2/17/2019

DATA698 - Capstone

**Team: Ali Harb, Dilip Ganesan & Raghunathan Ramnath**

**Predict Hospital Readmissions in Diabetes Patients**

**Abstract:**

Today’s health care is moving towards value based care. With this proposition in mind, CMS (Center for Medicare & Medicaid Services) came up with the concept of Hospital Readmissions Reduction Program (HRRP). This program is a Medicare value-based purchasing program that reduces payments to hospitals with excess readmissions. The program supports the national goal of improving healthcare for Americans by linking payment to the quality of hospital care. Based on this program the Department of Health and Human Services (HHS) reduce the payments to Inpatient Prospective Payment System (IPPS) hospitals for excess readmissions. Some of the diseases that are classified under the excess readmissions are listed below.

* Acute Myocardial Infarction (AMI)
* Chronic Obstructive Pulmonary Disease (COPD)
* Heart Failure (HF)
* Pneumonia
* Coronary Artery Bypass Graft (CABG) Surgery
* Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA/TKA)

As of FY2018, 6 more new disease conditions have got added to the HRRP program, but as of 2018 the HRRP does not consider diabetes mellitus part of the program. Although diabetes is not yet included in the penalty measures, American hospitals spent over $41 billion on diabetic patients who got readmitted within 30 days of discharge in the year 2011.

**Problem Statement:**

Predict whether a patient diagnosed with diabetes will be readmitted to hospital within 30 days of discharge.

**Research Question:**

The focus of this research is to answer the following questions.:

1. Is DRG (Diagnosis Related Group) as parameter a positive predictor for hospital readmission in diabetes mellitus patients?
2. What are the strongest predictors that lead to hospital readmission in diabetes mellitus patients?

**Literature Review:**

For this project, we have focused our literature review around the parameters that to be used for diabetes hospital readmission prediction. Since our research question is finding the strongest predictor for diabetes hospital readmission, we want to make sure the parameters are qualified for research.

We have grouped the parameters in the data set in to following categories:

**Patient Demographics** (Age, Race and Sex)

**Payment Methodologies** (DRG)

**Medical Condition and Medications** (HbA1C, Insulin and Sulfonylurea et.al)

**Patient Demographics**: While starting this research we strongly believe race and sex as the most important predictor for diabetes. According to [Elias K Spanakis et.al] in the U.S., 8.3% of the population or 25.8 million individuals have diabetes. The prevalence of diabetes is highest among Native Americans (33%) and lowest among Alaska natives (5.5%**)**. Non-Hispanic Whites and Asian Americans have similar prevalence rates of 7.1% and 8.4%, respectively, where Non-Hispanic Blacks and Hispanic Americans overall have higher prevalence rates of 11.8% and 12.6%, respectively. In the article, they went even deeper in their analysis stating, among Hispanic Americans, diabetes varied among their countries of origin. South Americans had one of the lowest prevalence rates (10.1 % in men and 9.8% in women). Similarly, low rates were found among Cuban men and women--13.2% and 13.9%, respectively. The prevalence of diabetes was the highest in those of Mexican, Puerto Rican, Central American, and Dominican descent, with rates of 16.2% to 19.3% for men and 18% to 19.4% for women. This holds good even in Asian American race with Asian Indians have the highest diabetes prevalence whereas Koreans and Japanese have the lowest diabetes rates. Another important parameter along with race is the age of the patient. Per [Elias K Spanakis et.al], The prevalence of diabetes was highest in NHWs in the U.S. between the ages of 0-9 and 10-19. NHB children between the ages of 0-9 and 10-19 years have prevalence, where Hispanic American children have high prevalence’s between the ages of 0-9 and 10-19, respectively. All in all, it makes clear that combination of age, race and sex plays a crucial role in diabetes. The above research analysis made us to pick the three parameters for our prediction.

**Payment Methodologies:**  Though many studies have been done on hospital readmission, the parameters that were used are more clinical or patient centric. Though research were made on the dollar impact because of readmission, very less analysis was done on the parameter of how hospitals were reimbursed. This brought us to the important parameter of DRG (Diagnosis Related Group). Hospital admissions are reimbursed based on DRG. [Joseph Futoma et.al] did the comparison of hospital readmission models based on DRG cohorts. For each visit they have a single [Diagnosis Related Group](https://www.sciencedirect.com/topics/medicine-and-dentistry/diagnosis-related-group) (DRG) code, selected from a set of 815 unique DRGs which break down admissions into broader diagnoses classes than the highly specific [ICD codes](https://www.sciencedirect.com/topics/medicine-and-dentistry/international-classification-of-diseases). They tested a variety of statistical models on 280 different patient-visit cohorts as determined by the DRGs. In the context of regression, this is equivalent to the inclusion of an interaction effect between disease groups and every predictor. According to [M W Rich et. al] there is more financial advantage to hospitals to code patients into more lucrative DRGs, so that patients with more severe disease could conceivably be “promoted” to higher paying DRGs, so the reimbursement increases. From the above research, we feel DRG could be one of the crucial parameter for readmission prediction. Though we do not have the parameter in our dataset, we have got the DRG for the clinical claims data from web scraping.

**Medical Condition and Medications:** As far as diabetes is concerned, one of the important parameter is H1A1C test, which measures whether a patient is diabetes or prediabetes or normal. According to [Beata Strack et.al] the decision to obtain a measurement of HbA1c for patients with diabetes mellitus is a valuable predictor for readmission. In their analysis, it showed that the profile of readmission differed significantly in patients where HbA1c was checked in the setting of a primary diabetes diagnosis, when compared to those with a primary circulatory disorder. While readmission rates remained the highest for patients with circulatory diagnoses, readmission rates for patients with diabetes appeared to be associated with the decision to test for HbA1c, rather than the values of the HbA1c result. So, the combination of HbA1c along with primary diagnosis plays are very important role in hospital readmission. Along with other medical condition other important factor is the type of medication which was administered to the patient. According to [Pamela C Heaton et al.] administration of SU[Sulfonylurea] drugs to patient with Type 2 diabetes is associated with an 30% increased risk of readmission compared to other drugs. According to [N. J. Wei] Diabetes medical regimen intensification during hospitalization was not associated with early readmission. Among patients with elevated HbA1c, glucose therapy intensification[Insulin] was associated with a decreased 30-day readmission/emergency department admission risk and lower outpatient HbA1c levels.

Apart from the above researched parameters, we also have other parameters which are available as part of clinical data set.

**Modeling:** Apart from the parameters of the dataset we also did some research on the modeling perspective, [Damian Mingle] has done the hospital readmission modeling based on the Extreme Gradient Boosted Tree. Where in the AUC of the machine learning model in greater than that of the LACE score used by the hospitals to determine hospital readmission risks. We will be using the gradient booster for our analysis. Another interesting approach is the use of deep learning to readmission prediction. Per [Ahmad Hammoudeh et.al] the Convolutional neural networks have provided higher AUC compared to other machine learning algorithms. This is another area of interest we are thinking to explore as part of this research thesis.

**Data Source:**

For this research, we used the Health Facts database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States. Health Facts is a voluntary program offered to organizations which use the Cerner Electronic Health Record System. The database contains data systematically collected from participating institutions electronic medical records and includes encounter data (emergency, outpatient, and inpatient), provider specialty, demographics (age, sex, and race), diagnoses and in-hospital procedures documented by ICD-9-CM codes, laboratory data, pharmacy data, in-hospital mortality, and hospital characteristics. All data were de identified in compliance with the Health Insurance Portability and Accountability Act of 1996 before being provided to the investigators.

The Health Facts data that is used was an extract representing 10 years (1999–2008) of clinical care at 130 hospitals and integrated delivery networks throughout the United States: Midwest (18 hospitals), Northeast (58), South (28), and West (16). Most of the hospitals (78) have bed size between 100 and 499, 38 hospitals have bed size less than 100, and bed size of 14 hospitals is greater than 500. Because this data represents integrated delivery network health systems in addition to stand-alone hospitals, the data contains both inpatient and outpatient data, including emergency department, for the same group of patients.

This dataset is available online at UCI Machine Learning Repository.

Along with the clinical claims data set that is available in UCI, we will also be using the diagnosis cross DRG data set that is available in CMS data source and introduce DRG as a parameter in our model. Deriving of appropriate DRG is complicated because it needs lot more claims information than what is available in the UCI data set, so for our analysis we will be deriving a more appropriate DRG for that episode of care.

**Methodology**

The project will begin with some exploratory analysis to identify statistical properties of the dataset and to try to evaluate variables of interest. It will attempt to investigate relationship between diagnosis, disease and readmission.

In the project, we will also bring in DRG information that is available for us online from various CMS sources. This is a pretty challenging task, we have identified many ways to extract the DRG information. One is to use the cross walk provided by CMS or to scrap the data from online calculator website which gives DRG based on the clinical claims information keyed in.

The idea is to find a pattern based on the patients getting readmitted to create awareness and identify the changes needed on the quality of healthcare. The models will be evaluated and will be compared to other research in the areas of interest. We would like to do various modelling like Logistic regression, Decision tree, Random forests and other classification algorithms to see which of the models has better accuracy in predicting the readmission of patient.

The goal of this project is to provide models to identify the patient population, cause of readmission, help reduce costs and optimize the healthcare system.

**Literature Review:**

1. <https://bmcendocrdisord.biomedcentral.com/articles/10.1186/s12902-016-0084-z>
2. <https://www.sciencedirect.com/science/article/pii/S1532046415000969#t0005>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3552066/>
4. <https://www.sciencedirect.com/science/article/pii/S1877050918317873>
5. <https://www.researchgate.net/publication/315640596_Predicting_Diabetic_Readmission_Rates_Moving_Beyond_HbA1c>
6. <https://www.hindawi.com/journals/bmri/2014/781670/>
7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1350282/?page=3>

**References:**

1. <https://clindiabetesendo.biomedcentral.com/articles/10.1186/s40842-016-0040-x>
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4439931/>
3. <https://www.mdmag.com/medical-news/patients-with-diabetes-often-readmitted-for-hypo-and-hyperglycemia>
4. <https://link.springer.com/article/10.1007/s11892-018-0989-1>
5. <https://www.healthleadersmedia.com/clinical-care/diabetes-complications-increase-readmission-risk>