Received: 14 September 2023 | Accepted: 5 December 2023

DOI: 10.1002/ksa.12031



# External validation of the Norwegian anterior cruciate ligament reconstruction revision prediction model using patients from the STABILITY 1 Trial

**R. Kyle Martin**[1,2,3] | **Hana Marmura**[4] | **Solvejg Wastvedt**[5] | **Ayoosh Pareek**[6] | **Andreas Persson**[3,7] | **Gilbert Moatshe**[3,7] | **Dianne Bryant**[8] | **Julian Wolfson**[5] | **Lars Engebretsen**[3,7] | **Alan Getgood**[4]

[1]Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, Minnesota, USA

[2]Department of Orthopaedic Surgery, CentraCare, Saint Cloud, Minnesota, USA

[3]Oslo Sport Trauma Research Center, Norwegian School of Sports Science, Oslo, Norway

[4]Department of Orthopaedic Surgery, University of Western Ontario, London, Ontario, Canada

[5]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, USA

[6]Department of Orthopedic Surgery, Hospital for Special Surgery, New York, New York, USA

[7]Orthopaedic Clinic, Oslo University Hospital Ullevål, Oslo, Norway

[8]School of Physical Therapy, University of Western Ontario, London, Ontario, Canada

**Correspondence**

R. Kyle Martin, Department of Orthopedic Surgery, University of Minnesota, 2512 South 7th Street, Suite R200, Minneapolis, MN 55455, USA.
Email: rkylemmartin@gmail.com

**Funding information**

Norwegian Centennial Chair Seed Grant

## Abstract

**Purpose:** A machine learning-based anterior cruciate ligament (ACL) revision prediction model has been developed using Norwegian Knee Ligament Register (NKLR) data, but lacks external validation outside Scandinavia. This study aimed to assess the external validity of the NKLR model (https://swastvedt.shinyapps.io/calculator_rev/) using the STABILITY 1 randomized clinical trial (RCT) data set. The hypothesis was that model performance would be similar.

**Methods:** The NKLR Cox Lasso model was selected for external validation owing to its superior performance in the original study. STABILITY 1 patients with all five predictors required by the Cox Lasso model were included. The STABILITY 1 RCT was a prospective study which randomized patients to receive either a hamstring tendon autograft (HT) alone or HT plus a lateral extra-articular tenodesis (LET). Since all patients in the STABILITY 1 trial received HT ± LET, three configurations were tested: 1: all patients coded as HT, 2: HT + LET group coded as bone-patellar tendon-bone (BPTB) autograft, 3: HT + LET group coded as unknown/other graft choice. Model performance was assessed via concordance and calibration.

**Results:** In total, 591/618 (95.6%) STABILITY 1 patients were eligible for inclusion, with 39 undergoing revisions within 2 years (6.6%). Model performance was best when patients receiving HT + LET were coded as BPTB. Concordance was similar to the original NKLR prediction model for 1- and 2-year revision prediction (STABILITY: 0.71; NKLR: 0.68–0.69). Concordance 95% confidence interval (CI) ranged from 0.63 to 0.79. The model was well calibrated for 1-year prediction while the 2-year prediction demonstrated evidence of miscalibration.

**Conclusion:** When patients in STABILITY 1 who received HT + LET were coded as BPTB in the NKLR prediction model, concordance was similar to

**Abbreviations:** ACL, anterior cruciate ligament; AUC, area under the curve; BPTB, bone-patellar tendon-bone; CI, confidence interval; HT, Hamstring tendon; IRB, Institutional Review Board; KOOS-QOL, Knee Injury and Osteoarthritis Outcome Score Quality of Life subscale; LET, lateral extra-articular tenodesis; NKLR, Norwegian Knee Ligament Register; RCT, randomized clinical trial; REK, Regional Ethics Committee; STABILITY 1, randomized clinical trial investigating outcomes of hamstring tendon autograft (HT) ACLR with or without lateral extra-articular tenodesis (LET).

the index study. However, due to a wide 95% CI, the true performance of the prediction model with this Canadian and European cohort is unclear and a larger data set is required to definitively determine the external validity. Further, better calibration for 1-year predictions aligns with general prediction modelling challenges over longer periods. While not a large enough sample size to elicit the true accuracy and external validity of the prediction model when applied to North American patients, this analysis provides more support for the notion that HT plus LET performs similarly to BPTB reconstruction. In addition, despite the wide confidence interval, this study suggests optimism regarding the accuracy of the model when applied outside of Scandinavia.

**Level of Evidence:** Level 3, cohort study.

**KEYWORDS**
ACL, external validation, machine learning, outcome prediction

## INTRODUCTION

Anterior cruciate ligament (ACL) reconstruction (ACLR) is a commonly performed procedure aimed at reducing instability and restoring normal knee biomechanics after injury. Unfortunately, graft rupture and subsequent revision surgery remains an issue of concern—especially among young, active patients [1, 2]. By now, several risk factors for ACLR failure have been identified and include both modifiable and non-modifiable traits [3–6]. Recognition of these factors enables the clinician to coarsely risk-stratify patients who can influence surgical decision-making and outcome expectations. However, due to the sheer number of potential risk factors and the complex interactions between them, fine-level risk estimation remains challenging.

The emergence of machine learning applications into the orthopaedic literature has been heralded as a potential adjunctive tool capable of improving outcome prediction accuracy [7, 8]. These advanced statistical techniques can identify and interpret complex and non-linear interactions between variables leading to a more accurate understanding of how risk factors may affect surgical outcomes, both together and in isolation. This opens the door to the possibility of patient-specific risk estimation, surgical discussion and outcome optimization.

Preliminary machine learning-derived models for the prediction of ACLR outcome, including revision surgery, have recently been developed based on patients in the Norwegian Knee Ligament Register (NKLR) [9, 10]. The revision surgery prediction model includes an open-access online clinical calculator (https://swastvedt.shinyapps.io/calculator_rev/) and has undergone further external validation using patients from the Danish Knee Ligament Registry [11]. In general, external validation of clinical machine learning models in orthopaedic surgery is uncommonly performed and represents a crucial step prior to widespread adoption and implementation in clinical practice. External validation is valuable for several reasons, including assessment of model generalizability, minimizing bias and model overfitting, and increasing trust and acceptance among patients and clinicians regarding the utility of the model.

The purpose of this study was to determine the external validity of the previously published ACL revision prediction model when applied to patients enrolled in the STABILITY 1 randomized clinical trial investigating outcomes of hamstring tendon autograft (HT) ACLR with or without lateral extra-articular tenodesis (LET) [12]. The hypothesis was that model performance would be similar to the index study, suggesting validity of the algorithm. This study represents the first attempt to assess external validation using patients from outside of Scandinavia. If successful, this algorithm may be used to estimate revision risk at a patient-specific level and help guide discussion surrounding outcome expectations preoperatively.

## MATERIALS AND METHODS

Development of the Norwegian ACL revision surgery prediction algorithm was based on data from the NKLR. At the time of enrolment in this national registry, all patients provide informed consent and the Norwegian Data Inspectorate granted permission to the register for collection, analysis and publication on this health-related data. All data were de-identified prior to retrieval for model development and no further ethical approval is required from the Regional Ethics Committee (REK) for NKLR-based studies [13]. The original prediction model development was performed at the University of Minnesota and the respective Institutional Review Board similarly concluded that the study was exempt from full review (#00012552). The STABILITY 1 trial was approved by the Western Ontario Health Sciences Research Ethics Board (#104524).

## Index model development

The original prediction model was developed through machine learning analysis of all patients contained within the NKLR who underwent primary ACLR [9]. This national knee ligament registry prospectively enrols patients undergoing cruciate ligament reconstruction preoperatively and records demographic, injury, surgical and follow-up outcome details including subsequent revision surgery. The NKLR was established in 2004 and reporting has been mandatory since 2017. Overall compliance with the NKLR approximates 88% [14]. Patients are registered using their unique Norwegian national identification number which links identification of subsequent revision surgery performed within Norway, regardless of the provider.

In the index study of NKLR patients [9], four machine learning prediction models were assessed for the ability to predict subsequent revision ACLR after primary surgery and the primary outcome was probability of revision ACLR within 1, 2 and/or 5 years. The four models tested were Cox Lasso, survival random forest, generalized additive model and gradient boosted regression. These four models are among the most commonly used for this type of analysis. The patients in the NKLR were randomly split into training (75%) and test (25%) sets, whereby the algorithm was developed using the training set of patients, and the performance of the algorithm was assessed with the hold-out test set, previously unseen by the models. The Cox Lasso model was the best-performing of the four tested models and was used for the development of an in-clinic revision-risk calculator.

Regarding outcome prediction, the four models considered all the available data in the NKLR to 'learn' which factors are associated with—and can be used to predict—which patients will eventually undergo revision surgery. Starting with the 24 total predictor variables in the NKLR, the models eliminated variables which do not significantly contribute to prediction ability, without sacrificing accuracy. The result was an algorithm developed using the Cox Lasso model that only required five variables (out of the 24) for outcome prediction. The model was generally well calibrated and demonstrated moderate discriminative ability in predicting revision surgery after primary ACLR [9].

## Data source

The validation data for this study were extracted from the STABILITY 1 study, a randomized clinical trial conducted across nine sites (seven in Canada and two in Europe) [12]. This study investigated the 2-year outcomes of patients under 25 years of age who were undergoing a primary ACLR. Patients were randomized to undergo a HT ACLR either with or without an LET. The patients in this trial were classified as being at high risk of re-injury and/or surgical failure based on meeting at least two of the following criteria: a pivot shift Grade 2 or higher, a desire to return to high-risk/pivoting sports, and/or generalized ligamentous laxity. Outcome data for these patients were obtained at 3, 6, 12, and 24 months postoperatively.

## Participants and predictors

This current study sought to validate the previously developed Cox Lasso model from the NKLR. The Cox Lasso model was selected for validation since it was the best performing model. The Cox Lasso is a penalized regression model, which selects a subset of available variables for inclusion [15]. Thus, while a more extensive set of patient characteristics were assessed in development of the model, only the five predictors required for the NKLR Cox Lasso model were used in this validation analysis. The five variables required for outcome prediction using the Cox Lasso model were: patient age at primary surgery, Knee Injury and Osteoarthritis Outcome Score Quality of Life subscale (KOOS-QOL) score at primary surgery, graft choice, femur fixation method, and time between injury and ACLR.

The graft choice options in the NKLR model are HT, bone-patellar tendon-bone (BPTB), or other/unknown. STABILITY 1 participants were randomized to have an HT ± LET; therefore, all patients had an HT for the graft choice. The two STABILITY 1 graft type groups were coded in three different ways (1: all patients coded as HT, 2: HT + LET group coded as BPTB, 3: HT + LET group coded as unknown/other) to understand which would be most appropriate and which group the HT + LET group behaved most similar to in the model. This approach was chosen based on a previous study that demonstrated the addition of LET to an HT behaved similarly to a BPTB [16]. Since the STABILITY 1 trial followed patients for 2 years post-operatively, 5-year data and predictions were not included. Two patients without a documented revision date were included, with their graft rupture date substituted for the revision date (both 21 months postoperative). Patient characteristics for the STABILITY 1 validation data set are shown in Table 1 [12].

## Model performance

Performance of the model was assessed using the same metrics as the NKLR study: calibration and concordance (discrimination) at each follow-up time. Performance evaluation included censoring of the time-to-event outcome. 'Censoring' refers to the fact that, at any given follow-up time, complete information

**TABLE 1** Characteristics of patients in validation data set.

| Characteristics | All patients, *N* = 618 (mean ± SD) or *n* (%) | Patients with complete model data, *N* = 591 (mean ± SD) or *n* (%) |
|---|---|---|
| Age, years | 18.9 ± 3.2 | 19.0 ± 3.2 |
| Missing | 1 (0.2) | 0 (0) |
| Sex | | |
| Male | 298 (48.2) | 287 (48.6) |
| Female | 319 (51.6) | 304 (51.4) |
| Missing | 1 (0.2) | 0 (0) |
| BMI, kg/m$^2$ | 24.1 ± 3.8 | 24.1 ± 3.8 |
| Missing | 8 (1.3) | 2 (0.3) |
| KOOS-QOL, 0-100 | 33.2 ± 17.8 | 33.2 ± 17.8 |
| Missing | 16 (2.6) | 0 (0) |
| Graft | | |
| HT | 311 (50.3) | 296 (50.1) |
| BPTB | 0 (0) | 0 (0) |
| Other/Unknown | 0 (0) | 0 (0) |
| HT + LET | 307 (49.7) | 295 (49.9) |
| Femur fixation method | | |
| Interference screw | 0 (0) | 0 (0) |
| Suspension or cortical device | 618 (100) | 591 (100) |
| Unknown or other | 0 (0) | 0 (0) |
| Years between injury & surgery | 0.72 ± 1.49 | 0.72 ± 1.50 |
| Missing | 19 (3.1) | 0 (0) |
| Revision | | |
| Yes | 40 (6.5) | 39 (6.6) |
| Within 1 year | 9 (1.5) | 8 (1.4) |
| Between 1 and 2 years | 22 (3.6) | 22 (3.7) |
| After 2 years | 9 (1.5) | 9 (1.5) |
| No | 570 (92.2) | 552 (93.4) |
| Missing | 9 (1.5) | 0 (0) |

Abbreviations: BMI, body mass index; BPTB, bone-patellar tendon-bone; HT, hamstring tendon; LET, lateral extra-articular tenodesis; KOOS-QOL, knee osteoarthritis outcome score quality of life scale.

on outcome is not known for all patients. Some patients have not been followed in the study for the requisite number of years, while others have not yet experienced revision and it is unknown when or if they ultimately will.

## Statistical analysis

The program R (RStudio 2022.07.1) was used to calculate predicted survival probabilities for all patients in the validation data set. Calibration refers to the accuracy of the risk estimates. In the NKLR study, calibration was calculated using a version of the Hosmer–Lemeshow statistic [17]. This statistic sums average misclassification in each predicted risk quantile and converts the result into a chi-squared statistic. However, for the validation analysis, the low number of revisions in the validation data necessitated a slightly different approach. Rather than divide the data into risk quantiles, it was divided into three groups as follows: 0–25th percentile, 26–50th percentile, and 51–100th percentile of predicted survival probability. This change

ensured an adequate number of revisions in each group while retaining the statistical validity of the original method. For both calibration methods, a larger calibration statistic indicates worse calibration, and statistical significance means the null hypothesis of perfect calibration is rejected. Concordance was computed using Harrell's C-index [18] at 1- and 2-year follow-up times. The C-index is a generalization of area under the curve (AUC) that measures the proportion of ranked pairs of observations in which the predicted ranking corresponds with true outcomes. As with AUC, the C-index ranges from 0 to 1 with 1 indicating perfect concordance.

# RESULTS

## Participants

Of the 618 participants randomized in the STABILITY 1 study, 591 (95.6%) had complete data on Norwegian Cox lasso variables (five predictor variables and outcome). Of note, there were only 39 (6.6%) revision events in the analysed data set, 30 of which occurred by 2-year follow-up (5.1% of analysed sample). Eight patients had revision surgery within 1-year of their primary surgery while another 22 patients had surgery between the first- and second-year time points. An additional nine patients underwent revision after the 2-year follow-up timepoint.

In contrast to the original NKLR study cohort, which included all patients undergoing ACLR in Norway, the STABILITY 1 trial patients had a narrow age range (14–25 years old) and all patients received HT with suspensory fixation on the femur. Further, time from injury to surgery was shorter and more consistent for the STABILITY 1 patients relative to the NKLR data set.

## Model performance

Model performance was best for both 1- and 2-year revision prediction when patients randomized to HT + LET in the STABILITY 1 trial were coded as BPTB in the prediction calculator (Table 2). The model concordance (discriminative ability) was similar in the validation data set (0.71 and 0.71) compared to the development data set (0.69 and 0.68) at 1- and 2-year follow-up, respectively, with 95% confidence intervals (CI) ranging from 0.63 to 0.79. The concordance values were slightly better in the STABILITY 1 data set compared to the Norwegian registry; however, the associated 95% CI were much wider (Table 2). The calibration statistic for the model predicting 1-year outcomes was adequately low (2.6) and the associated non-significant $p$ value (0.10) indicates the model is well calibrated (no significant difference between observed and predicted probabilities of revision). The 2-year prediction model demonstrated evidence of miscalibration in the validation data set (high calibration statistic [11.7] and significant $p < 0.01$), similar to the results seen in the Norwegian and Danish data [11].

# DISCUSSION

The most important finding of this study was that when patients in the STABILITY 1 trial who received HT + LET were coded as BPTB in the Norwegian prediction calculator, the model concordance was similar to the index study. However, the 95% CI for the validation set was wider than the original model, suggesting that more data are required to definitively determine the external validity. Further, model calibration was better for predicting revision surgery within 1-year and worse when predicting 2-year outcomes. This finding is in keeping with the original study and with prediction modelling in general, as predicting outcomes over longer time-periods is typically more challenging due to the increased outcome variability observed over time.

At first glance, the performance of the NKLR model presented in Table 2 may appear impressive—the discriminative ability of the algorithm to properly order patients regarding their revision risk (concordance) was higher in the external validation set relative to both the initial model validation and the external validation using Danish patients [9, 11]. Closer inspection, however, reveals a wide CI that extends beyond both ends of the NKLR model performance CI. This is an important distinction as it suggests the true accuracy of the model for the STABILITY 1 patient population remains unknown. A larger sample size would be necessary to narrow this CI and more clearly ascertain the performance of the NKLR model on this different patient data set.

Model performance was worse when the patients receiving HT + LET were coded as having received either HT or 'Unknown/Other' graft choices. The finding that the failure rate for HT + LET ACLR is most similar to BPTB ACLR is consistent with the literature [16]. In a previous study, the STABILITY 1 data set was used for external validation of the Multicenter Orthopaedic Outcomes Network (MOON) autograft risk calculator, which included either HT or BPTB as the graft type [16]. The validation analysis was run once with the HT + LET group coded as HT only and once coded as BPTB. Mirroring results of the present study, the risk calculator was most predictive (AUC = 0.73) when the LET group was coded as the BPTB group. Hamstring tendon and BPTB autograft both have a long history of use for ACLR, but several studies have found higher failure rates among patients receiving HT [19–21]. This difference is especially apparent in young active patients and has influenced both clinical practice and innovation within the specialty. In Norway, HT was the graft of choice until approximately 2015 when a NKLR-based study found higher failure rates versus BPTB [20]. In 2012, 79% of patients received HT

**TABLE 2** Model performance.

| Probability of revision | Model | Concordance | Calibration statistic (quintile method) | Calibration *p* value |
|---|---|---|---|---|
| 1 year | Original Norwegian Algorithm Performance[a] | 0.686 (0.652–0.721) | 4.9 | n.s. |
| | STABILITY data | 0.713 (0.634–0.791) | 2.6 | n.s. |
| | HT = HT | | | |
| | HT + LET = BPTB | | | |
| | STABILITY data | 0.609 (0.528–0.691) | 10.6 | <0.01* |
| | HT = HT | | | |
| | HT + LET = Unknown | | | |
| | STABILITY data | 0.674 (0.597–07.51) | 8.7 | <0.01* |
| | All patients = HT | | | |
| 2 years | Original Norwegian Algorithm Performance[a] | 0.684 (0.650–0.718) | 11.3 | 0.01* |
| | STABILITY data | 0.713 (0.637–0.789) | 11.7 | <0.01* |
| | HT = HT | | | |
| | HT + LET = BPTB | | | |
| | STABILITY data | 0.608 (0.530–0.688) | 8.9 | <0.01* |
| | HT = HT | | | |
| | HT + LET = Unknown | | | |
| | STABILITY data | 0.673 (0.598–0.747) | 10.2 | <0.01* |
| | All patients = HT | | | |

Abbreviations: BPTB, bone-patellar tendon-bone autograft; HT, hamstring tendon autograft; LET, lateral extra-articular tenodesis; n.s., not statistically significant.
[a]See Martin et al. [9].
*Statistical significance, $p \leq 0.05$.

autograft while in 2016 that number dropped to 32%, with BPTB representing 61% of all ACLR grafts that year. The study of LET as an augment to HT ACLR represents an important innovation in response to inferior outcomes with HT alone and has consistently demonstrated lower failure rates than HT alone [12, 22, 23]. These results reflect why HT + LET acts more like BPTB than HT alone or other types of grafts in predictive models of graft failure/revision surgery.

The orthopaedic literature has seen an exponential increase in studies applying a machine learning approach to data analysis. Most of these studies have sought automatic radiologic diagnostics (computer vision), language interpretation (natural language processing) or outcome prediction. While the number of novel machine learning and deep learning models has proliferated, very few have completed the important step of external validation.

While it is crucial to perform prior to widespread adoption and implementation of these models, external validation can present several challenges for clinician scientists. First, a large volume of data is required for model validation, and it can be difficult to find a suitably large study population with the necessary variables required for external validation. Ideally, patient populations should be similar with regard to the nature of data collection and tracking of outcomes for appropriate model evaluation. Another limitation is the possibility of data transfer barriers between nations or health regions due to local legislation and privacy concerns. For this reason, it is often easiest to share machine learning algorithms rather than patient data and requires collaborative efforts between study groups. Finally, most machine learning algorithms demonstrate a drop-off in performance during external validation, which raises the question of what constitutes acceptable model performance in this setting.

The debate regarding acceptable model performance is especially pertinent when evaluating models to predict outcomes such as ACL revision surgery which will likely never achieve excellent or perfect performance. These designations have historically been reserved for models with discrimination values greater than 0.90. Given the randomness associated with subsequent ACL graft rupture and the multiple variables which may contribute to the decision to proceed with revision surgery, the traditional interpretation of model performance is not realistic for clinical models like this one. There is also concern that models that *do* demonstrate 'excellent'

discriminative ability may be the product of model overfitting, limiting their real-world performance. Ultimately, the development of clinically useful outcome prediction models relies on a three-step approach consisting of model development, external validation and comparison with expert human prediction performance. It is this final step which establishes the baseline above which prediction models must perform in order to be clinically relevant. At this time, no such comparison exists for ACL outcome prediction and represents the next step in ACL outcome prediction research.

There were some limitations to the current study. First, the patient populations were different in several ways. Due to the standardization in the STABILITY 1 randomized trial, the data set includes a narrow age range from 14 to 25 years old compared to a much wider age range represented in the NKLR. Since all patients had HT autograft ACLR, the femoral fixation was universally suspensory/cortical leading to no variation in this predictor. Further, due to the nature of the STABILITY 1 trial, the chronicity of injury (time from injury to surgery in years) is much smaller and more standardized than observed in the NKLR. Another limitation is the fact that the sample size used for external validation was small, with few observed revision surgery events within 2 years ($n = 30$, 5%). This required a change in methodology for the calculation of model calibration compared with the index study. The calculation of model calibration was further complicated by the fact that patients in the STABILITY 1 trial were enrolled based on inclusion criteria that identified them as being particularly high-risk for ACLR failure.

Although the results of this study did not confirm the external validity of the NKLR revision prediction model, model performance with this separate cohort of patients was encouraging and will prompt further evaluation using larger patient data sets. Additionally, LET was only added as a variable collected by the NKLR in June 2019 and therefore was not considered during the original prediction model development. As the rate of LET during ACLR increases, it is important to consider how this adjunctive procedure may affect outcome predictions. This study, in keeping with the findings of Marmura et al. [16], suggests that outcome prediction for patients receiving LET in addition to HT ACLR may be more accurate if they are coded as BPTB in the revision prediction tool.

## CONCLUSION

When patients in STABILITY 1 who received HT + LET were coded as BPTB in the NKLR prediction model, concordance was similar to the index study. However, due to a wide 95% CI, the true performance of the prediction model with this Canadian and European cohort is unclear and a larger data set is required to definitively determine the external validity. Further, better calibration for 1-year predictions aligns with general prediction modelling challenges over longer periods. While not a large enough sample size to elicit the true accuracy and external validity of the prediction model when applied to North American patients, this analysis provides more support for the notion that HT plus LET performs similarly to BPTB reconstruction. In addition, despite the wide confidence interval, this study suggests optimism regarding the accuracy of the model when applied outside of Scandinavia.

### AUTHOR CONTRIBUTIONS

R. Kyle Martin was involved in study conception, design, interpretation of data analysis, and manuscript preparation and editing. Hana Marmura was involved in study design, carried out the data acquisition, analysis, and interpretation, and was involved in manuscript preparation and editing. Solvejg Wastvedt was involved in study design, data analysis and interpretation, and was involved in manuscript preparation and editing. Ayoosh Pareek was involved in study conception, design, interpretation of data analysis, and manuscript editing. Andreas Persson was involved in study conception, design, interpretation of data analysis, and manuscript preparation and editing. Gilbert Moatshe was involved in study conception, design, interpretation of data analysis, and manuscript preparation and editing. Dianne Bryant was involved in design and manuscript editing. Julian Wolfson was involved in study design, data interpretation, and was involved in manuscript editing. Lars Engebretsen was involved in study conception, design, interpretation of data analysis, and manuscript preparation and editing. Alan Getgood was involved in study conception, design, interpretation of data analysis, and manuscript preparation and editing. All authors have given final approval of the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### ACKNOWLEDGEMENTS

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

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

### ETHICS STATEMENT

The original prediction model development was performed at the University of Minnesota and the respective Institutional Review Board similarly concluded that the study was exempt from full review (#00012552). The STABILITY 1 trial was approved by the Western Ontario Health Sciences Research Ethics Board (#104524). Approval not required as informed consent was obtained by all patients at the time of enrolment in the national knee ligament register.

## ORCID

*R. Kyle Martin* http://orcid.org/0000-0001-9918-0264

## REFERENCES

1. Webster, K.E. & Feller, J.A. (2016) Exploring the high reinjury rate in younger patients undergoing anterior cruciate ligament reconstruction. *The American Journal of Sports Medicine*, 44(11), 2827–2832.
2. Webster, K.E., Feller, J.A., Leigh, W.B. & Richmond, A.K. (2014) Younger patients are at increased risk for graft rupture and contralateral injury after anterior cruciate ligament reconstruction. *The American Journal of Sports Medicine*, 42(3), 641–647.
3. Cristiani, R., Forssblad, M., Edman, G., Eriksson, K. & Stålman, A. (2021) Age, time from injury to surgery and quadriceps strength affect the risk of revision surgery after primary ACL reconstruction. *Knee Surgery, Sports Traumatology, Arthroscopy*, 29(12), 4154–4162.
4. Gifstad, T., Foss, O.A., Engebretsen, L., Lind, M., Forssblad, M., Albrektsen, G. et al. (2014) Lower risk of revision with patellar tendon autografts compared with hamstring autografts: a registry study based on 45,998 primary ACL reconstructions in Scandinavia. *The American Journal of Sports Medicine*, 42(10), 2319–2328.
5. Jaecker, V., Drouven, S., Naendrup, J.-H., Kanakamedala, A.C., Pfeiffer, T. & Shafizadeh, S. (2018) Increased medial and lateral tibial posterior slopes are independent risk factors for graft failure following ACL reconstruction. *Archives of Orthopaedic and Trauma Surgery*, 138(10), 1423–1431.
6. Kaeding, C.C., Pedroza, A.D., Reinke, E.K., Huston, L.J., Spindler, K.P., Amendola, A. et al. (2015) 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. *The American Journal of Sports Medicine*, 43(7), 1583–1590.
7. Martin, R.K., Ley, C., Pareek, A., Groll, A., Tischer, T. & Seil, R. (2022) Artificial intelligence and machine learning: an introduction for orthopaedic surgeons. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(2), 361–364.
8. Pruneski, J.A., Pareek, A., Kunze, K.N., Martin, R.K., Karlsson, J., Oeding, J.F. et al. (2023) Supervised machine learning and associated algorithms: applications in orthopedic surgery. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31(4), 1196–1202.
9. Martin, R.K., Wastvedt, S., Pareek, A., Persson, A., Visnes, H., Fenstad, A.M. et al. (2021) Predicting anterior cruciate ligament reconstruction revision: a machine learning analysis utilizing the Norwegian Knee Ligament Register. *Journal of Bone and Joint Surgery*, 104(2), 145–153.
10. Martin, R.K., Wastvedt, S., Pareek, A., Persson, A., Visnes, H., Fenstad, A.M. et al. (2022) Predicting Subjective Failure of ACL Reconstruction: A Machine Learning Analysis of the Norwegian Knee Ligament Register and Patient Reported Outcomes. *Journal of ISAKOS*, 7(3), 1–9.
11. Martin, R.K., Wastvedt, S., Pareek, A., Persson, A., Visnes, H., Fenstad, A.M. et al. (2022) Machine learning algorithm to predict anterior cruciate ligament revision demonstrates external validity. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(2), 368–375.
12. Getgood, A.M.J., Bryant, D.M., Litchfield, R., Heard, M., McCormack, R.G. & Rezansoff, A. et al. (2020) Lateral extra-articular tenodesis reduces failure of hamstring tendon autograft anterior cruciate ligament reconstruction: 2-year outcomes from the STABILITY Study randomized clinical trial. *The American Journal of Sports Medicine*, 48(2), 285–297.
13. Granan, L.-P., Bahr, R., Steindal, K., Furnes, O. & Engebretsen, L. (2008) Development of a national cruciate ligament surgery registry: the Norwegian National Knee Ligament Registry. *The American Journal of Sports Medicine*, 36(2), 308–315.
14. Norwegian Knee Ligament Register: Annual Report 2022. p. 283. Accessed June 2023. https://www.helse-bergen.no/nasjonalt-kvalitets-og-kompetansenettverk-for-leddproteser-og-hoftebrudd/arsrapporter
15. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. (2011) Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5), 1–13.
16. Marmura, H., Getgood, A.M.J., Spindler, K.P., Kattan, M.W., Briskin, I. & Bryant, D.M. (2021) Validation of a risk calculator to personalize graft choice and reduce rupture rates for anterior cruciate ligament reconstruction. *The American Journal of Sports Medicine*, 49(7), 1777–1785.
17. Vock, D.M., Wolfson, J., Bandyopadhyay, S., Adomavicius, G., Johnson, P.E., Vazquez-Benitez, G. et al. (2016) Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. *Journal of Biomedical Informatics*, 61, 119–131.
18. Harrell, F.E. (1982) Evaluating the yield of medical tests. *JAMA*, 247(18), 2543–2546.
19. Spindler, K.P., Huston, L.J., Zajichek, A., Reinke, E.K., Amendola, A., Andrish, J.T. et al. (2020) Anterior cruciate ligament reconstruction in high school and college-aged athletes: does autograft choice influence anterior cruciate ligament revision rates? *The American Journal of Sports Medicine*, 48(2), 298–309.
20. Persson, A., Fjeldsgaard, K., Gjertsen, J.-E., Kjellsen, A.B., Engebretsen, L., Hole, R.M. et al. (2014) 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. *The American Journal of Sports Medicine*, 42(2), 285–291.
21. Schuette, H.B., Kraeutler, M.J., Houck, D.A. & McCarty, E.C. (2017) Bone–Patellar tendon–bone versus Hamstring tendon autografts for primary anterior cruciate ligament reconstruction: a systematic review of overlapping meta-analyses. *Orthopaedic Journal of Sports Medicine*, 5(11), 232596711773648.
22. Borque, K.A., Jones, M., Laughlin, M.S., Balendra, G., Willinger, L., Pinheiro, V.H. et al. (2022) Effect of lateral extra-articular tenodesis on the rate of revision anterior cruciate ligament reconstruction in elite athletes. *The American Journal of Sports Medicine*, 50(13), 3487–3492.
23. Park, Y.-B., Lee, H.-J., Cho, H.-C., Pujol, N. & Kim, S.H. (2023) Combined lateral extra-articular tenodesis or combined anterolateral ligament reconstruction and anterior cruciate ligament reconstruction improves outcomes compared to isolated reconstruction for anterior cruciate ligament tear: a network meta-analysis of randomized controlled trials. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 39(3), 758–776.

**How to cite this article:** Martin, R.K., Marmura, H., Wastvedt, S., Pareek, A., Persson, A., Moatshe, G. et al. (2024) External validation of the Norwegian anterior cruciate ligament reconstruction revision prediction model using patients from the STABILITY 1 Trial. *Knee Surgery, Sports Traumatology, Arthroscopy*, 32, 206–213. https://doi.org/10.1002/ksa.12031