Machine Learning Methodology Predicts Comorbidities are Associated with Increased Total Healthcare Costs among Patients with Severe Peripheral Artery Disease

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

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

Introduction

- Peripheral artery disease (PAD) is manifested over a continuum of severity, with severe PAD patients at greater risk for comorbid coronary artery disease, [1] diabetes, and renal disease and related complications.[2]
- Increased number of comorbidities contributes to greater healthcare resource utilization (HRU)[3] and healthcare costs (HC).[4] An aging population, with increased prevalence of severe PAD and associated comorbidities, may increase these costs.
- Applying machine learning methods to identify clinical factors, from hundreds to thousands of covariates, associated with healthcare costs among severe PAD patients can lead to more accurate cost predictions.

Objectives

- The purpose of this study was to understand HRU and annual total all-cause HC in a severe PAD population, in the 12 months following PAD diagnosis.
- This goal was accomplished using Reverse Engineering Forward Simulation (REFS™), GNS Healthcare's novel machine learning platform.

Methods

Data Source

- Optum + Humedica, administrative claims and EMR databases, were used in this retrospective study (January 1, 2007 - September 30, 2015).
- This integrated database, spanning 20 states, is comprised of comprehensive data for more than 7 million patients.

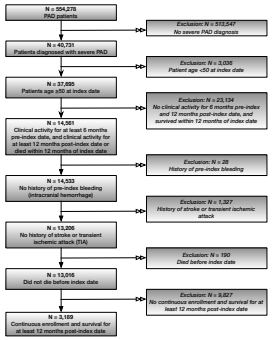
Inclusion Criteria

- 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; Figure 1).
- Patients age ≥50 at the index date.
- Patients with at least 6 months of clinical activity in the pre-index period.
- Patients with 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

- Patients with a history of intracranial bleeding, stroke, or transient ischemic attack (TIA) in the pre-index period.

Figure 1. Sample Selection of Severe PAD Patients



Study Outcome

- All-cause total HC were summarized over the 12-month post-index period and included the following categories of care: inpatient, outpatient, emergency department (ED), and pharmacy.

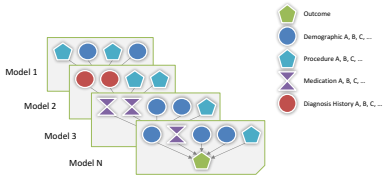
Study Covariates

- Demographic characteristics included gender, age, plan type, geographic location, race, and ethnicity.
- Clinical characteristics:
  - Comorbidities, risk factors, and specialist visits: 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. Specialist visits to cardiology and internal medicine were assessed during the 6-month pre-index period.
  - Severity score: The factors used in this score were rest pain, gangrene, ulceration, and amputation. The severity score was created for this study and defined as the total number of pre-index severe PAD factors.
  - Charlson Comorbidity Index (CCI): Patient comorbidities were used to calculate the CCI score, which classifies the severity of a patient's health.[5] CCI was calculated for each patient during the 6-month pre-index period.
- Healthcare resource use (HRU):
  - All-cause hospitalizations, ED visits, and length of stay (LOS) were assessed in the 12-month post-index period.

Statistical Analysis

- Demographics, clinical characteristics, HRU, and HC were summarized using descriptive statistics.
- A proprietary machine learning method, REFS™, was used to identify predictors of HC among patients with severe PAD.
  - REFS™ uses Bayesian scoring algorithms to reverse-engineer an ensemble of individual prediction models empirically from data, without a priori hypotheses (Figure 2).
  - REFS™ can explore relationships between large numbers of variables (including outcomes, demographics, procedures, medications, and diagnosis history) and the interactions between them.
  - REFS™ prediction ensembles provide information about the relative impact of specific patient factors on costs and the predictive value of available patient factors in a dataset.

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



This study contained 256 multiple linear regression models; Figure 2 represents our ensemble of models. Each model contains a different set of covariates, and we included the "best fit" multiple linear regression models as selected by a machine learning algorithm.[6]

Results

Demographics

- The study sample contained 3,189 eligible severe PAD patients (Figure 1).
- Overall mean age was 71.9 years, and females comprised 47.5% (Table 1).
- Most severe PAD patients were from the Midwest (36%) and South (34%) (Table 1).

Table 1. Severe PAD Demographics

Demographic Characteristic	Severe PAD Patients
Number of Patients (N)	3,189
Age group (%)	
50 - 54	5.02
55 -64	17.25
65-74	30.20
≥75	47.54
Age: Mean (SD)	71.90 (9.02)
Gender (%)	
Male	52.43
Female	47.51
Unknown	0.06
Plan type (%)	
Commercial	25.15
Medicare	21.45
Medicaid	0.75
Other	1.07
Uninsured	1.07
Unknown	50.52
Geographic location (%)	
Midwest	35.65
South	33.90
Northeast	14.46
West	13.14
Unknown	2.85
Race (%)	
Caucasian	74.04
African American	9.31
Asian	0.78
Other/Unknown	15.87
Ethnicity (%)	
Hispanic	2.70
Not Hispanic	73.91
Unknown	23.39

Data Source: Optum+Humedica, 01/01/2007 - 09/30/2015

Abbreviations: PAD, peripheral artery disease; SD, standard deviation.

Clinical Characteristics

- The most common PAD-related comorbidities among severe PAD patients in the pre-index period were neuropathy (59%), coronary artery disease (43%), and rest pain (39%). The mean pre-index period CCI was 6.91 (SD 2.78) (Table 2).
- The majority of severe PAD patients had at least one of the severity score factors in their pre-index period (71%).
- Approximately 32% of severe PAD patients had a cardiologist visit in the pre-index period.

HRU

- At least 18% of patients had at least one ED visit and 26% had two or more ED visits in the post-index period. The mean number of ED visits was 1.28 (SD 2.60) (Figure 3).
- Approximately 22% of severe PAD patients had at least one hospitalization and 32% had more two or more hospitalizations in the post-index period. The mean number of all-cause hospitalizations was 1.29 (SD 1.81), with an average LOS of 8.02 days (SD 18.21) (Figure 3).

HC

- The mean annual total all-cause HC per patient was $56,973 (SD $91,523; Table 3).
- Inpatient treatment incurred the greatest spending, followed by outpatient and pharmacy costs.

Top Predictors of HC Using REFS™

- Highly predictive factors associated with increased costs were other diseases of kidney and ureters (mean percentage change in costs (MPCC), [SD]: 2.20, [0.17]), chronic ulcer of skin (1.95, [0.13]), and chronic ulcer of leg or foot (1.93, [0.11]) (Table 4).
- Chronic kidney disease (CKD; 1.9, [0.2]), and cellulitis and abscess (1.8, [0.2]) were also predictive of increased HC.
- The mean 5-fold cross-validated R-squared was 0.12.

Table 2. Pre-Index Comorbidities, CCI, Severity Score, and Specialist Visits of Severe PAD Patients

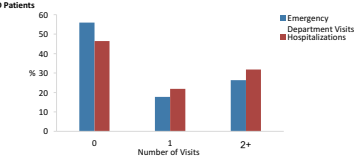
Clinical Characteristics	Severe PAD Patients
Number of Patients (N)	3,189
PAD-Related Comorbid Conditions (%)	
Neuropathy	58.95
Coronary Artery Disease	42.55
Rest Pain	38.98
Ulceration	29.88
Chronic Kidney Failure/End-Stage Renal Disease	28.66
Heart Failure	23.74
Atrial Fibrillation and Atrial Flutter	19.76
Renal Disease	17.06
Gangrene	12.51
Amputation All	2.60
Amputation Minor	1.82
Amputation Major	0.78
Type 1 Diabetes	10.44
CCI Score (Mean, SD)[2]	6.91 (2.78)
CCI Score (%)	
0	0.00
1	0.13
2	1.94
3	5.21
4+	92.72
Severity Score (%)[1]	
0	28.54
1	60.49
2	9.44
3	1.54
Any specialist visits (%)	
Internal medicine	31.92
Cardiologist	18.25

Data Source: Optum+Humedica, 01/01/2007 - 09/30/2015

Abbreviations: PAD, peripheral artery disease; CCI, Charlson Comorbidity Index; SD, standard deviation

[1] 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 these four comorbidities a person had in the pre-index period.

Figure 3. All-Cause Post-Index Emergency Department Visits and Hospitalizations among Severe PAD Patients



Note: Mean number of emergency department visits was 1.28 (standard deviation 2.60). Mean number of hospitalizations was 1.29 (SD 1.81), and mean length of stay was 8.02 (SD 18.21) days. Abbreviations: PAD, peripheral artery disease.

Table 3. Post-Index Annual Total Healthcare Costs among Severe PAD Patients

Costs	Mean (SD)
Number of Patients (N)	3,189
All cause ($)	
Total costs	56,972.71 (91,523.35)
Inpatient costs	29,215.83 (65,832.97)
Outpatient costs	22,612.03 (53,005.16)
Pharmacy costs	3,792.10 (7,787.51)
Emergency Department costs	1,352.75 (3,356.90)

Data Source: Optum + Humedica, 01/01/2007 - 09/30/2015

Abbreviations: PAD, peripheral artery disease; SD, standard deviation

[1] PAD-related costs were drawn from combined costs of antiplatelet, oral anticoagulation therapy (OAC), lipid-lowering, anti-heart failure, anti-angina, and anti-hypertensive medications.

Medication use was assessed during the 12-month post-index period, starting one day after the index date from patients with confirmed PAD diagnosis using ICD-9-CM codes 440.22 – 440.24, 444.01, 444.09, or 445.02.

Table 4. Top Predictors of HRU and HC among Severe PAD Patients

Description	Input Covariate	Selection Frequency	Mean β	SD
Demographic	Region			
	Midwest	75.39	0.84	0.03
	South	75.39	1.16	0.03
	West	75.39	0.42	0.02
	Unknown	75.39	1.38	0.05
Comorbidity	Other Diseases of Kidney and Ureters	64.45	2.20	0.17
Comorbidity	Blindness and Vision Defects	35.55	0.91	0.05
Comorbidity	Chronic Ulcer of Skin	32.81	1.95	0.13
Comorbidity	Chronic Ulcer of Leg or Foot	28.91	1.93	0.11
Comorbidity	Varicose Veins of Lower Extremity	26.17	0.92	0.04
Comorbidity	Peri-; Endo-; and Myocarditis; Cardiomyopathy	25.00	1.22	0.07
Comorbidity	Chronic Kidney Disease	25.00	1.86	0.21
Interaction	Blindness and Vision Defects*Peri-; Endo-; and Myocarditis; Cardiomyopathy	23.83	0.27	0.03

Data Source: Optum+Humedica, 01/01/2007 - 09/30/2015

Note: Example interpretation of edge frequency (EF, Selection Frequency) findings: If a covariate, input 1, has 20% edge frequency this means that the covariate showed up in 20% of the models selected by REFS to best predict the outcome.

The mean β 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.

Abbreviations: PAD, peripheral artery disease; SD, standard deviation

Limitations

- This study used an observational design, presenting a risk of confounding from unmeasured risk factors. This may impact accurate measurement of cost and identification of appropriate risk factors.
- Loss to follow up models using administrative claims and EMR data require assumptions for censoring patients, which may add bias to model results.
- Integration of claims and EMR data may yield a final study sample not representative of the general US population.

Conclusions

- REFS™ identified the presence of chronic ulcers in the lower extremities and CKD as two factors most predictive of all-cause total HC in a geographically diverse population of severe PAD patients.
- Additional comorbidities that were predictors of increased costs included other diseases of the kidneys and ureters, and vision defects including blindness.
- Severe PAD patients who have one or more of the most predictive comorbidities identified should be targeted by providers to reduce overall HRU and HC in this population.

References

1. Baser O et al. Prevalence, incidence, and outcomes of critical limb ischemia in the US Medicare population. Vasc Dis Manag 2013;10:E26-E36.
2. Nehler MR et al. Epidemiology of peripheral arterial disease and chronic limb ischemia in an insured national population. J Vasc Surg 2014;60:686-95.
3. Nguyen LL et al. Resource utilization in the treatment of critical limb ischemia: the effect of tissue loss, comorbidities, and graft-related events. J Vasc Surg 2008;44:971-5.
4. Dua A et al. Preventable complications driving rising costs in management of patients with critical limb ischemia. Ann Vasc Surg 2016;33:144-8.
5. 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.
6. Anderson JP et al. Reverse engineering and evaluation of prediction models for progression to type 2 diabetes: An application of machine learning using electronic health records. J Diabetes Sci Technol. 2016;10:6-18.

Disclosures

- Sponsorship: *Janssen Scientific Affairs, LLC*
- J.S. Berger: Consultant/Advisory Board; Significant; Janssen Pharmaceuticals, AstraZeneca.
- L. Haskell: Employment; Significant; Janssen Pharmaceuticals.
- W. Ting: Consultant/Advisory Board; Significant; Janssen Pharmaceuticals.
- F. Lurie: Consultant/Advisory Board; Modest; Janssen Pharmaceuticals.
- 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 QUALITY OF CARE AND OUTCOMES RESEARCH SCIENTIFIC SESSIONS; APRIL 2 – 3, 2017; ARLINGTON, VIRGINIA