Ceiling Effect of the Combined Norwegian and Danish Knee Ligament Registers Limits Anterior Cruciate Ligament Reconstruction Outcome Prediction

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: Clinical tools based on machine learning analysis now exist for outcome prediction after primary anterior cruciate ligament reconstruction (ACLR). Relying partly on data volume, the general principle is that more data may lead to improved model accuracy.

Purpose/Hypothesis: The purpose was to apply machine learning to a combined data set from the Norwegian and Danish knee ligament registers (NKLR and DKRR, respectively), with the aim of producing an algorithm that can predict revision surgery with improved accuracy relative to a previously published model developed using only the NKLR. The hypothesis was that the additional patient data would result in an algorithm that is more accurate.

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

Methods: Machine learning analysis was performed on combined data from the NKLR and DKRR. The primary outcome was the probability of revision ACLR within 1, 2, and 5 years. Data were split randomly into training sets (75%) and test sets (25%). There were 4 machine learning models examined: Cox lasso, random survival forest, gradient boosting, and super learner. Concordance and calibration were calculated for all 4 models.

Results: The data set included 62,955 patients in which 5% underwent a revision surgical procedure with a mean follow-up of 7.6 ± 4.5 years. The 3 nonparametric models (random survival forest, gradient boosting, and super learner) performed best, demonstrating moderate concordance (0.67 [95% CI, 0.64-0.70]), and were well calibrated at 1 and 2 years. Model performance was similar to that of the previously published model (NKLR-only model: concordance, 0.67-0.69; well calibrated).

Conclusion: Machine learning analysis of the combined NKLR and DKRR enabled prediction of the revision ACLR risk with moderate accuracy. However, the resulting algorithms were less user-friendly and did not demonstrate superior accuracy in comparison with the previously developed model based on patients from the NKLR alone, despite the analysis of nearly 63,000 patients. This ceiling effect suggests that simply adding more patients to current national knee ligament registers is unlikely to improve predictive capability and may prompt future changes to increase variable inclusion.

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

There has been an increased focus on outcome prediction using machine learning in orthopaedic surgery recently.[22] The primary goal of these early clinical predictive models was to enable patient-specific risk estimation to guide management discussions and expectations. Clinical tools based on machine learning analysis now exist for outcome prediction after anterior cruciate ligament reconstruction (ACLR) including revision surgery[30] and inferior patient-reported outcomes.[31] These models were developed from analyses of the Norwegian Knee Ligament Register (NKLR), and the revision prediction model has also been externally validated using the Danish Knee Ligament Reconstruction Registry (DKRR).[32]

The accurate prediction of outcomes after ACLR holds value for both the patient and surgeon. However, with so many interrelated variables contributing to the risk of a poor outcome, it can be challenging for a clinician to quantify that risk for the patient in the office, regardless of his or her experience level. Machine learning represents a novel approach to this problem and can facilitate patient-specific risk quantification through the analysis and interpretation of large volumes of data in ways that were previously unrealistic.

The American Journal of Sports Medicine
2023;51(9):2324–2332
DOI: 10.1177/03635465231177905

Relying partly on data volume to develop predictive algorithms, the general principle is that more data may lead to improved model accuracy. The rationale for this is that more data present more opportunity for the models to "learn" the association between predictors and outcomes. Therefore, the purpose of this study was to apply machine learning to a combined NKLR and DKRR data set, with the aim of predicting revision surgery with improved accuracy relative to a previously published model.[30] The original NKLR model was developed using machine learning analysis of approximately 25,000 patients, whereas the combined NKLR and DKRR data set includes nearly 63,000 patients. The hypothesis was that the additional patient data would result in a more accurate prediction of the revision ACLR risk.

METHODS

This article was written in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement.[6] The statement includes a 22-item checklist, with the goal of improving the transparency of prediction model studies through full and clear reporting.

Ethics

All patients provided informed consent for the NKLR, and the Norwegian Data Protection Authority granted permission for the register to collect, analyze, and publish health data. Data registration was performed confidentially according to European Union data protection rules, with all data de-identified before retrieval. The regional ethics committee stated that it was not necessary to obtain further ethical approval.[11] Similarly, the DKRR obtained informed consent at the time of enrollment, and patient data were de-identified before retrieval with no further ethical approval required.

Data Compilation

Patients who underwent primary ACLR between June 2004 and December 2020 were included. Patients missing data for graft choice, those with a graft choice recorded as "direct suture," and those missing data for the indicator of revision surgery were excluded. Variables considered for analysis are shown in Table 1.

A predictor indicating if a patient scored below the median score in the respective registry for all preoperative Knee injury and Osteoarthritis Outcome Score (KOOS) subscales was created. Patients who underwent revision ACLR before the follow-up time were considered to have experienced the event.

Machine Learning Modeling

NKLR and DKRR data were combined and then randomly split into training (75%) and test (25%) sets used to fit and evaluate the models, respectively. The primary outcome was the probability of revision ACLR within 1, 2, and 5 years. R (Version 4.1.11; R Core Team) was used to fit machine learning models that were adapted for censored time-to-event data. "Censoring" refers to the fact that patients who have not yet reached a given follow-up time point may still contribute partial information toward that endpoint. For example, a patient who has been revision-free for 4 years has not yet reached the 5-year selected outcome time point, but his or her revision-free time can still be considered in the analysis for the 5-year revision risk. Censoring also accounts for the fact that patients who have not yet undergone a revision procedure may ultimately undergo revision surgery in the future.

Four models intended for this type of data were used: Cox lasso, random survival forest, gradient boosting, and super learner. These models represent a range of approaches regarding the flexibility of model fitting and the number of variables incorporated. Cox lasso is a semiparametric, penalized regression model that selects a subset of the most important predictor variables for inclusion.[41] Random survival forest is a nonparametric model, meaning that it does not require prespecification of a model structure, and uses all available variables; this model is an adaptation of the widely used tree-based random forest method for censored data.[17] Gradient boosting is also a tree-based, nonparametric model adapted for censored data; this model iteratively updates to improve the fit using all available variables.[9] Super learner is an

*Address correspondence to R. Kyle Martin, MD, Department of Orthopedic Surgery, University of Minnesota, 2512 South 7th Street, Suite R200, Minneapolis, MN 55455, USA (email: rkylemmartin@gmail.com) (Twitter: @RKMartin6).
†Department of Orthopedic Surgery, University of Minnesota, Minneapolis, Minnesota, USA.
‡Department of Orthopedics, CentraCare, St Cloud, Minnesota, USA.
§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.
¶Department of Orthopaedic Surgery, Oslo University Hospital Ullevål, Oslo, Norway.
#Oslo Sports Trauma Research Center, Norwegian School of Sport Sciences, Oslo, Norway.
**Norwegian Knee Ligament Register, Haukeland University Hospital, Bergen, Norway.
††Aarhus University Hospital, Aarhus, Denmark.
Submitted October 3, 2022; accepted April 11, 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 consulting fees from Smith & Nephew and support for education from Gemini/Arthrex. 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 South-Eastern Norway and royalties from Arthrex and Smith & Nephew. 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 and Surgical Characteristics[a]

	Value (N = 62,955)
Revision	3205 (5)
Follow-up time or time to revision, mean ± SD, y	7.6 ± 4.5
Age at surgery, median (IQR), y	26 (20-36)
Age at injury, median (IQR), y	24 (18-34)
Missing, n	1870
Sex	
Male	36,509 (58)
Female	26,446 (42)
Preoperative KOOS–Quality of Life score (of 10), mean ± SD	3.63 ± 1.80
Missing, n	29,512
Preoperative KOOS-Sport score (of 10), mean ± SD	4.12 ± 2.69
Missing, n	29,708
All preoperative KOOS scores below median	6372 (19)
Missing, n	29,323
Activity that led to injury	
Nonpivoting	20,391 (32)
Pivoting	35,851 (57)
Other	6162 (10)
Missing, n	551
Meniscal injury	
Injury without repair	20,328 (32)
Injury with repair	10,554 (17)
None	32,061 (51)
Missing, n	12
Cartilage injury	
Grade 1-2	8766 (14)
Grade 3-4	3223 (5)
None	50,878 (81)
Missing, n	88
Graft choice	
Bone–patellar tendon–bone	15,639 (25)
Hamstring tendon	43,518 (69)
Quadriceps tendon	2520 (4)
Other	1278 (2)
Tibial fixation device	
Interference screw	55,792 (89)
Suspension/cortical device	3643 (6)
Other	2356 (4)
Missing, n	1164
Femoral fixation device	
Interference screw	16,434 (26)
Suspension/cortical device	39,742 (63)
Other	4822 (8)
Missing, n	1957
Fixation device combination	
2 interference screws	15,865 (25)
Interference screw (femur) and suspension device (tibia)	236 (0.4)
2 suspension/cortical devices	2994 (5)
Suspension device (femur) and interference screw (tibia)	34,895 (55)
Other	6529 (10)
Missing, n	2436
Injured side	
Right	32,147 (51)
Left	30,807 (49)
Missing, n	1
Previous surgery on opposite knee	4839 (8)
Missing, n	2946
Previous surgery on same knee	10,312 (16)
Missing, n	673
Time from injury to surgery, median (IQR), y	0.61 (0.33-1.32)
Missing, n	2083
Registry	
DKRR	34,554 (55)
NKLR	28,401 (45)

[a]Data are reported as n (%) unless otherwise indicated. DKRR, Danish Knee Ligament Reconstruction Registry; IQR, interquartile range; KOOS, Knee injury and Osteoarthritis Outcome Score; NKLR, Norwegian Knee Ligament Register.

"ensemble" model that creates a weighted average of other machine learning techniques, combining them into 1 overall fit and thereby providing an even more flexible approach[46]; the super learner model combines the random survival forest and gradient boosting models. Further descriptions of each model are included in the Appendix (available in the online version of this article).

Variables with nonzero coefficients were selected using the L1-regularized Cox model ("Cox lasso"; package *glmnet*; lambda value selected via cross-validation), retaining the variables shown in the top panel of Figure 1.

For the random survival forest, gradient boosting, and super learner models, a grid search method was used to determine hyperparameters (package *MachineShop*). This method compares all combinations of a range of possible hyperparameter values and chooses the optimal combination based on a performance metric: in this case, the C-index, described below. The random survival forest model (package *randomForestSRC*) was trained using the following hyperparameters: node size of 300, 10 variables per split, and 500 trees. The gradient boosting model (package *gbm*) was trained using a shrinkage parameter of 0.01, interaction depth of 3, minimum node size of 100, and 1,000 trees. The super learner model was trained using the same hyperparameter values for the random survival forest and gradient boosting models and utilizing the *SuperModel* function (package *MachineShop*) to determine, via cross-validation, the optimal weighting of the component models. All 4 models were restricted to patients with complete data for the predictors used (see Table 1 and Missing Data section).

Model Evaluation

Model performance was evaluated by calculating survival probabilities with each model for observations in the hold-out test set. Concordance and calibration were then

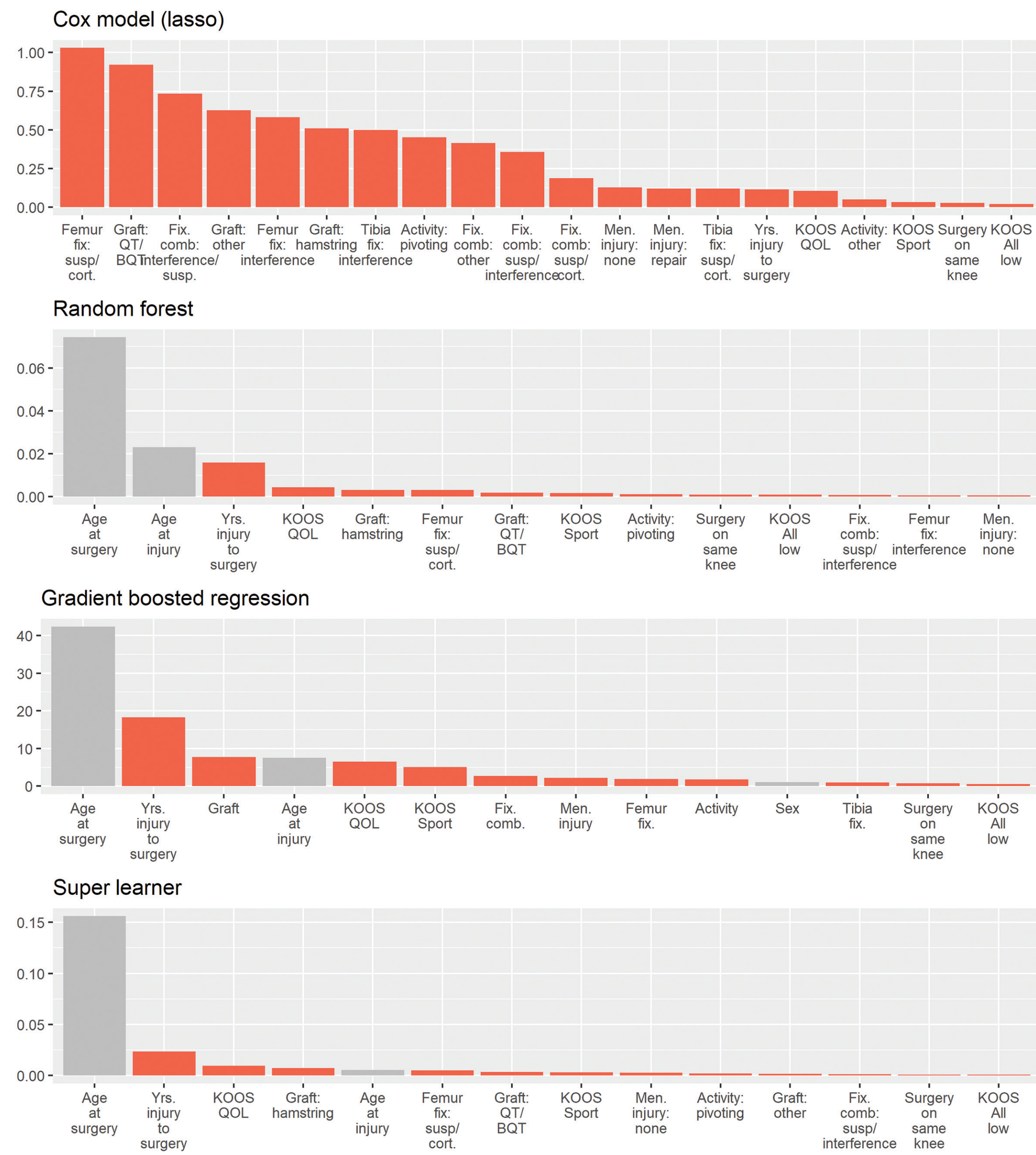

Figure 1. The 4 plots show the relative feature importance in each of the machine learning models. The highlighted bars indicate features selected for the Cox lasso model. The random survival forest, gradient boosting, and super learner plots show features in the top half according to the importance score for readability. Feature importance is measured on a different scale for each model, and thus, only rankings of features, rather than scores, should be compared among the models. The Cox lasso model measures feature importance by absolute effect size. The random survival forest and super learner models use permutation-based importance, which measures the relative change in model performance after randomly permuting values of the given feature. The gradient boosting model uses the difference in the error rate if the feature was to be removed, normalized to a total sum of 100. BQT, quadriceps tendon autograft with bone; comb, combined; cort, cortical; fix, fixation; KOOS, Knee injury and Osteoarthritis Outcome Score; Men, meniscus; QOL, Quality of Life; QT, quadriceps tendon autograft; Sport, Sport and Recreation Subscale; susp, suspension; Yrs, years.

calculated using methods adapted for censored data. Concordance was determined using the Harrell C-index at 1-, 2-, and 5-year follow-up. The C-index is a generalization of the common area under the receiver operating characteristic curve metric. As with the area under the curve, it ranges from 0 to 1, with 1 indicating perfect concordance. The C-index measures the proportion of pairs of observations in which predicted rankings of survival

TABLE 2
Model Performance With Complete Case Training Data

	Concordance (95% CI)	Calibration Statistic	Calibration P Value
1 y			
Cox lasso	0.59 (0.56-0.61)	7.19	.066
Random survival forest	0.67 (0.64-0.69)	5.54	.136
Gradient boosting	0.67 (0.65-0.70)	7.48	.058
Super learner	0.67 (0.65-0.69)	8.67	.034
2 y			
Cox lasso	0.58 (0.56-0.61)	8.17	.043
Random survival forest	0.67 (0.64-0.69)	6.42	.093
Gradient boosting	0.67 (0.64-0.69)	4.53	.210
Super learner	0.67 (0.64-0.69)	4.10	.250
5 y			
Cox lasso	0.58 (0.56-0.61)	11.37	.010
Random survival forest	0.67 (0.65-0.69)	9.27	.026
Gradient boosting	0.67 (0.64-0.69)	11.07	.011
Super learner	0.67 (0.64-0.69)	11.82	.008

probabilities correspond to actual rankings.[14] Furthermore, calculation of the C-index is limited to pairs of patients with sufficient information to determine the true ordering: either both patients must have known times to revision or one has undergone revision surgery and the other is censored (no revision yet, with the time since surgery at least as long as the other patient's time to revision). For example, a concordance of 0.80 would mean that for a random pair of patients, risk estimates match the true ordering of times to revision approximately 80% of the time.

Calibration is a measure of the accuracy of predicted probabilities that compares expected outcomes with actual outcomes. We calculated calibration using a version of the Hosmer-Lemeshow test that accounts for censoring.[47] This statistic sums the average misclassification in each predicted risk quintile and converts the sum into a chi-square statistic. Larger values of calibration indicate worse accuracy and correspond to smaller P values, with statistical significance indicating a rejection of the null hypothesis of perfect calibration.

Missing Data

Models were trained using observations from the training set with complete data on all variables. The models were then evaluated using observations from the test set with complete data on all variables needed for a given model. To assess the effect of restricting data to complete cases, we re-trained and re-evaluated the models using multiple imputation. 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. Multiple imputation by chained equations was conducted with 5 imputations on training and test data (package *mice*). The variables with nonzero coefficients for the Cox lasso model with complete cases were used to refit the model with each imputed training data set, averaging predictions over the 5 imputations. The random survival forest, gradient boosting, and super learner models were similarly refit. A bootstrap procedure was used to compare the calibration between the complete case and multiply imputed models.

RESULTS

Patient Data

Table 1 details the characteristics of the population at the time of surgery and shows all variables included for the analysis. After data cleaning, the combined registries' population consisted of 62,955 patients, with 55% from the DKRR and 45% from the NKLR. The primary outcome, revision surgery, occurred in 5% of patients with a mean follow-up of 7.6 ± 4.5 years. The population was 58% male, with a median age at the time of the primary injury of 24 years (interquartile range, 18-34 years) and a median age at the time of surgery of 26 years (interquartile range, 20-36 years).

Model Performance

The 3 nonparametric models—random survival forest, gradient boosting, and super learner—had moderate concordance (0.67) at all follow-up times, with 95% CIs ranging from 0.64-0.69 to 0.65-0.70 (Table 2).

The Cox lasso model performed more poorly, with a concordance of 0.58-0.59. The Cox lasso model showed moderate evidence of miscalibration (P = .01-.043) at 2 and 5 years. The other 3 models were better calibrated, with the exception of the super learner model at 1 year (P = .034) and 5 years (P = .008). The random survival forest and gradient boosting models also demonstrated moderate evidence of miscalibration at 5 years. Model performance for the original NKLR algorithm demonstrated similar concordance (0.67-0.69) and calibration.[30]

Model performance with imputation is presented in Table 3.

TABLE 3
Model Performance With Multiply Imputed Training Data

	Concordance (95% CI)	Calibration Statistic	Calibration *P* Value
1 y			
Cox lasso	0.59 (0.56-0.61)	8.35	.039
Random survival forest	0.66 (0.64-0.69)	4.17	.244
Gradient boosting	0.68 (0.65-0.70)	7.57	.056
Super learner	0.67 (0.65-0.70)	7.99	.046
2 y			
Cox lasso	0.59 (0.56-0.61)	8.81	.032
Random survival forest	0.67 (0.65-0.70)	8.96	.030
Gradient boosting	0.67 (0.65-0.70)	8.98	.030
Super learner	0.67 (0.65-0.70)	8.34	.039
5 y			
Cox lasso	0.58 (0.56-0.61)	8.30	.040
Random survival forest	0.67 (0.65-0.70)	8.95	.030
Gradient boosting	0.67 (0.65-0.69)	11.53	.009
Super learner	0.67 (0.65-0.69)	14.05	.003

Multiply imputed data did not show notable differences from the complete case analysis. The concordance 95% CIs were nearly identical in all cases. Observed calibration ratios from all 4 models were compared with the bootstrap distribution, and all the observed ratios were within the 95% CI. This suggests that there was no significant difference in calibration between the complete case and multiply imputed models.

Factors Predicting Outcome

The most important factors predicting revision surgery, according to the 3 best-performing models, included age at the time of surgery and injury, years between injury and surgery, graft choice, and preoperative KOOS–Quality of Life and KOOS-Sport and Recreation scores. Variables in approximately the top half by feature importance in the random survival forest, gradient boosting, and super learner models are shown in the bottom 3 panels of Figure 1. Variables with nonzero coefficients in the Cox lasso model are shown in the top panel of Figure 1. The Cox lasso model quantifies feature importance in terms of the absolute value of the associated effect size. The gradient boosting model uses the difference in the error rate if the feature was to be removed. The random survival forest and super learner models use permutation-based variable importance, measuring the relative change in model performance after randomly permuting values of the given variable.

DISCUSSION

Machine learning analysis of the combined NKLR and DKRR enabled the prediction of revision surgery after primary ACLR with moderate accuracy. The most important finding of this study, however, was that this analysis of nearly 63,000 patients yielded similar prediction accuracy as a previous study of approximately 25,000 patients.[30,32] This suggests that the ceiling effect of the registries has been reached, and the addition of more patients is unlikely to appreciably improve prediction accuracy. This information can be used to further the evolution of national ACLR registries regarding variable inclusion and data collection.

Machine learning applications within orthopaedic surgery have been increasing at an exponential rate in recent years.[22] These advanced statistical techniques can evaluate large data sets and recognize complex interactions between variables.[28] “Learning” from these interactions, machine learning models can create algorithms capable of predicting outcomes for patients, often at a level of accuracy superior to expert humans.[3,8,37,39,40,45,50]

Similar to how humans learn through repetition and experience, machine learning algorithms often require large volumes of data to optimize model accuracy. Data volume, however, is not the only factor that contributes to the accuracy of a model. Just as important is the quality of the data. If the data set used for model creation does not consider variables that are associated with the outcome of interest, then the full potential of the model may not be reached. Poor data quality can also manifest as substantial missing or incomplete data, which affects the ability of the model to learn and form accurate associations between predictors and outcomes. Techniques such as imputation can address some data quality inadequacies, but there are limits to what may be overcome.[2]

After nearly 20 years of data collection by the NKLR and DKRR, data quantity is superb, with satisfactory completeness and data accuracy.[7,34-36] However, the present study suggests that for an improvement in our ability to predict outcomes based on registry data, an evolution in the variables collected is required. This represents a significant challenge, as the balance between optimal variable collection and surgeon compliance is a delicate one.[11,29] Data collection must be streamlined to avoid survey fatigue, and the addition of variables to the registry must be carefully considered, weighing the added value against the additional onus on the surgeon, which may affect compliance.

Factors that may improve prediction accuracy and could be considered for supplementation in national registers include data regarding radiographic findings,[4,12,13,18,23,33,48] adjunctive surgical procedures, clinical examination results, rehabilitation details,[38] and alternative patient-reported outcome measures such as psychological factors.[5] Preoperative and postoperative radiographic indices could be manually captured, for example, tibial slope and coronal alignment, or included as raw image files that could then be evaluated using computer vision machine learning techniques.[21] The recording of additional surgical details such as graft diameter/size, ligament augmentation, lateral extra-articular tenodesis, or anterolateral ligament reconstruction may also be of value, given their recent association with outcomes.[1,10,15,16,24,26,42,52] Clinical examination and rehabilitation information such as preoperative knee laxity grade[25,43] could be obtained via third-party sources such as physical therapists or via natural language processing of patient chart notes.[49] Finally, the KOOS may not be the most appropriate patient-reported outcome tool for the patient population, and an alternative measurement of patient function, such as the baseline Marx activity level, could be considered for inclusion in registries moving forward.[19,27]

It is worth mentioning that an algorithm for the prediction of revision surgery after primary ACLR will likely never achieve perfect or even excellent performance in the traditional sense. There are 2 main reasons for this. First, reinjury events leading to revision surgery may occur randomly, such as after a slip on ice or a collision on the playing field. That randomness, combined with the variance related to uncollected variables, limits the predictive capability of ACLR failure models. The second reason is that the outcome, in this case, revision surgery, is itself imperfect; that is, not everyone who has experienced a failure will undergo revision surgery. This is a major consideration for most clinical predictive models, which are limited by the chosen endpoint. Although discrimination has often been interpreted as performance >0.9 being excellent, >0.8 being good, >0.7 being fair, and <0.7 being poor,[44] most clinically useful algorithms demonstrate performance in the range of 0.65 to 0.80.[51] In fact, discrimination >0.8 for clinical predictive models may represent data mismanagement or model overfitting.[20]

Modeling using combined DKRR and NKLR data revealed some notable differences between the 2 registries. The poor performance of the Cox lasso model is, in part, caused by the fact that when modeled separately, the 2 registry populations led to the selection of different variables and different effect sizes for the selected variables. The model fit to the combined data, therefore, is unable to achieve either of these individually optimal fits and thus performs more poorly. The nonparametric models did not have this limitation because they were able to fit the data with more flexibility. This observation helps explain the fact that although the Cox lasso model was the best model in the previous study of the NKLR,[30] here, the more flexible models performed better.

The present study has some limitations. First, even though several machine learning methods were considered, it is possible that another model may have performed differently. Second, there was a high proportion of missing preoperative KOOS data (47%, Table 1), and most patients with this missing variable were from the DKRR. Because preoperative KOOS data have been important in predicting outcomes based on previous studies, this substantial missingness likely contributed to the limited improvement in outcome prediction accuracy. In addition, patients were pooled across the entire time period from 2004 to 2020. Therefore, this analysis may inherit bias related to temporal changes in the revision surgery risk, as surgical indications, techniques, and trends have evolved over time. These changes were not directly accounted for in the present study but likely represent a low risk of bias, given the stable revision surgery rate observed in the registries.

Regarding clinical limitations of this study, more variables are required for revision prediction using this algorithm than the previously published NKLR calculator, which only required the input of 5 variables. This means that the present algorithms are more onerous to use in the office setting, with no appreciable improvement in prediction accuracy compared with the NKLR model. It therefore is likely of limited clinical value unless future external validation demonstrates superiority with different patient populations.

CONCLUSION

Machine learning analysis of the combined NKLR and DKRR enabled prediction of the revision ACLR risk with moderate accuracy. However, the resulting algorithms were less user-friendly and did not demonstrate superior accuracy in comparison with the previously developed model based on patients from the NKLR alone, despite the analysis of nearly 63,000 patients. This ceiling effect suggests that simply adding more patients to current national knee ligament registers is unlikely to improve predictive capability and may prompt future changes to increase variable inclusion.

ORCID iDs

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

REFERENCES

1. Beckers L, Vivacqua T, Firth AD, Getgood AMJ. Clinical outcomes of contemporary lateral augmentation techniques in primary ACL reconstruction: a systematic review and meta-analysis. *J Exp Orthop*. 2021;8(1):59.
2. Buuren SV, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1-67.
3. Choi JW, Cho YJ, Lee S, et al. Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. *Invest Radiol*. 2020;55(2):101-110.
4. 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.
5. Christino MA, Fleming BC, Machan JT, Shalvoy RM. Psychological factors associated with anterior cruciate ligament reconstruction recovery. *Orthop J Sports Med*. 2016;4(3):2325967116638341.
6. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
7. *Dansk Korsbånds Rekonstruktions Register Årsrapport 2020/2021*. Dansk Korsbåndsregister, 2021.
8. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542(7639):115-118.
9. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367-378.
10. 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.
11. 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.
12. 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.
13. Grassi A, Signorelli C, Urrizola F, et al. Patients with failed anterior cruciate ligament reconstruction have an increased posterior lateral tibial plateau slope: a case-controlled study. *Arthroscopy*. 2019;35(4):1172-1182.
14. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546.
15. Heusdens CHW, Blockhuys K, Roelant E, Dossche L, Van Glabbeek F, Van Dyck P. Suture tape augmentation ACL repair, stable knee, and favorable PROMs, but a re-rupture rate of 11% within 2 years. *Knee Surg Sports Traumatol Arthrosc*. 2021;29(11):3706-3714.
16. Hopper GP, Aithie JMS, Jenkins JM, Wilson WT, Mackay GM. Combined anterior cruciate ligament repair and anterolateral ligament internal brace augmentation: minimum 2-year patient-reported outcome measures. *Orthop J Sports Med*. 2020;8(12):2325967120968557.
17. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-860.
18. Jaecker V, Drouven S, Naendrup JH, Kanakamedala AC, Pfeiffer T, Shafizadeh S. Increased medial and lateral tibial posterior slopes are independent risk factors for graft failure following ACL reconstruction. *Arch Orthop Trauma Surg*. 2018;138(10):1423-1431.
19. 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.
20. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling, part II: generalization and overfitting. *Acta Neurochir Suppl*. 2022;134:15-21.
21. Ko S, Pareek A, Ro DH, et al. Artificial intelligence in orthopedics: three strategies for deep learning with orthopedic specific imaging. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(3):758-761.
22. Kunze KN, Krivicich LM, Clapp IM, et al. Machine learning algorithms predict achievement of clinically significant outcomes after orthopaedic surgery: a systematic review. *Arthroscopy*. 2022;38(6):2090-2105.
23. Lee CC, Youm YS, Cho SD, et al. Does posterior tibial slope affect graft rupture following anterior cruciate ligament reconstruction? *Arthroscopy*. 2018;34(7):2152-2155.
24. Magnussen RA, Lawrence JTR, West RL, Toth AP, Taylor DC, Garrett WE. Graft size and patient age are predictors of early revision after anterior cruciate ligament reconstruction with hamstring autograft. *Arthroscopy*. 2012;28(4):526-531.
25. Magnussen RA, Reinke EK, Huston LJ, et al. Effect of high-grade preoperative knee laxity on 6-year anterior cruciate ligament reconstruction outcomes. *Am J Sports Med*. 2018;46(12):2865-2872.
26. Mariscalco MW, Flanigan DC, Mitchell J, et al. The influence of hamstring autograft size on patient-reported outcomes and risk of revision after anterior cruciate ligament reconstruction: a Multicenter Orthopaedic Outcomes Network (MOON) cohort study. *Arthroscopy*. 2013;29(12):1948-1953.
27. Marmura H, Tremblay PF, Getgood AMJ, Bryant DM. The Knee Injury and Osteoarthritis Outcome Score does not have adequate structural validity for use with young, active patients with ACL tears. *Clin Orthop Relat Res*. 2022;480(7):1342-1350.
28. Martin RK, Ley C, Pareek A, Groll A, Tischer T, Seil R. Artificial intelligence and machine learning: an introduction for orthopaedic surgeons. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(2):361-364.
29. Martin RK, Persson A, Visnes H, Engebretsen L. Registries. In: Musahl V, Karlsson J, Hirschmann MT, et al, eds. *Basic Methods Handbook for Clinical Orthopaedic Research*. Springer; 2019:359-369.
30. 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.
31. 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.
32. 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.
33. 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*. 2020;28(4):1119-1129.
34. Midttun E, Andersen MT, Engebretsen L, et al. Good validity in the Norwegian Knee Ligament Register: assessment of data quality for key variables in primary and revision cruciate ligament reconstructions from 2004 to 2013. *BMC Musculoskelet Disord*. 2022;23(1):231.
35. *Norwegian Arthroplasty Register, Norwegian Cruciate Ligament Register, Norwegian Hip Fracture Register, and Norwegian Paediatric Hip Register 2020 Annual Report*. Norwegian National Advisory Unit on Arthroplasty and Hip Fractures; 2020:376.
36. Rahr-Wagner L, Thillemann TM, Lind MC, Pedersen AB. Validation of 14,500 operated knees registered in the Danish Knee Ligament Reconstruction Register: registration completeness and validity of key variables. *Clin Epidemiol*. 2013;5:219-228.
37. Rouzrokh P, Wyles CC, Philbrick KA, et al. A deep learning tool for automated radiographic measurement of acetabular component inclination and version after total hip arthroplasty. *J Arthroplasty*. 2021;36(7):2510-2517.e6.
38. Samitier G, Marcano AI, Alentorn-Geli E, Cugat R, Farmer KW, Moser MW. Failure of anterior cruciate ligament reconstruction. *Arch Bone Jt Surg*. 2015;3(4):220-240.
39. 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(2):e200198.
40. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020; 577(7792):706-710.
41. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw*. 2011;39(5):1-13.
42. Snaebjörnsson T, Hamrin-Senorski E, Svantesson E, et al. Graft diameter and graft type as predictors of anterior cruciate ligament revision: a cohort study including 18,425 patients from the Swedish and Norwegian National Knee Ligament Registries. *J Bone Joint Surg Am*. 2019;101(20):1812-1820.
43. Sonnery-Cottet B, Saithna A, Cavalier M, et al. Anterolateral ligament reconstruction is associated with significantly reduced ACL graft

rupture rates at a minimum follow-up of 2 years: a prospective comparative study of 502 patients from the SANTI Study Group. *Am J Sports Med*. 2017;45(7):1547-1557.
44. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240(4857):1285-1293.
45. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol*. 2019;48(2):239-244.
46. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6(1):25.
47. Vock DM, Wolfson J, Bandyopadhyay S, et al. Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. *J Biomed Inform*. 2016;61:119-131.
48. Webb JM, Salmon LJ, Leclerc E, Pinczewski LA, Roe JP. Posterior tibial slope and further anterior cruciate ligament injuries in the anterior cruciate ligament–reconstructed patient. *Am J Sports Med*. 2013;41(12):2800-2804.
49. Wyatt JM, Booth GJ, Goldman AH. Natural language processing and its use in orthopaedic research. *Curr Rev Musculoskelet Med*. 2021;14(6):392-396.
50. Yamada Y, Maki S, Kishida S, et al. Automated classification of hip fractures using deep convolutional neural networks with orthopedic surgeon-level accuracy: ensemble decision-making with anteroposterior and lateral radiographs. *Acta Orthop*. 2020;91(6):699-704.
51. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol*. 2014;39(2):204-221.
52. Zhao D, Pan JK, Lin FZ, et al. Risk factors for revision or rerupture after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Am J Sports Med*. Published online October 3, 2022. doi: 10.1177/03635465221119787