# Analysis of Medicare Inpatient Hospitals by Provider and Service

## Unreimbursed Amount towards Medicare

## Average Medicare Payments

**ISM 6137 Statistical Data Mining**

**Dr. Anol Bhattacherjee**

**University of South Florida**

**Tejas Kulkarni**

**Venkat Mohith Pathipati**

**Prasad Ganesh Hegde**

**Adil Mohd. Hasan Shaikh**

**Contents:**

## 1. Executive Summary

### 1.1 Medicare

Medicare is a federal government health care program that provides health insurance plans to people aged 65 and above, young people with disabilities and people with end stage renal diseases. Medicare was formulated in the year 1966 under the Social Security Administration (SSA) but is presently administered by the Centre for Medicare and Medicaid services (CMS).

As of 2018, Medicare provided health insurance to about 60 million individuals – more than 52 million people aged 65 and above and about 8 million younger people with disabilities. This number is expected to reach about 80 million by the year 2030 as the US population ages. In 2020, the US federal government spent about $776.2 billion on Medicare.

Medicare is divided into 4 parts: Part A, B, C, D.

**Part A** covers inpatient care in hospitals, skilled nursing, facility care, hospice care and home health care

**Part B** covers services from doctors, outpatient care, preventive services and most professionally administered prescription drugs

**Part D** covers most self-administered prescription drugs

**Part C** is an alternative managed plan for Parts A and B, called Managed Medicare or Medicare Advantage, which allows patients to opt for health plans with the same of more coverage as Parts A and B. It also helps cover benefits of Part D and an annual out – of – pocket expense limit lacked by Parts A and B. A beneficiary must enroll in Parts A and B before signing up for Part D.

### 1.2 Inpatient Prospective Payment System (IPPS)

This is a payment system set up by the Social Security Act for the operating costs of acute care hospital inpatient stays under the Medicare Part A based on certain preset rates. Under the IPPS, every patient's medical diagnosis in categorized into a diagnosis-related group (DRG). The base payment is divided into a labor-related and nonlabor share. The labor-related share is adjusted by the wage index applicable to the area where the hospital is located, and if the hospital is located in Alaska or Hawaii, the nonlabor share is adjusted by a cost-of-living adjustment factor. This base payment rate is multiplied by the DRG relative weight.

### 1.3 Total Performance Score (TPS) of hospitals

The TPS of any hospital is computed as a weighted sum of scores from the following four domains:

- **Domain Score** –
  This is computed based on a 30-day mortality rate for Acute Myocardial Infarction and Pneumonia.
- **Efficiency and Cost Reduction Score** –
  This awarded based on the Medicare spend per beneficiary for the hospital.
- **Safety Score** –
  This is awarded based on 27 measures which are defined by CMS. It consists of two types which are process oriented and outcome-oriented measures and process oriented measures.
- **Community and Engagement Score –**
  Community and engagement score is awarded based on the quality of communication between hospital staff and patients.

## 2. Problem Definition and Significance

Under the Inpatient Prospective Payment System (IPPS), Medicare payments are supposed to be uniform across all DRGs across the states. But we need to understand the fact that Medicare payments for diagnosis related groups across all hospitals in USA are not uniform. We verified it by doing preliminary data exploration.

Therefore, we decided to analyze unreimbursed amount and Medicare payment for each hospital. Our primary motive is to assess the factors which affect unreimbursed amount and Medicare amount and to find the possible solutions to minimize unreimbursed amount and maximize Medicare payment.

We also want to find top 2 and bottom 2 diagnosis related groups across USA in terms of unreimbursed amount and Medicare payments and try to find the reasons for that.

## 3. Prior Literature

In the literature survey, we would like to emphasize more on the following papers which are relevant to our analysis.

"Medicare and Cost Effectiveness Analysis" is a paper written by Dr. Neumann, Dr. Rosen, Dr. Milton, and Dr. Weinstein, published in the New England Journal of Medicine. From this paper, we understood the importance of cost effectiveness analysis to obtain health care decisions. It has certainly helped us to understand the fact that health care resources for Medicare can be utilized more efficiently. By understanding this paper, we decided to perform analysis on unreimbursed amount.

"Hospital Reimbursement under Medicare" is a paper written by Dr. Lave in Milbank memorial fund Quarterly. Dr. Lave is a professor of health informatics in the University of Pittsburg and her article helped us to get the basic idea about Medicare and to understand the process of Medicare prospective payment system in a lucid language. She has written about the Medicare payment disparities between hospitals for same diagnosis related group. We also have understood that Medicare payments is the dominant factor for hospitals to work with financial stability.

"Alternatives for using multivariate regression to adjust prospective payment rates" is a paper written by Dr. Sheingold, published in the Health Care Financing Review. In this paper, regression for understanding Medicare amount is illustrated which in turn is useful for calculating average unreimbursed amount. Also, the effect of teaching hospitals and number of beds in the hospital on payment variable is discussed.

## 4. Data Source and Preparation

We have collected data from four different sources, three of them being sourced from the Center for Medicare and Medicaid Services(CMS) and one dataset from census.gov.

We excluded variables from one big, combined dataset which are non-relevant to our dependent variable. Our final dataset consists of 22 features

Variables which we have considered from CMS dataset 1 are – city, state, rural urban commuting area codes(RUCA), Diagnosis related group(DRG) description, drg_desc, total discharges, Average Medicare payments, hospital cost.

Variables which we fetched from CMS dataset 2 are – domain score, engagement score, safety score, cost reduction score, teaching hospitals,

Variables which were taken from CMS dataset 3 are – Employees on payroll, number of interns and residents, number of beds, uncompensated care, contract labor, cost_charge_ratio.

Variable which we considered from census dataset – median income

We have joined all these datasets together with *python and foreign key for joining was facility ID i.e., unique ID for hospitals across USA.

We have a total of 3032 hospitals in the data across all states of USA. Total diagnosis related groups associated with the given dataset are 532.

For the calculation of total unreimbursed amount for the hospital for the specific diagnosis related groups is formulated as follows –

Average unreimbursed amount = (Average total cost - Average Medicare amount )

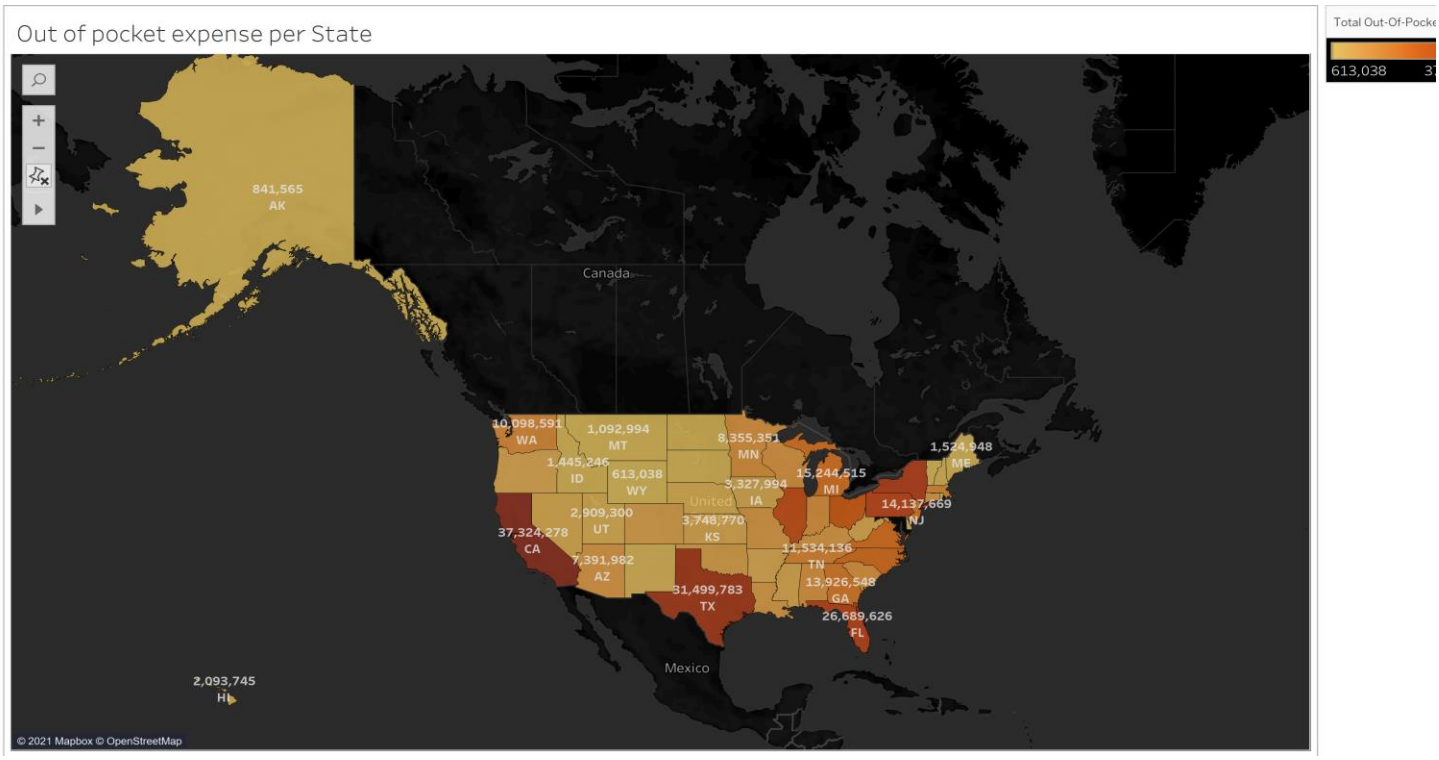
## 5. Variable Choice

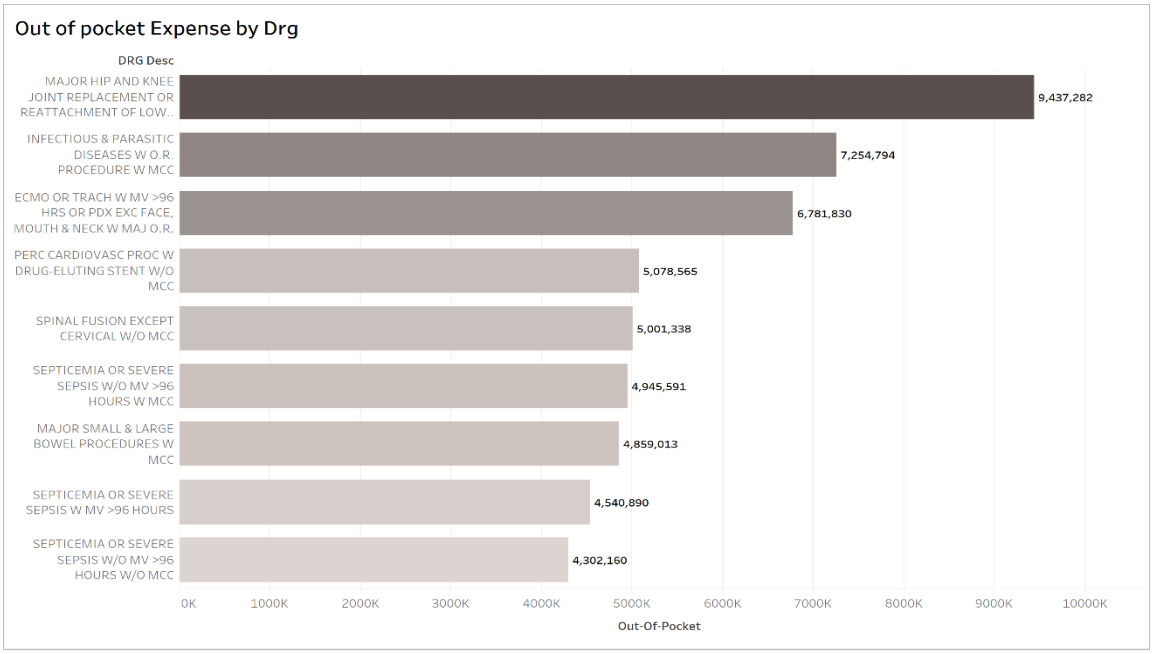
**Dependent Variables:** Unreimbursed Amount, Average Medicare Payments

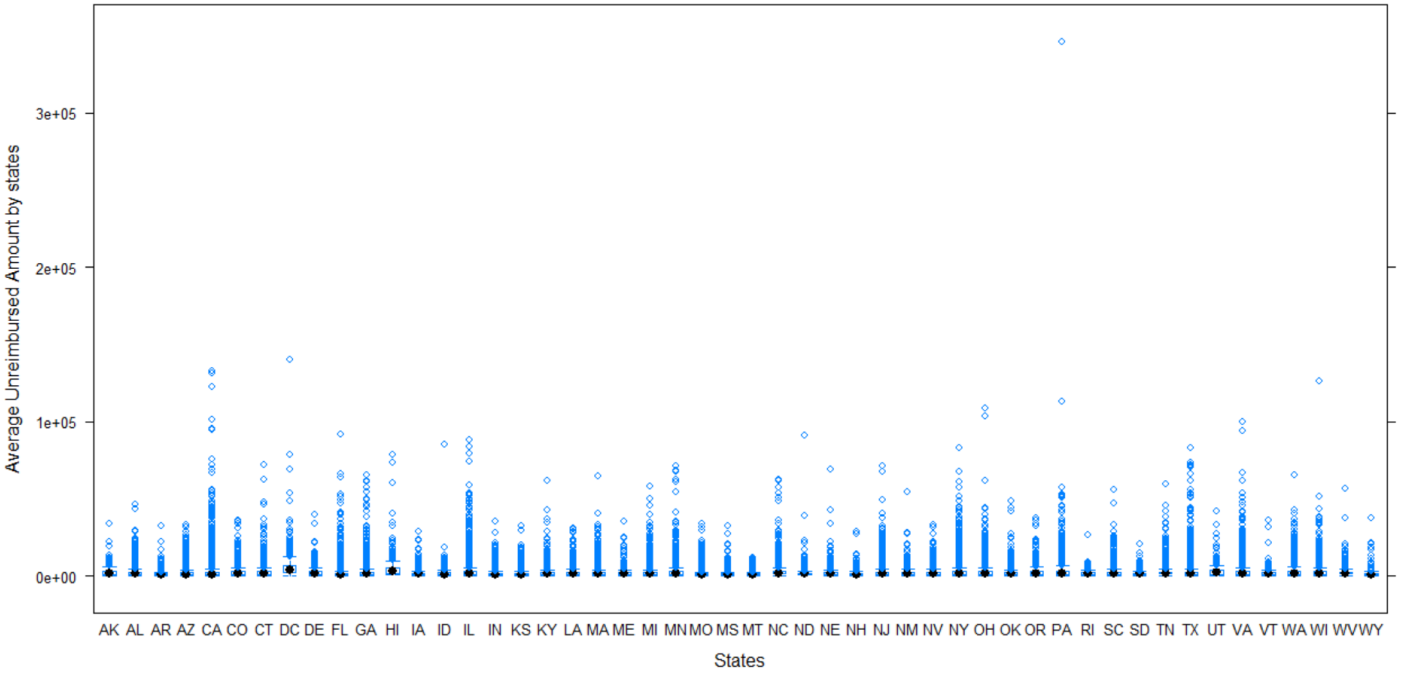
| Predictor | Effect | Rationale |
|---|---|---|
| hospital_cost | + | If the hospital costs are more, the unreimbursed amount maybe more for the hospital since Medicare will only pay a fixed part of the total costs |
| domain_score (clinical outcome score) | - | If the Clinical Outcome score is higher, average unreimbursed amount tend to be low since these hospitals may have effectively reduced the mortality rates for diagnoses that are used to determine this score. |
| engagment_score | - | If engagement between customers and staff is good then there will be less medicare amount paycuts which means we will have less unreimbursed amount. |
| safety_score | - | Safety score is calculated based on 27 measures consisting of process oriented and outcome-oriented measures. CMS reimburses higher portion of the total amount to hospitals with high safety score, so the average unreimbursed amount maybe less. |
| cost_reduction_score | - | This score is calculated based on average Medicare spend per patient per hospital. If the cost reduction score is more, average unreimbursed amount maybe less |
| no_of_employees | + | Higher the number of beds and employees, higher maybe the unreimbursed amount since cost to hospital will be more. |
| Cost to charge ratio | + | Higher CCR may drive up the average unreimbursed amount |
| contract_labor | + | Higher the total salaries paid to the employees, higher maybe the unreimbursed amount since overall costs will be more |

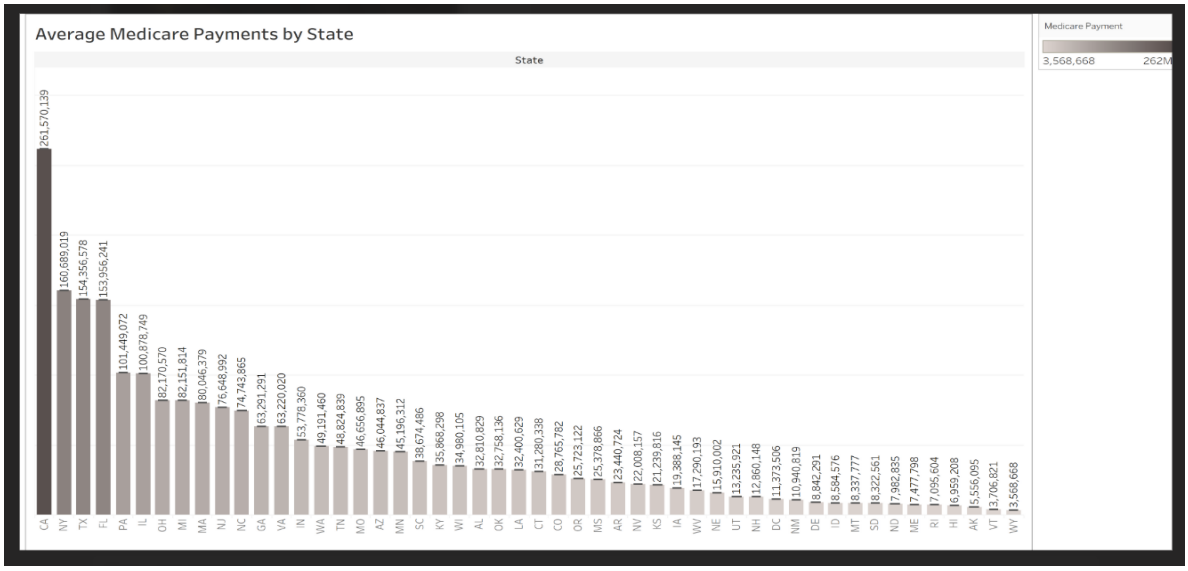
| | | |
|---|---|---|
| Teaching_hospital | + | Teaching hospital may have higher unreimbursed amount due to extra expenses incurred towards maintenance of additional staff like interns and residents. |
| Median income | + | Hospitals in areas with high median income tend to charge more per DRG, driving up the unreimbursed amount |
| Total Salaries | + | If a hospital pays more towards the total salaries of its employees, its average unreimbursed amount will be more. |

## 6. Descriptive Analysis and Data Visualization




Out of pocket Expense by Drg
DRG Desc
MAJOR HIP AND KNEE JOINT REPLACEMENT OR REATTACHMENT OF LOW..
INFECTIOUS & PARASITIC DISEASES W O.R. PROCEDURE W MCC
ECMO OR TRACH W MV >96 HRS OR PDX EXC FACE, MOUTH & NECK W MAJ O.R.
PERC CARDIOVASC PROC W DRUG-ELUTING STENT W/O MCC
SPINAL FUSION EXCEPT CERVICAL W/O MCC
SEPTICEMIA OR SEVERE SEPSIS W/O MV >96 HOURS W MCC
MAJOR SMALL & LARGE BOWEL PROCEDURES W MCC
SEPTICEMIA OR SEVERE SEPSIS W MV >96 HOURS
SEPTICEMIA OR SEVERE SEPSIS W/O MV >96 HOURS W/O MCC
9,437,282
7,254,794
6,781,830
5,078,565
5,001,338
4,945,591
4,859,013
4,540,890
4,302,160
0K
1000K
2000K
3000K
4000K
5000K
6000K
7000K
8000K
9000K
10000K
Out-Of-Pocket
Total Out-of-pocket Expense
4,302,160
9,437,282


Average Unreimbursed Amount by states
3e+05
2e+05
1e+05
0e+00
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
States




## 7. Models

For the analysis of unreimbursed amount and Medicare payment across all states and for all diagnosis related groups, we have formulated following models –

### Model 1

Model_unreimbursedAmt =
lmer(log(unreimbursed_amount)~hospital_cost+domain_score+engagment_score+safety_score+
cost_reduction_score+no_of_employees+cost_charge_ratio+
contract_labor+total_salaries+teaching_hospital+
median_income+(1|State/drg_description), data = df, REML = FALSE)

--For analyzing multilevel effect with random effect for state/drg description for unreimbursed amount

### Model 2

mode_medicare = lmer(log(medicare_coverage)~domain_score+engagment_score+safety_score+
cost_reduction_score+median_income+Total_discharge+
(1|State), data = df, REML = FALSE)

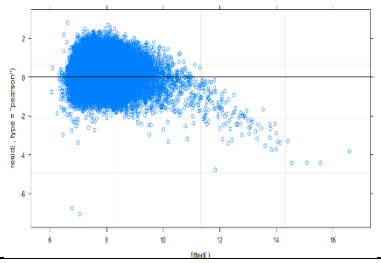
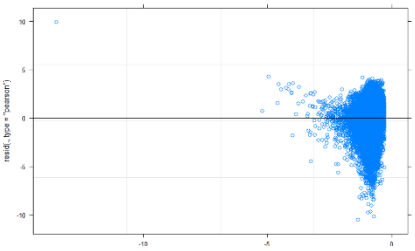
--For analyzing multilevel effect for Medicare payment

## 8. Quality Checks

For checking the quality of our multilevel models, we have looked for following assumptions –

1. Homoskedasticity

2. Multicollinearity
3. Independence

| Assumption | Model 1 | Model 2 |
|---|---|---|
| Homoskedasticity |  |  |
| Multicollinearity | vif test<br>hospital_cost 1.078035<br>domain_score 1.152474<br>engagment_score 1.173500<br>safety_score 1.113762<br>cost_reduction_score 1.158938<br>cost_charge_ratio 1.004136<br>contract_labor 1.152723<br>teaching_hospital 1.244459<br>median_income 1.137863 | vif test<br>domain_score 1.141316<br>engagment_score 1.143917<br>safety_score 1.023689<br>cost_reduction_score 1.074315<br>median_income 1.117602<br>Total_discharge 1.005140 |
| Independence | durbinWatsonTest(resid(model2))<br>1.47856 | durbinWatsonTest(resid(mode_medicare))<br>1.620408 |

- From above assumptions, we can concur that residual are heteroscedastic but we went ahead with the analysis of this data as most real world datasets are heteroscedastic.
- Acceptable in terms of multicollinearity
- Acceptable in terms of independence

## 9. Interpretations

### 9.1 Model 1

**Fixed marginal effects interpretations for Unreimbursed Amount**

| | |
|---|---|
| Average Total Payment | Every $10,000 increase in Average Total Payment will drive up the unreimbursed amount by $2662 |
| Domain Score | Every unit increase in Domain Score will drive up the unreimbursed amount by $88.74 |
| Engagement Score | Every unit increase in Engagement Score will drive up the unreimbursed amount by $133 |
| Safety Score | Every unit increase in Safety Score will drive down the unreimbursed amount by $3000 |
| Cost Reduction Score | Every unit increase in Cost Reduction Score will drive down the unreimbursed amount by $110 |
| No. of Employees | Every unit increase in the No. of Employees will drive up the unreimbursed amount by $614 |
| Cost to Charge Ratio | Every unit increase in CCR will drive up the unreimbursed amount by $59 |
| Contract Labour Charges | Every $1,000 increase in Contract Labour Charges will drive up the unreimbursed amount by $451 |
| Total Salaries | Every $1,000 increase in Total Salaries will drive up the unreimbursed amount by $317 |
| Teaching Hospital (Yes) | Teaching Hospitals incur an additional unreimbursed amount of $1265 with respect to non-teaching hospitals |
| Median Area Income | Every $1,000 increase in Total Salaries will drive up the unreimbursed amount by $37 |

**Top 2 DRG Codes with highest Unreimbursed Amounts**

| DRG Description | Average Unreimbursed amount above the mean |
|---|---|
| KIDNEY TRANSPLANT | $22,176 |
| MAJOR MALE PELVIC PROCEDURES | $21,415 |

**Bottom 2 DRG Codes with highest Unreimbursed Amounts**

| DRG Description | Average Unreimbursed amount below the mean |
|---|---|
| HEART TRANSPLANT OR IMPLANT OF HEART ASSIST SYSTEM | $8668 |
| ECMO OR TRACH W MV >96 HRS OR PDX EXC FACE, MOUTH & NECK | $9831 |

**9.2 Model 2**

**Fixed marginal effects interpretations for Average Medicare Payments**

| | |
|---|---|
| Domain Score | Every unit increase in Domain Score will increase the Average Medicare Payment amount by $459.5 |
| Engagement Score | Every unit increase in Engagement Score will decrease the Average Medicare Payment amount by $6.4 |
| Safety Score | Every unit increase in Safety Score will decrease the Average Medicare Payment amount by $744 |

| | |
|---|---|
| Cost Reduction Score | Every unit increase in Cost Reduction Score will decrease the Average Medicare Payment amount by $643 |
| Total Discharges | Every unit increase in the Total Discharges will decrease the Average Medicare Payment amount by $92 |
| Median Area Income | Every $10,000 increase in Total Salaries will decrease the Average Medicare Payment by $46 |

**2 States with highest Medicare Payments**

| DRG Description | Average Medicare Payments above the mean |
|---|---|
| ARKANSAS | $9512 |
| OREGON | $9422 |

**2 States with lowest Medicare Payments**

| DRG Description | Average Medicare Payments below the mean |
|---|---|
| MAINE | $6973 |
| CONNECTICUT | $7020 |

## 10. Recommendations

- If the safety score is increased by one unit, the unreimbursed amount goes down by 3000. So, hospitals must strive to improve their safety score in order for the unreimbursed amount to decrease.
- Medicare must increase payments for Kidney Transplant and Major Male Pelvic Procedures as these have turned out to be the most frequent procedures with the highest unreimbursed amounts.
- Every $1000 increase in the Contract Labor Charges increases the Unreimbursed amount by $459. Hospitals must have a control on their contract labor charges and look for more efficient ways to hire and utilize their contract labor. Decrease in the total man hours dedicated towards labor charges may drive down the unreimbursed amount.
- Teaching Hospitals do not incur a significant unreimbursed amount over non-teaching hospitals. Hospitals must dedicate more of their resources towards teaching and research.
- Maine and Connecticut have the lowest Medicare payments below the mean across the US, while Arkansas and Oregon have the highest. Maine and Connecticut must investigate the overall quality metrics of hospitals in Arkansas and Oregon and try to find ways to increase their Average Medicare Payment per DRG.

## 11. References


Medicare Inpatient Hospitals by Provider and Service Data Dictionary - Centers for Medicare & Medicaid Services Data (cms.gov)

Medicare Inpatient Hospitals - by Provider and Service - Centers for Medicare & Medicaid Services Data (cms.gov)

Hospital Provider Cost Report Data Dictionary - Centers for Medicare & Medicaid Services Data (cms.gov)

Census.gov

## 12. Appendix A: R Code For Analysis

```
library(rio)

library(moments)

library(car)

library(PerformanceAnalytics)

library(bit64)

setwd("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT")

#Laoding the merged Dataset

df = import("FiNAL.csv")


View(df)

library(stringr)


#Removing $ and , sysmbol from amount column to convert it to numerical

num_data <- str_replace_all(df$`Total Liabilities`, "[^[:alnum:]]", "")

df$`Net Income` <- str_replace_all(df$`Net Income`, "[^[:alnum:]]", "")

df$`Total Salaries (adjusted)` <- str_replace_all(df$`Total Salaries (adjusted)`,
"[^[:alnum:]]", "")

df$`Wage-Related Costs (Core)` <- str_replace_all(df$`Wage-Related Costs (Core)`,
"[^[:alnum:]]", "")

df$`Total Unreimbursed and Uncompensated Care` <- str_replace_all(df$`Total Unreimbursed
and Uncompensated Care`, "[^[:alnum:]]", "")

#Checking for NA values or missing values

lapply(df,function(x) { length(which(is.na(x)))})


#feature engineering

df$`Number of Interns and Residents (FTE)`[is.na(df$`Number of Interns and Residents
(FTE)`)] <- 0

df$teaching_hospital = ifelse(df$`Number of Interns and Residents (FTE)`==0,"no","yes")

drop = c("Number of Interns and Residents (FTE)","Total Bed Days Available + Total for
all Subproviders")

df = df[,!(names(df) %in% drop)]



#Median imputation
```

```
df$`Number of Beds`[is.na(df$`Number of Beds`)] = median(df$`Number of Beds`,na.rm = TRUE)

df$`Cost To Charge Ratio`[is.na(df$`Cost To Charge Ratio`)] = median(df$`Cost To Charge Ratio`,na.rm = TRUE)

df$`Contract Labor`[is.na(df$`Contract Labor`)] = median(df$`Contract Labor`,na.rm = TRUE)

df$`Net Income`[is.na(df$`Net Income`)] = median(df$`Net Income`,na.rm = TRUE)

df$`Total Salaries (adjusted)`[is.na(df$`Total Salaries (adjusted)`)] = median(df$`Total Salaries (adjusted)`,na.rm = TRUE)

df$domain_score[is.na(df$domain_score)] = median(df$domain_score,na.rm = TRUE)

df$engagment_score[is.na(df$engagment_score)] = median(df$engagment_score,na.rm = TRUE)

df$safety_score[is.na(df$safety_score)] = median(df$safety_score,na.rm = TRUE)

df$cost_reduction_score[is.na(df$cost_reduction_score)] = median(df$cost_reduction_score,na.rm = TRUE)

df$median_income[is.na(df$median_income)] = median(df$median_income,na.rm = TRUE)


lapply(df,function(x) { length(which(is.na(x)))})

sum(is.na(df))


ceiling(df$`FTE - Employees on Payroll`)

df = df[df$State!="",]


#Descriptive Analysis


#Which Drug has highest expense in any hospital

df_higest_expense_drug = df[which.max(df$`out-of-pocket`),c("Rndrng_Prvdr_Org_Name","DRG_Desc","Avg_Tot_Pymt_Amt","City_x","State")]


#Which Drug has minimum expense in any hospital

df_lowest_expense_drug = df[which.min(df$`out-of-pocket`),c("Rndrng_Prvdr_Org_Name","DRG_Desc","Avg_Tot_Pymt_Amt","City_x","State")]


#Calculating the Sum of hospital expense by Drug

df_sum_drug = aggregate(df$`unreimbursed_amount`, by=list(df$DRG_Desc),FUN =sum)

df_sum_drug_ordered = df_sum_drug[order(-df_sum_drug$x),]


#Calculating hospital expense by drug and by state
```

```
df_sum_drug_state = aggregate(df$` unreimbursed_amount`,
by=list(df$DRG_Desc,df$State),FUN =sum)


#Which hospital expense is more after combining all the drug

df_sum_hospital = aggregate(df$` unreimbursed_amount`,
by=list(df$Rndrng_Prvdr_Org_Name),FUN =sum)

df_sum_hospital_ordered = df_sum_hospital[order(-df_sum_hospital$x),]



#Renaming the column name

colnames(df) = c("hospital_name","zipcode","ruca","drug_description",
                 "Total_discharge","submitted_charge","avg_tot_payment",
                 "avg_medicare_coverage","city","State","domain_score",
                 "engagment_score","safety_score","cost_reduction_score",
                 "tps","rural_vs_urban","provider_type","type_of_control",
                 "no_of_employees","beds",
                 "uncomopnseted_care",
                 "total_salaries","contract_labor",
                  "net_income","cost_charge_ratio",
                 "median_income"," unreimbursed_amount",
                 "teaching_hospital")
length(unique(df$drug_description))


#converting to correct data type

df$domain_score = as.numeric(df$domain_score)

df$engagment_score = as.numeric(df$engagment_score)

df$safety_score = as.numeric(df$safety_score)

df$cost_reduction_score = as.numeric(df$cost_reduction_score)

df$median_income = as.numeric(df$median_income)


num_data <- str_replace_all(df$median_income, "[^[:alnum:]]", "")

df$median_income = as.numeric(df$median_income)



#Checking for Correlation for numerical variables
```

```
library(PerformanceAnalytics)
library(corrplot)


attach(df)
dfcorr = subset(df,select = c(domain_score, engagment_score, safety_score,
cost_reduction_score,
                              tps, no_of_employees, beds, uncomopnseted_care,
                              total_salaries, contract_labor,
                              net_income, cost_charge_ratio, median_income))
chart.Correlation(dfcorr)



dfcorr$uncomopnseted_care = as.numeric(dfcorr$uncomopnseted_care)
dfcorr$total_salaries = as.numeric(dfcorr$total_salaries)

#Downloading the pre-processed Dataset to local system
write.csv(df,"final_csv.csv", row.names = FALSE)

#Loading the Final Data set
setwd("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT")
df = import("final_csv.csv")

#Creating unreimbersed column
df$unreimbersed_amount = (df$avg_tot_payment - df$avg_medicare_coverage)



#Scaling all numerical values
df$submitted_charge = scale(df$submitted_charge)
df$avg_tot_payment = scale(df$avg_tot_payment)
df$Total_discharge = scale(df$Total_discharge)
df$avg_medicare_coverage = scale(df$avg_medicare_coverage)
df$domain_score = scale(df$domain_score)
df$engagment_score = scale(df$engagment_score)
df$safety_score = scale(df$safety_score)
df$cost_reduction_score = scale(df$cost_reduction_score)
```

```
df$no_of_employees = scale(df$no_of_employees)

df$beds = scale(df$beds)

df$uncomopnseted_care = scale(df$uncomopnseted_care)

df$contract_labor = scale(df$contract_labor)

df$net_income = scale(df$net_income)

df$median_income = scale(df$median_income)

df$cost_charge_ratio = scale(df$cost_charge_ratio)

df$total_salaries = scale(df$total_salaries)

df$unreimbersed_amount = ifelse(df$unreimbersed_amount==0,1,df$unreimbersed_amount)


#Converting columns to factor variables data type

cols <- c("hospital_name",
"ruca","city","rural_vs_urban","State","drug_description","provider_type","type_of_contro
l","teaching_hospital")

df[cols] <- lapply(df[cols], factor)


#Checking the histogram of output variable

library(ggplot2)

qplot(df$unreimbursed_amount, geom="histogram",xlab = "Unreimbersed Amount")

qplot(log(df$unreimbursed_amount), geom="histogram",xlab = "log(Unreimbersed
Amount)",main = "Log of Unreimbersed Amount")

qplot(log(df$medicare_coverage), geom="histogram",xlab = "Average Medicare Payment")


#boxplot

library(lattice)

bwplot(unreimbursed_amount~State, data = df, xlab = "States", ylab = "Average
Unreimbursed Amount by states", las=2)

#Loading lmer libraray

library(lme4)

#First model

model1 =
lmer(log(unreimbursed_amount)~avg_tot_payment+domain_score+engagment_score+safety_score+

             cost_reduction_score+no_of_employees+cost_charge_ratio+

               contract_labor+total_salaries+teaching_hospital+

               median_income+(1|State)+(1|drug_description), data = df, REML = FALSE)

summary(model1)
```

```
ranef(model1)


#second model

model2 =
lmer(log(unreimbursed_amount)~avg_tot_payment+domain_score+engagment_score+safety_score+

                cost_reduction_score+no_of_employees+cost_charge_ratio+

                contract_labor+total_salaries+teaching_hospital+

                median_income+(1|State/drug_description), data = df, REML = FALSE)

summary(model2)

dfs1 = ranef(model2)

df1.coef = coef(model2)

df_drg_state=df1.coef$`drug_description:State`

df_state = df1.coef$State


#model with interaction

model3 =
lmer(log(unreimbursed_amount)~avg_tot_payment+domain_score+engagment_score+safety_score+

                cost_reduction_score+no_of_employees+cost_charge_ratio+

                contract_labor+total_salaries+teaching_hospital+


median_income+no_of_employees*engagment_score+(1|df$State)+(1|drug_description), data =
df, REML = FALSE)

summary(model3)


dfsss1 = ranef(model2)

dfsss11 = dfsss$`drug_description:State`

dfsss12 = dfsss$State


#model for average medicare payment

mode_medicare = lmer(medicare_coverage~domain_score+engagment_score+safety_score+

                        cost_reduction_score+median_income+Total_discharge+

                        (1|drug_description), data = df, REML = FALSE)

summary(mode_medicare)

ranef(mode_medicare)

library(stargazer)

stargazer(model1, model2, model3,mode_medicare, type = "text", single.row = TRUE)
```

## Appendix B - Python Code for Data Merge

```
#Data merging is done in python

import pandas as pd

import os

os.chdir("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT")

df_left =
pd.read_csv("Medicare_Inpatient_Hospitals_by_Provider_and_Service_2019_12_31.csv"
)

df_right = pd.read_csv("hvbp_tps.csv")

df_final =df_left.merge(df_right, left_on='Rndrng_Prvdr_CCN', right_on='Facility
ID', how="left")

df_final.to_csv("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT/final1.csv", index =
False)

df = pd.read_csv("final1.csv")

df2 = pd.read_csv("Hospital_Provider_Cost_Report_Characteristics_2017_11_02.csv")

df_final2 = df.merge(df2, right_on='Provider CCN', left_on='Rndrng_Prvdr_CCN',
how="left")

df_final2.to_csv("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT/final2.csv", index
= False)

df_final2 = pd.read_csv("final2.csv")

df_income = pd.read_csv("Income.csv")

df_final =df_final2.merge(df_income, left_on='Rndrng_Prvdr_Zip5',
right_on='Area', how="left")

df_final.to_csv("C:/Users/prasa/OneDrive/Desktop/SDM_PROJECT/final3.csv", index =
False)
```