Developing a Model to Predict Hospital Readmissions

Background

High hospital readmission rates are associated with poor patient outcomes and high financial costs for both private and government-sponsored insurance programs. Hospital readmissions to the US healthcare system account for an estimated $17.4 billion in spending annually by Medicare. In an effort to push healthcare systems to decrease preventable readmissions, in 2012, the Hospital Readmissions Reduction Program (HRRP) was established by the Affordable Care Act. This program authorized Medicare to reduce payment to hospitals with excess readmission rates based on a calculation of the risk-standardized 30-day readmission rate for the previous three years for Medicare beneficiaries hospitalized with specific diagnoses.

Objective

With the objective to update healthcare system recommendations and best practices, our client, the Center for Medicare and Medicaid Services (CMS), has asked us to perform an initial analysis using updated US population data to develop a model that predicts hospital readmissions.

The Data

The National Health and Nutrition Examination and Survey (NHANES) provides a cross-sectional survey of health status of the US population each year. NHANES is made up of data collected from a series of health and nutrition surveys, physical examination and laboratory tests conducted by the US Centers for Disease Control and Prevention (CDC).

Every year, approximately 5,000 individuals of all ages from 15 counties throughout the US are included in NHANES. For our training data, we used the dataset from NHANES 2015-2016. In 2015-2016, 15,327 persons were selected for NHANES from 30 different survey locations. Of those selected, 9,971 completed the interview and 9,544 were examined. The data is publicly available on the CDC's website, at https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2015. In the 2015-2016 survey, non-Hispanic Asians, Hispanics, non-Hispanic blacks, older adults, and low income whites/others were over-sampled.

Target Variables

There are several datasets available for the 2015-2016 Survey. All datasets can be linked by the SEQN variable which is tagged to each individual survey participant. We obtained data from the following portions of the survey: Demographics, Laboratory, Examination, and Questionnaire. Target variables were renamed during data preparation for ease of use. A few of the features examined included:

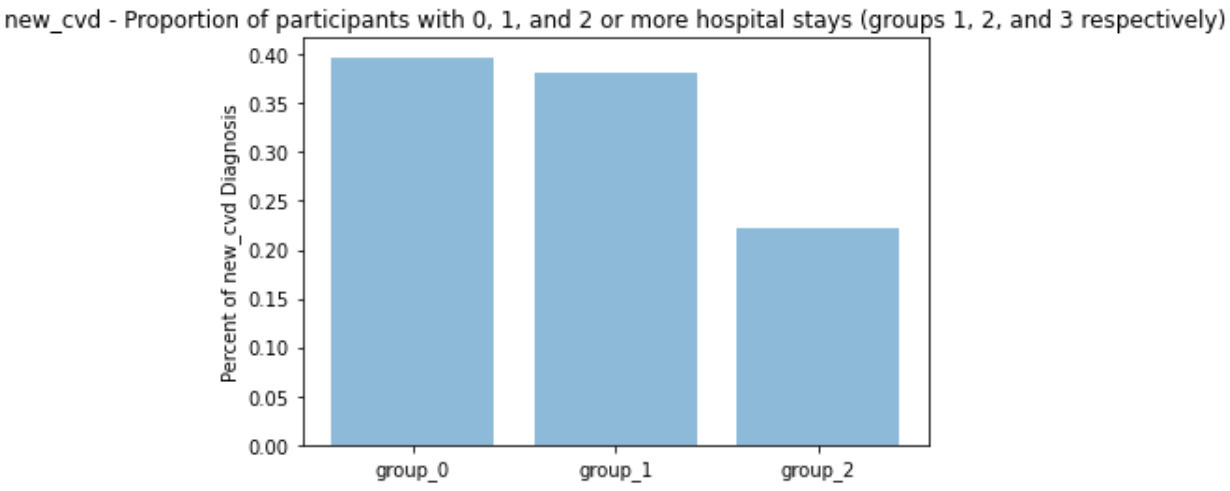
Feature	Variable Type	Description
numhosp	Continuous (float)	Number of hospitalizations
newcvd	Categorical(bool)	New onset of cardiovascular disease
depr_score	Continuous (float)	Depression score
ethnicity	Categorical (category)	Ethnicity
age	Continuous (float)	Age
hgb	Continuous (float)	blood hemoglobin (g/dL)
hct	Continuous (float)	blood hematocrit (%)
foodsec	Categorical (category)	Food security score
chf	Categorical (bool)	Congestive heart failure
kidney	Categorical(bool)	Kidney disease

Initial Findings

During the data preparation phase, we removed participants under the age of 18 to focus the project on adults. The total number of participants included in our project was 5592. Out of this sample, 466 had a single hospital stay during the year of study and 204 had two or more hospital stays.

CHF and Other Diagnoses

Because congestive heart failure is a common predictor of hospital readmissions, our initial plan was to examine only readmissions related to new onset congestive heart failure (CHF). However, we found that although the number of participants with CHF alone seemed adequate (n=498), only 49 reported hospital readmissions, and many fewer reported onset of disease during the year of the study (n=30 total new onset with n=11 readmissions). In an attempt to increase the sample size, we included new onset cardiovascular disease in general, and created a new diagnosis feature ‘newcvd’ which included new onset ‘chf’, ‘chd’, ‘mi’,’angina’,and ‘cva’. The sample size of newcvd was larger than that of ‘chf’ (n=63) with 14 readmissions .



We reject the null hypothesis that the proportion of participants in the 3 groups are equal ($p<0.01$).

Although the 'newcvd' proportion of readmissions was compelling, due to the rich dataset, we also decided to investigate other diagnosis features (which included a compilation of previous and new onset of disease). Other features whose proportions within the 3 groups were significant at the $p<0.01$ level included 'chf', 'kidney', 'copd', 'stroke', 'emphysema', and 'cancer'.

Depression Scores

Participants were divided into two groups based on depression scores:

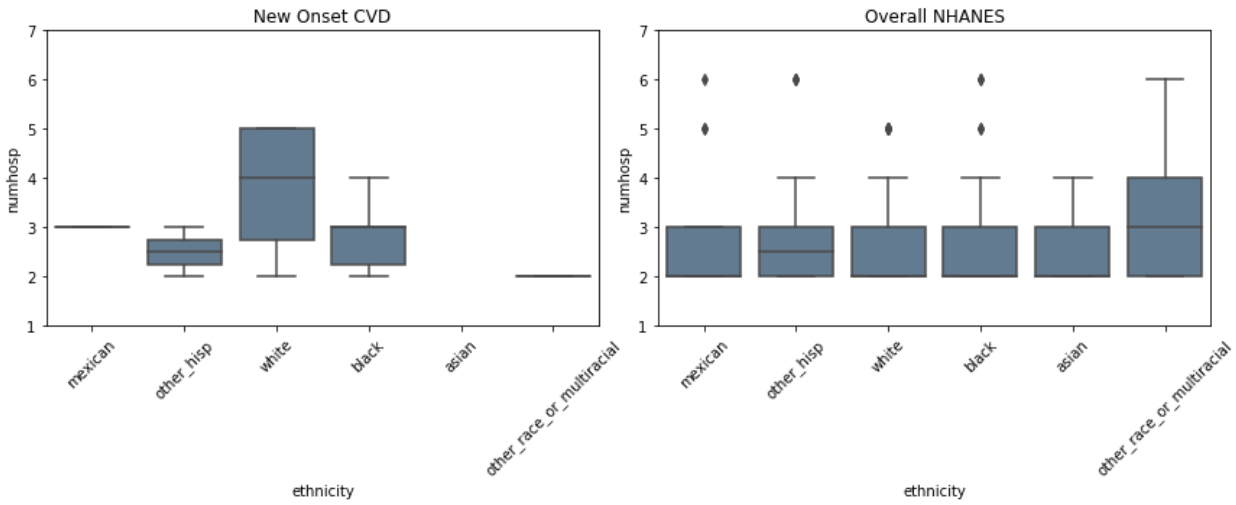
Group 1: depr_score < 5 # little/no depression detected

Group 2: depr_score >= 5 #mildly depressed or worse

A one-tailed, 2 proportion z-test was performed to detect the potential effect of depression score on readmission. Based on the two-proportion z-test, those with depr_score>5 had a higher proportion of hospital readmissions when compared with those with depr_score<5. ($p=0.03$)

Ethnicity

For the overall nhanes group, the Chi square test implied that that ethnicity did not affect readmission status ($p=0.65$. However, when we examined new onset CVD, ethnicity does seem to have an effect on the proportion of hospital readmissions ($p=0.03$).



Laboratory Samples

A 2-sided t-test allowed us to to compare the means of each of various laboratory values between the two groups of interest: readmission == 1 (positive readmission) and readmission == 0 (negative readmission.

We found that mean BUN ($p<0.01$), creatinine ($p<0.05$), and osmolality ($p<0.01$) were significantly varied significantly between the two groups. These lab values relate to kidney function and fluid balance, so it could be speculated that those with readmission. Additionally, other lab values approached significance: hgb ($p=0.052$) - which is an indicator of anemia; and tchol($p=0.055$) - an indicator of heart health.