Statistical Analysis Report

Keith Engwall

4/8/2018

Project Objectives

Create a predictive model for identifying patients with Diabetes

Project Dataset

Practice Fusion De-Identified Data Set containing EHR data for approximately 10,000 de-identified patients, including data points for diagnoses, medication, transcript data, and lab observations.

Initial Findings

Age & Gender

The diabetic population trends male and older than the non-diabetic population.

Weight, Height & BMI

Diabetic patients have a slightly higher weight and shorter height than non-diabetic patients. Thus, their BMI trends higher.

Allergies

Among non-medical allergy types, there appears to be a large proportion of diabetic patients with egg and peanut allergies in comparison to the general population. The number of diabetic patients with egg allergies actually outnumbers non-diabetic patients.

Among the allergies to medicine, there appears to be an large proportion of diabetic patients with an allergy to Lisinopril compared to the general population.

Blood Pressure

The pulse pressure (Systolic BP - Diastolic BP) of diabetic patients is slightly higher than non-diabetic patients.

Diagnosis Categories

To simplify analysis of diagnoses, the various diagnoses were divided into categories based on the ICD9 Codes. The diagnosis categories with the highest ratio between diabetic and non-diabetic patients are Circulatory and Endocrine/Nutritional/Metabolic.

Lab Result Analysis

The lab results were limited to those for which there were sufficient abnormal readings data for the diabetic population. For each lab result, an overall measure of central tendency was analyzed, as well as an analysis of abnormal lab result status.

Hemoglobin Lab Results

Hemoglobin levels have a lower central tendency in diabetic patients than in non-diabetic patients. When results recorded as abnormal are separated out, the diabetic population has a larger percentage of below normal readings. The central tendency of above normal readings were higher and of below normal readings were lower among the diabetic population. The central tendency of normal readings was slightly lower in the diabetic population.

Hematocrit Lab Results

The central tendency of Hematocrit percentage was lower in the diabetic population than in the non-diabetic population. The ratio of above normal readings was lower and of below normal readings was higher among the diabetic population. The central tendency of above normal readings was higher and of below normal readings was lower in the diabetic population. The central tendency of normal readings was lower in the diabetic population.

Triglyceride Lab Results

The central tendency of triglyceride levels is higher in the diabetic population than in the non-diabetic population. There is a higher ratio of above normal triglyceride readings in the diabetic population, and both the above normal and normal triglyceride levels are higher in the diabetic population.

Platelets Lab Results

The central tendency of Platelet levels is lower in the diabetic population than in the non-diabetic population, particularly in the normal set. The ratio of both above and below normal readings are greater in the diabetic population, but the abnormal levels aren't as severe.

Prescription Data

There are 18 medications with at least 300 prescriptions and at least 60% use by diabetic patients.