One-year Mortality Following Hip Fracture Surgery: Development and External Validation of a Clinical Prediction Tool

R. Kyle Martin MD PhD FRCSC[1,2,3], Bryant Oliphant MD MBA MSc[4], Raymond Finzel[5], Brooke Patterson MPH[5], Tyler Loftus MD PhD[6], Die Hu MS[6] Christopher Tignanelli MD MS[5,7,8]
Email of Presenting Author: rkylemmartin@gmail.com

[1]University of Minnesota, MN, USA [2]CentraCare M Physicians Orthopedics, MN, USA [3]Norwegian School of Sport Sciences [4]Department of Orthopedic Surgery, University of Michigan, MI [5]Center for Learning Health System Sciences, University of Minnesota, MN [6]Department of Surgery, University of Florida, FL [7]Department of Surgery, University of Minnesota, MN [8]Institute for Health Informatics, University of Minnesota, MN,

Disclosures: R. Martin: Unpaid consultant: Smith and Nephew; Committee member with AOSSM and ISAKOS.

INTRODUCTION: Hip fractures are common injuries among the elderly and one-year mortality rates following a hip fracture have been reported to be 25-30%[1]. The ability to accurately estimate post-operative risk of mortality would support informed decision making and help patients and their families to have realistic expectations. Machine learning analysis of existing large institutional electronic health record (EHR) databases has the potential to improve our predictive capability. Although several machine learning models have been developed to predict hip fracture outcome, there remains a paucity of externally validated models for the North American population[1]. The purpose of this study was to 1) develop a one-year mortality prediction model for patients undergoing hip fracture surgery through analysis of an institutional EHR (Institution A) and 2) externally validate the model using the EHR of Institution B.

METHODS: Ethics approval was obtained from both institutions. Patients were identified in the two EHR databases that met the inclusion criteria: Age ≥ 50 and both hip fracture diagnostic (ICD-CM) and surgical management (CPT or ICD-PCS) codes in the same 30-day encounter from 2011 to 2023. Hip fracture was defined as a fracture to the neck, intertrochanteric, or subtrochanteric regions of the proximal femur. Exclusion criteria included peri-prosthetic or pathological hip fracture, prophylactic nailing, and fractures to the pelvis or other regions of the femur. Approximately 250 readily available variables were extracted from the EHR of Institutions A and B for each patient and the data from Institution A was split into training (80%) and test (20%) sets. Six different machine learning models were tested: Gradient Boosting, Random Forest, Support Vector Machine, Neural Network, Logistic Regression (LR), and XGBoost. The primary outcome for the models was probability of mortality within one year. Model calibration was assessed using a version of the Hosmer-Lemeshow statistic with decile calibration curves and discrimination was computed as the area under the receiver operator characteristic curve (AUC). The best performing model was evaluated using the same performance metrics when applied to the external validation dataset from Institution B.

RESULTS: The dataset from Institution A included 8,224 records (29.7% male) with an average age of 81.3 ± 10.6 years. The one-year mortality rate was 18.1%. All six models were generally well-calibrated, with concordance in the moderate range (0.73-0.78). The LR model performed best and required only 16 variables for risk estimation (Table 1). These variables included: age, sex, weight, alcohol use, smoking status, ASA score, comorbidities (type I diabetes, hypertension, lymphoma, chronic kidney disease, and hypothyroidism), hemoglobin, platelet count, previous hospital admission (within 7 days and 30 days), and glucocorticoid prescription prior to admission. There were 1,974 records from Institution B (35.3% male) with an average age of 75.7 ± 11.7 years and a one-year mortality rate of 16.9%. The LR model was well-calibrated and demonstrated an AUC of 0.73 (95% CI: 0.70-0.76) when evaluated on the patients from Institution B. Performance of the LR model including AUC and calibration is presented in Figures 1 and 2, respectively.

DISCUSSION: Machine learning analysis of institutional EHR data can predict the risk of one-year mortality with moderate accuracy. Importantly, this model has undergone the crucial yet rarely performed step of external validation using data from an outside institution. This algorithm supports the creation of a clinical calculator for point-of-care risk stratification based on the input of 16 variables.

SIGNIFICANCE/CLINICAL RELEVANCE: This represents the first machine learning driven model for predicting mortality risk following hip fracture surgery that has undergone external validation using a population from the United States of America. An accurate predictive model for clinical outcome following hip fracture surgery is beneficial as it allows patient and surgical information to guide clinical decision-making regarding patient-specific management. However, the true clinical utility of this model remains unclear until it has been compared with clinician risk stratification[2].

REFERENCES:

1. Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C, Ravi B. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2023;6(3):e233391. doi:10.1001/jamanetworkopen.2023.3391
2. Martin R. *Predicting Anterior Cruciate Ligament Reconstruction Outcome: Machine Learning Analysis of National Knee Ligament Registries*. Norwegian School of Sport Sciences; 2025. https://hdl.handle.net/11250/3198948

TABLES AND FIGURES:

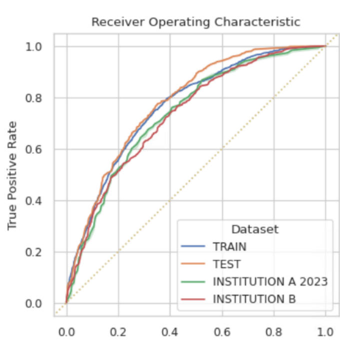

Figure 1.

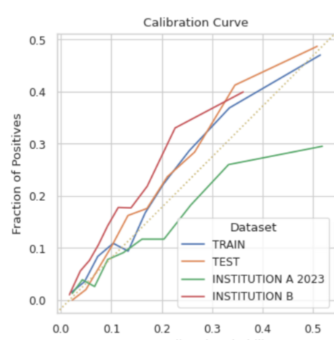

Figure 2.

Model	AUC-ROC	AUPRC	Accuracy
Gradient Boosting	0.77	0.40	0.71
Random Forest	0.74	0.38	0.82
SVM	0.77	0.40	0.68
Neural Network	0.76	0.40	0.66
Logistic Regression	0.78	0.41	0.68
XGBoost	0.73	0.34	0.74

Table 1.