Hospital Inpatient Charges for top 10 DRGs in the U.S.

Background

Medicare is a federally-funded health insurance program that pays for health care (hospital and medical) expenses, typically for senior citizens aged 65 and older, adults with certain approved medical conditions or qualifying permanent disabilities. The program is administered by the Centers for Medicare and Medicaid Services (CMS), which is a division of the U.S. Department of Health and Human Services.

For payment purposes, inpatient stays for patients are divided into groups by means of a statistical classification system known as Diagnosis-related Groups(DRG). All possible diagnoses are divided into more than 20 major body system groups, which are subdivided into almost 500 groups for the purpose of Medicare reimbursement. In addition to the diagnoses involved, factors such as hospital resources necessary to treat the condition, are used to determine the DRG payment amount. For every DRG assigned to a given patient, the hospitals are paid a fixed amount for the inpatient services involved.

Client

As the agency monitoring payments to hospitals for medicare eligible inpatient expenses, the CMS would potentially be interested in knowing whether certain hospitals are overcharging for the same diagnoses as compared to others. Charges and payments may differ based on provider state, number of discharges, income groups of the patients as well as whether the hospital is a teaching hospital or not. Based on the analysis of billing and payment information, the CMS would determine whether the payment rates need to be modified for particular providers.

Data

The hospital inpatient charges dataset for the top 10 DRGs in the U.S. is owned by the US government and is freely available on Data.gov. It is also available on Kaggle.com.
The dataset consists of the following information:
- DRG Code and definition

- Provider ID and Name
- Provider Street Address
- Provider City
- Provider State
- Provider Zip Code
- Hospital Referral Region Description
- Total Discharges
- Average Covered Charges
- Average Total Payments
- Average Medicare Payments

The DRG Codes, Average Covered Charges and Average Medicare Payments are the most important variables in the dataset in order to understand the billing and payment differences across the providers. These may differ across individual providers as well as across states.

The lack of time series data and limited variables are some of the limitations of this dataset. Due to the lack of this information, we are unable to make any predictions regarding charges and payments. It also lacks information regarding different characteristics of the listed institutions such as number of people served, whether they are teaching hospitals, etc. This limits analysis of factors influencing the charges and payments.

Goal and Objectives

The goal of the project is to gain insights into the inpatient billing and payment information so as to aid the CMS in making policy decisions related to payment rates assigned per DRG for each provider.

The specific objectives of the analysis are:
1. To determine the distribution of charges for particular DRGs across states
2. Visualize the differences in charges for DRGs across states
3. To assess the differences in Medicare reimbursement for the top 10 billed DRGs across states

Data will be accessed and cleaned. Data visualization techniques will be utilized for data exploration. Clustering will be performed in order to identify patterns within the dataset. Linear regression will be performed to discover associations between charges and other variables in the dataset.

Data Wrangling

The dataset was fairly clean with no missing variables or extreme outliers. The variable names were long and were trimmed for ease of coding. The DRG variable contained both DRG code as well as definition. For example, "176 - PULMONARY EMBOLISM W/O MCC". These we split into two separate variables, "code" and "definition".
Three variables(Average.Covered.Charges, Average.Total.Payments, Average.Medicare. Payments) were character variables and the observations contain "$". These were converted to numeric and the "$" was removed.

Data Exploration

The first step in data exploration was to examine the dimensions and structure of the dataset.
This revealed that the dataset has 163,065 observations and 12 variables.
The following table provides the summary information for the numeric variables:

Table 1: Summary information

Variable Name	Mean	Minimum Value	Maximum Value
Average Covered Charges	$36,134	$2,459	$929,119
Average Medicare Payments	$8,494	$1,149	$154,621
Average Total Payments	$9,707	$2,673	$156,158
Total Discharges	43	11	3383

Based on the average covered charges, the costliest DRGs were determined (Table 2).

Table 2: Costliest DRGs

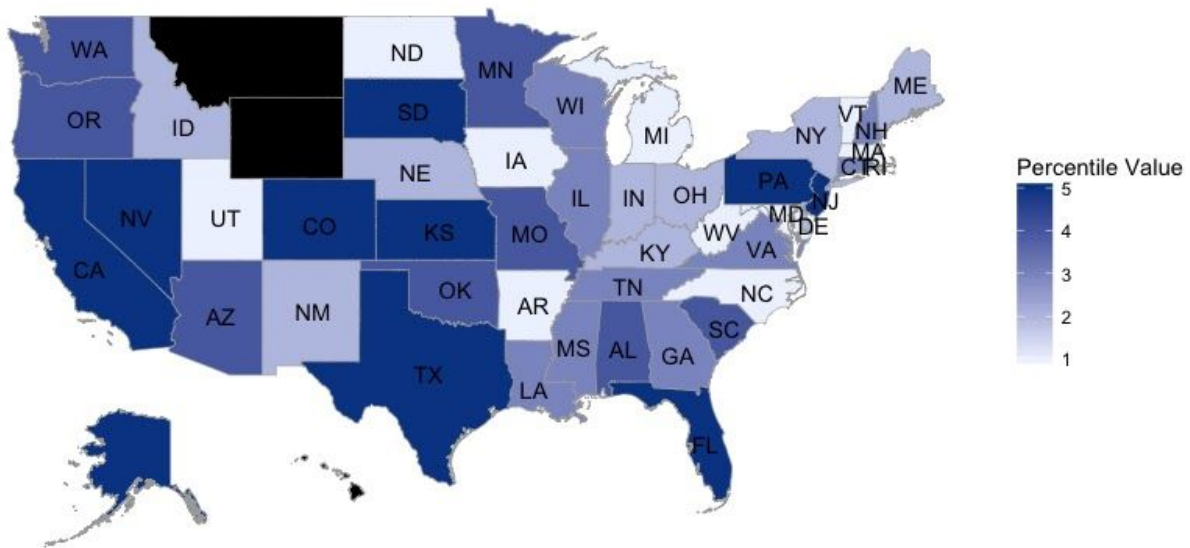| DRG	Average Covered Charges
870 - SEPTICEMIA OR SEVERE SEPSIS W MV 96+ HOURS	$163227.33
207 - RESPIRATORY SYSTEM DIAGNOSIS W VENTILATOR SUPPORT 96+ HOURS	$143428.05
853 - INFECTIOUS & PARASITIC DISEASES W O.R. PROCEDURE W MCC	$139186.35
329 - MAJOR SMALL & LARGE BOWEL PROCEDURES W MCC	$135330.94
246 - PERC CARDIOVASC PROC W DRUG-ELUTING STENT W MCC OR 4+ VESSELS/STENTS	$96348.81
460 - SPINAL FUSION EXCEPT CERVICAL W/O MCC	$95568.28
238 - MAJOR CARDIOVASC PROCEDURES W/O MCC	$85010.45
252 - OTHER VASCULAR PROCEDURES W MCC	$83502.96
469 - MAJOR JOINT REPLACEMENT OR REATTACHMENT OF LOWER EXTREMITY W MCC	$81704.23
480 - HIP & FEMUR PROCEDURES EXCEPT MAJOR JOINT W MCC	$75339.61

The average covered charges can be visualized with the help of the following heat map, which illustrates the difference in billed amount between states for the same DRG. The average covered charges for each state include the charges for individual hospitals/providers within each state. It suggests that some states, such as California, New Jersey and Nevada, charge more for the same DRG as compared to other states, whereas Maryland and Montana seem to charge a lot less for the same diagnoses.

Furthermore, the same states seem to be consistently bill higher amounts across all DRGs as compared to other states.

Figure 1: Heat map displaying average covered charges by state for top billed DRGs

In order to explore differences in charges and payments across states for a particular DRG, the top billed DRG, "870 - SEPTICEMIA OR SEVERE SEPSIS W MV 96+ HOURS", was selected arbitrarily and analyzed further.

Figure 2: Mapping average covered charges by state for DRG 870 (Septicemia or severe sepsis)

In addition to the heatmap, figure 2 is another way to visualize differences in average covered charges by state for an individual DRG. The average covered charges were classified into quintiles and assigned values from 1 to 5 (5 being highest) for mapping. The category 5 contains states with the top 20 percentile of average covered charges.

Figure 3: Mapping average medicare payments for top billed DRG (870)
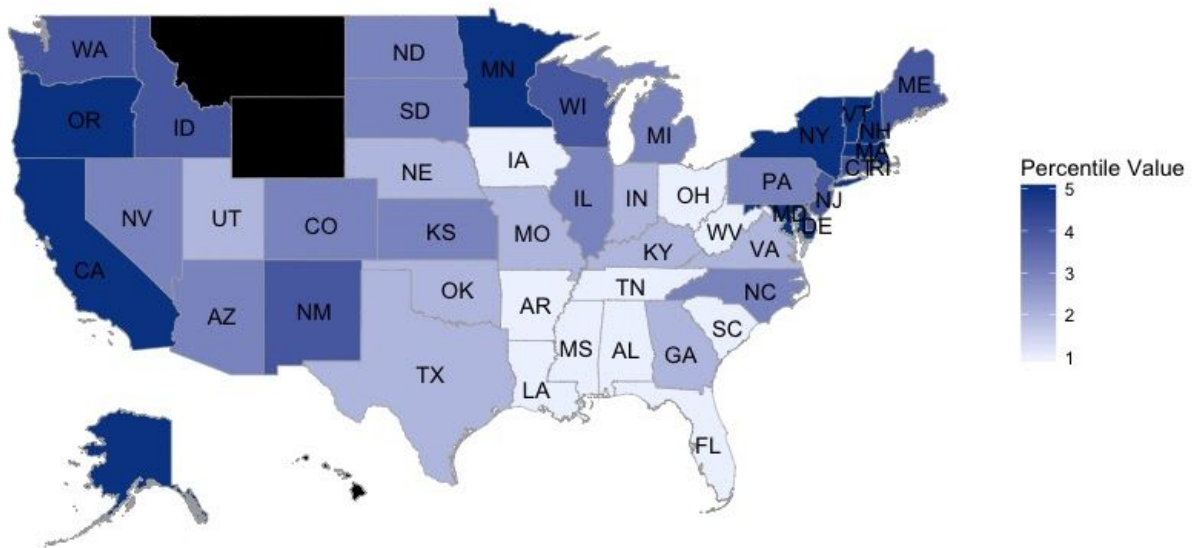
The average medicare payments were similarly classified into quintiles and assigned values from 1 to 5 (5 being highest) for mapping in figure 3. The category 5 contains states with the top 20 percentile of average medicare payments by CMS. This map is useful to assess discrepancies in billed amount vs payments. For example, the state of Nevada is in the top 20 percentile (category 5) for amount billed but receives average medicare payments in the (60-80 percentile) category 4. The reverse is true for Utah which bills in category 1 (bottom 20 percentile), but gets paid in category 2 (20-40 percentile).

Figure 3: Mapping percentage difference in covered charges and medicare payments for top billed DRG (870)

Figure 3 maps the difference between billed charges and medicare payments, as a percentage, i.e. (Average Covered Charges - Average Medicare Payments)/Average Covered Charges * 100. This is another way of assessing the difference in payments by state, where the magnitude discrepancy is controlled for the billed amount.

Limitations

Although this dataset has hospital level information for billed as well as medicare payments, it does not contain additional information about the characteristics of these hospitals, such as teaching hospital/ community hospital, primary/secondary/tertiary center, population served, etc., which could potentially be used to explain the variance in billed charges and reimbursement.
This is a cross-sectional dataset that does not allow us to trend the reimbursement discrepancy (or other outcomes) over time or the factors affecting such trends.

Conclusions

As the regulatory authority responsible for health care payments, CMS would be interested in the regional differences as well as differences across DRGs in billed charges and reimbursements. Mapping these differences using the above techniques provide an intuitive way of assessing such discrepancies and distributing resources appropriately. Such data can also inform policy decisions when deciding reimbursements for particular DRGs as well as healthcare budgets.

Future Work

Additional data related to provider characteristics and trends will enable us to explain the variability in reimbursement using linear regression and time series analysis. This will help the CMS make more informed policy decisions.

Acknowledgements

Mr. Joel Nicholas Bangalan - My Mentor