size). For the random forest and GBM, the scale is the decrease in model error rate if the variable were to be removed from the model. The highlighted bars indicate variables that were selected using the lasso and included in the final model used for the in-clinic calculator. GAM, generalized additive model; GBM, gradient boosted regression model.

Table 2a
Lasso logistic regression/generalised additive model complete/incomplete case comparison.

Variable*	Incomplete N = 14,810	Complete N = 6,008	Total N = 20,818	P-value**
Years: surgery to data current date (2020-01-12)	9.1 (4.2)	7.6 (2.5)	8.6 (3.9)	<0.001
KOOS QoL <44 at 2 years	1,270 (23%)	1,286 (21%)	2,556 (22%)	0.13
Missing	9,188	0	9,188	
Age at injury	26 (10)	27 (11)	26 (10)	0.006
Missing	1,072	0	1,072	
Pre-surgery BMI	25.1 (3.7)	24.8 (3.7)	25.0 (3.7)	<0.001
Missing	7,244	0	7,244	
Pre-surgery KOOS QoL score (out of 10)	3.48 (1.83)	3.55 (1.85)	3.50 (1.83)	0.016
Missing	4,022	0	4,022	
Pre-surgery KOOS Sports score (out of 10)	4.29 (2.71)	4.42 (2.70)	4.33 (2.71)	0.002
Missing	4,162	0	4,162	
Below median on all pre-surgery KOOS scores	2,199 (20%)	1,086 (18%)	3,285 (19%)	0.001
Missing	3,893	0	3,893	
Activity that led to injury				<0.001
Non-pivoting	2,784 (19%)	1,325 (22%)	4,109 (20%)	
Pivoting	8,433 (59%)	3,574 (59%)	12,007 (59%)	
Other	3,122 (22%)	1,109 (18%)	4,231 (21%)	
Missing	471	0	471	
Cartilage injury				0.015
ICRS 1-2	2,648 (18%)	977 (16%)	3,625 (17%)	
ICRS 3-4	692 (4.7%)	301 (5.0%)	993 (4.8%)	
None	11,470 (77%)	4,730 (79%)	16,200 (78%)	
Previous surgery on same knee	2,824 (19%)	960 (16%)	3,784 (18%)	<0.001

*Statistics presented: Mean (SD); n (%).
**Statistical tests performed: t-test, chi-square test.

To assess the impact of restricting the analysis to patients with complete KOOS QoL score data, covariate distributions between patients with complete outcomes and the full dataset were compared (Table 1). Covariate distributions between the complete cases for each model and the full dataset were also compared (Table 2a and Table 2b). Due to the large sample sizes, some comparisons produce p values below the significance threshold: those with complete data were newer to the register, had their surgeries at higher-volume hospitals, and were more likely to be female. However, these differences were in general small and of limited clinical significance. An inverse-probability-weighted analysis and an analysis imputing missing covariate data was also performed. Neither alternative analysis showed meaningfully different results from the complete case models (Table 3 and Table 4).

Model performance

The lasso logistic regression, GBM, and GAM all demonstrated AUC in the moderate range (0.67–0.68), with the GAM performing best at 0.68 (95% CI 0.64–0.71). Lasso logistic regression, gradient boosted regression, and the GAM were well-calibrated, and the random forest showed evidence of miscalibration (Table 5).

Factors predicting outcome

The most important predictors of subjective failure at 2 years following primary surgery in the lasso logistic regression model in order were below the median on all KOOS subscale scores at the time of surgery, cartilage injury at the time of surgery, activity leading to injury, previous surgery on the same knee, KOOS Sports and QoL scores at surgery, body mass index (BMI) at surgery, and age at injury. In the random forest, predictors in the top third by variable importance score also included age at surgery, graft choice, years between injury and surgery, fixation device combination, and femur fixation. The GAM and GBM produced similar rankings of feature importance (Fig. 1). The lasso logistic regression and GAM measure feature importance by effect size associated with the variable. The other models use the difference in model error rate where the feature is to be removed.

Risk-prediction calculator

The GAM was selected to create an easy-to-use in-clinic calculator to predict the risk of a patient experiencing a subjective failure at 2 years of follow-up after primary ACL reconstruction (https://swastvedt.shinyapps.io/calculator_koosqol/and Fig. 3). The GAM was chosen out of the four models because it combines performance with simplicity, using fewer predictor variables than the similarly performing GBM. Whereas the overall risk of failure in the register was 22%, this calculator can quantify the risk at a patient-specific level (Video 1).

Discussion

The most important finding of this study was that machine learning analysis of a knee ligament register allows the creation of a validated algorithm to predict a patient's risk of experiencing subjective failure of ACL reconstruction with fair accuracy. Additionally, despite having 20 possible prognostic variables contained within the NKLR, the algorithm required only eight factors for the prediction of 2-year risk. Variables required for risk prediction include age at injury, pre-operative KOOS subscale scores, activity leading to an ACL injury, concomitant cartilage injury, history of previous surgery on the same knee, and pre-operative BMI. Using this algorithm, we developed an in-clinic calculator was developed that can estimate the risk of subjective failure.

This represents the first machine learning model for predicting the subjective outcome of ACL reconstruction at a patient-specific level. Estimation of revision risk has been developed previously [9], and together, these two prediction tools can be used to guide the discussion surrounding the surgical options and realistic outcome goals at a patient-specific level. For the clinician, this represents a valuable adjunct to the assessment of patients with ACL deficiency desiring surgical management.

Similar to the previous study of revision risk [9], four models were used to analyse the NKLR and create algorithms predicting the risk of subjective failure after ACL reconstruction. Discrimination (AUC) was similar for the prediction of subjective outcome evaluated with this study

Table 2b
Random forest/gradient boosted regression complete/incomplete case comparison.

Variable*	Incomplete N = 15,040	Complete N = 5,778	Total N = 20,818	P-value**
Years: surgery to data current date (2020-01-12)	9.0 (4.2)	7.5 (2.5)	8.6 (3.9)	<0.001
KOOS QoL <44 at 2 years	1,329 (23%)	1,227 (21%)	2,556 (22%)	0.058
Missing	9,188	0	9,188	
Age at surgery	28 (10)	28 (11)	28 (10)	0.19
Age at injury	26 (10)	27 (11)	26 (10)	0.006
Missing	1,072	0	1,072	
Sex				<0.001
Male	8,890 (59%)	2,779 (48%)	11,669 (56%)	
Female	6,150 (41%)	2,999 (52%)	9,149 (44%)	
Pre-surgery BMI	25.1 (3.7)	24.8 (3.7)	25.0 (3.7)	<0.001
Missing	7,244	0	7,244	
Pre-surgery KOOS QoL score (out of 10)	3.48 (1.82)	3.56 (1.85)	3.50 (1.83)	0.006
Missing	4,022	0	4,022	
Pre-surgery KOOS Sports score (out of 10)	4.28 (2.71)	4.43 (2.71)	4.33 (2.71)	0.001
Missing	4,162	0	4,162	
Below median on all pre-surgery KOOS scores	2,244 (20%)	1,041 (18%)	3,285 (19%)	0.001
Missing	3,893	0	3,893	
Activity that led to injury				<0.001
Non-pivoting	2,846 (20%)	1,263 (22%)	4,109 (20%)	
Pivoting	8,564 (59%)	3,443 (60%)	12,007 (59%)	
Other	3,159 (22%)	1,072 (19%)	4,231 (21%)	
Missing	471	0	471	
Meniscus injury	7,908 (53%)	3,034 (53%)	10,942 (53%)	0.940
Cartilage injury				0.031
ICRS 1-2	2,683 (18%)	942 (16%)	3,625 (17%)	
ICRS 3-4	710 (4.7%)	283 (4.9%)	993 (4.8%)	
None	11,647 (77%)	4,553 (79%)	16,200 (78%)	
Graft choice				<0.001
BPTB autograft	5,454 (36%)	1,880 (33%)	7,334 (35%)	
Hamstring autograft	9,358 (62%)	3,839 (66%)	13,197 (63%)	
Other	228 (1.5%)	59 (1.0%)	287 (1.4%)	
Tibia fixation device				<0.001
Interference screw	12,494 (87%)	5,399 (93%)	17,893 (89%)	
Suspension/cortical device	1,700 (12%)	373 (6.5%)	2,073 (10%)	
Other	146 (1.0%)	6 (0.1%)	152 (0.8%)	
Missing	700	0	700	
Femur fixation device				<0.001
Interference screw	4,671 (32%)	1,654 (29%)	6,325 (31%)	
Suspension/cortical device	7,817 (53%)	3,812 (66%)	11,629 (57%)	
Other	2,172 (15%)	312 (5.4%)	2,484 (12%)	
Missing	380	0	380	
Fixation device combination				<0.001
Interference screw x2	4,391 (31%)	1,637 (28%)	6,028 (30%)	
Interference/suspension	40 (0.3%)	11 (0.2%)	51 (0.3%)	
Suspension/interference	6,177 (43%)	3,458 (60%)	9,635 (48%)	
Suspension/cortical device x2	1,292 (9.1%)	354 (6.1%)	1,646 (8.2%)	
Other	2,316 (16%)	318 (5.5%)	2,634 (13%)	
Missing	824	0	824	
Injured side				0.18
Right	7,711 (51%)	2,902 (50%)	10,613 (51%)	
Left	7,329 (49%)	2,876 (50%)	10,205 (49%)	
Previous surgery on opposite knee	1,157 (7.7%)	369 (6.4%)	1,526 (7.3%)	0.001
Previous surgery on same knee	2,856 (19%)	928 (16%)	3,784 (18%)	<0.001
Time injury to surgery (years)	1.72 (3.31)	1.68 (3.50)	1.71 (3.36)	0.42
Missing	1,076	0	1,076	
Systemic antibiotic prophylaxis	14,897 (99%)	5,772 (100%)	20,669 (100%)	<0.001
Missing	51	0	51	

*Statistics presented: Mean (SD); n (%).
**Statistical tests performed: t-test, chi-square test.

(0.65–0.68) compared with the revision risk prediction (0.67–0.69), and all models except the random forest demonstrated appropriate calibration. It is interesting to note that while the factors used for predicting revision risk included modifiable surgical details (graft choice, femoral fixation device, and length of time between injury and surgery) [9], the prediction of subjective failure appears to be static. That is, most of the variables used to predict subjective outcome are based on patient-driven factors that are present prior to surgery (age, concomitant chondral injury, history of previous surgery, and activity leading to injury) and may not be amenable to optimisation.

Of the variables identified by the algorithm as important for predicting the risk of subjective failure, the only truly modifiable factor was patient BMI at the time of surgery. The extent to which efforts to decrease BMI prior to surgery may influence the risk of poor functional outcomes is unclear and raises an interesting area for future study. Similarly, given the impact of the pre-surgical KOOS scores on the eventual post-operative subjective outcome, efforts to optimise functional outcomes prior to surgery through physiotherapy or cognitive behavioural coaching may also be beneficial. Regarding variable relative importance (Fig. 1), BMI was the least important variable in the GAM, while KOOS QoL had the

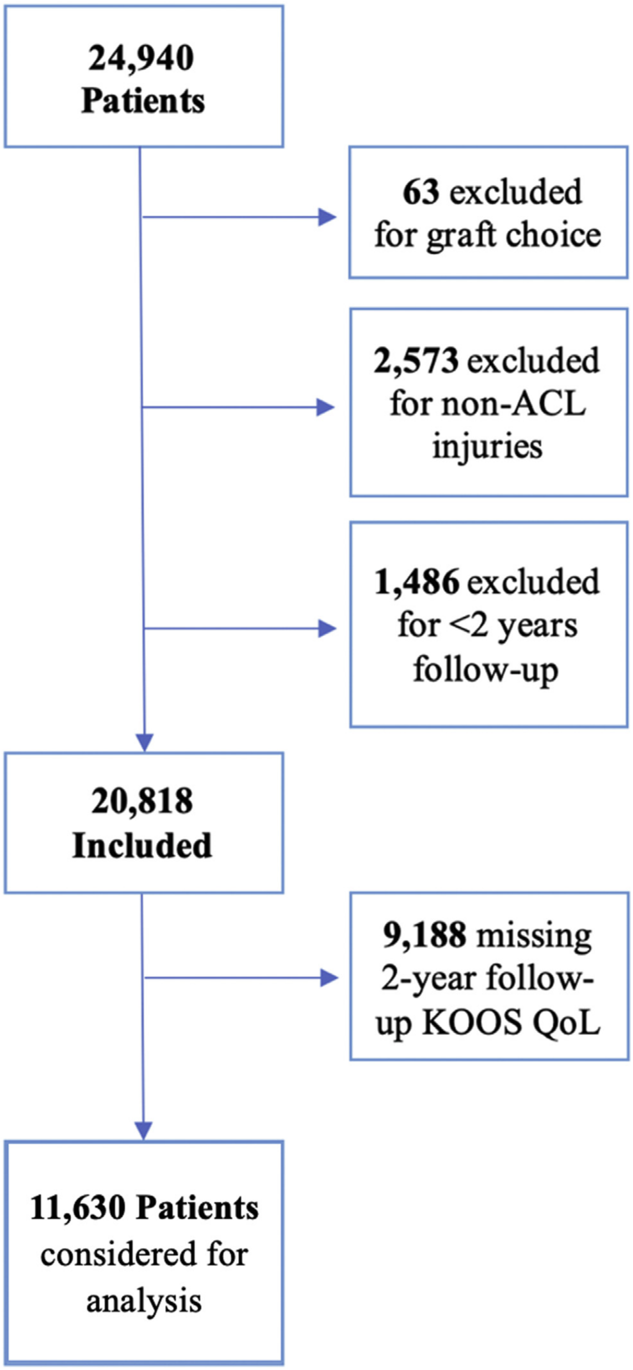

Fig. 2. Patient inclusion flowchart.

highest relative importance. It should be noted, however, that the present study was designed to predict subjective failure risk and does not represent a comparative study to determine the effect of risk factor modification.

Table 3
Inverse probability weighted model performance.

Model	AUC	Weighted calibration statistic	Unweighted calibration	Calibration p-value (unweighted)
Logistic regression (lasso)	0.67	0.020	4.33	0.228
Random forest	0.65	0.054	24.65	<0.001
Gradient boosted regression	0.67	0.017	6.65	0.084
Generalised additive model	0.67	0.019	7.45	0.059

Table 4
Multiple imputation model performance.

Model	AUC	Calibration statistic	Calibration p-value
Logistic regression (lasso)	0.68	2.54	0.468
Random forest	0.67	21.30	0.006
Gradient boosted regression	0.69	1.62	0.656
Generalised additive model	0.68	2.46	0.482

Table 5
Model performance.

Model	AUC	AUC confidence interval	Calibration statistic	Calibration p-value
Logistic regression (lasso)	0.67	(0.64, 0.71)	4.57	0.206
Random forest	0.65	(0.62, 0.69)	26.83	<0.001
Gradient boosted regression	0.68	(0.64, 0.71)	4.03	0.258
Generalised additive model	0.68	(0.64, 0.71)	4.74	0.192

The primary outcome of the subjective failure of ACL reconstruction was defined as a KOOS QoL score of <44. Other possible measures of subjective outcome include, but are not limited to, the minimal clinically important difference (MCID) or Patient Acceptable Symptom State and may use other assessment tools such as a visual analogue scale or the International Knee Documentation Committee questionnaire. While there are advantages and disadvantages to each measure of functional outcome, KOOS QoL was selected for this study since it has previously been validated as a measure of inadequate knee function associated with prospective ACL reconstructed graft failure and represents a poor outcome after surgery [11]. Further, the prevalence of a KOOS QoL score of <44 was 22%, which suggests that the outcome is clinically relevant across the population.

SCAN ME

Fig. 3. QR Code for 2-year subjective failure point-of-care risk stratification at the time of primary ACL reconstruction. ACL, anterior cruciate ligament.

Limitations

The most significant limitation of this study is the missing follow-up KOOS data. Whereas overall compliance with the NKLR is 86% for tracking revision surgery following ACL reconstruction [10], follow-up KOOS scores were only available for 56% of patients at 2 years. While we cannot determine that data were missing completely at random, the inverse probability weighted analysis does provide evidence that the group of patients with complete KOOS follow-up data was not meaningfully different from the group with missing data based on recorded characteristics. Complete PROM follow-up represents a challenge for all national knee ligament registers since patients are typically young and reside throughout the country. Patient compliance is typically higher when research teams and surgeons are actively engaged in the data collection [2], which is not feasible for a large national register like the NKLR. Second, although several machine learning models were evaluated, a model that not considered may have performed better. A third limitation is the fact that the analysis was limited to the variables contained within the register. Although these variables included several known risk factors for ACL reconstruction failure, there are also many other factors that may be associated with the poor outcome that are not recorded in the NKLR. Examples include radiographic variables such as tibial slope and coronal alignment [24–28], physical examination and rehabilitation details [29,30], and surgical technique factors such as tunnel position [31] and graft size [32,33]. Further, while meniscus and chondral injuries were recorded, the surgical treatments employed at the time of surgery were not included as variables and may represent a source of exclusion bias.

There are also limitations regarding the clinical utility of this analysis. The machine learning models use several variables for outcome prediction. To account for this, the GAM was selected for the in-clinic calculator due to its simplicity, requiring fewer input variables without a significant decrease in performance versus the more complex models. Further, this study included patients from a single national register, and the results may not be applicable to other populations. External validity could be established through the evaluation of model performance when applied to patients from other registers or databases. While an advantage of registers like the NKLR is the generalisability and real-world applicability [34], the inclusion of all Norwegian surgeons in the data collection may result in wide variability. Finally, while the machine learning algorithm was well calibrated, the AUC was fair. The accuracy of the model may be improved if radiographic, rehabilitation, and/or other variables not included in the model were assessed.

Conclusion

Machine learning analysis of a national knee ligament register can predict subjective failure risk following ACL reconstruction with few factors required for outcome prediction and moderate accuracy overall. This algorithm supports the creation of an easy-to-use in-clinic calculator for point-of-care risk stratification. Clinicians can use this calculator to estimate subjective failure risk at a patient-specific level when discussing outcome expectations pre-operatively.

Institutional review board

Approval not required as consent was obtained by all patients at time of enrolment in the national knee ligament register.

Sources of funding

This study was funded by a Norwegian Centennial Chair Seed Grant.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jisako.2021.12.005.

References

[1] An VV, Scholes C, Mhaskar VA, et al. Limitations in predicting outcome following primary ACL reconstruction with single-bundle hamstring autograft — a systematic review. Knee 2017;24:170–8. https://doi.org/10.1016/j.knee.2016.10.006.

[2] Marx Robert G, Wolfe Isabel A, Turner Brooke E, et al., MOON Knee Group. MOON's strategy for obtaining over eighty percent follow-up at 10 years following ACL reconstruction. J Bone Joint Surg Am 2021 Aug 23. https://doi.org/10.2106/JBJS.21.00166.

[3] Nguyen JT, Wasserstein D, Reinke EK, et al. Does the chronicity of anterior cruciate ligament ruptures influence patient-reported outcomes before surgery? Am J Sports Med 2017;45:541–9. https://doi.org/10.1177/0363546516669344.

[4] Brophy RH, Huston LJ, Briskin I, et al. Articular cartilage and meniscus predictors of patient-reported outcomes 10 years after anterior cruciate ligament reconstruction: a multicenter cohort study. Am J Sports Med 2021;49:2878–88. https://doi.org/10.1177/03635465211028247.

[5] Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. Front Bioeng Biotechnol 2018;6:75. https://doi.org/10.3389/fbioe.2018.00075.

[6] Van Eetvelde H, Mendonça LD, Ley C, et al. Machine learning methods in sport injury prediction and prevention: a systematic review. J Exp Orthop 2021;8:27. https://doi.org/10.1186/s40634-021-00346-x.

[7] Schock J, Truhn D, Abrar DB, et al. Automated analysis of alignment in long-leg radiographs by using a fully automated support system based on artificial intelligence. Radiol Artif Intell 2021;3:e200198. https://doi.org/10.1148/ryai.2020200198.

[8] Krogue JD, Cheng KV, Hwang KM, et al. Automatic hip fracture identification and functional subclassification with deep learning. Radiol Artif Intell 2020;2:e190023. https://doi.org/10.1148/ryai.2020190023.

[9] 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 2021;104(2):145–53. https://doi.org/10.2106/JBJS.21.00113.

[10] Annual report. Bergen, Norway: Norwegian national advisory unit on arthroplasty and hip fractures. 2021.

[11] Granan L-P, Baste V, Engebretsen L, et al. Associations between inadequate knee function detected by KOOS and prospective graft failure in an anterior cruciate ligament-reconstructed knee. Knee Surg Sports Traumatol Arthrosc 2015;23:1135–40. https://doi.org/10.1007/s00167-014-2925-5.

[12] LaPrade CM, Dornan GJ, Granan L-P, et al. Outcomes after anterior cruciate ligament reconstruction using the Norwegian knee ligament registry of 4691 patients: how does meniscal repair or resection affect short-term outcomes? Am J Sports Med 2015;43:1591–7. https://doi.org/10.1177/0363546515577364.

[13] Persson A, Fjeldsgaard K, Gjertsen J-E, 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;42:285–91. https://doi.org/10.1177/0363546513511419.

[14] Persson A, Kjellsen AB, Fjeldsgaard K, et al. 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;43:2182–8. https://doi.org/10.1177/0363546515584757.

[15] Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.

[16] Granan L-P, Bahr R, Steindal K, et al. Development of a national cruciate ligament surgery registry: the Norwegian National Knee Ligament Registry. Am J Sports Med 2008;36:308–15. https://doi.org/10.1177/0363546507308939.

[17] R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/. [Accessed 19 May 2020].

[18] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Software 2010;33(1):1–22. pmid: 20808728.

[19] Breiman L. Random forests. Mach Learn 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

[20] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29. https://doi.org/10.1214/aos/1013203451.

[21] Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal 2002;38: 367–78. https://doi.org/10.1016/S0167-9473(01)00065-2.
[22] Wood SN. Generalized additive models: an introduction with R. 2nd ed. Chapman and Hall/CRC; 2017. https://doi.org/10.1201/9781315370279.
[23] Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Commun Stat Theor Methods 1980;9:1043–69. https://doi.org/10.1080/03610928008827941.
[24] Bernholt DL, Dornan GJ, DePhillipo NN, et al. High-grade posterolateral tibial plateau impaction fractures in the setting of a primary anterior cruciate ligament tear are correlated with an increased preoperative pivot shift and inferior postoperative outcomes after anterior cruciate ligament reconstruction. Am J Sports Med 2020;48:2185–94. https://doi.org/10.1177/0363546520932912.
[25] Bayer S, Meredith SJ, Wilson KW, et al. Knee morphological risk factors for anterior cruciate ligament injury: a systematic review. J Bone Jt Surg 2020;102:703–18. https://doi.org/10.2106/JBJS.19.00535.
[26] Li Y, Hong L, Feng H, et al. Posterior tibial slope influences static anterior tibial translation in anterior cruciate ligament reconstruction: a minimum 2-year follow-up study. Am J Sports Med 2014;42:927–33. https://doi.org/10.1177/0363546514521770.
[27] Bernhardson AS, Aman ZS, Dornan GJ, et al. Tibial slope and its effect on force in anterior cruciate ligament grafts: anterior cruciate ligament force increases linearly as posterior tibial slope increases. Am J Sports Med 2019;47:296–302. https://doi.org/10.1177/0363546518820302.
[28] Mehl J, Otto A, Kia C, et al. Osseous valgus alignment and posteromedial ligament complex deficiency lead to increased ACL graft forces. Knee Surg Sports Traumatol Arthrosc Off J ESSKA 2020;28:1119–29. https://doi.org/10.1007/s00167-019-05770-2.
[29] Roe C, Jacobs C, Kline P, et al. Correlations of single-leg performance tests to patient-reported outcomes after primary anterior cruciate ligament reconstruction. Clin J Sport Med Off J Can Acad Sport Med 2021;31:e265–70. https://doi.org/10.1097/JSM.0000000000000780.
[30] Grindem H, Snyder-Mackler L, Moksnes H, et al. Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study. Br J Sports Med 2016;50:804–8. https://doi.org/10.1136/bjsports-2016-096031.
[31] Liu A, Sun M, Ma C, et al. Clinical outcomes of transtibial versus anteromedial drilling techniques to prepare the femoral tunnel during anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthrosc Off J ESSKA 2017;25:2751–9. https://doi.org/10.1007/s00167-015-3672-y.
[32] Conte EJ, Hyatt AE, Gatt CJ, et al. Hamstring autograft size can be predicted and is a potential risk factor for anterior cruciate ligament reconstruction failure. Arthrosc J Arthrosc Relat Surg 2014;30:882–90. https://doi.org/10.1016/j.arthro.2014.03.028. Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc.
[33] Magnussen RA, Lawrence JTR, West RL, et al. Graft size and patient age are predictors of early revision after anterior cruciate ligament reconstruction with hamstring autograft. Arthrosc J Arthrosc Relat Surg 2012;28:526–31. https://doi.org/10.1016/j.arthro.2011.11.024. Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc.
[34] Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The Evidence-Based Medicine Working Group. JAMA 1996;275(7):554–8. https://doi.org/10.1001/jama.275.7.554.