Final Project – Data Visualizations

MIS 505 - Professor Christopher Sibona

May 2, 2020

Gil Graybill

**Executive Summary**

Patient Copays can vary greatly for hospital procedures covered by Medicare in North Carolina. The differences can occur within an individual procedure, by region, and even within the same city. There is a handful of procedure categories that are more likely to have a high copay. There is less variance for the procedures that are performed the most often. The reason for the differences cannot be explained by the data that was examined.

**Data Overview and Exploratory Data Analysis**

The data I chose for analysis shows the Medicare charges and payments for the top 100 procedures performed in the United States in 2017 - <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3> . This topic interests me because I believe there is a lot of waste in America’s health care system. I had hoped that by analyzing this data I could find proof that a lack of competition or the inability to compare prices leads to higher prices.

This dataset has over 162K rows with 12 columns. The columns consist of the DRG procedure code (A standardized classification system hospitals use for categorizing procedures - <https://coder.aapc.com/drg-codes-range>) , provider (hospital) information (ID, name, address), a region that provider is assigned to, discharges (the number of times that procedure was performed), Average Covered Charges, Average Total Payments, and Average Medicare Payments.

The last three columns are where the interesting data was. “Average Covered Charges” is what the hospital charges Medicare for a procedure. Briefly looking at the data, I could see that for the same procedure, some hospitals were charging 2-3 times more. Jackpot! Let’s find those crooked hospitals and put them on “60 Minutes”. However, the description of the field has the following: “These will vary from hospital to hospital because of differences in hospital charge structures.” Further research indicated that this field was like the quoted retail price for a car, diamond, or a knife set advertised on late-night TV…nobody ever pays that. It’s the number that the hospitals use to start negotiating prices, whether it’s with Medicare, a private insurer, or someone paying out-of-pocket. Since there are two levels of randomness involved there, I decided not to analyze based on this field.

“Average Medicare Payments” is what Medicare actually gives pays the provider for the procedure, and these looked to vary a lot. Perhaps this was ripe for analysis. More research showed that hospitals have a different “base rate” for Medicare payment. It makes sense that a hospital in New York City should get paid more for the same procedure than one in Lumberton, NC. All I needed to do was find the base rate for each hospital, join that to the dataset and divide to get what the standardized Medicare payment should be. Alas, the base rate is not public information. Could I calculate it? Seeing this document discouraged it - <https://oig.hhs.gov/oei/reports/oei-09-00-00200.pdf> .

“Average Total Payments” is the Medicare payments plus the co-payment and deductible amounts that the patient is responsible for. Could the copay and deductible be worthy of analysis? Creating a new calculated field with a simple subtraction showed that this could be an interesting direction to take my research.

The more I thought about it, the more it made sense. I’m supposed to make the data interesting. Does a patient really care if Medicare pays a hospital X for a procedure? No, it’s all funny money until it comes down to the copay. Then, as a consumer, it gets your attention.

Health care is not a classic “supply-and-demand” dynamic. In many instances as a patient there is no choice about where to have a procedure done, either because of timing, insurance plan affiliations, or specialties. For an individual to shop around for services by copay is a difficult, if not impossible task. My wife is a nurse at a hospital, and she says that quickly determining the copay is not a service that is provided. I think that it should be, and if this information was available, the “Invisible Hand” of capitalism would bring down copays, if not the entire actual cost of the procedure.

I did choose to filter down the data to use only North Carolina providers. I had flashbacks to my Predictive Analytics class where analyzing large datasets could take all night and I didn’t want my computer crashing while working with 160K rows. Filtering took it down to just over 5000 rows.

**Exploratory and Explanatory Analysis and Chart Choice**

Boxplot Patient Copay - A fine place to start is to create a boxplot for the copay of each procedure with the providers for the points. This showed me that there is quite a variance in the mean and standard deviation for each procedure, and the presence of outliers. I chose to make the dots bigger to highlight the outliers better. Also it seemed like it would be nice to know which providers were represented by the outliers, so I added the provider name to the label so it would show up in the tooltip. However, I wanted more detail in the tooltip for the copay ($1.4K instead of $1K) but that is tied to the format of the axis. Google helped me out with that one.

Stdev of Copay – The standard deviation of the copay with respect to the procedure (DRG Definition) would be the variable that I would do most of my analysis with. I created a set consisting of all procedures with a Stdev of more than $1000, which was 11 of them. In this chart I also added the number of providers that performed that procedure, to see if anything looked worthy of analysis.

Scatterplot of hospitals vs copay – I was intrigued enough by the number of providers performing a given procedure that I created a scatterplot vs. the StDev of Copay. There seems to be a logarithmic correlation between high StDev of Copay and a lower number of providers.

Dashboard: STDEV of Copay – The three preceding charts. I chose the coral color for StDev of Copay related analysis

Boxplot Patient Copay by DRG CLASS – Let’s do the same treatment for DRG classification. I wrote a calculation that grouped the DRG by their official classification (From <https://coder.aapc.com/drg-codes-range>, with the codes indexed below) and then did the same boxplot analysis. The same outliers show up, but that leads to the next chart.

stdev of Copay per DRG Class – It looks like much of the large variance is grouped into 5 different DRG Categories. Could this be a place to dive deeper to find out why Copays vary so much?

Better Heat Map – Is there variance in Copay within a region? I chose to concentrate on the hospitals in the regions of RTP (Raleigh/Durham/Chapel Hill), not because that’s where I’m from, but because it had a good cross-section of urban and rural hospitals, along with several hospitals close to each other. I mapped the StDev of Copay for the high variance procedures vs. the providers in the area. A standard heat map didn’t really highlight the outliers like I wanted them to, so I created an LOD calculation for a new measure to define the rankings I wanted to show. I feel the chart was worthy of its own dashboard.

Multiple Hospital Copays top variance – Is there variance within a city? I chose the cities that had more than one hospital performing a high variance procedure. Even within the same city the Copays vary. I debated between keeping the colors for the providers or not and I went with keeping them. There’s not too many that makes it look like a “My Little Pony” convention, but I think it helps when looking down the rows. Again, I deemed it worthy of its own dashboard.

Most Popular procedures and % - The same type of variance probably won’t occur for the most popular procedures, but will it be noticeable? First we need to know what the most popular procedures are (I would not have guessed “Major joint replacement or lower extremity reattachment”, but it seems like a catchall category whereas other things like Pneumonia and Chest Pain/Heart Failure are broken out more. But that’s an analysis for another day). I added the % of hospitals that performed that procedure just for fun. I created a set for “Top 10 performed procedures”.

City Popular Procedures – Similar to one of the previous charts, I charted the “Top 10 Performed Procedures against the big cities. The variance is certainly not as pronounced as the top variance procedures, but there are just a few cases where it seems relevant. Does competition work?

Patient copay and popular procedures (Dashboard) – Combine the preceding graphs together into a dashboard.

StD Copay per region top 10 procedures – Do Copays vary by region? The data has a “Hospital Referral Region” that is defined by a large city in the region, and each hospital is assigned to one. Comparing Stdev of Copay for each region for the “Top 10 Performed Procedures” shows there is not a lot of variance, with the Hickory region leading the way, and the Raleigh region being the smallest. “Top 10 Performed Procedures” color is set to gold for those who are noticing.

StD Copay per region all - When it’s expanded to all procedures, the variance is larger, and different regions take the prize. Highest Copay goes to Durham, with the lowest going to Wilmington.

StD Copay per region top StDev procedures – Might as well check it against the top StDev procedures. The results mirror the preceding chart for all procedures.

Regional Maps (Dashboard) – Put all the maps together and hope apologize for the distortion.

Histogram of copay – Is there a correlation between Copay and the reimbursed cost of the procedure? It seems like a logical place to look. Start off with a good old-fashioned histogram, which shows that most Copays are between $700 and $1200 for all procedures. I did choose to cut off the two big outliers so the graph would fit better.

Corr. Copay and med. Payments – Is there a correlation between the Average Medicare Payments and the Copay? Not really, since the Copays have a tight distribution compared to the Medicare payments.

Corr copay% vs med. Payments – If we calculate the percentage of a payment that the Copay makes up, is there a correlation between it and the Medicare payments? Yes there is. For a low-cost procedure, the proportion of the Copay is higher, and the reverse is also true. It’s sensible and good to know. The correlation appears to be logarithmic instead of linear.

Copay Distribution (Dashboard) – Put the three of them on a dashboard, and a story unfolds.

**Exploratory Analysis that didn’t make the cut**

Average Total Payment Map – When I first thought that Total payments actually meant something I was going to chart it per provider. Then I also realized that if there’s more than one provider in a city that it was hard to get it to show up.

Count of Hospitals per city – I did use this for analysis, but only to get the list of multiple hospitals in the city. Quick, what’s the “2nd Provider” in Wilmington? “WILMINGTON TREATMENT CENTER” makes the list for 99 discharges of “897 - ALCOHOL/DRUG ABUSE OR DEPENDENCE W/O REHABILITATION THERAPY W/O MCC”.

Bad Z Score and Bad z score 2 – I tried to normalize the variances and it seemed like I was getting close, but it just wasn’t working. When I tried to explain it to my cat she looked really confused and I realized that I was confused myself. Since this was a visualization class I decided to let that go for another day. Normalizing would have been nice, but I was still able to get some good visuals without it.

Bad patient copay 329 RTP – At one point I was going to plot each of the high variance procedures against the RTP providers. I decided against it because (1) it would take up too much space and (2) The outliers would mean a lot of wasted map space, and (3) a heat map would be better. The most prominent item on this chart is the huge cost in Sanford, but that draws attention from the high copay cost in 2 of the three Raleigh hospitals, which is what inspired the examinations at a city-wide level.

Bad Heat Map – I tried a few things to make it pop like different colors for the text, but it just wasn’t doing it for me. The result was turning to Google and coming up with the “Better Heat Map” chart.

Max Copay with label – I thought it would be neat to just show the huge outliers, but instead I realized I could modify “Boxplot Patient Copay”, by adding the label and making the dots bigger so they wouldn’t be mistaken for dirt on my screen. I did have the labels working correctly with some LOD magic but I deleted it when I thought I wouldn’t need it.

Raleigh CoPays top Variance and Charlotte Popular Procedures – I started doing the analysis for Raleigh, Durham, and Charlotte (3,3, and 4 hospitals), and then stacking the graphs on the dashboard, but then realized I could do it all in one. I added the other cities that had more than one provider in it for completeness.

**Appendix: DRG Codes**

PRE **001-017** [**Pre-MDC**](https://coder.aapc.com/drg-codes-range/1)

01 [**020-103**](https://coder.aapc.com/drg-codes-range/2) [**Diseases & Disorders of the Nervous System**](https://coder.aapc.com/drg-codes-range/2)

02 [**113-125**](https://coder.aapc.com/drg-codes-range/3) [**Diseases & Disorders of the Eye**](https://coder.aapc.com/drg-codes-range/3)

03 [**129-159**](https://coder.aapc.com/drg-codes-range/4) [**Diseases & Disorders of the Ear, Nose, Mouth & Throat**](https://coder.aapc.com/drg-codes-range/4)

04 [**163-208**](https://coder.aapc.com/drg-codes-range/5) [**Diseases & Disorders of the Respiratory System**](https://coder.aapc.com/drg-codes-range/5)

05 [**215-320**](https://coder.aapc.com/drg-codes-range/6) [**Diseases & Disorders of the Circulatory System**](https://coder.aapc.com/drg-codes-range/6)

06 [**326-395**](https://coder.aapc.com/drg-codes-range/7) [**Diseases & Disorders of the Digestive System**](https://coder.aapc.com/drg-codes-range/7)

07 [**405-446**](https://coder.aapc.com/drg-codes-range/8) [**Diseases & Disorders of the Hepatobiliary System & Pancreas**](https://coder.aapc.com/drg-codes-range/8)

08 [**453-566**](https://coder.aapc.com/drg-codes-range/9) [**Diseases & Disorders of the Musculoskeletal System & Connective Tissue**](https://coder.aapc.com/drg-codes-range/9)

09 [**570-607**](https://coder.aapc.com/drg-codes-range/10) [**Diseases & Disorders of the Skin, Subcutaneous Tissue & Breast**](https://coder.aapc.com/drg-codes-range/10)

10 [**614-645**](https://coder.aapc.com/drg-codes-range/11) [**Endocrine, Nutritional & Metabolic Diseases & Disorders**](https://coder.aapc.com/drg-codes-range/11)

11 [**652-700**](https://coder.aapc.com/drg-codes-range/12) [**Diseases & Disorders of the Kidney & Urinary Tract**](https://coder.aapc.com/drg-codes-range/12)

12 [**707-730**](https://coder.aapc.com/drg-codes-range/13) [**Diseases & Disorders of the Male Reproductive System**](https://coder.aapc.com/drg-codes-range/13)

13 [**734-761**](https://coder.aapc.com/drg-codes-range/14) [**Diseases & Disorders of the Female Reproductive System**](https://coder.aapc.com/drg-codes-range/14)

14 [**768-833**](https://coder.aapc.com/drg-codes-range/15) [**Pregnancy, Childbirth & the Puerperium**](https://coder.aapc.com/drg-codes-range/15)

15 [**790-795**](https://coder.aapc.com/drg-codes-range/16) [**Newborns & Other Neonates with Conditions Originating in Perinatal Period**](https://coder.aapc.com/drg-codes-range/16)

16 [**799-816**](https://coder.aapc.com/drg-codes-range/17) [**Diseases & Disorders of Blood, Blood Forming Organs, Immunologic Disorders**](https://coder.aapc.com/drg-codes-range/17)

17 [**820-849**](https://coder.aapc.com/drg-codes-range/18) [**Myeloproliferative Diseases & Disorders, Poorly Differentiated Neoplasms**](https://coder.aapc.com/drg-codes-range/18)

18 [**853-872**](https://coder.aapc.com/drg-codes-range/19) [**Infectious & Parasitic Diseases, Systemic or Unspecified Sites**](https://coder.aapc.com/drg-codes-range/19)

19 [**876-887**](https://coder.aapc.com/drg-codes-range/20) [**Mental Diseases & Disorders**](https://coder.aapc.com/drg-codes-range/20)

20 [**894-897**](https://coder.aapc.com/drg-codes-range/21) [**Alcohol/Drug Use & Alcohol/Drug Induced Organic Mental Disorders**](https://coder.aapc.com/drg-codes-range/21)

21 [**901-923**](https://coder.aapc.com/drg-codes-range/22) [**Injuries, Poisonings & Toxic Effects of Drugs**](https://coder.aapc.com/drg-codes-range/22)

22 [**927-935**](https://coder.aapc.com/drg-codes-range/23) [**Burns**](https://coder.aapc.com/drg-codes-range/23)

23 [**939-951**](https://coder.aapc.com/drg-codes-range/24) [**Factors Influencing Health Status & Other Contacts with Health Services**](https://coder.aapc.com/drg-codes-range/24)

24 [**955-965**](https://coder.aapc.com/drg-codes-range/25) [**Multiple Significant Trauma**](https://coder.aapc.com/drg-codes-range/25)

25 [**969-977**](https://coder.aapc.com/drg-codes-range/26) [**Human Immunodeficiency Virus Infections**](https://coder.aapc.com/drg-codes-range/26)

Ungroupable [**981-999**](https://coder.aapc.com/drg-codes-range/27)