Application of Survival Machine Learning Methodology to Predict Time-Dependent Risk of Revision Anterior Cruciate Ligament Reconstruction Graft Failure in the MARS Cohort

Kinjal Vasavada[1], Vrinda Devang Vasavada, Sai Devana, Changhee Lee, Sharon L Hame[2], Laith M Jazrawi[3], Rick W Wright[4], Laura Huston[5], Amanda Haas Braun[6], Christina R Allen, Daniel E Cooper, Thomas M DeBerardino, Brick A Lantz, Kurt P Spindler[7], Michael J Stuart, Annunziato Amendola[8], Christopher C Annunziata[9], Robert A Arciero[10], Bernard R Bach, Champ Baker, Arthur R Bartolozzi, Keith M Baumgarten[11], Jeffrey H Berg[12], Geoffrey A Bernas[13], Stephen F Brockmeier[14], Robert H Brophy[15], Charles A Bush-Joseph[16], Jay Brad Butler, James L Carey, James E Carpenter, Brian J Cole[17], Charles Leonard Cox, Robert Alexander Creighton[18], Tal S David[19], Warren Dunn[20], David Clint Flanigan[21], Robert W Frederick, Theodore J Ganley[22], Charles J Gatt[23], Steven R Gecha, Jo A Hannafin, N Lindsay Harris[24], Keith S Hechtman[25], Elliott Hershman[26], Rudolf Hoellrich[27], David Charles Johnson[28], Timothy S Johnson[29], Morgan H Jones, Christopher C Kaeding[21], Ganesh V Kamath[30], Bruce A Levy[31], ChunBong Benjamin Ma[32], G Peter Maiers[33], Matthew J Matava[34], Robert G Marx[35], Gregory M Mathien[36], David R McAllister[37], Bruce Scott Miller, Carl Wilson Nissen[38], Daniel Fulham O'Neill, Richard D Parker[39], Arun J Ramappa, Michael A Rauh[40], Jon K Sekiya[41], Kevin G Shea[42], James R Slauterbeck[43], Matthew V Smith[34], Jeffrey Tweed Spang[44], Steven James Svoboda, Timothy N Taft, Joachim Tenuta, Edwin M Tingstad[45], Armando Felipe Vidal, Darius Viskontas, Richard A White, James Simpson Williams, Michelle Wolcott[46], Brian R Wolf[47], James J York[48]

[1]NYU Langone Orthopedics, [2]UCLA, [3]Center For Musculoskeletal Care, [4]Vanderbilt Dept of Orthopaedics, [5]Vanderbilt University Medical Center, [6]Washington University St. Louis School of Medicinie, [7]Cleveland Clinic Foundation, [8]Duke University, [9]Orthovirginia, [10]Univ of CT Health Center, [11]Orthopedic Institute, [12]Town Center Orthopaedic Association, [13]UBMD Orthopaedics and Sports Medicine, [14]University of Virginia, [15]Washington University Orthopedics, [16]Midwest Orthopaedics at Rush, [17]Rush University Medical Center, [18]Dept of Orthopaedics, [19]Synergy Orthopedic Specialists, [20]Texas Orthopedic Hospital, [21]OSU Sports Medicine Ctr, [22]Children's Hospital of Philadelphia, [23]University Orthopaedic, [24]Grand River Health, [25]UHZ Sports Medicine Institute, [26]Lenox Hill Hospital/Ortho Surg, [27]Slocum Center, [28]The National Sports Medicine Institute, [29]National Sports Medicine Institute, [30]University of NC- Chapel Hill, [31]Mayo Clinic, [32]UCSF Med Ctr, [33]Fort Sports Medicine and Orthopedics, [34]Washington University, [35]Hosp for Special Surgery, [36]Knoxville Orthopaedic Clinic, [37]UCLA Department of Orthopaedic Surgery, [38]PRISM Sports Medicine, [39]Cleveland Clinic Hillcrest Hosp Admin Office, [40]UBMD Orthopaedics & Sports Medicine, [41]MedSport - Univ of Michigan Med Ctr, [42]Stanford, [43]University of South Alabama Orthopedics, [44]Univ of NC, Dept of Ortho, [45]Inland Orthopaedics, [46]CU Sports Medicine, [47]Univ of Iowa Hosps & Clinics, [48]AAMG Orthopedic and Sports Medicine Specialists

INTRODUCTION:

Previous multicenter anterior cruciate ligament revision study MARS machine learning (ML) analysis has shown that a novel ML model built using MARS cohort data can predict graft failure at 6 years postoperatively as a binary outcome with moderate ability. However, binary classification does not take into account the time-to-event component of graft failure as an outcome. In contrast, survival models are able to factor the time-to-event component of an outcome, incorporate time dependent variables, and appropriately process censored data to deliver more clinically meaningful outcome predictions. While the previously built binary classifier can provide information about whether or not the primary outcome graft failure will occur at 6 years, survival analysis models are able to provide information on when graft failure occurs (for example, within 2, 4, or 6 years). Therefore, this study sought to apply ML survival analysis methodology to MARS cohort data to produce a model that can most accurately estimate the probability of revision anterior cruciate ligament reconstruction (rACLR) graft failure within 2, 4, and 6 years postoperatively, and identify factors that impact probability estimates at each of these time horizons.

METHODS:

rACLR patients were prospectively recruited by MARS group. Preoperative radiographs, surgeon-reported intraoperative findings, and 2 and 6-year follow-up data on patient-reported outcomes (PROs), additional surgeries, and graft failure were obtained. ML models including Cox Proportional Hazards (CoxPH), XGBoost Survival Embeddings (XGBoost), Random Survival Forest (RSF), Extra Survival Tree (EST), and CoxBoost, were built to predict the probability of graft failure within 2, 4, and 6 years postoperatively. Validated performance metrics and feature importance measures were used to evaluate model performance.

RESULTS:

The cohort included 1,142 patients; 4.8% (n=55) experienced graft failure during the 6-year postoperative period. CoxBoost demonstrated the highest discriminative power across all studied time points **(**Timepoint: Time Dependent C-index | 2-year: 0.725 ± 0.127 | 4-year: 0.669 ± 0.057 | 6-year: 0.645 ± 0.054), with well-calibrated scores (Timepoint: Time dependent Brier Score | 2-year: 0.023 ± 0.012 | 4-year: 0.035 ± 0.010 | 6-year: 0.046 ± 0.012) as listed in Table 1 and Table 2 respectively. Based on the Kaplan Meier survival plot, the survival probability (probability of remaining graft failure

free) at 2, 4, 6 years was 0.977, 0.965, and 0.954 respectively. Within the context of the predictive models, features deemed important for CoxBoost differed from those deemed important for CoxPH predictive ability.

DISCUSSION AND CONCLUSION:

Survival machine learning models can predict the risk of rACLR graft failure up to 6 years postoperatively, with the most robust predictions at the 2 years postoperative timepoint. While external validation with further combined registry datasets is required, this present study builds on prior machine learning analyses towards the development of a bedside calculator for rACLR outcome prediction and risk stratification.

Table 1. Discrimination statistics at 2, 4, and 6 years postoperatively

Method	2-year C-index	4-year C-index	6-year C-index
CoxBoost	0.725 ± 0.127	0.669 ± 0.057	0.645 ± 0.054
EST	0.666 ± 0.097	0.596 ± 0.103	0.609 ± 0.090
CoxPH	0.638 ± 0.126	0.593 ± 0.074	0.603 ± 0.074
RSF	0.638 ± 0.100	0.619 ± 0.071	0.611 ± 0.083
XGBoost	0.611 ± 0.094	0.626 ± 0.045	0.624 ± 0.031

Table 2. Calibration statistics at 2, 4, and 6 years postoperatively

Method	2-year C-index	4-year C-index	6-year C-index
CoxBoost	0.023 ± 0.012	0.035 ± 0.010	0.046 ± 0.012
EST	0.023 ± 0.012	0.035 ± 0.010	0.046 ± 0.012
CoxPH	0.027 ± 0.013	0.044 ± 0.007	0.056 ± 0.009
RSF	0.023 ± 0.012	0.035 ± 0.010	0.046 ± 0.012
XGBoost	0.027 ± 0.011	0.040 ± 0.007	0.051 ± 0.010