Bin Lin

02/19/2020

DATA 698 Capstone Project Proposals: Does Higher Medicare Hospital Spending by Claim Lead to Better Outcome and Which Drug Utilization Model is More Cost Saving – FFS or MCO?

Introduction:

According to an article published by CMS (Centers for Medicare & Medicaid Services), America spends around 3.5 trillion dollars on healthcare, which is approximately $10,739 per capita in 2017. The healthcare cost is approximately 17.9% of GDP. This figure is considerably higher than that of most other developed countries (around 10%). In addition, the cost keeps increasing as those baby boomers start hitting their retirement age. If America spends too many resources on healthcare, there will be less money and resources we can spend on elsewhere because the budget is always limited.

What I would like to study is to investigate the breakdown of spending of the two largest federal entitlement programs Medicare and Medicaid on hospital spending by claim and prescription drug spending. My hypothesis is that for those hospitals spend more on prior-to-index admission, index admission, and post-index admission do not necessarily have better quality of care. For my research in particular, I am using hospital readmission as my criteria for measuring the quality of care.

Furthermore, I would want to know if Managed Care Organization (MCO) drug utilization that Affordable Care Act advocates has relatively lower cost than the traditional Medicare drug utilization model-Fee-for-Service (FFS) , whether the result is statistically significant. The objective is significant because it is widely known that US spend more money on healthcare but doesn't end up with better health outcome compared to other major countries. This research project will give us an insight if the assumption is true or not from hospital spending by claim perspective. The second objective will give decision maker suggestion about which drug utilization model to adopt in order to save healthcare cost on prescription drugs.

Literature Review:

One of the article I found is called ‘Factors Associated With Increases in US Health Care Spending, 1996-2013’ written by [Joseph L. Dieleman, PhD](https://jamanetwork.com/searchresults?author=Joseph+L.+Dieleman&q=Joseph+L.+Dieleman); [Ellen Squires, MPH](https://jamanetwork.com/searchresults?author=Ellen+Squires&q=Ellen+Squires); [Anthony L. Bui, MPH](https://jamanetwork.com/searchresults?author=Anthony+L.+Bui&q=Anthony+L.+Bui). As summarized in the article, there are five key factors associated with increase in healthcare. They are population growth, population aging, disease prevalence or incidence, medical service utilization and service price and intensity. In general we are not able to control the population growth and aging population. While the service price and intensity is very difficult to measure using the dataset that I am able to obtain. The scope of disease prevalence or incidence is too broad for this course. Therefore, my research project will focus solely on medical service utilization.

# According to Centers for Medicare and Medicaid, in 2012, Affordable Care Act established Hospital Readmissions Reduction Program (HRRP) to reduce payments to hospitals for excess readmissions. Per Mayo Clinic, hospital readmission is patient admission to a hospital within 30 days after being discharged from an earlier hospital stay. The reason that ACA wants to reduce hospital readmissions is because it is one of the critical category of data used to evaluate the quality of hospital care. There are many studies investigating the way to reduce readmission. For example, the study called “Predictive Modeling of Hospital Readmitting Rates Using Electronic Medical Record-Wide Machine Learning: A Case-Study Using Mount Sinai Heart Failure Cohort” was conducted in Mount Sinai back in 2016. Those studies including this one include a lot of variables relevant to patient population, demographics, disease states, procedures, medications et cetera. Most of the variables are attributable to PHI and EMR. However, instead of taking the clinicians’ perspective, I would like the tackle the issue using administrative perspective with particular focus on money. Using Medicare Hospital Spending by Claim to predict hospital readmission, I am able to keep my research project original. The methodologies offered by other studies will still be helpful for my study.

Before speaking about drug utilization, we have to know the definition of FFS and MCO. Using definition from Healthcare.gov, FFS is a method in which doctors and other health care providers are paid for each service performed. Examples of services include tests and office visits. MCO means managed care organization. They accept a set per member per month (capitation) payment to provide for delivery of Medicaid health benefits and additional services. In order to reign the cost of prescription drugs, organization such as Missouri Department of Social Services conduct a study to compare the cost between two payment method. The result showed that Managed Care Organization saves approximately 1.7% of the healthcare cost compared with FFS. However, the study is more of a descriptive analysis of the summary statistics. It does not offer p-value, therefore, we cannot generalize what has been studied. However, this issue will be addressed in my research project.

Hypothesis:

First Issue:

Null Hypothesis (H0): Higher Medicare Hospital Spending by Claim does not lead to lower hospital readmission

Alternative Hypothesis (HA): Higher Medicare Hospital Spending by Claim leads to lower hospital readmission

Second Issue:

Null Hypothesis (H0): FFS and MCO prescription drug utilizations are equally cost saving

Alternative Hypothesis (HA): FFS and MCO prescription drug utilizations are not equally cost saving

Data and Variables:

I will obtain most of my data from [www.data.gov](http://www.data.gov), [www.cms.gov](http://www.cms.gov), and [www.healthcare.gov](http://www.healthcare.gov). For the first problem that I want to address, I will use Medicare Hospital Spending by Claim data as long as the Hospital Readmission Reduction Program data. The variables that are available are the six claim types (Home Health Agency, Hospice, Inpatient, Outpatient, Skilled Nursing Facility, Durable Medical Equipment, Carrier) over three admission periods (1 to 3 days Prior to Index Hospital Admission, During Index Hospital Admission, and 1 through 30 days After Discharge from Index Hospital Admission) for each individual hospital. After data tidying and transformation, I will have 18 explanatory variables. The outcome variable I can use the actual number of readmission or readmission rate. For the second problem that I want to address, I will get the State Drug Utilization Data 2017. Two categorical variables are FFS and MCO, while the Units Reimbursed and Medicaid Amount Reimbursed will both be numerical variables.

Statistical Methods:

To investigate the relationship between hospital claims and readmission, I will randomly divide the datasets into the training and testing dataset (80/20). Using the training dataset, I can build a multivariable linear regression model to see the correlation between hospital claims and readmission. I would have come up with the parameters such as the correlation coefficient – R as well as correlation of determination – R^2 to further explain my findings. I would probably perform both forward selection and backward selection to build two different models which include the most significant valuables. I will also build a model that includes all the variables. These three models will be evaluated using Mean Absolute Error, Mean Square Error, Root Mean Square Error et cetera. In the end, I will pick the best model to make predictions on my testing dataset. I am also intended to build a logistic regression model. I would like to know if certain hospital will have higher than average national hospital readmission or lower. I will calculate the average national hospital readmission rate as the decision boundary to determine the classification. Any value equal or higher than the average will receive classification of 1, while any value smaller that the average will receive classification of 0. Confusion matrix will be built to help calculate sensitivity, specificity, precision, FNR, FDR etcetera. ROC curve will be drawn to illustrate the accuracy of the model. The AIC, F1 score and area under the curve (AUC) are the parameters that I will use to evaluate the accuracy of the model. If time permits, I can try different type of classification models such as Random Forest, Support Vector Machine, and Neural Networks et cetera.

For prescription drug utilization study, I will perform Two Sample T-Test to check if the drug reimbursement will have any differences between MCO utilization and FFS utilization. The differences will be measure in terms of T-score. Then it can be translated into p-value to tell us how significant is the differences. Confidence interval will also be calculated. In this case, if the interval crosses zero, then the result is not significant and vice versa. Another approach is I will resample the data points and calculate its bootstrap statistics.