

## **PPO 16-308: Natural Language Processing to Develop a Peripheral Artery Disease Registry in the Veterans Affairs Healthcare System**

**Saket Girotra, MBBS**

**Iowa City VA Health Care System, Iowa City, IA**

**Funding Period: January 2018 - December 2018**

### **Background / Rationale:**

Lower extremity peripheral artery disease (PAD), affects an estimated 8 million people in the United States. Nearly 50% of patients with PAD die within a 10-year period. A significantly larger proportion experience disability and poor quality of life due to a combination of limb-related events (leg pain, non-healing ulcers, and amputation) and cardiovascular (CV) events (myocardial infarction [MI] and stroke). While risk factors for PAD, which include smoking, diabetes, hypertension and dyslipidemia are common in veterans, contemporary data regarding the epidemiology, treatment patterns and clinical outcomes in PAD patients remain sparse. A critical barrier to the progress of PAD-related outcomes research is the poor sensitivity and specificity of International Classification of Diseases (ICD) codes in identifying PAD. In contrast, an ankle-brachial index (ABI, ratio of systolic blood pressure in the ankle to the arm) has a high degree of accuracy (sensitivity >80%), specificity >95%) and is routinely performed in the VHA. Currently, the ABI test results reside within dedicated test reports in the VHA's electronic health record as semi-structured text, and therefore not available for research purposes.

### **Objective(s):**

The overall objective of this pilot application is to develop and validate an automated algorithm to identify patients with PAD in administrative data with a high degree of accuracy, which would facilitate the development of a nationwide PAD registry in the Veterans Health Administration (VHA).

### **Methods:**

We propose to update and implement an existing natural language processing (NLP) system to extract information on key variables from the ABI test reports, evaluate its performance on national VHA data, and determine the diagnostic accuracy of an automated algorithm based on NLP-extracted values to identify PAD.

### **Findings / Results:**

A key step in this project is to identify ABI report documents from all VHA facilities. Currently, our team has successfully identified and annotated ABI report documents from 84 out of 121 VHA facilities, and are in the process of updating the search.

### **Status:**

Analyses are in progress to develop the NLP algorithm to extract ABI values, and confirm its validity in identifying PAD among veterans.

### **Impact:**

Developing a national registry of PAD patients in the VHA could promote high-quality outcomes research and would allow our team to systematically study epidemiology of PAD in veterans, examine current risk modification strategies and treatment gaps, compare the effectiveness of different treatments, as well as serve as a platform for conducting pragmatic clinical trials in this population.