

Electronic Health Record-Based Machine Learning Models for Predicting Revascularization, Amputation, and Cardiovascular Mortality in At-Risk Patients

Sydney E. Browder

sydney_browder@med.unc.edu

UNC Gillings School of Global Public Health

Mirjami Laivuori

Helsinki University Hospital

Tomas McIntee

North Carolina Translational and Clinical Sciences (TraCS) Institute

Peter J. Leese

North Carolina Translational and Clinical Sciences (TraCS) Institute

Olamide Alabi

Emory School of Medicine

Katharine L. McGinagle

UNC School of Medicine

Article

Keywords: Peripheral artery disease (PAD), chronic limb-threatening ischemia (CLTI), diabetic foot ulcer (DFU), Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), machine learning (ML), prediction

Posted Date: October 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10977718/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Patients with peripheral artery disease (PAD), chronic limb-threatening ischemia (CLTI), and/or diabetic foot ulcer (DFU) are at increased risk for limb loss and death. Existing risk calculators lack consistent clinical adoption; therefore, we aimed to identify vascular and non-vascular predictors of revascularization, amputation, and cardiovascular death using data from the OMOP CDM plus vascular extension. We assembled a cohort of adults with PAD, CLTI, or DFU receiving care within the UNC Health System from 2014–2024 (n = 178,995). Using 127 vascular-specific and thousands of non-vascular HER variables, we used logistic regression with inverse probability weighting and MCMC-based dimensionality reduction to predict revascularization, amputation, and cardiovascular death. The mean age was 72.5 ± 13.9 years; 50.9% were female and 23.5% Black. The revascularization and amputation models performed well (AUC = 0.984 and 0.926, respectively), while cardiovascular mortality prediction was modest (AUC = 0.772). Revascularization predictors were dominated by arterial disease, amputation prediction was driven by treatment interactions with dementia and frailty, and cardiovascular events and risk factors were the strongest predictors of cardiovascular mortality. Prediction models derived from harmonized EHR data demonstrated strong performance for revascularization and amputation. These findings support harmonized EHR-based vascular prediction modeling and warrant multi-institutional validation.

Introduction

Limb loss due to chronic illness is a devastating outcome, irreversibly altering the ability to mobilize unassisted and increasing the risk of contralateral amputation and death.[1] Patients with peripheral artery disease (PAD), especially in its most severe form, chronic limb-threatening ischemia (CLTI), and patients with diabetic foot ulcer (DFU) constitute the majority of patients at risk of limb loss.[2] The progression of these conditions is often the combination of years of poorly controlled medical comorbidities such as hypertension, hyperlipidemia, and diabetes, as well as the cumulative effects of clinical, social, and environmental factors. There is considerable heterogeneity in individual comorbidity and risk factor profiles, multiple treatment options, and variation in treatment response. This complexity complicates clinical decision-making, disease management, and prognostic evaluations for these patients. Thus, preserving life while simultaneously optimizing limb outcomes presents a significant clinical challenge with potentially serious repercussions for both the patient and the healthcare system.

To aid the clinician in navigating treatment and outcome possibilities, risk calculators and staging systems have been developed. While staging systems like Rutherford and Wifl (Wound, Ischemia, and foot Infection) have been adopted, none of the existing risk calculators for PAD have gained widespread acceptance.[3] This is partly due to the complexity of disease, making it impossible to narrow down to a few risk assessment variables. Recently, advancements in data processing capabilities and the use of machine learning (ML) have made it possible to analyze large scale datasets to discover previously unknown patterns of disease presentation and risk factor combinations that could aid in clinical decision-making and patient-level prognosis evaluation. The application of ML approaches to analyze

sets of predefined risk factors for limb loss in PAD patients has proven the method applicable in reinforcing the importance of known risk profiles. For example, CLTI is the strongest predictor of limb loss or death in patients undergoing revascularization for PAD.[4, 5]

Electronic health records (EHRs) contain rich longitudinal patient data.[6] Historically, vascular EHR datasets have been small, single-institution collections assembled for specific projects and typically require extensive manual data abstraction due to lack of uniform structure and varying clinician-specific documentation practices. To address these limitations, we developed a peripheral vascular extension to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), a widely adopted data architecture designed to standardize electronic health data across institutions.[7, 8] The vascular extension adds 127 disease-specific variables not available in the base OMOP CDM, which already includes more than 12,000 distinct conditions and observations, thereby enabling more comprehensive characterization of peripheral vascular disease and patients at risk of limb loss.

While there is broad consensus regarding vascular-specific variables associated with treatment decisions and outcomes, far less is known about the impact of traditionally non-vascular variables on vascular outcomes. Historically, the time-intensive nature of manual data collection has limited vascular datasets to a narrow set of variables, leaving the relationship between non-vascular variables and vascular outcomes largely unexplored. The integration of common vascular variables with thousands of non-vascular variables within the OMOP CDM plus vascular extension provides a unique opportunity to address this gap. Accordingly, our objective was to identify predictors of vascular-related treatments and outcomes in a large population of patients at risk of limb loss. To accomplish this, we developed predictive models for revascularization, amputation, and cardiovascular mortality using data derived from the OMOP CDM plus vascular extension.

Methods

This study was approved by the University of North Carolina (UNC) Institutional Review Board and the need for informed consent for participation was waived. All methods were carried out in accordance with relevant guidelines and regulations, including the Declaration of Helsinki.

Study Population

We previously identified a broad cohort of patients at risk of limb loss, including all adults with PAD, CLTI, or DFU who sought care within the UNC Health System between April 2014 and July 2024 to create a peripheral vascular-specific extension to the OMOP CDM. The ICD-10 codes used to identify these patients are outlined in Supplemental Table 1. Full details on how we built the OMOP peripheral vascular extension can be found in Leese et al.[7] For the peripheral vascular extension, we leveraged data from core OMOP CDM tables, which included demographic, diagnosis, procedure, medication, laboratory, encounter, vital status, and patient reported data, and then supplemented that data with peripheral vascular laboratory (PVL; ankle-brachial index and toe pressure) and external death data. These source

tables were then converted to intermediate tables for common multidimensional vascular-related areas (e.g. numerical PVL, smoking status, dialysis status, revascularization, amputation, etc.), and then to flat analytic datasets.

Outcomes

Revascularization and amputation procedures were identified using corresponding CPT and ICD-10 procedure codes and SNOMED concepts outlined in Supplemental Table 2. All-cause mortality was determined from EHR death status supplemented by NC state death certificate data and cardiovascular mortality was further determined by the presence of an “I” ICD code as the primary cause of death plus a second “I” code as the second or third cause of death.[9, 10]

Statistical Analysis

We built three separate logistic regression-based prediction models for revascularization, amputation, and death in patients at risk for limb loss. Utilizing the full OMOP source tables and intermediate tables, a patient summary table was developed that included relevant patient and disease features including vascular risk features across the OMOP content including revascularization, amputation, and mortality. In addition to literature and clinical informed variable selection, we also employed large-scale dimensionality reduction to the pool of approximately 12,000 distinct OMOP conditions and observations.

Feature selection

The primary source of potential features for the model were the OMOP-structured **condition_occurrence** and **observation** tables.[11] There were 12,508 conditions and observations represented in our data as features. Conditions occurring in fewer than 1,800 patients were consolidated into more prominent, related features that were, at most, 4 ancestor connections up in the **concept_ancestor** relation table. Preliminary OMOP features numbered 1,290 after this first consolidation, or rollup, stage.

A second rollup stage considered the immediate ancestor concepts of the 1,290 preliminary features. We compared logistic regressions of amputation and death using the ancestor concept, the collection of its descendants, and treatment variables. Quality of fit losses were calculated as the difference between quality of fit of the consolidated (ancestor + treatment) model and the unconsolidated (descendants + treatment) model. If this loss was < 0.001 for both death and amputation outcomes, the ancestor replaced its descendants, leaving 1,196 OMOP features.

The three main rollup parameters (1,800 patients, 4 levels of maximum initial rollup, 0.001 maximum acceptable quality of fit loss) were tuned based on final feature count. Physician review identified remaining clusters of medically closely related features. A final manual rollup consolidated 53 features into 6 diseases: chronic kidney disease, end stage kidney disease, diabetes, DFU, PAD, and peripheral

neuropathy. These were combined with 41 manually selected features, including the added vascular measurements, age, sex, prior revascularizations, and a window size proxy to adjust for time window size effects. After filtration, this resulted in a final feature pool of 1,276 variables usable by the models (Fig. 1).

Building the full model

With the dataset processed with a collection of features, it was then divided into a training set (fixed at $n = 100,000$) and a validation set (remaining eligible non-censored patients, $n = 70,590$).

The model for each outcome (amputation and CV-death) was a two-level model, consisting of a prediction first of the probability of revascularization (treatment), and then of the inverse propensity for treatment (IPTW)-corrected probability of subsequent death or amputation with the inclusion of revascularization as a predictor variable. For patients with at least one revascularization, a randomly selected revascularization was used as the canonical decision point. For patients without any revascularizations, a synthetic decision date was generated by random draw with replacement from the set of time periods from known initial vascular diagnoses to known revascularizations among treated patients; thus, both treated and untreated patients had similar distributions of time windows used for model evaluation.

The first stage of the model generated a prediction of the probability that the patient would receive revascularization at the given decision point based on features dated before that decision point. In the second stage of the model, the population was weighted based on IPTW to control for selection bias. Both death and amputation models shared a common first stage treatment model, as the calculation of the probability of treatment was not dependent on the final outcome analyzed.

Initial weights were calculated based on the Bayesian odds ratio of each binary feature as a predictor in isolation. Variable weights were updated by selecting 100 variables at random utilizing the current weights, adding the corresponding treatment interaction variables, generating a logistic regression using the **glmnet** library, and updating the weights based on the absolute value of the coefficient of the best fit logistic regression model. Coefficients of each variable in each model iteration were logged as an output for analysis purposes. The process of iterating over weights was a Markov chain Monte Carlo (MCMC) process used as a dimension selection (reduction) technique.

The final models were then built using a logistic regression over the 100 most consistently important variables selected via the MCMC process, along with their accompanying interaction variables. The eliminated variables, consisting of roughly 90% of the variable pool, were generally zero in most MCMC iterations.

The final models were evaluated using a receiver operator characteristic curve, taking the measurement of area under the curve (AUC) as an estimate of overall model quality. AUC for predictive models generally ranges from 0.5 (random assignment of scores) to 1.0 (perfect classification).

Results

We identified 178,995 adult patients at risk of limb loss during our study period. These patients had an average age of 72.5 ± 13.9 years, were 50.9% female, 23.5% Black, 23.3% were current or former smokers, 46.0% had diabetes mellitus, and 30.1% had chronic kidney disease.

Important Predictive Variables

Table 1 contains a condensed list of variables with model coefficients > 0.5 or ≤ -0.5 for each model; however, a complete list of variables and model coefficients can be found in Supplemental Table 3. The final treatment model had 158 non-zero coefficients (Fig. 1). The top three predictive variables for receiving a revascularization were atherosclerosis of the arteries of the lower limb, thrombosis of the arteries of the lower extremity, and peripheral vascular disease. Overall, the most important predictive variables for receiving treatment were related to arterial disease. Coefficients for variables predicting *not* getting treatment were mostly weaker in magnitude and suggestive of frailty and immobility, malignancies, and high burden of comorbidities. For the final amputation model, there were 145 non-zero coefficients. The top coefficients *predictive* of amputation were treatment interaction terms with dementia, open wound of the lower limb, and cachexia. The coefficients most *protective* against amputation were similar in magnitude and were also interaction terms with treatment. There were 94 non-zero coefficients in the final cardiovascular mortality model. The top three coefficients predictive of cardiovascular mortality were endocarditis, disorder due to type 2 diabetes mellitus, and hyperglycemia due to type 1 diabetes mellitus, which are broadly associated with cardiovascular disease, including peripheral vascular disease. Notably, treatment interaction weights for cardiovascular death were almost all zero, indicating that revascularization is neither predictive of nor protective against cardiovascular death.

Table 1

Predictive variables with model coefficients > 0.5 or < -0.5 for treatment, amputation, and cardiovascular mortality

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Atherosclerosis of artery of lower limb	2.01
Thrombosis of arteries of lower extremity	1.60
Peripheral vascular disease	1.56
Arterial embolus and thrombosis	1.55
Thrombosis	0.98
Chronic occlusion of artery of extremity	0.88
Stricture of artery	0.78
Disorder of cardiovascular system	0.73
Follow-up encounter	0.73
Osteodystrophy	0.69
Postoperative care	0.68
Coronary occlusion	0.65
Peripheral angiopathy due to diabetes mellitus	0.63
Requires vaccination	0.62
Deep venous thrombosis of lower extremity	0.56
Family history of cardiac disorder	0.53
Disorder of vein of lower extremity	0.53
Intermittent claudication	0.52
Chronic kidney disease stage 1	0.51
Cervical spine ankylosis	-0.50
Ascites	-0.52
Malignant neoplasm of urinary bladder	-0.53
Physician or other qualified health care professional services for outpatient cardiac rehabilitation	-0.56
Chronic mucositis	-0.57
Not for resuscitation	-0.57

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Orthotic(s) management and training (including assessment and fitting when not otherwise reported), upper extremity(ies), lower extremity(ies) and/or trunk initial orthotic(s) encounter, each 15 minutes	-0.58
Cerebral hemorrhage	-0.63
Bacterial pneumonia	-0.65
Pressure injury of sacral region of back	-0.69
Anemia in neoplastic disease	-0.94
Physician or allowed practitioner re-certification for medicare-covered home health services under a home health plan of care (patient not present), including contacts with home health agency and review of reports of patient status required by physician	-1.15
Ulnar neuropathy	-1.86
Model 2: Amputation	
Treatment*Cerebral degeneration presenting primarily with dementia	3.59
Treatment*Sequelae of open wound of lower limb	2.66
Treatment*Cachexia	2.55
Treatment*Intestinal obstruction co-occurrent and due to decreased peristalsis	2.33
Treatment*Undernutrition	2.19
Thickening of pleura	2.16
Decreased range of shoulder movement	2.11
Treatment*Open wound of toe without complication	2.06
Treatment*Foot pain	2.03
Pressure injury of back stage II	2.03
Treatment*Stricture of artery	1.95
Cholelithiasis and cholecystitis without obstruction	1.91
Treatment*Peripheral vascular disease	1.91
Chewing tobacco user	1.87
Treatment*Pain	1.85
Treatment*Gout	1.84
Treatment*Prediabetes	1.81

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Requires vaccination	1.80
Anxiety state	1.74
Disorder of fracture healing	1.68
Psychoactive substance withdrawal	1.61
Treatment*Family history of metabolic disorder	1.59
Ischemic foot ulcer due to atherosclerosis of artery of lower limb	1.58
Treatment*Lymphadenopathy	1.46
Treatment*Hyperglycemia due to type 1 diabetes mellitus	1.38
Treatment*Seizure	1.34
Treatment*Fluid volume disorder	1.23
Treatment*Arteriosclerotic vascular disease	1.21
Neurogenic dysfunction of urinary bladder	1.21
Absence of lower limb	1.11
Treatment*Dialysis finding	1.09
Disorder due to type 2 diabetes mellitus	1.08
Treatment*Arthritis of spine	1.05
Elevated blood-pressure reading without diagnosis of hypertension	1.03
Chronic osteomyelitis of ankle and/or foot	1.02
Treatment*Pericardial finding	1.02
Urine microscopy: RBC's present	0.99
Pain in lower limb	0.96
Tobacco dependence syndrome	0.93
Treatment*Disorder of fracture healing	0.92
Treatment*Absence of lower limb	0.87
Treatment*Lymphedema	0.87
Stiffness	0.86
Foot ulcer due to type 2 diabetes mellitus	0.84

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Gender identity	0.83
Disorder of muscle	0.79
Chronic ulcer of foot	0.78
Complication following amputation of limb	0.77
Aftercare	0.72
Treatment*Disorder of carotid artery	0.71
Chronic ulcer of ankle	0.68
Treatment*Disorder of amputation stump	0.66
Treatment*Rheumatic disease of heart valve	0.66
Treatment*Hypercholesterolemia	0.62
Treatment*Decreased range of shoulder movement	0.58
Hyperglycemia due to type 1 diabetes mellitus	0.58
Disorder of nervous system	0.55
Benign hypertensive renal disease	0.54
Type 2 diabetes mellitus	0.53
Gangrenous disorder	0.53
Treatment*Tobacco dependence syndrome	0.50
Treatment*Dermopathy due to type 2 diabetes mellitus	-0.51
Fluid volume disorder	-0.51
Long-term current use of anticoagulant	-0.52
Lymphadenopathy	-0.54
Hypercholesterolemia	-0.66
Treatment*Psychoactive substance withdrawal	-0.67
Treatment*Infection by methicillin sensitive Staphylococcus aureus	-0.69
Pressure injury of sacral region of back	-0.70
Treatment*Pressure injury of back stage II	-0.72
Treatment*Stiffness	-0.77

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Prediabetes	-0.79
Treatment*Pressure injury of sacral region of back	-0.85
Treatment*Disorder of muscle	-0.89
Disorder of phosphorus metabolism	-0.89
Asthenia	-0.91
Treatment*Disorder of nervous system	-0.95
Pain	-0.96
Treatment*Open wound of foot	-0.97
Treatment*Disorder due to type 2 diabetes mellitus	-1.06
Treatment*Benign hypertensive renal disease	-1.08
Treatment*Chronic ulcer of ankle	-1.09
Sequelae of open wound of lower limb	-1.11
Treatment*Ischemic foot ulcer due to atherosclerosis of artery of lower limb	-1.18
Seizure	-1.19
Dementia associated with another disease	-1.21
Gout	-1.23
Treatment*Pain in lower limb	-1.23
Treatment*Chronic ulcer of foot	-1.24
Foot pain	-1.32
Treatment*Aftercare	-1.33
Rheumatic disease of heart valve	-1.38
Stricture of artery	-1.49
Pericardial finding	-1.56
Treatment*Injury of ankle	-1.77
Cachexia	-2.04
Treatment*Pressure injury of foot	-2.15
Undernutrition	-2.20

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Open wound of toe without complication	-2.21
Treatment*Elevated blood-pressure reading without diagnosis of hypertension	-2.30
Treatment*Cholelithiasis AND cholecystitis without obstruction	-2.44
Treatment*Requires vaccination	-2.57
Treatment*Chewing tobacco user	-2.62
Intestinal obstruction co-occurrent and due to decreased peristalsis	-2.70
Treatment*Thickening of pleura	-2.71
Treatment*Anxiety state	-3.24
Model 3: Cardiovascular Mortality	
Endocarditis	1.54
Disorder due to type 2 diabetes mellitus	1.26
Hyperglycemia due to type 1 diabetes mellitus	1.10
Injury of ankle	1.02
Analgesic adverse reaction	0.99
Hearing loss of right ear	0.98
Pulmonary artery finding	0.82
Dysthymia	0.76
Burn any decree involving less than 10 percent of body surface	0.75
Cardiac arrest	0.59
Hypertensive crisis	0.57
Rectal hemorrhage	0.55
Cerebral edema	0.54
Bacterial infectious disease	-0.51
Cervical spine ankylosis	-0.51
Lung mass	-0.55
Traumatic injury of skeletal muscle	-0.58
Cataract	-0.69

Variable Name	Coefficient
Model 1: Treatment (Revascularization)	
Disorder of hematopoietic structure	-0.69
Clinical decision support mechanism national decision support company, as defined by the medicare appropriate use criteria program	-0.71
Benign neoplastic disease	-0.71
Pharyngitis	-0.73
Vaginitis	-0.74
Hypertension secondary to endocrine disorder	-0.79
Neoplasm of digestive system	-0.91
Osteoarthritis of wrist	-1.47
Contusion of hip	-1.59
Disorder of refraction AND/OR accommodation	-1.76

^Positive coefficients are predictive of the treatment or outcome, while negative coefficients are protective. Any variable beginning with "Treatment*" indicates an interaction between treatment and the following condition.

Prediction Model Performance

With 1,000 iterations of the MCMC cycle, the treatment model had exceptionally high performance (AUC = 0.984). A cross-validation check suggested that there was only modest overfitting, with model performance improving negligibly when the model was applied to its own training set (AUC = 0.984). This implies that revascularization treatments occur on a highly non-random basis, with revascularizations generally targeting patients strongly in need of revascularization. The amputation prediction model showed signs of strong performance with 1,000 MCMC iterations (AUC = 0.925). With 2,000 iterations, performance improved marginally (AUC = 0.926). The cardiovascular death prediction model had lower performance with 1,000 iterations of the MCMC cycle (AUC = 0.698) and improved significantly with 2,000 MCMC iterations (AUC = 0.772). Figure 2 displays the receiver operator curves for each model.

Discussion

In this study, we demonstrate a real-world application of longitudinal EHR data organized within the OMOP CDM plus vascular extension to predict revascularization, amputation, and cardiovascular death among patients at risk of limb loss. By leveraging harmonized, structured EHR data and ML methods, we provide an example of how large-scale clinical data can be used to model significant vascular outcomes.

All three prediction models converged on findings that are broadly consistent with established vascular disease biology, while also revealing where the relationship between revascularization and downstream outcomes becomes more clinically nuanced. The treatment model, which predicted receipt of revascularization, demonstrated exceptionally high performance, and its strongest individual predictive variables were uniformly related to arterial dysfunction. The amputation model also performed well with interaction terms between treatment and frailty-related variables being the strongest predictors of subsequent amputation. In the amputation model, revascularization was both predictive of and protective against amputation depending on the accompanying condition. In contrast, the cardiovascular death model had the lowest performance, even though its strongest predictors were unambiguously cardiovascular in nature. Taken together, these findings suggest that death, even when constrained to cardiovascular death, is inherently harder to predict than revascularization or amputation, due to the countless pathways by which a patient can die. Revascularization and amputation, by contrast, are outcomes highly correlated with the disease process itself, occurring as direct consequences of the same arterial and tissue pathology the models were built to capture.

Historically, studies leveraging EHR data were often limited to small, single institution cohorts and required labor-intensive manual data extraction from unstructured notes, frequently resulting in incomplete data for relevant patient characteristics. The increased adoption of artificial intelligence and ML approaches has enabled the analysis of substantially larger patient populations. Moreover, the use of a CDM like OMOP facilitates structural harmonization of data, allowing for reproducibility and comparison of results across institutions. Prior work using the general OMOP CDM has demonstrated its utility in predicting future cardiovascular events among patients with PAD. In one such study of 7,686 patients from two centers, coded information and unstructured text contributed more to model accuracy than laboratory values or prescriptions, achieving an AUC of 0.81 (95% CI, 0.80–0.83) using a random forest model.[12] Our work builds on this foundation by incorporating PAD-specific variables, including vascular laboratory measurements, thereby adding a new level of clinical specificity to the study of populations with vascular disease.

According to our results, the decision to revascularize patients appears to be driven by relatively uniform factors, primarily related to symptoms of arterial disease and related clinical findings such as atherosclerosis and thrombosis of the arteries in the lower extremities. This consistency may reflect the fact that, although the study population was large, it was drawn from a single healthcare system where clinical criteria and practice patterns for revascularization are likely similar across clinicians. Additionally, our finding that the strongest predictive variables in the treatment model are related to arterial dysfunction mirrors the emphasis on ischemia as a central driver of intervention.

Our amputation model found that revascularization could be either protective against or predictive of amputation depending on the accompanying condition, and the most influential treatment interaction terms were related to dementia and frailty. This pattern may be explained by treatment decision making among frail patients rather than by disease severity alone. Among patients who do not undergo revascularization, conditions signaling frailty or advanced illness may function as protective against

amputation, since patients judged too sick to tolerate surgery may be steered toward palliative end-of-life care instead of amputation. Among patients who do undergo revascularization despite significant frailty, dementia, or advanced CLTI, that same illness burden may instead become predictive of eventual amputation, since entering an active treatment pathway may increase the likelihood of complications.

Previously, attempts have been made to build prediction tools for amputation using clinically available variables. The most widely adopted of these is the Wifl scoring system, although it only applies to CLTI and DFU patients, not those with less severe versions of limb-based disease.[13] Other attempts at creating a prediction tool to assist clinical and patient decision-making include AMPREDICT and SHOW-ME PAD, neither of which have been widely accepted in clinical use.[14, 15] However, the fact that the amputation prediction model performed so well implies that amputation can be predicted with the correct clinical variables. Thus, future work should focus on delineating the correct coefficients available in clinical practice to incorporate in a prediction tool since, based on our results, predicting amputation using extracted EHR data is feasible.

Importantly, the three models in this study were not constructed using predefined risk factors or patient characteristics. Instead, the most influential predictors were identified algorithmically, leading to several key observations. Firstly, although ML approaches offer the potential to identify novel or previously underappreciated risk factors, this was not significantly observed in our analysis focused largely on SNOMED concepts. The most important variables for predicting revascularization, amputation, and cardiovascular death were clinically intuitive. This finding supports the robustness of existing clinical understanding of vascular disease, as reflected in tools such as Wifl classification system.[13] The absence of newly identified risk factors is consistent with prior studies, which have similarly found that limb-based disease severity, age, and comorbidity burden are the dominant predictors of amputation free survival.[16, 17]

Secondly, our findings illustrate that identifying clinically appropriate predictors does not guarantee strong model performance, particularly for an outcome like death that can be reached through numerous physiologic routes.[12] The cardiovascular death model's leading predictors were biologically plausible drivers of cardiovascular mortality, yet the model's overall performance remained low. This mirrors broader experience with ML-based cardiovascular risk prediction. A recent systematic review and meta-analysis of ML models built on EHR data reported pooled AUCs in roughly the 0.77–0.87 range, even for random forest and deep learning approaches, alongside substantial heterogeneity across studies and populations.[17] Similarly, in a PAD-specific cohort, an interpretable ML mortality model achieved an AUC of only 0.814 despite relying on well-established predictors such as age and Rutherford stage.[18] This may also explain why, unlike in the amputation model, treatment interaction terms in the cardiovascular death model were almost uniformly zero. Revascularization addresses regional arterial insufficiency but appears to have little bearing on the systemic cardiovascular processes that ultimately determine cardiovascular mortality in this population. Taken together, these results reinforce the need for cautious interpretation of mortality prediction models in complex vascular populations, even when their top predictors align with established clinical knowledge.

This is the first study to leverage the OMOP CDM plus vascular extension to predict vascular-related treatments and outcomes, demonstrating the feasibility of applying this standardized framework to clinically meaningful questions in vascular disease. However, the data were derived from a single health system in North Carolina, which may limit generalizability, particularly with respect to revascularization and amputation decisions. In addition, this cohort only included patients already known to have increased risk of limb loss due to chronic illness limiting generalizability in other patient cohorts with different risk profiles. However, the use of the OMOP CDM plus vascular extension facilitates reproducibility and provides a clear pathway for external validation across institutions and health systems. Future studies should extend this work by applying similar models in multicenter settings and by evaluating their performance in broader populations, including patients not currently at high risk for limb loss, to better understand the generalizability and clinical utility of these approaches.

Conclusions

In patients at risk of limb loss, prediction models derived from EHR datasets organized within the OMOP CDM plus vascular extension demonstrated strong performance for revascularization and amputation. Vascular-specific features were the dominant drivers of revascularization, treatment interaction terms with frailty-related variables were the strongest predictors of amputation, and cardiovascular-related mortality was largely influenced by common cardiovascular events and risk factors. These findings support the utility of harmonized EHR data for modeling vascular outcomes and highlight the need for broader, multi-institutional validation and application earlier in the disease course.

Declarations

Author Contributions:

Conceptualization: SB, ML, TM, PL, OA, KM

Data Curation: TM, PL

Formal Analysis: TM

Funding Acquisition: KM

Investigation: TM

Methodology: TM, PL

Project Administration: SB

Resources: N/A

Software: TM

Supervision: KM

Validation: TM

Visualization: TM, SB

Writing – original draft: SB, ML, KM

Writing – review & editing: SB, ML, TM, PL, OA, KM

Data Availability Statement:

The datasets generated and analyzed during the current study are not publicly available due to institutional data governance restrictions and the presence of protected health information. Study data were extracted from electronic health records and are housed on the University of North Carolina's Secure Health Informatics Research Environment (SHIRE) server under an IRB-approved protocol and institutional data use agreement; consistent with the OMOP Common Data Model's design for access via secure, distributed research networks rather than open public release, these data cannot be shared outside the institution. The datasets are, however, available from the corresponding author on reasonable request, subject to institutional and IRB approval.

Conflicts of Interest: In the last 24 months, Katharine McGinagle received research support from Orthofix; consulting fees from Medtronic, Vascular Technologies, Humacyte, and Orthofix; advisory board for Solventum.

Funding: This work was supported by the Foundation to Advance Vascular Cures

References

1. Glaser, J. D. et al. Fate of the contralateral limb after lower extremity amputation. *J. Vasc Surg.* **58** (6). 10.1016/j.jvs.2013.06.055 (2013). PubMed PMID: 23921246; PubMed Central PMCID: PMC3844063.
2. Conte, M. S. et al. Global vascular guidelines on the management of chronic limb-threatening ischemia. *J. Vasc Surg.* **69** (6S), 3. 10.1016/j.jvs.2019.02.016 (2019). PubMed PMID: 31159978; PubMed Central PMCID: PMC8365864S-125S.e40.
3. Preece, R. A. et al. A Systematic Review and Narrative Synthesis of Risk Prediction Tools Used to Estimate Mortality, Morbidity, and Other Outcomes Following Major Lower Limb Amputation. *Eur. J. Vasc Endovasc Surg. Off J. Eur. Soc. Vasc Surg.* ;**62**(1):127–135. doi:10.1016/j.ejvs.2021.02 (2021). .038 PubMed PMID: 33903018; PubMed Central PMCID: PMC11847594.
4. Li, B. et al. Predicting outcomes following lower extremity open revascularization using machine learning. *Sci. Rep.* **14**, 2899. 10.1038/s41598-024-52944-1 (2024). PubMed PMID: 38316811; PubMed Central PMCID: PMC10844206.

5. Li, B. et al. Predicting outcomes following open revascularization for aortoiliac occlusive disease using machine learning. *J. Vasc Surg.* **78** (6), 1449–1460e7. 10.1016/j.jvs.2023.07.006 (2023). PubMed PMID: 37454952.
6. Zhao, J. et al. Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. *Sci. Rep.* ;**9**(1):717. doi:10.1038/s41598-018-36745-x (2019).
7. Leese, P. J. et al. Extending the OMOP Common Data Model to Support Observational Peripheral Vascular Disease Research [Internet]. medRxiv; 2026 [cited 2026 Feb 13]. p. 2026.02.01.26345276. Available from: <https://www.medrxiv.org/content/10.64898/2026.02.01.26345276v1> doi:10.64898/2026.02.01.26345276.
8. Stang, P. E. et al. Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership. *Ann. Intern. Med.* **153** (9), 600–606. 10.7326/0003-4819-153-9-201011020-00010 (2010).
9. Powers, J. P., Nandhakumar, S., Dard, S. Z., Kovach, P. & Leese, P. J. Recovering missing electronic health record mortality data with a machine learning-enhanced data linkage process. *J. Am. Med. Inf. Assoc.* **32** (6), 1061–1065. 10.1093/jamia/ocaf060 (2025).
10. Instructions for Classifying the Underlying Cause-of-Death. ICD-10, 2025 [Internet]. 2025 [cited 2026 Sep 2]. Available from: <https://www.cdc.gov/nchs/nvss/manuals/2025/2a-2025.html>
11. Informatics, O. H. D. S. and. The Book of OHDSI [Internet]. [cited 2026 Feb 13]. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>
12. Ross, E. G. et al. Predicting Future Cardiovascular Events in Patients With Peripheral Artery Disease Using Electronic Health Record Data. *Circ. Cardiovasc. Qual. Outcomes.* **12** (3), e004741. 10.1161/CIRCOUTCOMES.118.004741 (2019).
13. Mills, J. L. et al. The Society for Vascular Surgery Lower Extremity Threatened Limb Classification System: Risk stratification based on Wound, Ischemia, and foot Infection (Wlfi). *J. Vasc Surg.* **59** (1), 220–234e2. 10.1016/j.jvs.2013.08 (2014). .003 PubMed PMID: 24126108.
14. Czerniecki, J. M. et al. The development and validation of the AMPREDICT model for predicting mobility outcome after dysvascular lower extremity amputation. *J. Vasc Surg.* **65** (1), 162–171e3. 10.1016/j.jvs.2016.08.078 (2017). PubMed PMID: 27751738; PubMed Central PMCID: PMC7521625.
15. Smolderen, K. G. et al. Treatment decisions for patients with peripheral artery disease and symptoms of claudication: Development process and alpha testing of the SHOW-ME PAD decision aid. *Vasc Med. Lond. Engl.* **26** (3), 273–280 (2021). 20988780 PubMed PMID: 33627058; PubMed Central PMCID: PMC8169642.
16. Goffart, S. et al. Deep Learning Strategies for Predicting Amputation Free Survival in Patients with Peripheral Artery Disease. *Eur. J. Vasc Endovasc Surg. Off J. Eur. Soc. Vasc Surg.* **71** (2), 293–303. 10.1016/j.ejvs.2025.10.043 (2026). PubMed PMID: 41151636.

17. Liu, Y., Xue, J. & Jiang, J. Application of machine learning algorithms in electronic medical records to predict amputation-free survival after first revascularization in patients with peripheral artery disease. *Int. J. Cardiol.* **383**, 175–184. 10.1016/j.ijcard.2023.04.040 (2023).

18. Li, Y. et al. Prediction of mortality risk in patients with peripheral artery disease using interpretable machine learning. *Front. Cardiovasc. Med.* **13**, 1879240. 10.3389/fcvm.2026.1879240 (2026). PubMed PMID: 42638675; PubMed Central PMCID: PMC13500288.

Figures

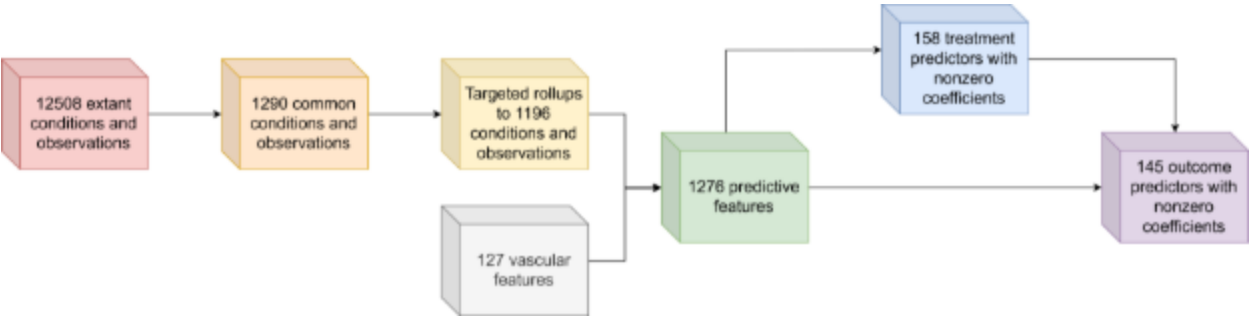


Figure 1

Feature selection for amputation prediction model

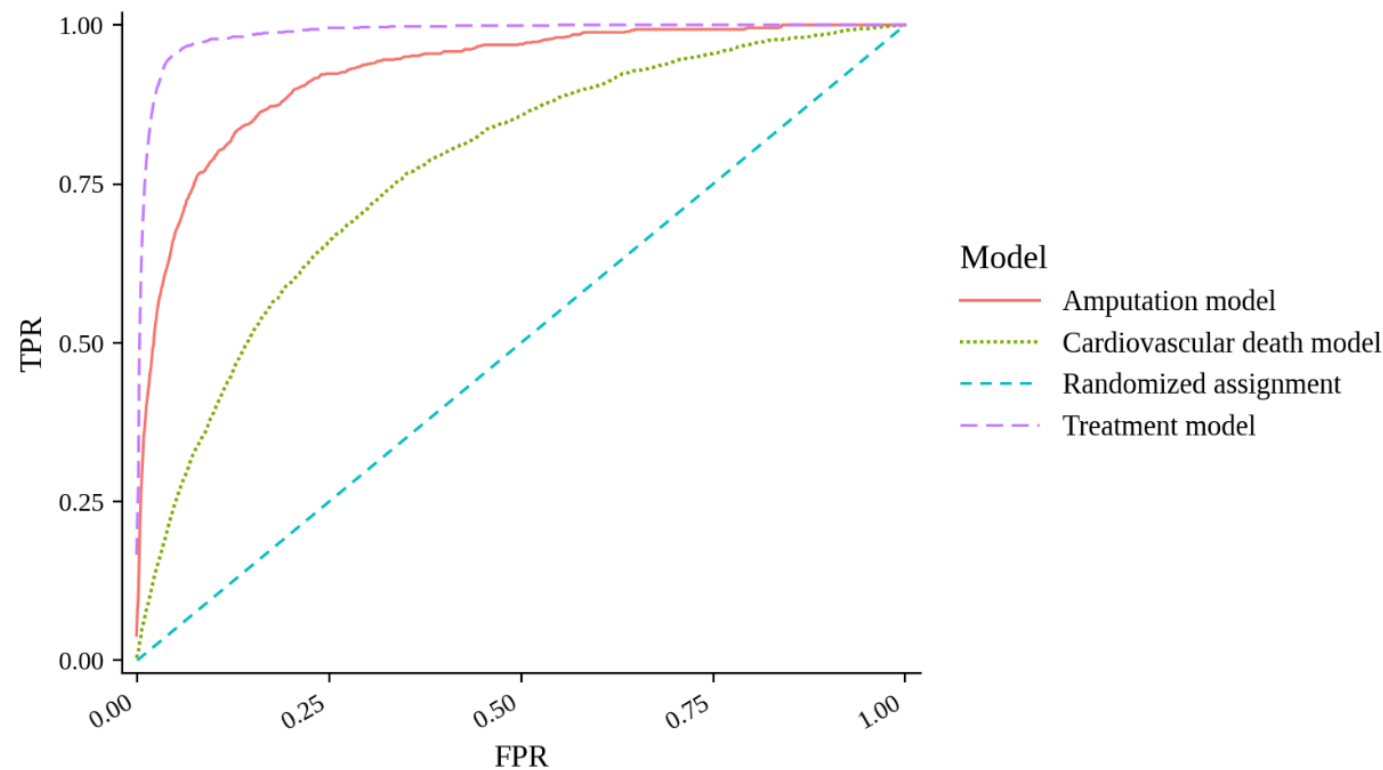


Figure 2

Receiver operator curves (ROCs) for our three prediction models

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalTables.docx](#)