Machine Learning Methodology Identifies Predictors of a Cardiovascular Composite Measure Among Severe Peripheral Artery Disease Patients

Windsor Ting[1], Lloyd Haskell[2], Fedor Lurie[3], Jeffrey S. Berger[4], Zubin Eapen[5], Matthew Valko[6], Veronica Alas[6], Kelly Rich[6], Concetta Crivera[2], Jeffrey R. Schein[2]

[1]Mt. Sinai Hospital, New York, NY, USA, [2]Johnson and Johnson, Raritan, NJ, USA,[3]ProMedica, Toledo, OH, USA, [4]New York University Langone Medical Center, New York, NY, USA, [5]Duke University Hospital, Durham, NC, USA, [6]GNS Healthcare, Cambridge, MA, USA

Introduction

- Patients with severe PAD who go untreated have a 20%-40% greater risk of death and/or amputation within 1 year of severe PAD diagnosis, compared with treated severe PAD patients. An aging population, with associated comorbidities, increases the likelihood of these risks as well as their subsequent interventions
- Diabetes significantly elevates the risk of morbidity and mortality in severe PAD patients,[2] and 27% to 76% of patients with chronic critical limb ischemia (severe PAD) have diabetes.[3]
- Use of machine learning methods, including a multivariate rather than univariate correction, can provide new or confirmatory insights into the factors that increase the risk of cardiovascular outcomes in severe PAD patients.
- Machine learning methods can identify risk factors from large numbers of known or unknown predictors, without bias. This allows for more robust and complete clinical findings based on the data itself.

Objectives

- The purpose of this study was to identify risk factors for major adverse cardiovascular events (MACE) in a severe PAD population, in the 12 months following diagnosis. MACE included stroke, myocardial infarction (MI), major lower extremity amputation, and all-cause death. We further examined the association of diabetes with MACE.
- GNS Healthcare's novel machine learning platform, Reverse Engineering Forward Simulation (REFS™), was used to accomplish this goal.

Methods

Data Source

- This retrospective study used a large, integrated US dataset from Optum + Humedica administrative claims and EMR databases (January 1, 2007 - September 30, 2015), representing a wide cross-section of severe PAD patients.
- This integrated database includes comprehensive data for over 5 million patients throughout 20 states.

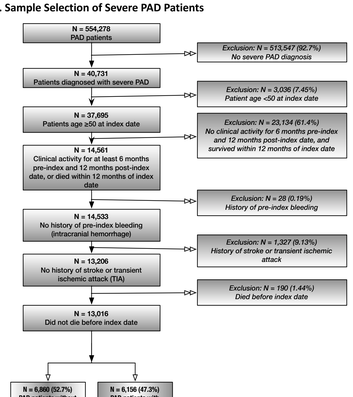
Inclusion Criteria

- Diagnosis of severe PAD, defined as rest pain, ulceration, or gangrene in extremities, in conjunction with PAD, between January 1, 2007 – September 30, 2015 (first available diagnosis was identified as index date).
- Age ≥50 at the index date.
- At least 6 months of clinical activity in the pre-index period.
- At least 12 months of clinical activity in the post-index period, or documentation of death during the 12-month post-index period.

Exclusion Criteria

- History of bleeding, stroke, or transient ischemic attack (TIA) in the pre-index period.

Figure 1. Sample Selection of Severe PAD Patients



Study Cohorts

- Patients were stratified into two cohorts:
  1. Diabetes: Identification of a diabetes diagnosis in the 6-month pre-index period
  2. No diabetes: No prior diagnosis of diabetes in the 6-month pre-index period

Study Outcome

- MACE was defined as the presence of at least one of the following cardiovascular events during the 12-month post-index period:
  1. Stroke
  2. MI
  3. Major lower extremity amputation (below knee, except toe)
  4. All-cause death

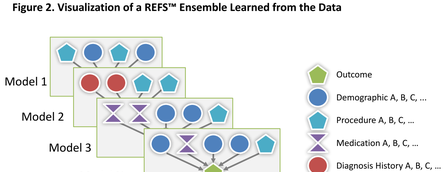
Study Covariates

- Demographic characteristics: age, gender, plan type, geographic location, race, and ethnicity.
- Clinical characteristics:
  - Comorbidities and risk factors: Pre-index period PAD-related comorbidities and symptoms were quantified by mapping the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes to Clinical Classification Software (CCS) level 2 and 3 codes. Smoking status and alcohol status were assessed closest to the index date.
  - Severity score: Created for this study and defined as the total number of severe PAD factors in the pre-index period. The factors used in the score were rest pain, gangrene, ulceration, and amputation.
  - Charlson Comorbidity Index (CCI): CCI is a method to classify the severity of a patient's health, based on comorbidities of the patient.[4] CCI was calculated for each patient during the 6-month pre-index period.
- Healthcare resource use (HRU):
  - Medication use: Cardiovascular and diabetic treatments within the pre-index period were quantified by mapping treatments from National Drug Code (NDC) Directory to the Medi-Span drug ontology (level 2 and 3 codes).
  - Provider visits: Pre-index period visit to cardiologist or internal medicine physician were evaluated.

Statistical Analysis

- Descriptive statistics were computed to summarize MACE, demographics, clinical characteristics, and HRU for the overall severe PAD population as well as both the diabetic and non-diabetic cohorts.
- F-test was used for continuous study measures. Chi-squared test was used to evaluate the statistical differences between MACE and known predictor, diabetes status.
- A proprietary machine learning method, REFS™, was used to identify covariates of MACE among severe PAD patients.
  - REFS™ learns an *ensemble*, a collection of Bayesian logistic regression models (n=256), from the data using a hypothesis-free methodology based on a Bayesian scoring algorithm (Figure 2).
  - REFS™ can explore relationships between large numbers of variables (including outcomes, demographics, procedures, medications, and diagnosis history) and the interactions between them.
  - Covariates that predicted the outcome of interest were selected by REFS™ based on the data.

Figure 2. Visualization of a REFS™ Ensemble Learned from the Data



This study contained 256 logistic regression models; Figure 2 represents our ensemble of models. Each model contains a different set of covariates, and we included the "best fit" logistic regression models as selected by a machine learning algorithm.

Results

Demographics

- The study sample contained 13,016 eligible severe PAD patients, and 47.3% had diabetes (Figure 1).
- Overall mean age was 70.2 years, and females comprised 44.5% (Table 1).

Table 1. Severe PAD Demographics by Cohort

Demographic Characteristic	Overall	No Diabetes Cohort	Diabetes Cohort[1]	P-Value
Number of Patients (N)	13,016	6,860	6,156	
Age group (%)				<0.001
50 - 54	6.91	6.33	7.55	
55 -64	22.66	20.87	24.66	
65-74	30.27	28.05	32.75	
>75	40.16	44.75	35.04	
Age: Mean (SD)	70.16 (9.40)	70.92 (9.46)	69.31 (9.27)	<0.001
Gender (%)				<0.001
Male	55.49	51.71	59.70	
Female	44.48	48.28	40.24	
Unknown	0.04	0.01	0.06	
Plan type (%)				<0.001
Commercial	21.53	21.76	21.26	
Medicaid	1.25	1.02	1.51	
Medicare	32.68	31.52	33.97	
Other	1.65	1.49	1.84	
Uninsured	2.40	1.94	2.91	
Unknown	40.50	42.27	38.52	
Geographic location (%)				<0.001
Northeast	13.68	14.36	12.93	
Midwest	36.59	36.91	36.22	
South	37.57	35.60	39.77	
West	9.52	10.35	8.59	
Unknown	2.64	2.78	2.49	
Race (%)[2]				<0.001
African American	10.87	8.66	13.34	
Asian	0.90	0.92	0.88	
Caucasian	75.98	78.72	72.94	
Other/Unknown	12.25	11.71	12.85	
Ethnicity (%)				<0.001
Hispanic	2.90	1.56	4.39	
Not Hispanic	77.68	79.31	75.86	
Unknown	19.42	19.13	19.75	

[1] Diabetes was evaluated for those patients with a history of recorded ICD-9 250.XX

[2] Race and ethnicity overlap and are not mutually exclusive; patients under one category may be reflected in the other.

Clinical Characteristics

- Overall, the most common PAD-related comorbidities among severe PAD patients were hypertension, dyslipidemia, and coronary artery disease. Patients with diabetes had a higher proportion of these comorbidities compared to patients without diabetes (Table 2).
- The occurrence of ulceration was greater in the leg or foot (35.48%) as compared to ankle or foot (27.99%) among severe PAD patients. Gangrene occurred in approximately 14% of patients. Compared to patients without diabetes, patients with diabetes had a higher proportion of both ulceration and gangrene (Table 2).
- Overall, the majority of severe PAD patients had at least one of the severity score factors (59.67%) in their pre-index period (Table 2).

MACE

- MACE occurred in approximately 28.5% of severe PAD patients (Figure 3). A greater proportion of patients with diabetes showed MACE as compared to patients without diabetes (33.6% vs. 24.0%, p<0.001).
- In general, the four conditions of MACE (stroke, MI, major lower extremity amputation, all-cause death) were found to occur more frequently in patients with diabetes as compared to patients without diabetes. All-cause death had the highest difference between patients with and without diabetes (15.7% vs. 11.4%, p<0.001) (Figure 3).

Cardiovascular Medications

- Approximately 29% of severe PAD patients were treated with statins in the pre-index period, and patients with diabetes had higher cardiovascular medication use compared to patients without diabetes (Figure 4).

Top Predictors of MACE Using REFS™

- REFS™ selected the following pre-index conditions to be highly predictive of MACE among severe PAD patients (selection frequency [SF], OR, SD): acute MI (100%, 3.5, 0.2), congestive heart failure (99.6%, 1.9, 0.1), acute and unspecified renal failure (93.4%, 1.8, 0.1), gangrene (79.3%, 2.3, 0.1), and diseases of the heart (57.4%, 1.5, 0.05) (Table 3).
- Although not a high SF, diabetes with neurological manifestations was the only diabetes-related diagnosis that was selected to be highly predictive of MACE among severe PAD patients (0.4%).

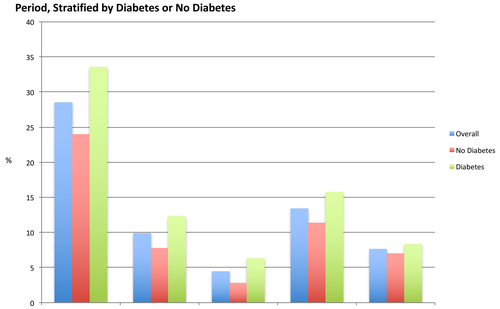
Table 2. Clinical Characteristics of Severe PAD in the Pre-Index Period by Cohort

Clinical Characteristic	Overall	No Diabetes Cohort	Diabetes Cohort[1]	P-Value
Number of Patients (N)	13,016	6,860	6,156	
PAD-Related Comorbid Conditions (%)				
Hypertension	73.56	63.63	84.62	<0.001
Dyslipidemia	56.78	47.39	67.24	<0.001
Coronary artery disease	41.08	33.73	49.27	<0.001
Coronary atherosclerosis and other heart disease	41.00	33.60	49.25	<0.001
Acute MI	4.96	3.56	6.51	<0.001
Acute and unspecified renal failure	12.65	7.80	18.05	<0.001
Congestive heart failure; nonhypertensive	21.23	14.65	28.56	<0.001
Cardiac dysrhythmias	27.85	26.21	29.68	<0.001
Atrial fibrillation and atrial flutter	18.15	16.60	19.87	<0.001
Amputation all	2.01	0.76	3.40	<0.001
Amputation minor	1.25	0.29	2.32	<0.001
Amputation major	0.81	0.50	1.17	<0.001
Type 1 diabetes	8.13	0.00	17.19	<0.001
Type 2 diabetes	39.17	0.00	82.81	<0.001
Diabetes with neurological manifestation	17.23	0.01	36.42	<0.001
PAD Symptomology				
Rest Pain	38.15	44.21	31.38	<0.001
Chronic ulceration of leg or foot	35.48	23.37	48.98	<0.001
Ulceration of ankle or foot	27.99	16.30	41.02	<0.001
Gangrene	13.54	8.73	18.89	<0.001
CCI Score: Mean (SD)	6.42 (2.71)	5.29 (2.20)	7.68 (2.68)	<0.001
Severity Score (%)[2]				<0.001
0	29.95	35.41	23.86	
1	59.67	59.52	59.84	
2	9.14	4.74	14.05	
3	1.23	0.34	2.23	
4	0.01	0.00	0.02	
Smoking in Pre-Index (%)				<0.001
Not Current	36.70	32.99	40.84	
Current	13.60	15.50	11.48	
Unknown	49.70	51.52	47.68	
Alcohol use in Pre-Index (%)				0.345
Not Current	0.44	0.39	0.49	
Current	87.80	88.16	87.39	
Unknown	11.76	11.44	12.12	

[1] Diabetes was evaluated for those patients with a history of recorded ICD-9 250.XX

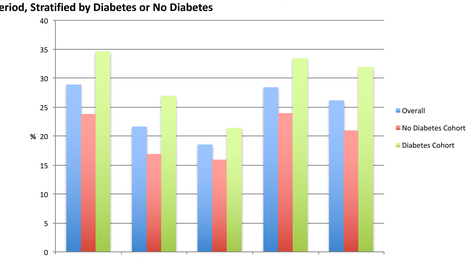
[2] Severity score was evaluated for those patients with a history of recorded amputation, gangrene, ulceration, and rest pain. The score was computed by summing the total number of comorbidities a person had in the pre-index period, specifically amputation, gangrene, ulceration, and rest pain.

Figure 3. Major Adverse Cardiovascular Events in Severe PAD Patients in Post-Index Period, Stratified by Diabetes or No Diabetes



* p< 0.05; ** p<0.001

Figure 4. Most Common Cardiovascular Medications Among Severe PAD Patients in Pre-Index Period, Stratified by Diabetes or No Diabetes



** p<0.001. Aspirin medication use was evaluated in the population; however, evidence of use could only be assessed via prescription claims and not over-the-counter treatment. Hence, we noted aspirin use was likely underreported in the population.

Table 3. Top Predictors of MACE from REFS™ Modeling

Description	Input Covariate	Selection Frequency	Odds Ratio Mean	Odds Ratio SD
Comorbidity	Acute MI	100.00	3.52	0.23
Comorbidity	Congestive Heart Failure; nonhypertensive	99.61	1.93	0.12
Comorbidity	Acute and Unspecified* Renal Failure	93.36	1.75	0.09
Comorbidity	Gangrene	79.30	2.26	0.10
Comorbidity	Diseases of the Heart	57.42	1.50	0.05
Comorbidity	Chronic Ulcer of Skin	41.02	1.53	0.07
Comorbidity	Chronic Ulcer of Leg or Foot	36.72	1.44	0.06
Demographic	Age^	30.47		
	55 – 64		1.18	0.01
	65 - 74		1.55	0.03
	>75		2.31	0.06
Comorbidity	Coronary Atherosclerosis and Other Heart Disease*	17.58	1.35	0.02
Comorbidity	Cardiac Dysrhythmias	14.06	1.49	0.05

Note: Example interpretation of selection frequency findings: If a covariate, input 1, has 20% selection frequency this means that the covariate showed up in 20% of the models selected by REFS™ to best predict the outcome. The mean OR (SD [sample SD]) will provide insight into the distribution of the effect of a covariate on an outcome when different sets of covariates are present and controlled for across models.

*Unknown or unspecified and other diagnoses of a given disease can be included based on the CCS ontology system.

OR Interpretation: The average odds ratios (AORs) are provided for all 134 variables across 256 predictive models to assess the overall impact of these covariates on the outcome. OR > 1 indicates a positive relationship between the variable and the probability of the outcome, and vice versa. OR = 1 indicates that the covariate has no impact on predicting the outcome. For example, acute MI has an AOR of 3.52. This means patients who have been diagnosed with acute MI are more likely to have a composite cardiovascular measure when other variables are the same. The small standard deviations associated with the AORs indicate that the resulting ORs associated with each variable from different models are very close together.

Overall Cross Validated AUC = 0.71

^Note: Age was selected as a single top predictor of MACE; within the age category, three unique age range groups, with distinct odds ratios, indicate the effect of the range on the outcome of interest.

Limitations

- This study used a non-randomized, observational design. The risk of confounding from unmeasured risk factors can impact accurately measuring MACE and finding appropriate risk factors.

Conclusions

- REFS™ found confirmatory risk factors that increase incidence of MACE among severe PAD patients in a geographically diverse study. Acute MI, gangrene, and congestive heart failure were found to be the strongest predictors.
- Descriptively, diabetes was a significant predictor of MACE, but when adjusted with other predictors, REFS™ did not select it with a high frequency.

References

1. Chung J, Modrall JG, Valentine RJ. The need for improved risk stratification in chronic critical limb ischemia. J Vasc Surg. 2014; 60:1677-1685.
2. Shammas NW. Epidemiology, classification, and modifiable risk factors of peripheral arterial disease. Vasc Health Risk Manag. 2007;3:229-234.
3. Dick F, Diehm N, Galimanis A, et al. Surgical or endovascular revascularization in patients with critical limb ischemia: Influence of diabetes mellitus on clinical outcome. J Vasc Surg. 2007; 45:751-761.
4. D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative databases. J Clin Epidemiol. 1996; 49:1429-1433.

Disclosures

- Sponsorship: *Janssen Scientific Affairs, LLC*
- W. Ting: Consultant/Advisory Board; Significant; Janssen Pharmaceuticals.
- L. Haskell: Employment; Significant; Janssen Pharmaceuticals.
- F. Lurie: Consultant/Advisory Board; Modest; Janssen Pharmaceuticals.
- J.S. Berger: Consultant/Advisory Board; Significant; Janssen Pharmaceuticals, AstraZeneca.
- Z. Eapen: Ownership Interest; Significant; Pattern Health Technologies. Consultant/Advisory Board; Modest; Amgen, Cytokinetics, Janssen, Medtronic, Myokardia, SHL Telemedicine. Consultant/Advisory Board; Significant; Novartis.
- M. Valko: Consultant/Advisory Board; Modest; Janssen Pharmaceuticals.
- V. Alas: Consultant/Advisory Board; Modest; Janssen Pharmaceuticals.
- K. Rich: Consultant/Advisory Board; Modest; Janssen Pharmaceuticals.
- C. Crivera: Employment; Significant; Janssen Pharmaceuticals.
- J. Schein: Employment; Significant; Janssen Pharmaceuticals.

PRESENTED AT THE AMERICAN HEART ASSOCIATION SCIENTIFIC SESSIONS; NOVEMBER 12 – 16, 2016; NEW ORLEANS, LOUISIANA