

Does Medicare Inequality Exist by State?

1/10/19

Chris Jiang

Table of Contents

- Page 3: **Introduction**
- Page 4: **Objective**
- Page 4: **Data Wrangling**
- Page 5: **Data Modeling**
- Page:13: **Conclusion**

Appendix: some abbreviations that are used and some references

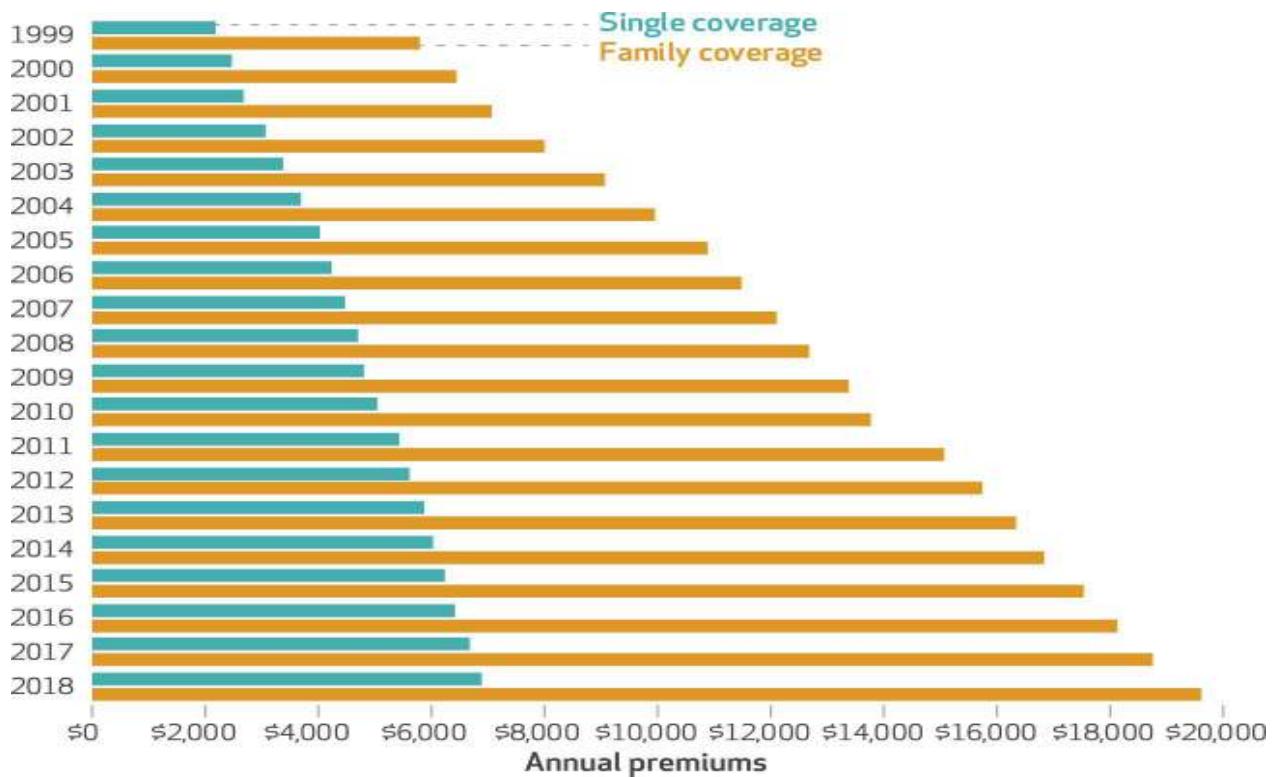
- **DRG definition (DRG)**- medical procedure
- **State**
- **Average Medicare coverage (AMC)**- average charges for services covered by Medicare
- **Average total payments (ATP)**– total payments to the provider for procedure, includes copayment and deductible which are paid by the patients
- **Average Medicare Payments (AMP)** – average amount that Medicare pays to the provider

Introduction

Healthcare has always been a serious topic and the issues that exist with it has plagued the United States for a long time. Between the rich and the poor, the healthcare difference is as plain as day. A key reason for this is the cost. Majority of Americans can barely afford quality healthcare due to low income and the price of the coverages. Many even go bankrupt due to the inability to pay for the bills.

The American healthcare system is mostly a private system where the customers shop around for different policies that are not managed by the federal or state government. Private insurance can be purchased from multiple entities: your employer, a private marketplace etc. Due to this, the policies differ greatly and the more expensive policies provide a wider arrange of coverages and the cheaper policies provide more limited coverages. The government does provide a governmental healthcare program, Medicare, for the elderly.

Below is a chart of the price in health care as the year progresses.



(https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.1001?utm_campaign=HASU%3A+10-05-18+%28Copy%29&utm_medium=email&utm_content=Health+Affairs++October+Issue%3A+Social+Determinants%2C+Drug+++Device+Prices+++More%3B+2018+Costs+For+Employer-Sponsored+Family+Health+Coverage%3B+Aftermath+Of+A+School+Shooting&utm_source=Newsletter) – link to graph

Objective

The main goal of this research is to determine if Medicare inequality exists between states. Solving healthcare inequality has always been a serious issue and Medicare has help alleviate some of the burden for the elderly. Therefore, providing a government funded program for the elderly should help bridge the gap between the difference in healthcare. The difference in healthcare is caused by the multitude of different coverages between different policies, but since Medicare is only one policy, there should be no difference in coverages that “Patient A” and “Patient B” receives.

Data Manipulation

The data set that will be used is the “Hospital Charges for Inpatients” that can be found on Kaggle (<https://www.kaggle.com/speedoheck/inpatient-hospital-charges>). This data set is owned by the US government and updated periodically. The file has 163066 rows and 12 columns with 12 unique variables.

DRG Defin... :	Provider Id :	Provider ... :	Provider S... :	Provider C... :	Provider S... :
039 - EXTRACR...	10001	SOUTHEAST AL...	1108 ROSS CL...	DO THAN	AL
039 - EXTRACR...	10005	MARSHALL ME...	2505 U S HIGH...	BOAZ	AL
039 - EXTRACR...	10006	ELIZA COFFEE ...	205 MARENGO...	FLORENCE	AL
039 - EXTRACR...	10011	ST VINCENT'S ...	50 MEDICAL P...	BIRMINGHAM	AL
039 - EXTRACR...	10016	SHELBY BAPTI...	1000 FIRST STR...	ALABASTER	AL
039 - EXTRACR...	10023	BAPTIST MEDI...	2105 EAST SO...	MONTGOMERY	AL

I will be concentrating on certain variables from the base data set for this project:

- **DRG Definition (DRG)** - medical procedure
- **State**
- **Average Medicare Coverage (AMC)**- average charges for services covered by Medicare
- **Average total payments (ATP)**– total payments to the provider for procedure, includes copayment and deductible which are paid by the patients
- **Average Medicare Payments (AMP)** – average amount that Medicare pays to the provider

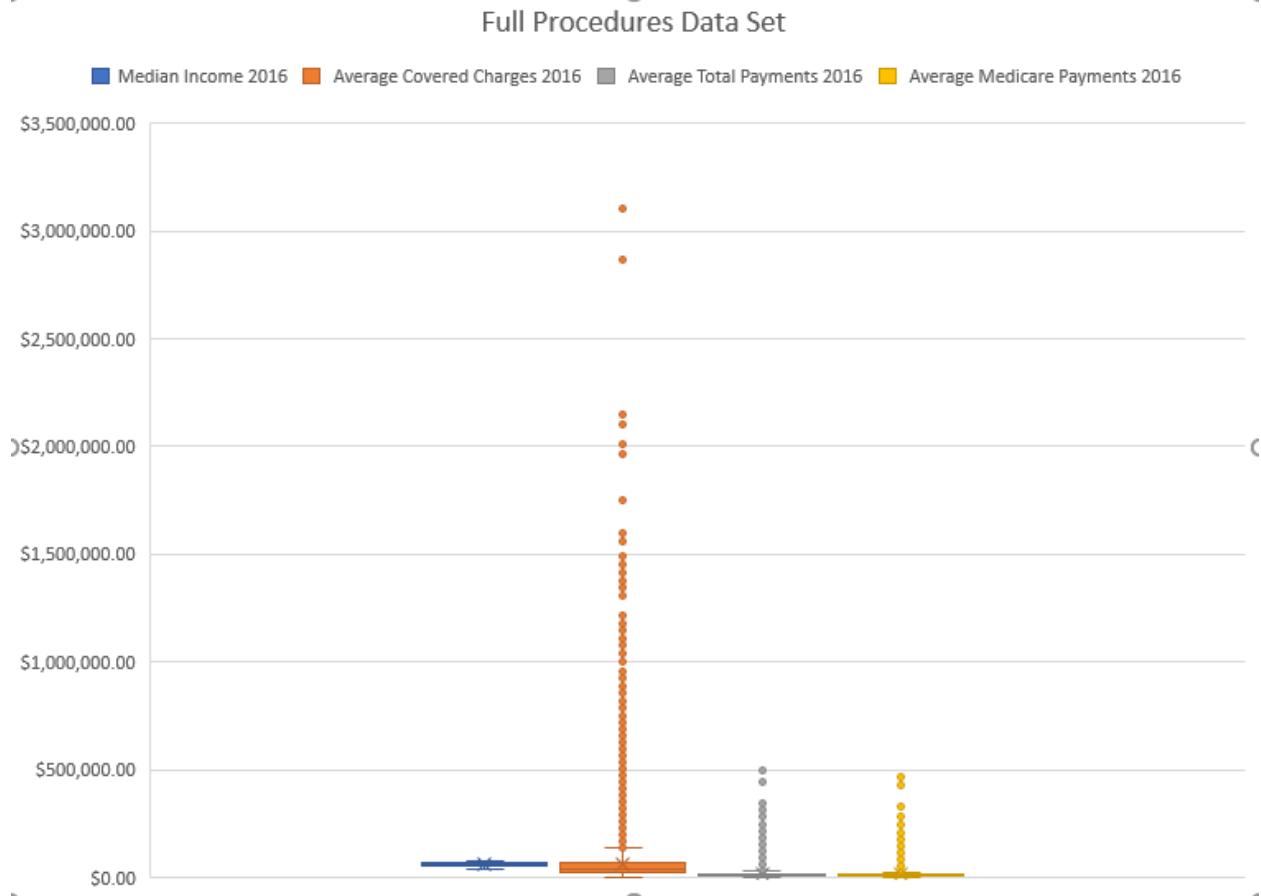
I have also added **Median Income** into the data for each state. This value is used to help differentiate between the different states.

To start with, I cross matched each state with the respective median income:

Poverty Rate	Median Income 2016	Total Discharges	Average Covered Charges 2016	Average Total Payments 2016	Average Medicare Payments 2016
0.132	66637	11	1108206.45	110174.82	104458.27
0.132	66637	11	166545.18	63184.36	54041.27
0.132	66637	11	106158.64	15111.64	13947.18
0.132	66637	11	132457.91	36407.36	32984.36
0.132	66637	11	70193.45	9348.82	6238.55
0.132	66637	11	50204.18	7790.27	6736.45
0.132	66637	11	49164.82	9547.27	6588.18
0.132	66637	11	33627.09	14440.45	13508.82
0.132	66637	11	42632.36	16043.36	9512.18
0.132	66637	11	84385.73	9351.45	4858.36
0.132	66637	11	37876.18	7068.82	4754.82

Data Modeling

In order to get a better understanding, general size, and distribution of the data, a box and whisker plot was plotted. This will help in observing the general spread and distribution of the data and show if any outliers exist within the overall data set:



Looking at what we are working with, we see that there are a lot of outliers, which makes sense due to the data set having multiple **DRGs** at varying costs. The median income range for the states seems to be compact, no outliers are present.

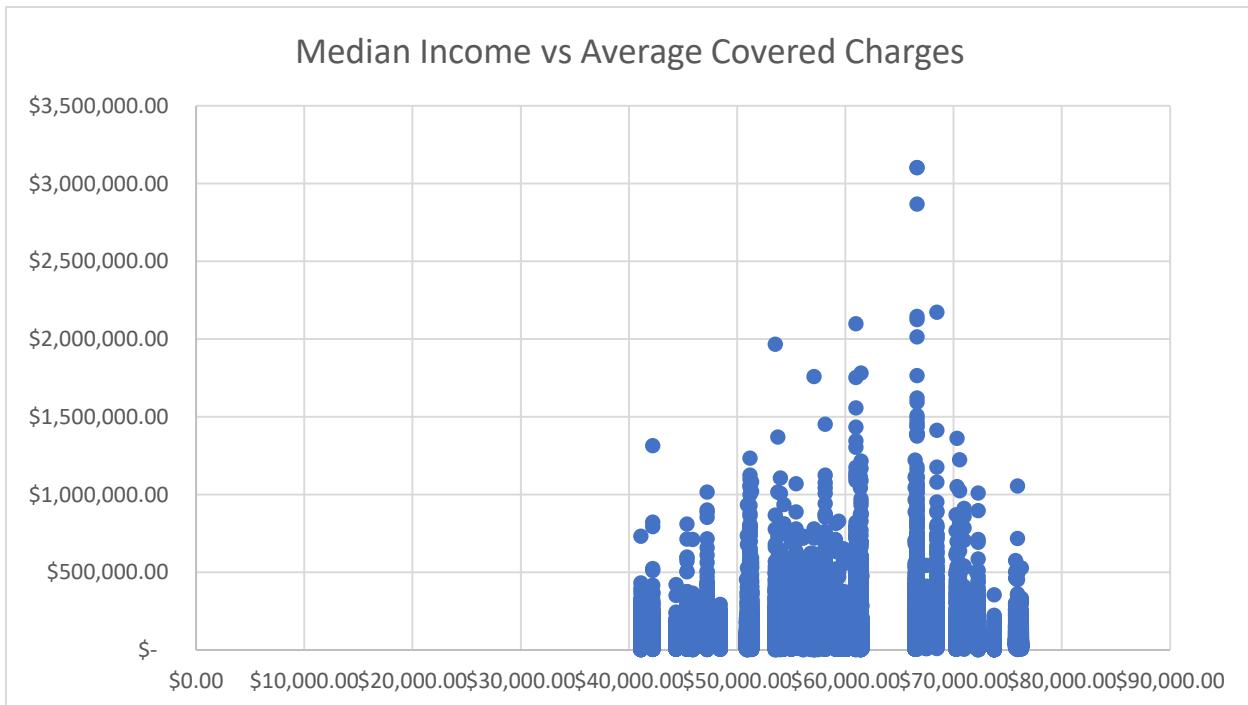
A correlation analysis is done to determine the correlation coefficients for each variable on each other – it can be seen below.

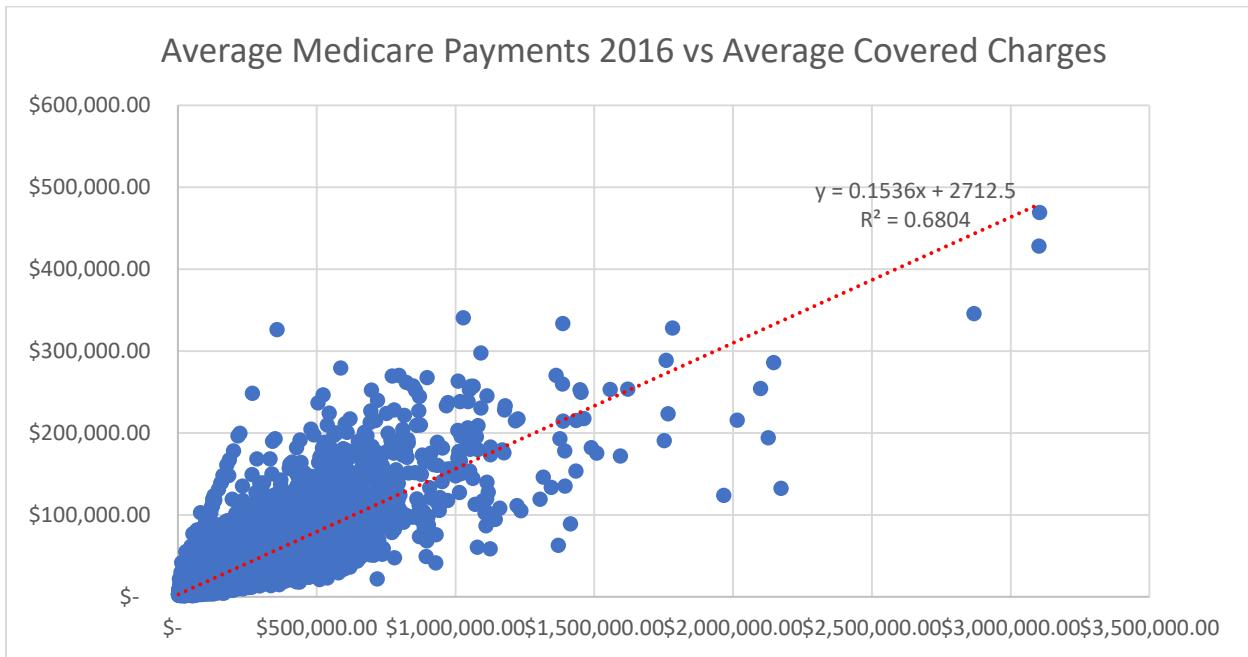
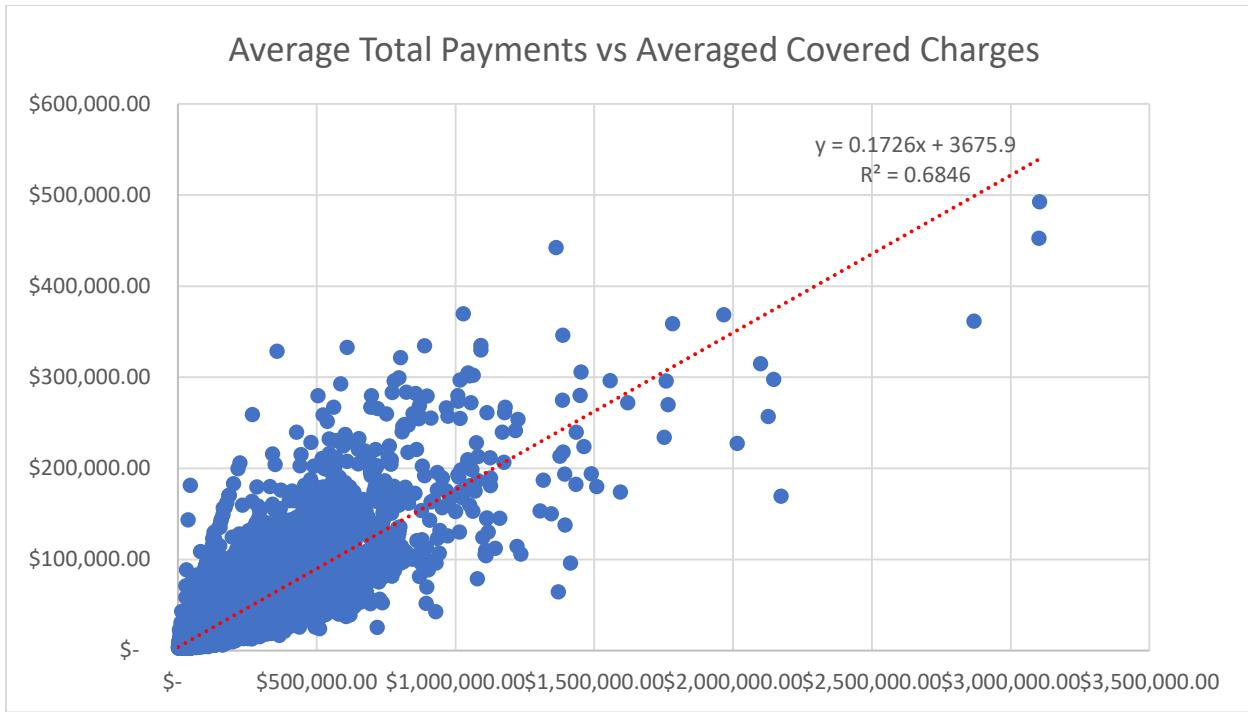
	<i>Median Income</i>	<i>Average Medicare Coverage</i>	<i>Average Total Payments</i>	<i>Average Medicare Payments</i>
<i>Median Income</i>	1			
<i>Average Medicare Coverage</i>	0.048717487	1		
<i>Average Total Payments</i>	0.112505654	0.827386742	1	
<i>Average Medicare Payments</i>	0.114710517	0.82483834	0.983775548	1

Looking at these results, we notice some observations that stand out. We want to concentrate on the **AMC** since that is the amount of money that is covered through Medicare for a **DRG**. Looking at **Median Income**, the correlation figures seems to be very low, 4%. So interpreting the overall data, the different states don't correlate to the amount of the coverage a patient receives per **DRG**.

We see that there are 2 variables that have high correlation with **AMC**: **ATP** and **AMP** with around 82% correlation Does that mean that the **ATP** and **AMP** cause determines the coverage? No, correlation isn't cause and effect. It can however be used as a predictive measure to predict coverages.

Going a bit further, a regression analysis is done for **Median Income**, **ATP** and **AMP** on **AMC** to see if a model can predict the amount of coverage base on **Median Income**(different states), **AMC**, and **ATP** In order to do this, a scatter plot is created to show the relationship between one variable and another. In this case, the variables will be plotted against **AMC**. After, a trend line will be created to show the relationship in a numeric format and with the R^2 . The Graphs are shown below:

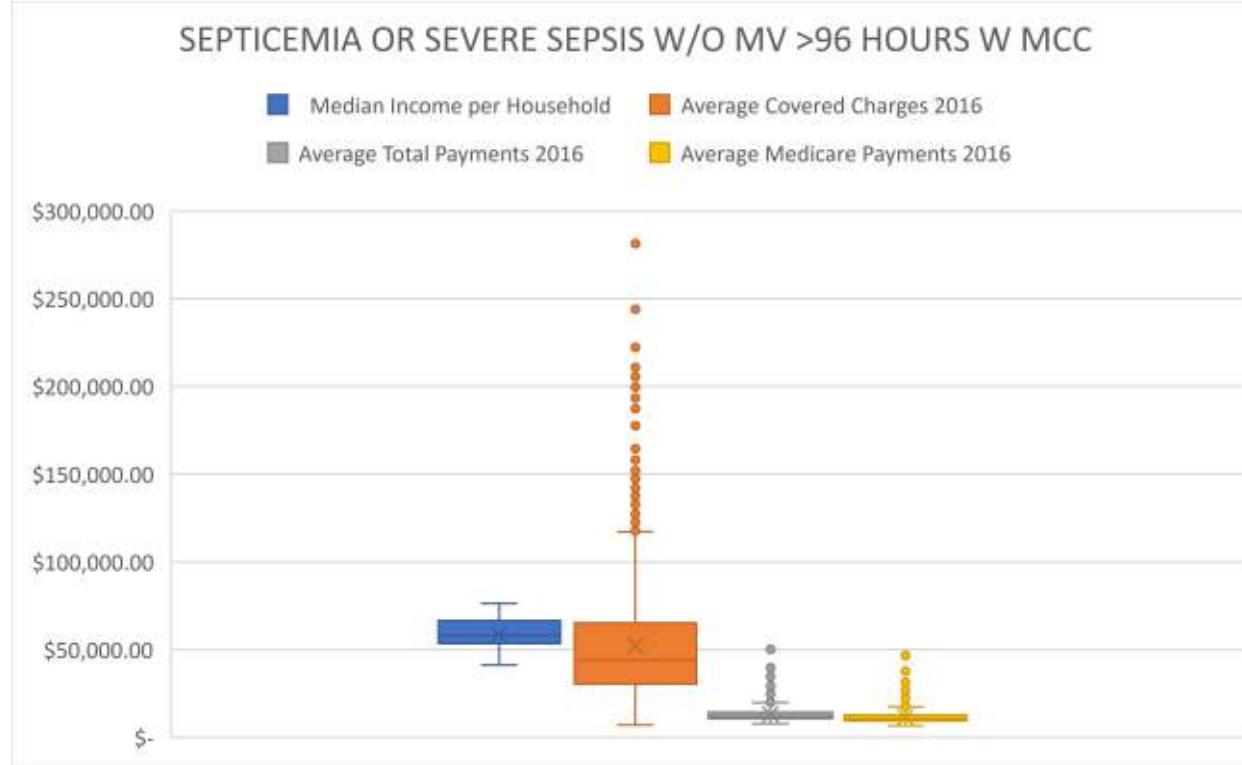




Immediately, the charts show that **Median Income** (different states) is a terrible predictor for the amount of coverage a person will receive from Medicare. The R^2 is at a measly .0024. The other charts however paint a different picture. We can see that for both **AMP** and **ATP**, there is some sort of relationship. Those variables are a better predictor for **AMC**.

In order to dig a bit deeper, I have split the data and now solely focusing on one **DRG, SEPTICEMIA OR SEVERE SEPSIS W/O MV >96 HOURS W MCC** because it is the most

frequent medical procedure from the original data set. The analysis will be done. Below, we have a box and whisker chart for the data:



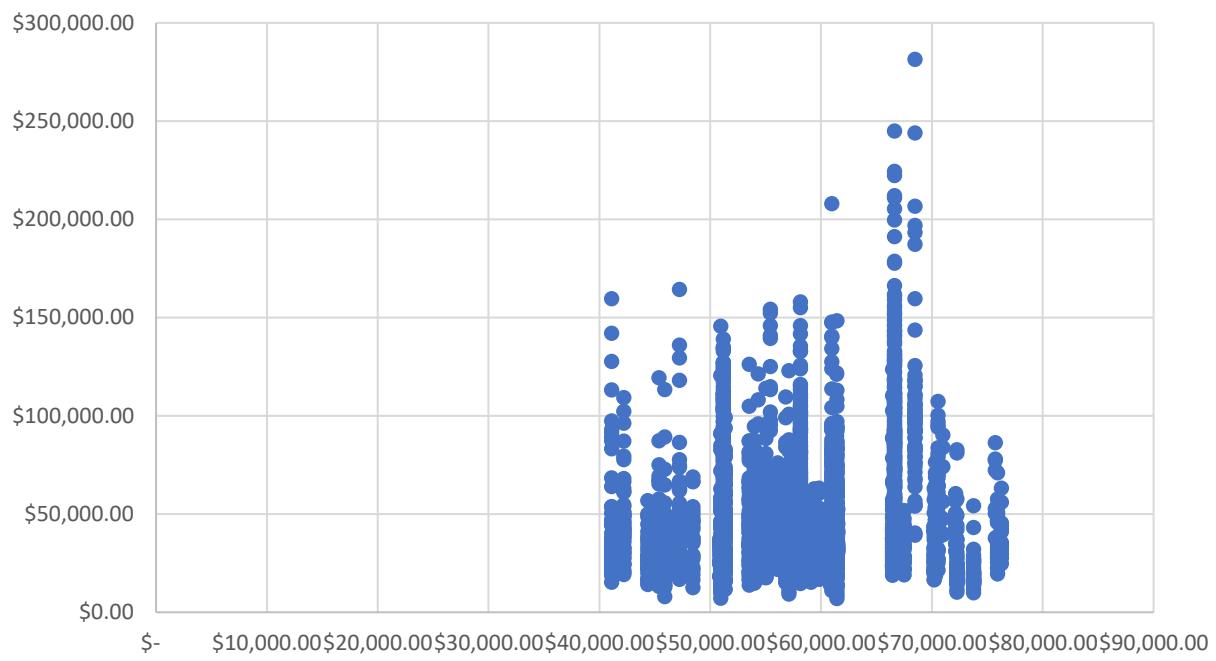
There is now a lot less outliers due to the data being more focused. Like before, a correlation analysis will also be run to determine the correlation coefficients for each of the variables:

	Median Income per Household	Average Covered Charges	Average Total Payments	Average Medicare Payments
Median Income per Household	1			
Average Covered Charges	0.123075964	1		
Average Total Payments	0.407990844	0.305376193	1	
Average Medicare Payments	0.420469604	0.282595456	0.937345725	1

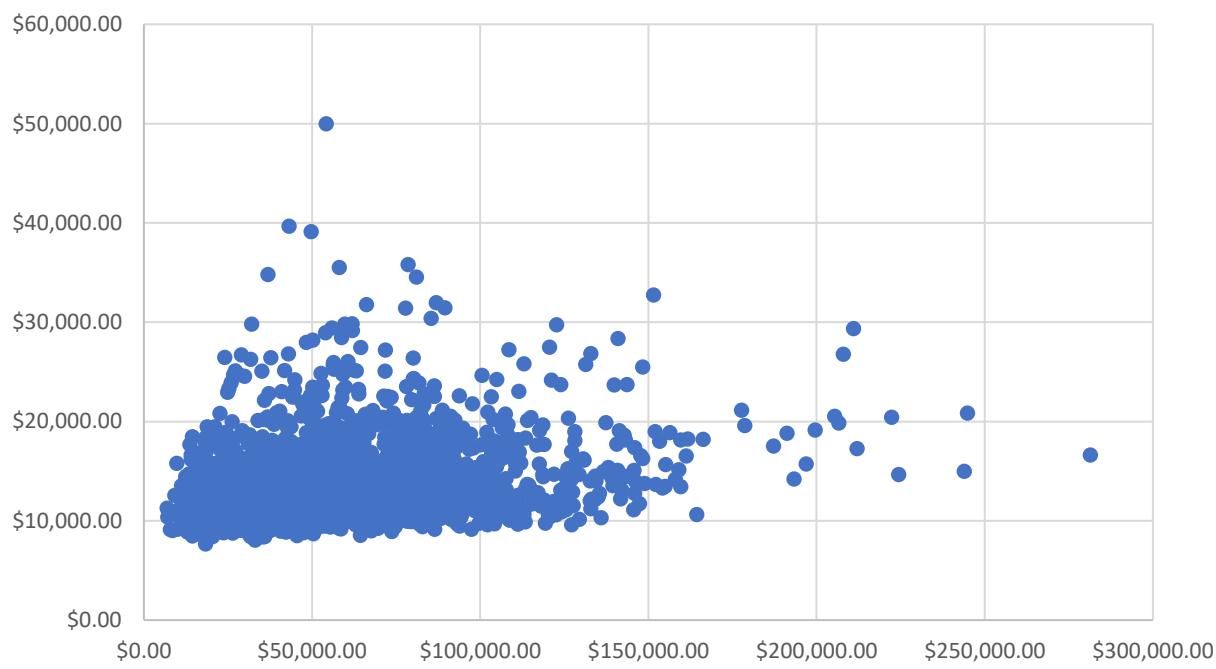
The correlation results are totally different from before. **The Median Income** (different states) has the same result as before, but now **ATP** and **AMP** have low correlation coefficients. This might be due to the data being more specialized rather than an overview of the different **DRGs**. The data being more limited, it can affect the overall correlation coefficients for the data.

Also like before, scatter plots were created and a best fit trendline will be generated along with the formula and the R^2 . Judging by a change in the correlation coefficients, we can assume that the R^2 's will be different as well. Below are the scatterplots:

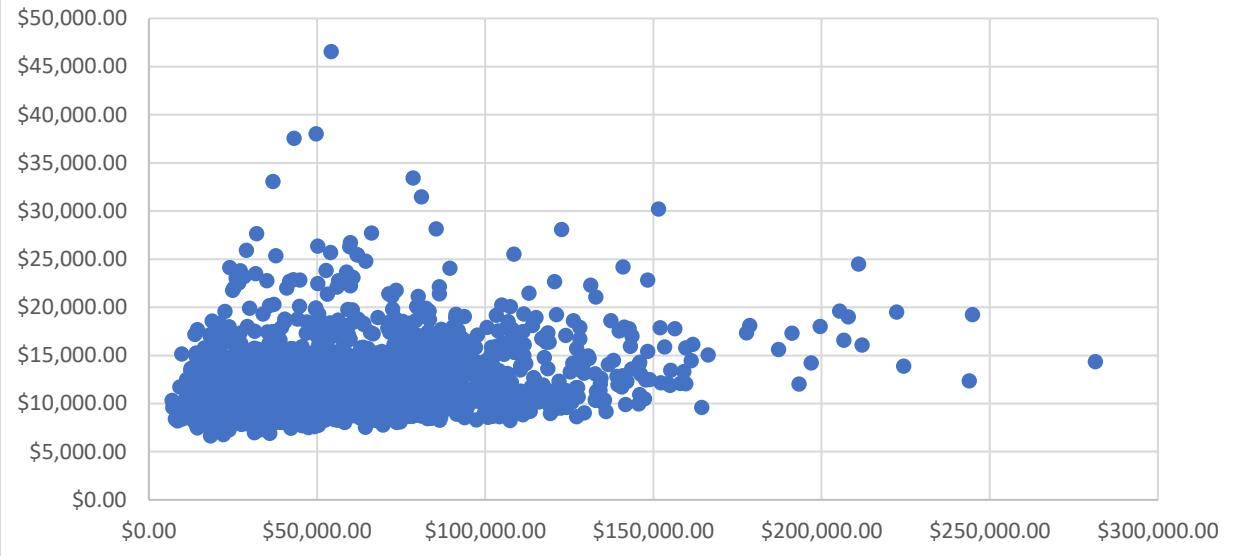
Median Income 2016 vs Averaged Covered Charges 2016



Average Total Payments 2016 vs Medicare Charges 2016



Average Medicare Payments 2016 vs Average Covered Charges 2016



Looking at the revised charts, we can clearly see some big changes. The R^2 's have drastically decreased for the **ATP** and **AMP**, at a measly 9% and 7.9% respectively. The **Median Income** (different states) went up a minuscule amount and is now at 1%. In terms of predicting coverage on a specific procedure, seems it is rather inaccurate to use **AMP** and **ATP**

A regression analysis was then done on the data set to determine if the different variables are a good predictor of the coverages that a patient receives from Medicare. The results are shown below:

Regression Statistics	
Multiple R	0.305555889
R Square	0.093364401
Adjusted R Square	0.092404321
Standard Error	29783.56386
Observations	2837

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	19089.53043	4181.375115	4.565371416	5.19868E-06	10890.68297	27288.37789	10890.68297	27288.37789
Average Total Payments	2.790136176	0.429870403	6.490644981	1.0054E-10	1.947245556	3.633026797	1.947245556	3.633026797
Average Medicare Payments	-0.290021597	0.501516115	-0.578289686	0.563114522	-1.273395251	0.693352057	-1.273395251	0.693352057
Median Income	-0.001794796	0.077905148	-0.023038218	0.98162141	-0.154551343	0.150961751	-0.154551343	0.150961751

The R^2 is very small at a 9% which shows the variation of Medicare coverage is explained 9% of the time with the independent variables, **ATP**, **AMC** and **Median Income** (different states). The only significant variable is **ATP** but the other variables are not significant at all with very high P-Values.

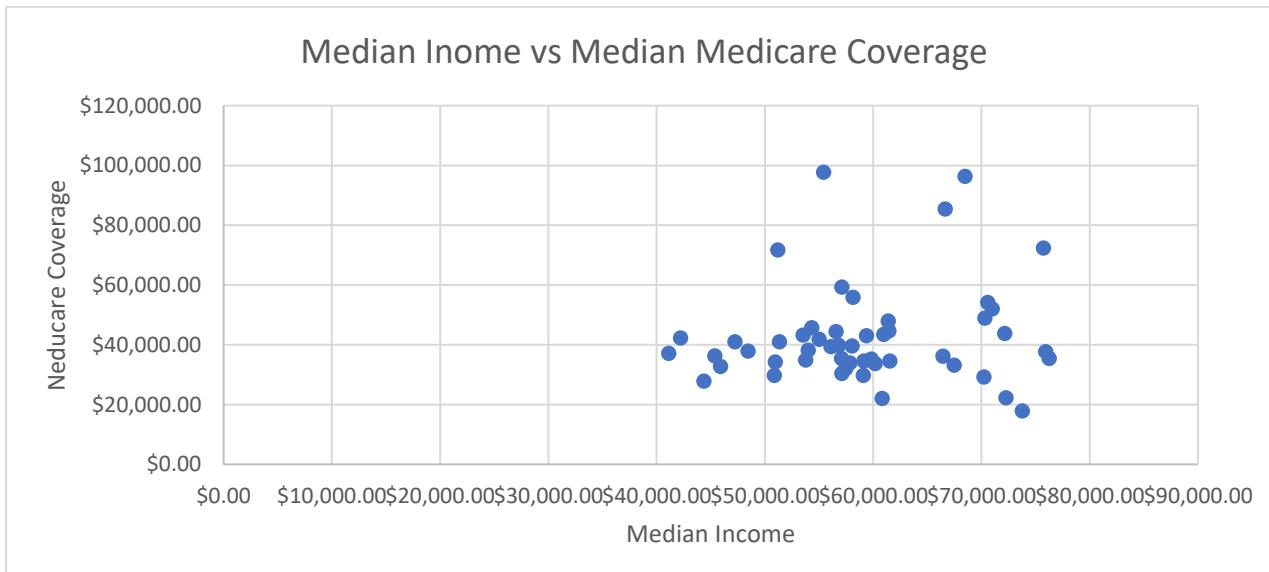
The final task now is to determine if the coverages are correlated by state and significantly different from one another by state. Average coverage wasn't used due to outliers skewing the data so instead, median coverage was used for one specific procedure for each state

State	Abbreviation	Median Income	Median Average Medicare Coverage	Median Average Total Payments
Alabama	AL	\$47,221.00	\$41,002.36	\$10,022.77
Alaska	AK	\$75,723.00	\$72,283.48	\$22,071.46
Arizona	AZ	\$57,100.00	\$59,266.25	\$12,623.81
Arkansas	AR	\$45,907.00	\$32,673.55	\$10,351.92
California	CA	\$66,637.00	\$85,422.81	\$16,160.29
Colorado	CO	\$70,566.00	\$54,183.18	\$12,370.57
Connecticut	CT	\$75,923.00	\$37,641.25	\$14,496.21
Delaware	DE	\$58,046.00	\$39,584.37	\$12,963.90
District of Columbia	DC	\$70,982.00	\$51,957.56	\$16,759.06
Florida	FL	\$51,176.00	\$71,730.06	\$10,709.24
Georgia	GA	\$53,527.00	\$43,258.58	\$10,940.64
Hawaii	HI	\$72,133.00	\$43,745.64	\$18,555.53
Idaho	ID	\$56,564.00	\$44,393.45	\$12,424.14

Correlation analysis was then done again on the new data set to determine if any meaningful correlation exists:

	Medicare coverage	Total Payments	Medicare Payments	Median Income
Medicare Coverage	1			
Total Payments	0.226179363	1		
Medicare Payments	0.210661902	0.98158474	1	
Median Income	0.124793055	0.592964007	0.543075375	1

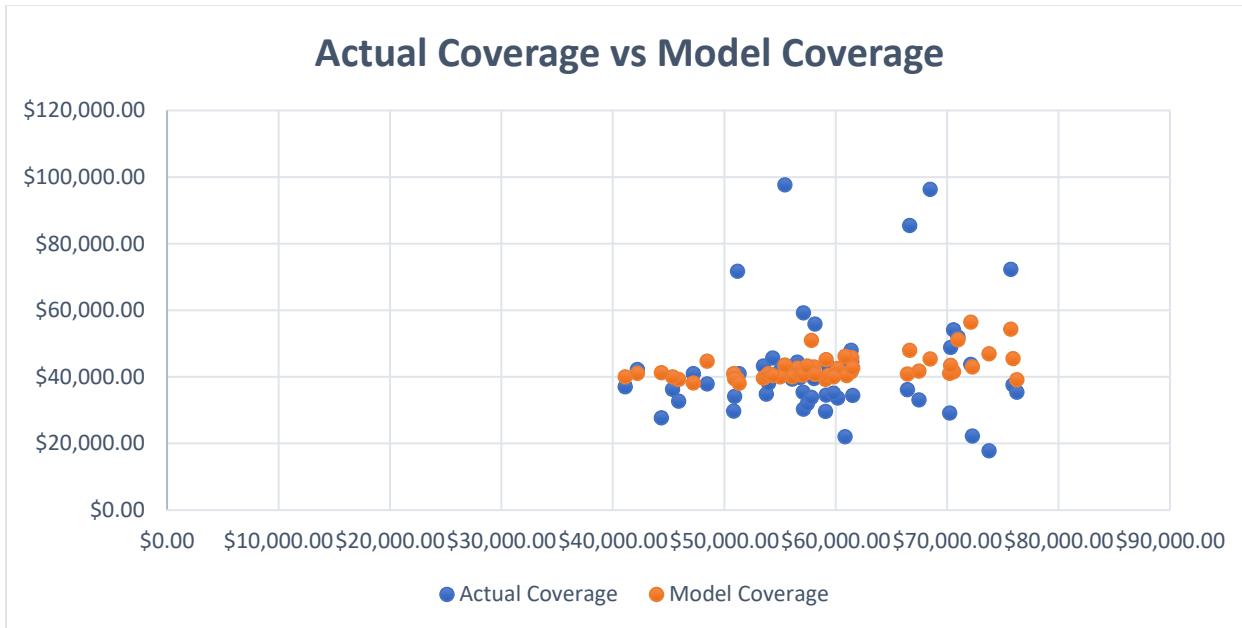
With our revised data set, it is at a 12% correlation, which is rather insignificant. The other variables are also a lot lower than before as well. Lastly, a regression analysis was then done to determine if the relationship between Medicare Coverage and the other variables are significant. The results along with the scatterplot are shown below:



Regression Statistics	
Multiple R	0.235471363
R Square	0.055446763
Adjusted R Square	-0.004843869
Standard Error	16575.38737
Observations	51

	Coefficients	Standard Error	t Stat	P-value	Lower 95%
Intercept	23877.07382	15947.32827	1.497246023	0.141015404	-8204.812536
Total Payments	3.953746925	5.364826218	0.736975709	0.464800245	-6.838891326
Medicare Payments	-2.455903834	5.400807321	-0.454729022	0.651397632	-13.32092673
Median Income	-0.064486622	0.332045449	-0.194210227	0.846848275	-0.732475905

Looking at our results for the revised data set, we see that the R^2 is very small and the P-Value for Median income is rather large so it isn't significant at all. The other variables, **AMP** and **ATP** are also very high so with this particular data set, they are also not significant. This data supports our analysis on the overall data set since the P-Values for **Median Income** is very large and the correlation coefficient is very tiny. With the support of the R^2 . This shows that for this particular model, the difference coverage between states isn't statistically significant in the 95% level. We can compare our model to the actual Medicare coverages. See below to how the Medicare coverages stack with our model:



We unfortunately aren't able to predict the coverages that well based the variables presented. This along with the very high p-value, low correlation and low R^2 shows that Medicare inequality does not exists between the states based on the data we are given.

Conclusion

The information unfortunately shows that coverage by state isn't too significantly different at the 95% level using median income to differentiate between the states. With the multiple low R^2 's, they aren't a good fit for our mathematical model. There are some limitations however due to the data

set not having enough information. When the data was split, some states were more represented than others. Because of this, some states were unfortunately undervalued. The results would make sense in the end due to how the healthcare system functions in the United States. Each hospital works differently in terms of coverages and because of that, coverages vary greatly between one another rather than each individual state. However, inequality still might exist per patient rather than state. Some of the correlations for the other variables are high along with low p-values. This is not the final conclusion on this topic and more studies still needs to be done.