Medicare Cost Saving Strategy

Suffyan Asad

Jessica Smith

Sonya Tahir

A thesis submitted to

The Faculty of

The School of Business

of The George Washington University

in partial fulfillment of the requirements

for the degree of Master of Science

August 31, 2016

Thesis directed by

Dr. Shivraj Kanungo

Associate Professor of Decision Sciences

Dr. Mehmet S. Altug

Assistant Professor of Decision Sciences

Ryan Hill

Senior Consultant Deloitte

Abstract

As computing and data management systems become increasingly cost-accessible and easy to maintain, many public institutions are making greater efforts to release data to the public. In the realm of health care, this means greater access to information pertaining to state and federal costs, quality of care, and patient characteristics. In collaboration with Deloitte and the faculty at The George Washington University, we sought to leverage these data to uncover cost-savings opportunities that could be applied to the United States government's health care institutions.

A comprehensive literature review revealed that Medicare represents a significant portion of government health care expenditures, accounting for 17.5% of US Gross Domestic Product (GDP) in 2014. Furthermore, with the baby boomer generation entering retirement, the Medicare cost growth is expected to accelerate in coming years, such that Medicare cost growth is projected to outpace GDP growth by approximately 1% each year. The timeliness and scope of Medicare makes this program a salient topic for investigation.

Existing research also revealed the persistence of extreme geographic variation in Medicare spending across the country. Researchers have yet to uncover the driving forces behind such variation, but believe a combination of patient population health, poverty, and medical practices are likely culprits. We were motivated to assist researchers in answering the question of Medicare cost variation by examining the impact of chronic conditions.

The Centers for Medicare & Medicaid Services (CMS) maintain a rich set of data resources freely available to the public, including several sources associated with Medicare. We applied several methodologies from the field of data science to these data, including regression, principal component analysis, and clustering. Through these explorations, we identified several key findings:

- At the state level, geographic cost variation is less extreme. Research conducted at the county level reveals greater cost variation that is otherwise lost in the state averages, and therefore county-level analyses are more likely to reveal potential cost-savings insights.
- Patient health is indeed a key driver of county-level Medicare cost. However, counties with similar patient populations may still demonstrate cost differences.
- Rare conditions are often costlier and more sensitive to slight changes in prevalence rates. Common conditions are less expensive and less sensitive. As a result, we cannot isolate one single condition as being the driver of overall cost.

Further study into factors such as medical infrastructure, administrative expenses, and medical practices may reveal the driving forces other than patient characteristics that drive cost differentiation. We hope that by providing the tools to identify counties with similar sickness profiles yet differing costs we have enabled the medical community to explore Medicare cost variation while keeping the factor of patient sickness profile constant.

Table of Contents

List of Figures

List of Tables

Introduction

The research and analysis presented in this report represent a collaboration between ourselves, several members of the George Washington University Department of Decision Sciences faculty, and advisors from the Deloitte Corporation. Our team was tasked with applying modern statistical and data science methodologies to a challenge affecting the United States health care system. In recent years, both federal and local government agencies have made greater efforts to share data with the public. As a result, a wealth of data exists for researchers who have the skills to effectively gather, synthesize, and utilize it. Our objective was to make use of these publicly available data in order to define a cost savings strategy for a subset of US government medical care.

Background Research

We began by conducting research into the current state of US government medical expenditures as well as future projections. The Centers for Medicare and Medicaid Services' National Health Expenditures Fact Sheet states that national health expenditures accounted for nearly 18% of the US Gross Domestic Product (GDP) and totaled $3.0 trillion in 2014 (NHE Fact Sheet, 2016). Medicare spending accounted for 20% of that figure, while Medicaid accounted for 16%. The Centers for Medicare and Medicaid Services (CMS) project that national health expenditures will grow at an average rate of 5.8% per year between 2014 and 2024, and will outpace GDP growth by approximately 1% (NHE Fact Sheet, 2016).

We investigated Medicare in greater depth since it accounted for such a significant portion of government medical expenditures. The Kaiser Family Foundation has conducted extensive research on topics relating to US health policy, and provided a wealth of data regarding Medicare. The following chart, taken from their 2015 Facts on Medicare Spending and Financing report, highlights the significant role of Medicare spending in the US Federal Budget.

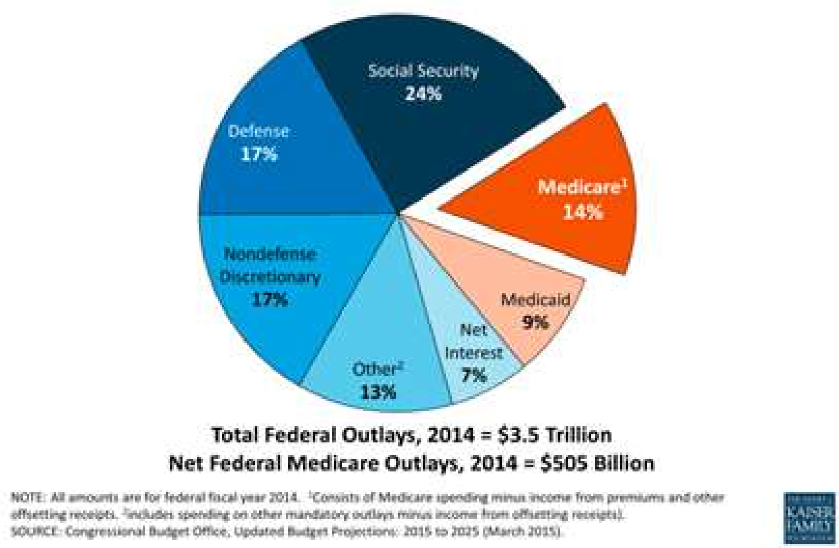

Figure 1: Medicare as a share of the Federal Budget

The Kaiser Family Foundation also provided data regarding future projected Medicare expenditures. The following graph shows Medicare spending following a steady growth pattern over the next few years, with growth accelerating after 2018. The report attributes the acceleration to the growing number of aging baby boomer generation individuals becoming eligible for Medicare (The Henry J. Kaiser Family Foundation, 2015).

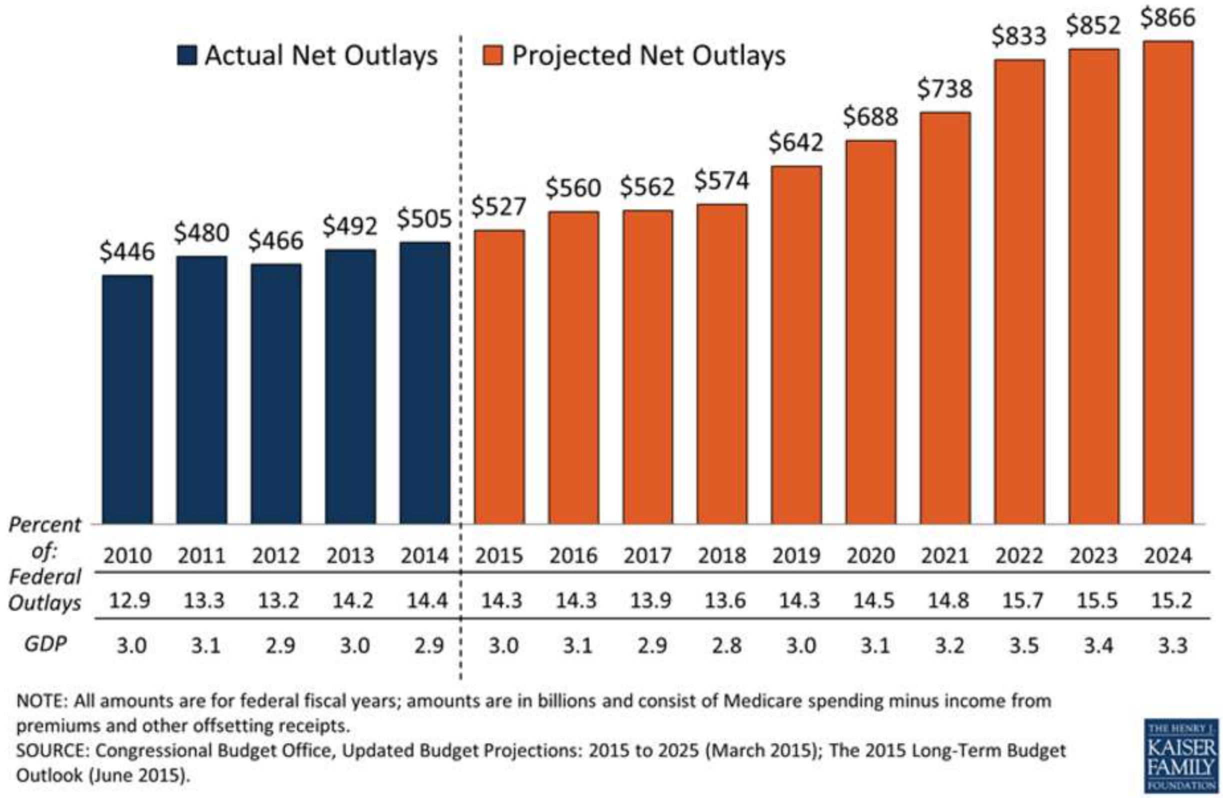

Figure 2: Actual and Projected Net Medicare Spending, 2010-2024

The impact of Medicare on the federal budget, as well as the relevance of this federal program in the lives of many Americans, made Medicare an appropriate topic for our research project. We set out to identify potential cost savings strategies to help alleviate the rising costs of Medicare in the future. Once again, we looked to the research already conducted by the Kaiser Family Foundation. In 2015, researchers at the foundation completed a report aimed at uncovering and explaining geographic variation in Medicare spending. The report explained that extreme variation in spending based on geography has long been observed in the Medicare program, but has never been adequately explained (Cubanski, Neuman, & White, 2015). Researchers have puzzled over these variations, wondering if the driving factors are related to patient characteristics, such as poverty or disease, or to medical practices, such as prices set by doctors or insurance policies. As of now, researchers have yet to draw a firm conclusion. However, the 2015 Kaiser Family Foundation report presented the following findings:

> *"Our analysis shows that geographic variation in Medicare per capita spending persists … deep differences in per capita Medicare spending in different parts of the country remain and are likely to persist due to underlying differences in beneficiary characteristics related to poverty and poor health, along with differences in the prices that Medicare pays for services, that contribute to variations in spending."*

The report found that beneficiary characteristics, particularly poverty and poor health, were important drivers of cost variation (Cubanski, Neuman, & White, 2015). However, the researchers stopped short of identifying a particular condition or illness that most contributed to rising costs.

Research Questions

We were motivated to build upon the Kaiser Family Foundation's research by attempting to identify specific conditions that lead to increases in Medicare spending. We developed the following guiding research questions:

1. Which conditions drive increases in Medicare spending?
2. Can we identify states or counties that treat certain conditions more efficiently than others?

Public policy aimed at curtailing the growth of Medicare spending cannot be developed unless the driving forces of geographic cost variation are understood. By answering these questions, we hoped to provide additional insight to the medical community regarding the drivers of Medicare cost in order to facilitate the creation of such policy and reduce the burden of Medicare on the federal government in the years to come.

Data Background

After extensive research into the heath care data in the public domain, we decided to utilize the data that have been made available by the Centers for Medicare and Medicaid Services (CMS). CMS has been collecting data relating to Medicare beneficiaries for many years, and has readily-accessible data files available for the years 2007-2014. The CMS files contain annual county-level statistics such as the number of Medicare beneficiaries, total Medicare spending, and the prevalence rates of several common chronic conditions. These data were gathered from the following CMS sources:

1. Public Use Geographic Variation Data Files
2. Beneficiary Enrollment and Characteristics Data Files
3. Chronic Conditions Data Files

There are several important considerations to keep in mind regarding the data. First, all three of the CMS data sources mentioned above pertain solely to individuals enrolled in the Medicare fee-for-service program (also known as Original Medicare). These are enrollees who receive both Part A (hospital insurance) and Part B (medical insurance) Medicare. Individuals who were enrolled in the Medicare Advantage program, or who were enrolled in only Part A or Part B were not included in the data files, and therefore are not a part of this analysis.

Second, the files present both an actual and a standardized measure of total Medicare spending. Actual spending represents the total payment amount for Medicare services covered in parts A and B. Standardized payments have been adjusted to help mitigate the effect of geographic differences in payment rates for services, such as local wages (Chronic Conditions among Medicare Beneficiaries: A Methodological Overview, 2016). As a result, the differences in the standardized payment amount is more reflective of variations such as local medical practices and the medical needs of the patient population. Since these were the effects we wished to study, we chose to focus exclusively on the standardized spending levels.

Finally, CMS suppressed certain data in order to protect the privacy of individual beneficiaries. If 11 or fewer beneficiaries were present, then the data were suppressed. For example, data regarding total Medicare cost was suppressed in counties where fewer than 11 Medicare fee-for-service beneficiaries required treatment in that year. In counties with more than 11 beneficiaries but fewer than 11 with a particular condition, the prevalence rate for that particular condition was suppressed. The data suppressions were indicated with an asterisk in the raw files. Any counties in which the total Medicare spending had been suppressed were excluded from our analysis. In situations where county-level data were available but the prevalence rate for a particular condition was suppressed, the data were included but the prevalence rate for that condition was set to zero. If fewer than 11 beneficiaries have a particular condition, we can be confident that the prevalence rate would have been close to zero, if not exactly zero. This assumption allowed us to include much more of the available data than we would otherwise have been able to include.

To conclude, throughout this report we will make use of the term "Cost per FFS Beneficiary." As stated above, this refers to the standardized Medicare spending per beneficiary enrolled in the fee-for-service program.

Exploratory Analysis

To start any analytical process, it is important to first get a basic understanding of the data. This includes getting familiar with the variables involved by looking at their values and their variations across different dimensions. Looking at the numbers visually helps in understanding the overall picture. The following exploratory analysis uses visual charts and maps to look at cost per FFS beneficiary over time and across geographical boundaries of state and counties. It also explores the prevalence of all 19 chronic conditions to see how they vary over time and across geographic regions.

Average Cost per FFS Beneficiary by Year ($)

The bar chart shows the change in average cost per FFS beneficiary (y-axis) over the years (x-axis). An overall increasing trend is observed from 2007 to 2014 except for a decrease in cost between 2010 and 2011.

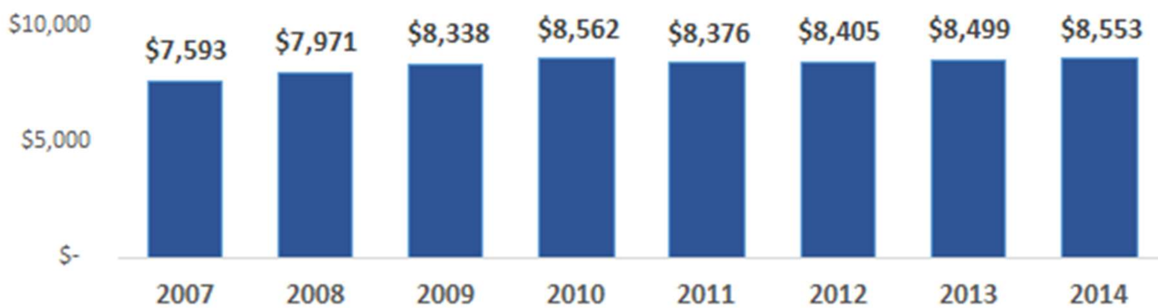

Figure 3: Average Cost per FFS Beneficiary (2007 - 2014)

Average Cost per FFS Beneficiary by County ($)

The following map demonstrates how the average cost per beneficiary is distributed geographically. There is a great variation in spending across the counties with some counties spending significantly more than others. Most counties in Texas and Florida seem to have a high cost. Most of the states on the west side have a low cost. This chart pertains to 2014 data only. However, the distribution is very similar across all years.

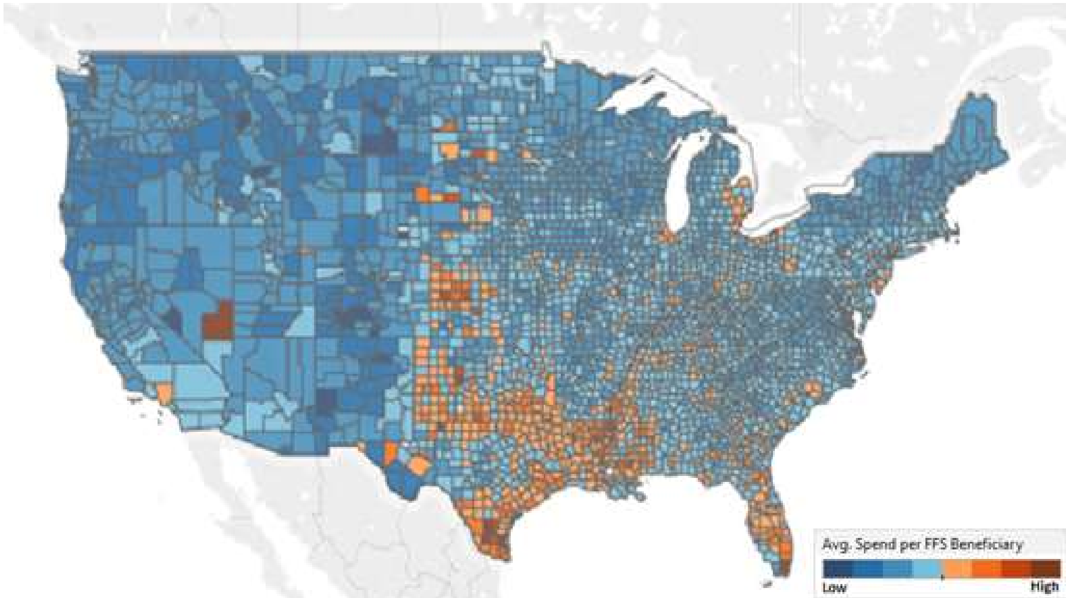

Figure 4: Average Cost per FFS Beneficiary by County

Average Condition Prevalence Rates by Year (%)

The line chart below shows years on the x-axis and the average condition prevalence rates across all counties on the y-axis. It allows for a comparison of prevalence rates of the conditions. It also assists in evaluating trends over the years. The chart shows that Hypertension is by far the most prevalent condition followed by Hyperlipidemia. HIV/AIDS and Autism are the least prevalent. For most of the conditions, the prevalence rates remain steady over the years. There is an increase in the average prevalence rates of Arthritis, Depression and Chronic Kidney Disease whereas the average prevalence rates of Ischemic Heart and Heart Failure decrease over the observed period.

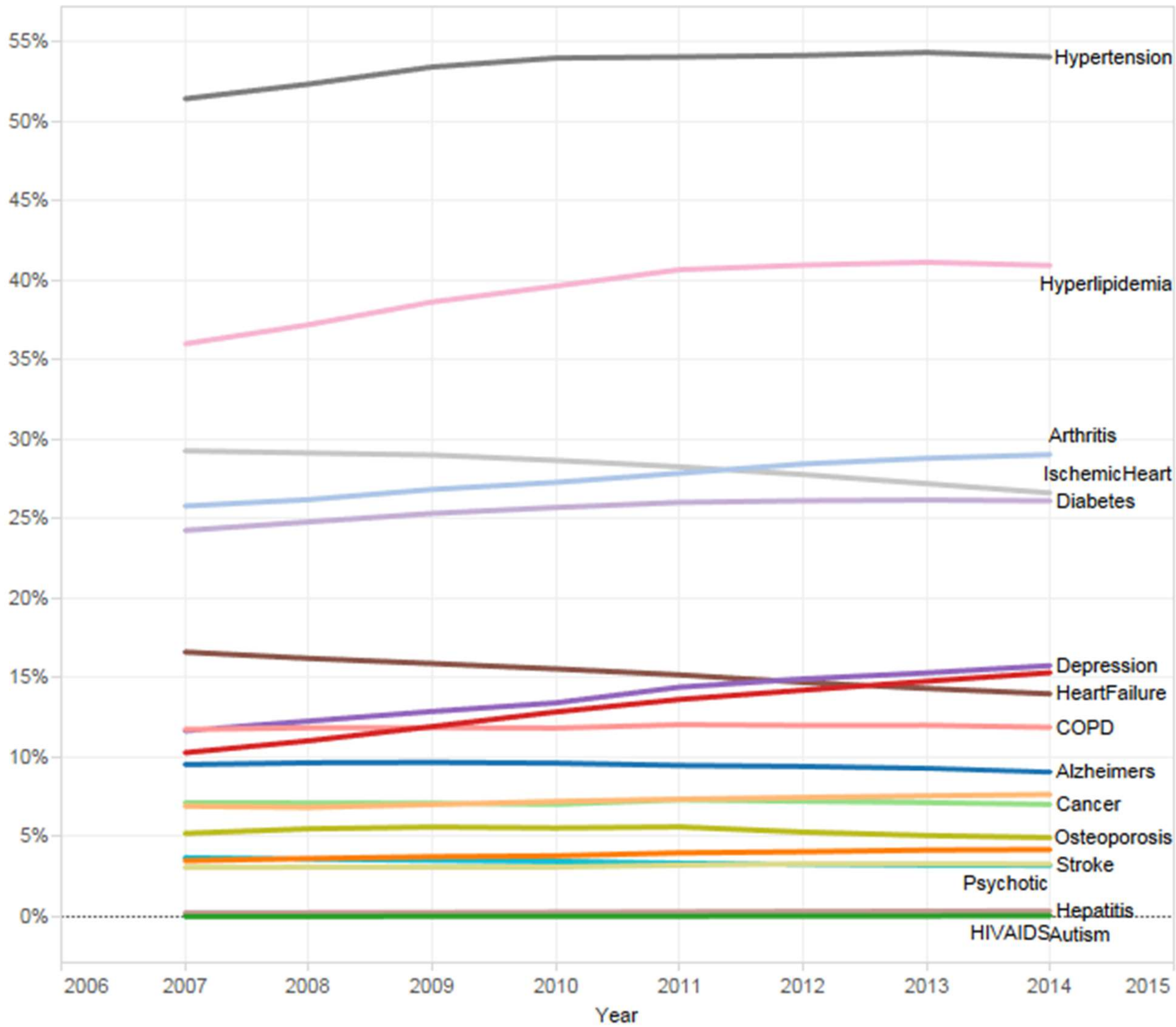

Figure 5: Average Condition Prevalence Rates by Year (%)

Spread of Average Condition Prevalence Rates (%)

The following chart shows box plots for the prevalence rates (y-axis) of the 19 chronic conditions (x-axis). Each dot represents the average prevalence rate in a county across all years. The variation in prevalence rates across counties can be observed for each condition. The plot shows that Hyperlipidemia and Hypertension have the widest spread whereas HIV/AIDS and Autism have the narrowest spread.

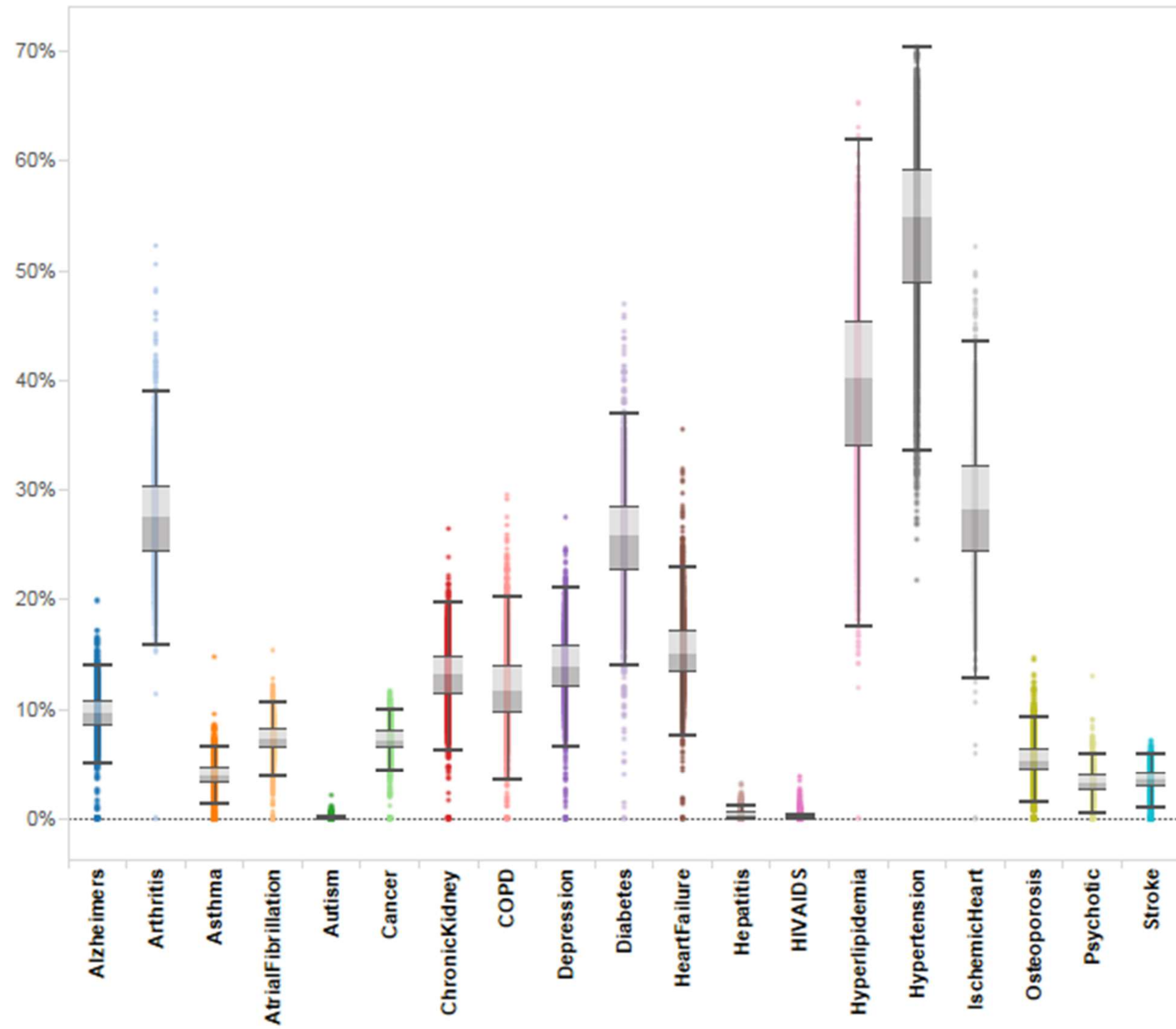

Figure 6: Average Condition Prevalence Rates by Year (%)

In evaluating figures 5 and 6, we notice Hypertension and Hyperlipidemia stand out among the other diseases in terms of high average prevalence and a wide variation amongst counties. On the other end of the spectrum are HIV/AIDS and Autism that have very low prevalence rates in almost all counties and therefore have the lowest average as well.

Condition Prevalence by County (%)

Plotting the condition prevalence rates on a map shows the geographical distribution for each condition. Comparing the maps of all the conditions shows that the geographical distribution varies greatly for the conditions. Below are maps demonstrating the prevalence of three conditions in 2014: Autism with very low prevalence across the country, Alzheimer's with medium to high prevalence in a few states, and Hypertension with high prevalence in a large number of states.

Autism

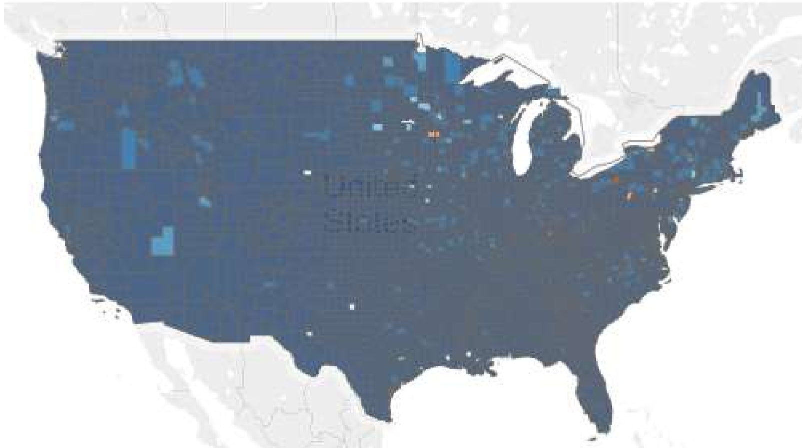

Figure 7: Prevalence of Autism by County in 2014

Alzheimer's

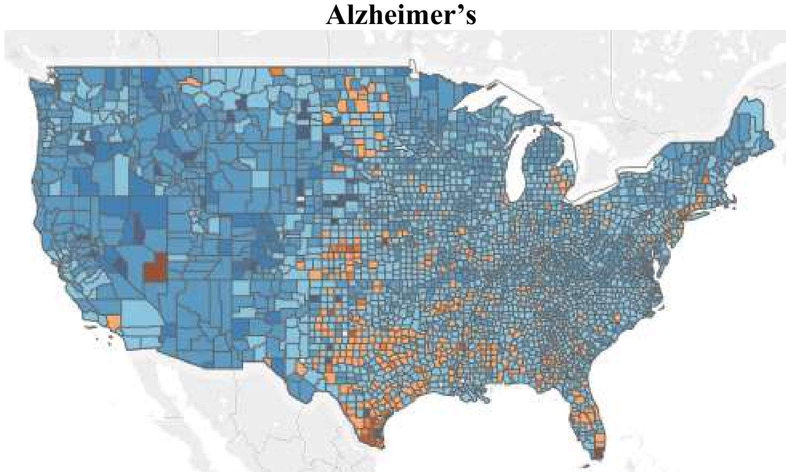

Figure 8: Prevalence of Alzheimer's by County in 2014

Hypertension

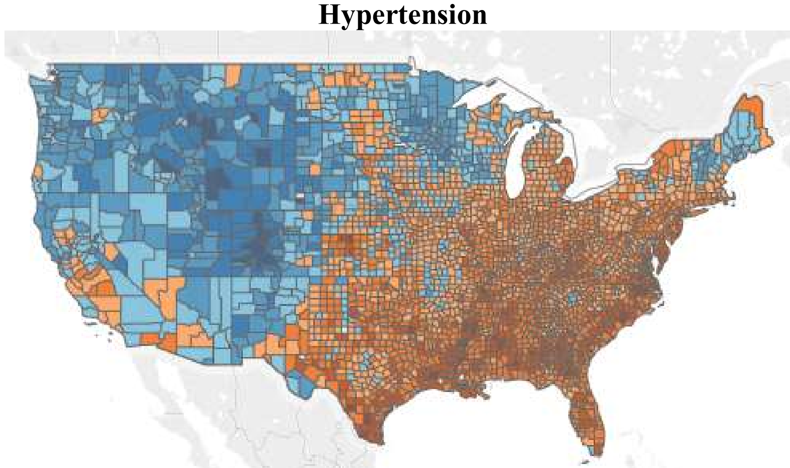

Figure 9: Prevalence of Hypertension by County in 2014

Average Cost per FFS Beneficiary by State ($)

The comparison of average cost per FFS beneficiary by state shows that although there are differences across the states, the variation is not very large. Louisiana, Mississippi, Texas and Florida have the highest average cost per FFS beneficiary. Oregon, Puerto Rico, Hawaii and the Virgin Islands have the lowest cost per FFS beneficiary.

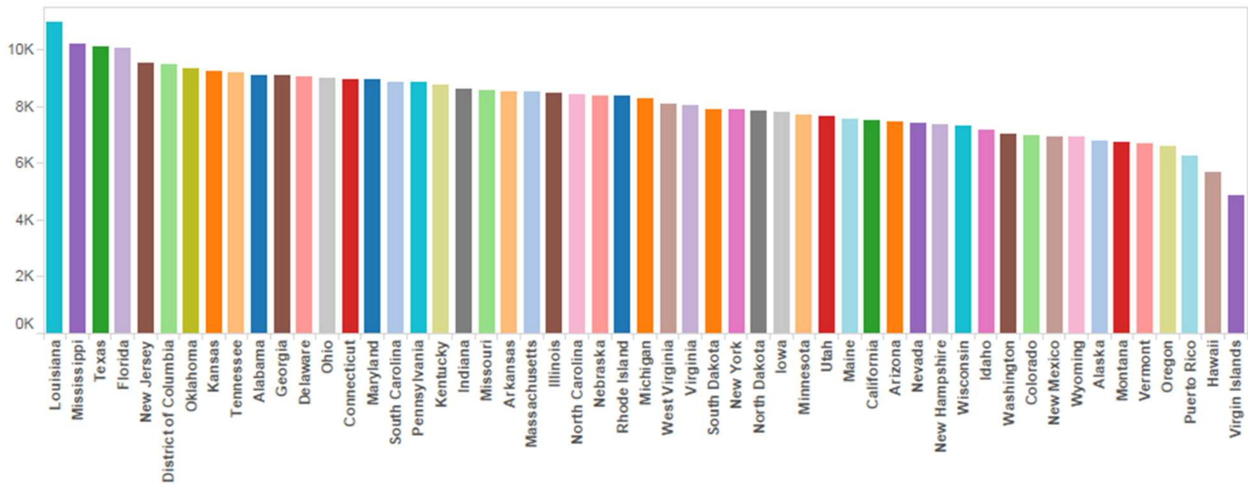

Figure 10: Average Cost per FFS Beneficiary by State

Average County-Level Cost per FFS Beneficiary by State ($)

Looking at the cost figures at county level shows a more interesting picture. There is variation in county-level spending within each state. Amongst the states with high cost, Texas shows the largest range in county-level spending. The states with the lowest average cost demonstrate a low variation in the costs at county-level.

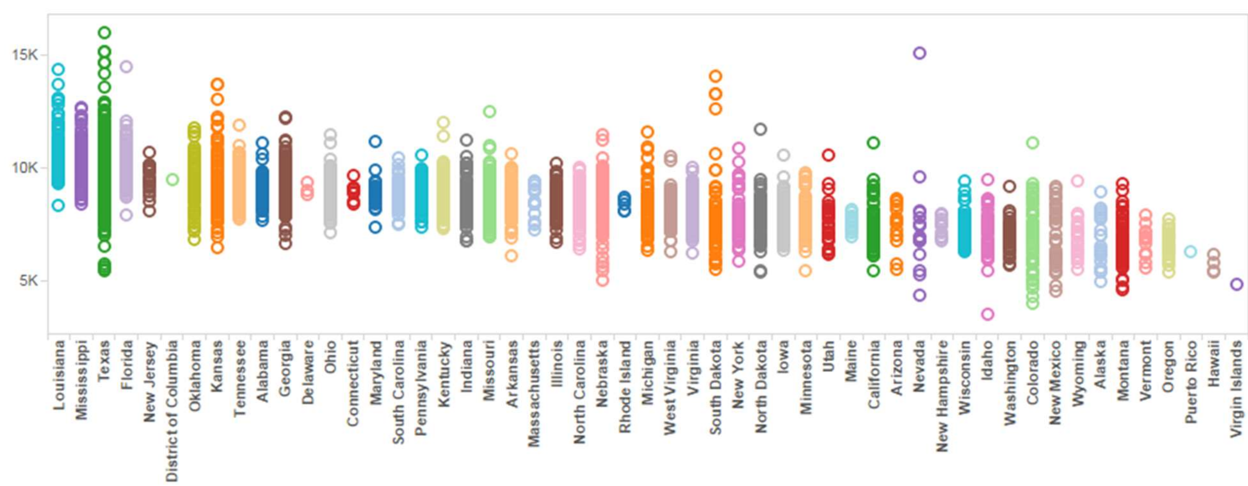

Figure 11: Average County-Level Cost per FFS Beneficiary by State

Regression Analysis

Methodology

A linear regression analysis was performed to model Standardized Cost Per FFS Beneficiary for each combination of Chronic Condition, State, Year, and Prevalence. The objective of regression analysis was to identify states, conditions and years that contributed the most to Standardized Cost Per FFS Beneficiary. Using the regression coefficients, the objective was to examine the following:

- Are there any states that stand out as being extremely expensive? Are there states that are performing exceptionally well or poorly in terms of cost per FFS beneficiary compared to other states, and require immediate attention?
- Are there any conditions that drive up the cost?
- How does cost change in different states as prevalence rates of different conditions change?

The purpose of using regression analysis was to observe the coefficients for each variable to see its contribution towards the cost. For categorical variables, one variable is taken as the base case and the coefficients of other variables are taken as deviations from the base. This makes it very easy to compare the effect of different values of a variable. We wanted to use this property of the regression method to compare different variables.

We needed to be able to compare the different states by looking at the categorical variable State. By sorting the regression coefficients of State in descending order, we would be able to identify the states contributing the most to cost. Year was another variable that we wanted to analyze. This was also treated as a categorical variable to get comparable coefficients. The last comparison we wanted to make was for the 19 chronic conditions. The original data contained separate columns for the prevalence rates of each condition. The names of the conditions appeared as column names rather than values in the data. Having each condition as a unique column would have made it impossible to compare the coefficients such that one would be the base and others would show deviations from it. In order to compare regression coefficients directly for the conditions, we created a categorical variable Condition so that the values were the names of the 19 chronic conditions. An additional numeric variable, Prevalence, contained the prevalence rates for each particular condition. This increased the number of rows in the dataset by a factor of 19. In this way, Condition was a more useful variable and the prevalence rates were given in a single column rather than 19. This enabled us to explore the effect of prevalence rates on cost as well.

The dataset, in its final form had the following variables:

- Year
- State
- County (not used in regression)
- Condition
- Prevalence
- Standardized Cost per FFS Beneficiary (Target)

Interaction terms were added to obtain separate coefficients to quantify the effect of changes in prevalence rates in each state and for each condition. Following is the set of variables that was used to train the model:

- Year
- State
- Condition

- State * Prevalence
- Condition * Prevalence
- Standardized Cost per FFS Beneficiary (Target)

The interaction of Year and Prevalence was also explored but was not added to the final model since the coefficients were very small compared to the other coefficients in the model. The model also showed that the variation by year was small, similar to our findings in the exploratory analysis. Therefore, the impact of prevalence and year together on cost was negligible, and was ignored to keep the model simpler and more interpretable.

The intercept was removed to make sure that the effect for each categorical variable value became visible as a coefficient. The resulting model is similar to a regression model with intercept, but the interpretation is much simplified. Following is the equation of the fitted regression model:

Cost per FFS Beneficiary ~ Year + State + Condition + (State * Prevalence) + (Condition * Prevalence)

Model Diagnostics and Coefficients

After fitting the regression model, several diagnostic measures were analyzed to ensure the appropriateness of the model. To see diagnostic measures such as R-squared, adjusted R-squared, residual squared error, Cook's Distance, and a heteroscedasticity check, please refer to Appendix 3.

After checking model diagnostics, the following coefficients were examined. The coefficients are presented graphically below:

Year

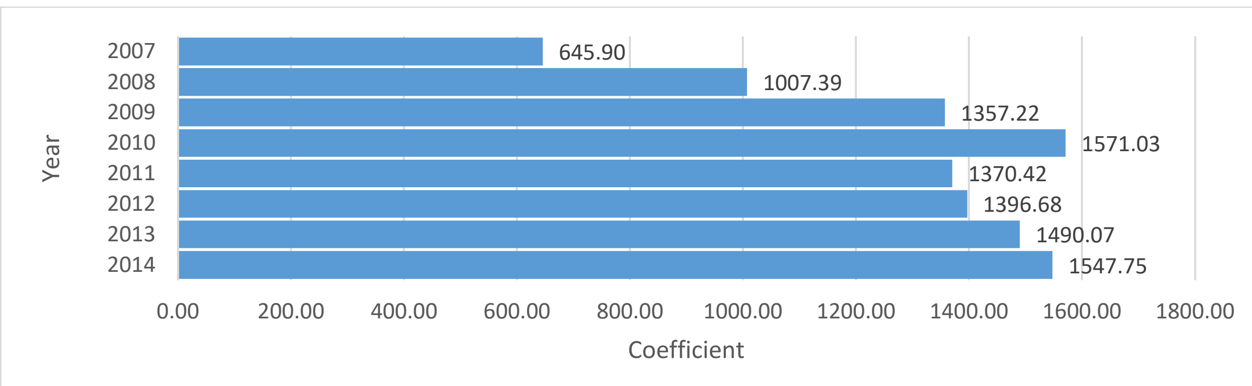

Figure 12: Regression Coefficients of the Year Variable

Coefficients of the Year variable followed the trend observed in the exploratory analysis: there is a rising trend except for a drop after 2010. Additionally, Year coefficients were small compared to State and Condition (and interaction) coefficients. Therefore, Year was not deemed to have a significant impact in comparison with States and Conditions.

Conditions

In order to determine the costliest conditions, the coefficients of the Condition variable and of the interaction term Condition * Prevalence were examined.

Condition coefficients describe the impact of a particular condition on Cost per FFS Beneficiary, whereas the interaction term demonstrates how Cost per FFS Beneficiary changes with change in the prevalence rate of each condition. Conditions with higher coefficients are more sensitive to change in prevalence rates. Therefore, a minor increase in their prevalence rates could result in a significant cost increase.

Below are the coefficients of the Condition variable and the Condition * Prevalence interaction terms:

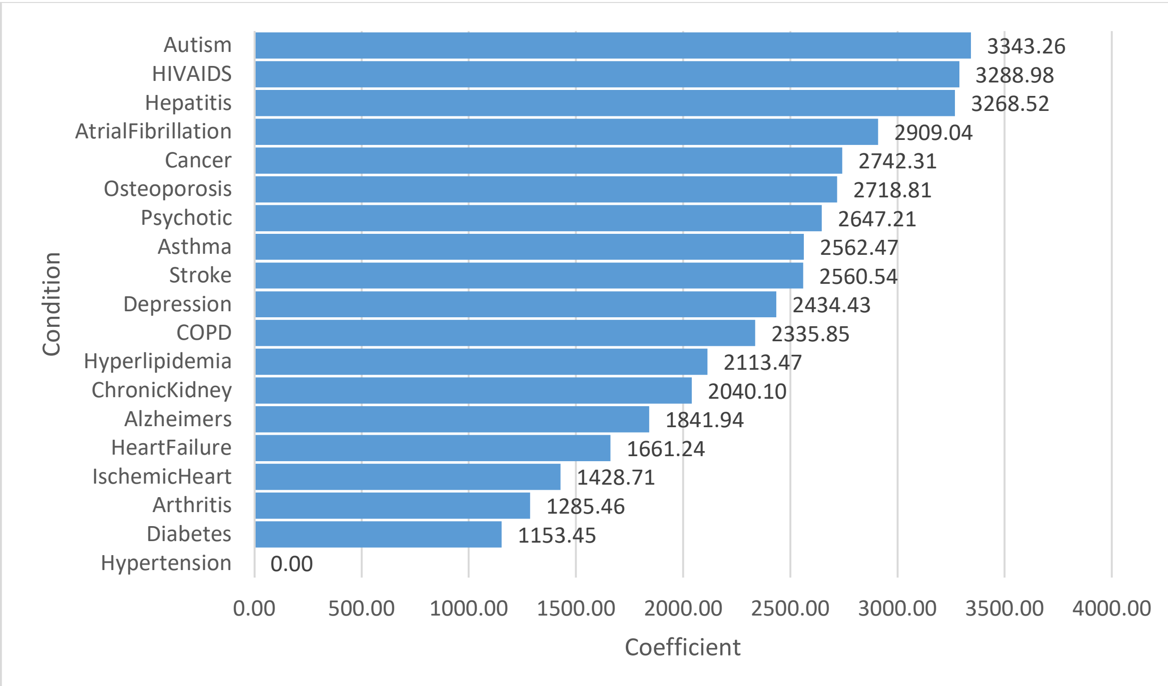

Figure 13: Regression Coefficients of the Condition Variable

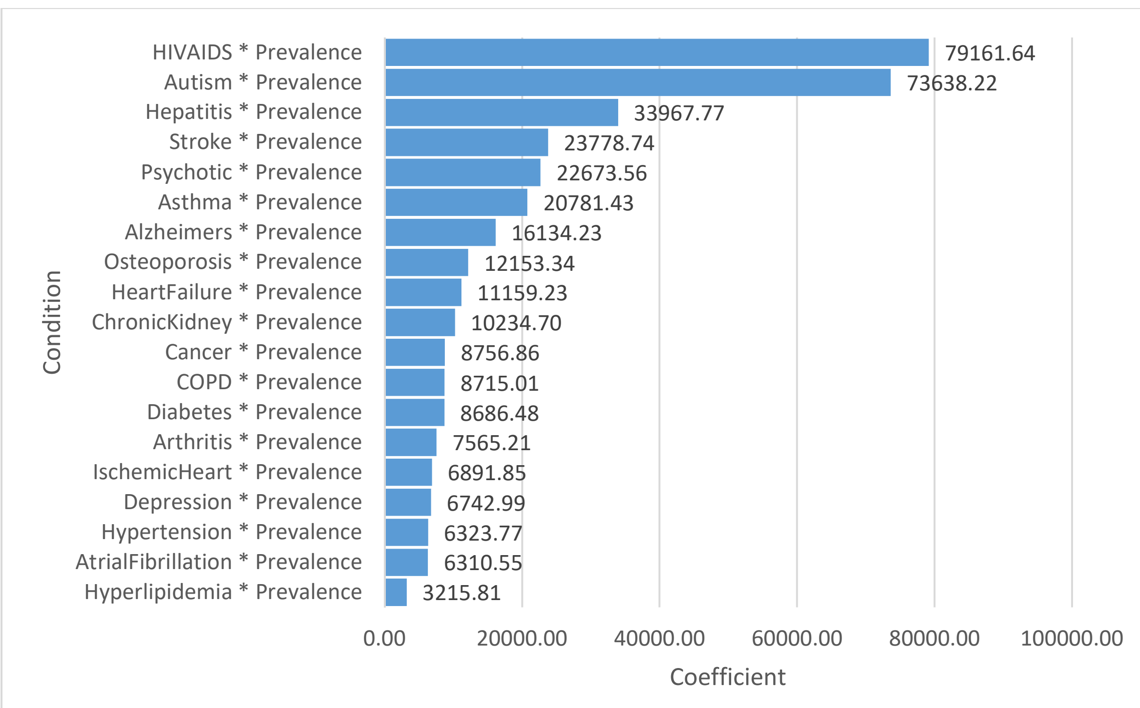

*Figure 14: Regression Coefficients of Condition * Prevalence Interaction Terms*

It is important to mention that prevalence rates take a value between 0 and 1, and therefore, the above coefficients represent an increase in prevalence rate from 0% to 100%. To see increase in Cost per FFS Beneficiary per 1% increase in prevalence rate, the coefficients have to be divided by 100.

It can be observed from the coefficients of Conditions that HIV/AIDS and Autism are the two most expensive conditions. They are also the most sensitive to changes in prevalence rates. Prevalence of HIV/AIDS in San Francisco in 2014 was 3.19%. If it increases by 0.1%, it would add $79.16 to the Cost per FFS, which, given the number of FFS Beneficiaries in San Francisco in 2014 (69,218), would add $5.48 million to the total cost in that county.

Conditions that have a high cost coefficient but a low interaction term coefficient indicate low sensitivity to change in prevalence rate. As a summary, the following table presents the coefficient of each chronic condition interaction term, its average prevalence rate across all counties in 2014, and the increase in Cost per FFS Beneficiary associated with a one percent increase in prevalence rate:

Condition	Interaction Coefficient ($)	Rise in cost per 1% increase in Prevalence Rate ($)	Average Prevalence Rate in 2014 (%)
Alzheimer's	16134.23	161.34	9.03
Arthritis	7565.21	75.65	28.96
Asthma	20781.43	207.81	4.19
Atrial Fibrillation	6310.55	63.11	7.63
Autism	73638.22	736.38	0.06
Cancer	8756.86	87.57	7.01
Chronic Kidney	10234.70	102.35	15.29
COPD	8715.01	87.15	11.86
Depression	6742.99	67.43	15.73
Diabetes	8686.48	86.86	26.07
Heart Failure	11159.23	111.59	0.10
Hepatitis	33967.77	339.68	13.96
HIVAIDS	79161.64	791.62	0.36
Hyperlipidemia	3215.81	32.16	40.83
Hypertension	6323.77	63.24	53.92
Ischemic Heart	6891.85	68.92	26.56
Osteoporosis	12153.34	121.53	4.94
Psychotic	22673.56	226.74	3.29
Stroke	23778.74	237.79	3.20

Table 1: Sensitivity of Cost to Changes in Prevalence Rates

State

Observing the coefficients of the State variable, the most expensive states are Louisiana, Mississippi, Texas, and Florida, all with a cost of $4,800 or more per beneficiary. This, multiplied by the number of beneficiaries in these states, accounts for billions of dollars to Medicare costs.

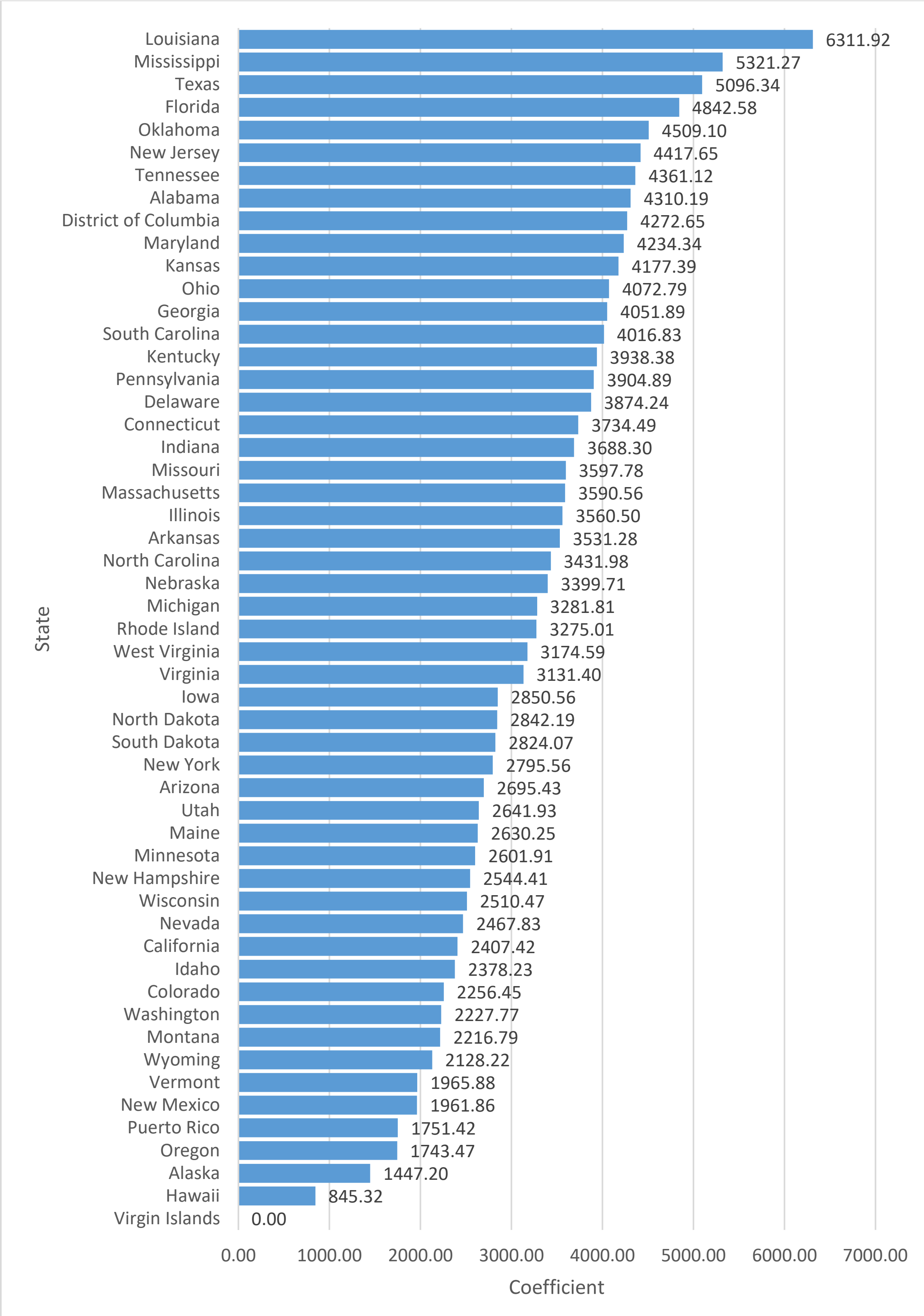

Figure 15: Regression Coefficients of the State Variable

The State * Prevalence interaction term describes the effect of changes in prevalence rates on Medicare costs in each state. Below are the coefficients:

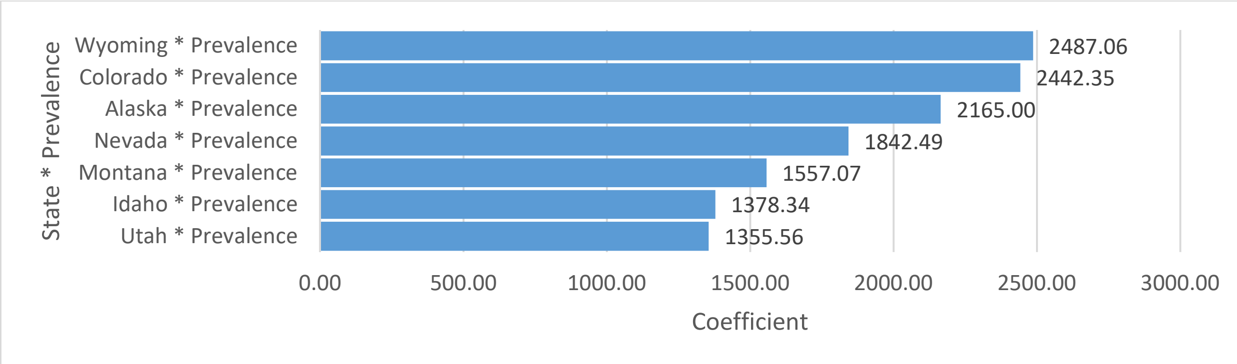

*Figure 16: Regression Coefficients of State * Prevalence Interaction Terms*

The only significant coefficients are those for Utah, Idaho, Montana, Nevada, Alaska, Colorado and Wyoming. The interaction coefficients of remaining states are not significant. This indicates that the cost in most of the states is not sensitive to changes in prevalence rates.

Additionally, the statistically significant interaction coefficients are very small. Wyoming has a coefficient of $2487.06, meaning the state will have to bear an additional cost of $24.87 per FFS beneficiary if the prevalence rate of a condition increases by 1%. This is small compared to other coefficients in the model.

Plotting the Results of Regression Analysis

The following figure shows the predicted cost per FFS beneficiary for each condition. The dots are colored based on the state. We can see that the variation in cost is similar across states as the state lines seem to be parallel. For conditions such as HIV/AIDS and Autism, the slopes are high but the prevalence rates are low. For other conditions like Hypertension and Hyperlipidemia, the prevalence rates have a wide range but the slopes are quite flat. Thus we conclude that the state level differences are insignificant. There is variation in the impact of conditions on cost but no single condition alone seems to be the major cost driver.

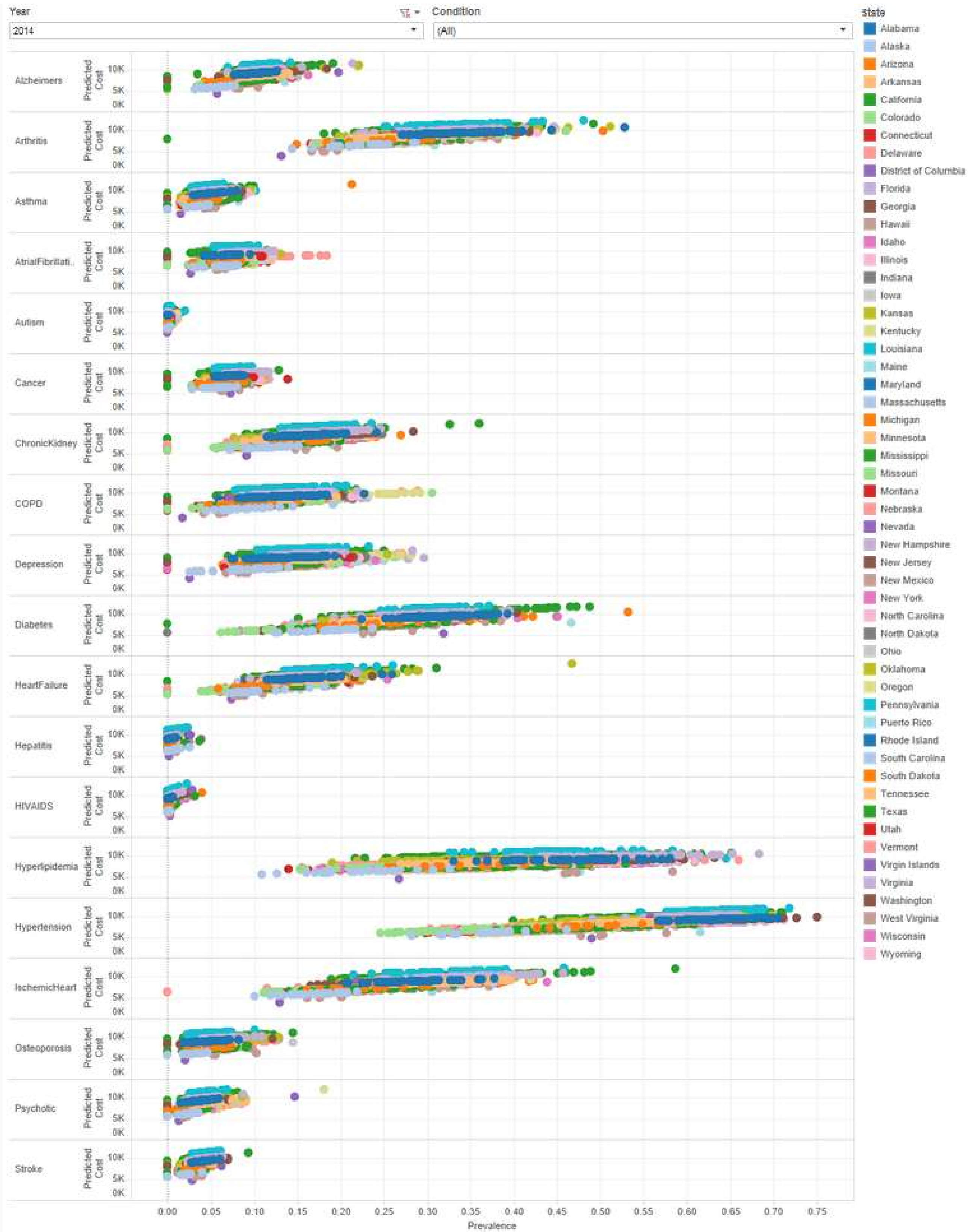

Figure 17: Predicted County-Level Cost per FFS Beneficiary by Chronic Condition

Below is a summary of the regression analysis results:

- The effect of the Year variable on cost is very small compared to the effects of the State and Condition variables.
- Prevalence of chronic conditions is a major driver of cost. However, no single condition can be isolated as the most influential driver of cost.
    - HIV/AIDS and Autism are very costly conditions, and very sensitive to change in prevalence. However, they have very low prevalence overall and are therefore not contributing too much to the current cost.
    - On the other hand, conditions like Hypertension and Hyperlipidemia have low sensitivity to changes in prevalence, meaning that a 1% increase in their prevalence rates does not add a great deal to the total cost. However, since they are so prevalent, their contribution to cost cannot be ignored.
- States in general have smaller coefficients and do not seem to drive the cost. They are also not sensitive to changes in prevalence rates.
    - Louisiana, Mississippi, Texas and Florida are high-cost states.
    - The Virgin Islands, Hawaii, Alaska, and Oregon are low-cost states/territories.

Principal Component Analysis

The regression analysis demonstrated that Medicare spending increased as prevalence increased for all 19 conditions. We wished to study the relationship between prevalence and spending in greater detail. Conducting the same analysis 19 times (once for each condition) would have been redundant. Instead, we performed a principal component analysis with the objective of developing a single sickness index that would serve as a proxy for the general level of sickness seen in a county.

To develop such an index, the annual county-level prevalence rates for each condition were standardized (scaled and centered). Then, a principal component analysis was conducted on the standardized data. The first principal component explained 34% of the overall variation in the data. The following table provides the loading values for the first principal component.

Alzheimers	Arthritis	Asthma	Atrial Fibrillation	Autism	COPD	Cancer	Chronic Kidney	Depression	Diabetes
0.254	0.256	0.253	0.127	0.000	0.245	0.136	0.265	0.228	0.288
HIV/AIDS	**Heart Failure**	**Hepatitis**	**Hyperlipidemia**	**Hypertension**	**Ischemic Heart**	**Osteoporosis**	**Psychotic**	**Stroke**	
0.118	0.206	0.143	0.295	0.327	0.264	0.205	0.205	0.278	

Table 2: First Principal Component Loadings of Chronic Conditions

The loadings revealed that the first principal component provided a general summary of the 19 conditions. All loadings except for Autism fell in the 0.11 – 0.33 range, indicating that no single condition was overwhelmingly driving the first principal component, and that most conditions contributed relatively equally to the component. Based on the results of the loadings as well as the percent of total variation explained, we chose to use the first principal component as a sickness index. Counties with higher sicknesses indexes had higher condition prevalence rates, while those with lower sickness indexes demonstrated lower prevalence rates. In other words, counties with low index scores were generally healthier, and counties with high index scores were generally sicker. This distinction allowed us to separate regions based on a general sickness indicator, rather than limiting ourselves to studying only one condition at a time.

Armed with the new sickness index, we moved on to exploring the relationship between sickness indexes and Medicare spending. Our background research suggested that sicker patient populations generally spent more on Medicare. Therefore, we hypothesized that counties with high sickness indexes would display higher Medicare spending. Heat maps were used to visualize the sickness indexes across the country as well as the Medicare spend across the county. The two heat maps below represent data from 2014.

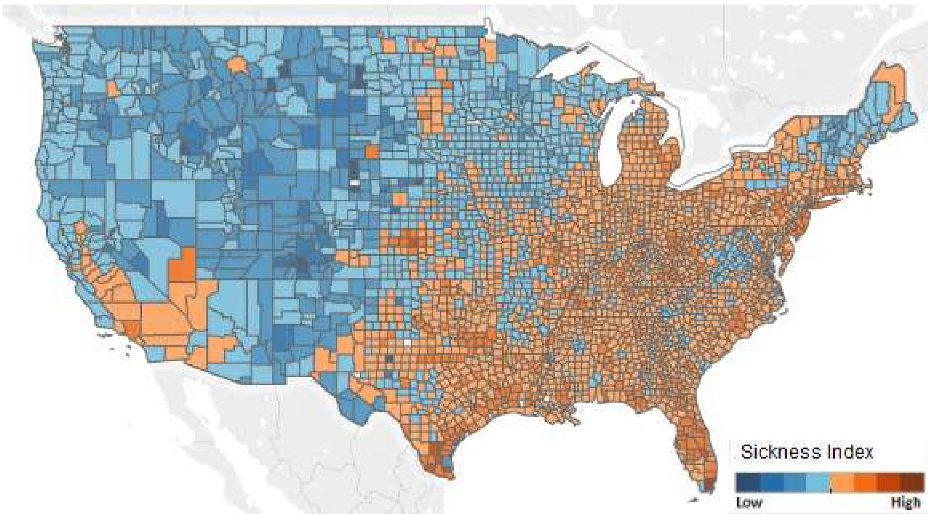

Figure 18: Sickness Index by County

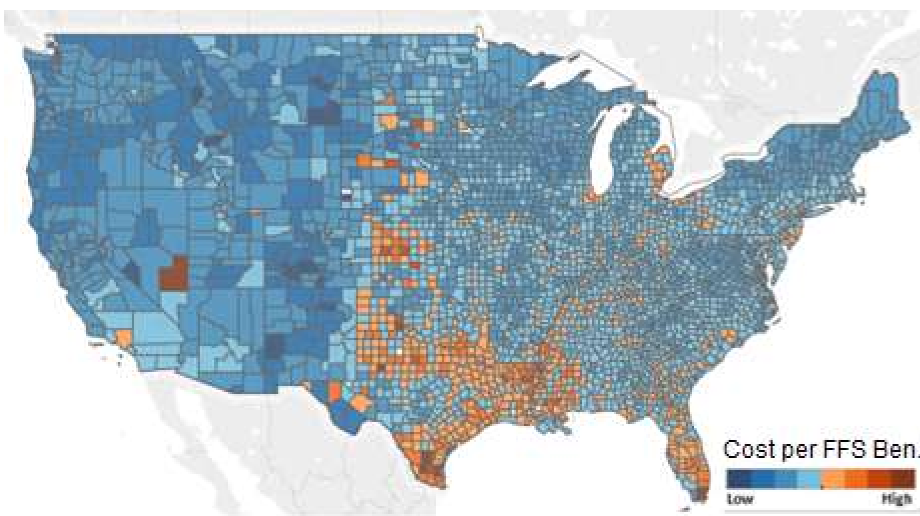

Figure 19: Cost per FFS Beneficiary by County

The heat maps tell an interesting story. Counties in the southern and eastern regions of the country have generally higher sickness indexes, while counties to the west have lower sickness indexes. In the south, many of the sicker counties also demonstrate high spending. However, in the eastern and northern regions, several of the sicker counties demonstrate low spending. While some of the highest cost counties have high sickness indexes, there are also several counties with high sickness indexes that demonstrate low costs.

Therefore, the presence of a high sickness index in a particular area does not necessarily indicate that the area will have a high level of Medicare spending.

To explore this variation in greater detail, we examined eight states whose counties demonstrated high sickness indexes. Four of these states had high levels of spending, and four had low levels of spending. The results are summarized below:

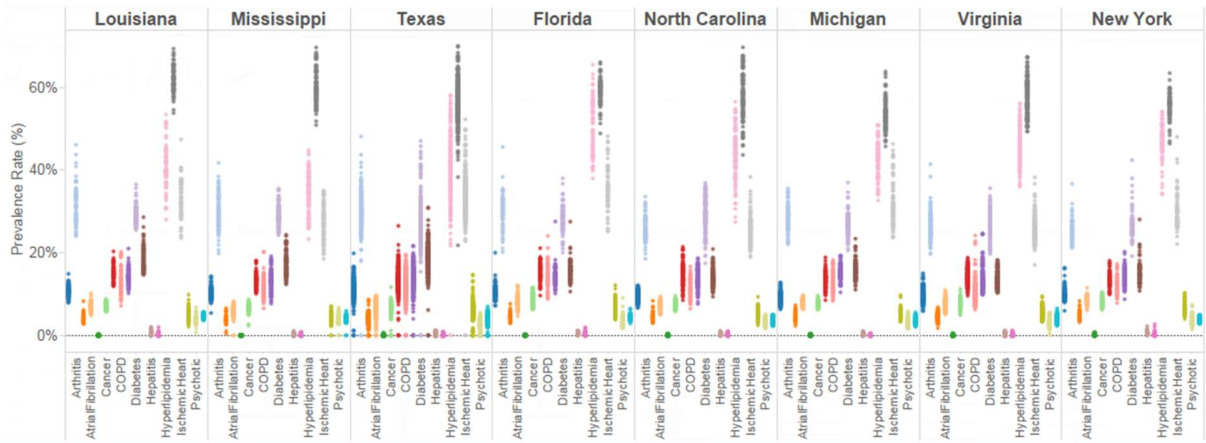

Figure 20: Condition Prevalence Rates by State

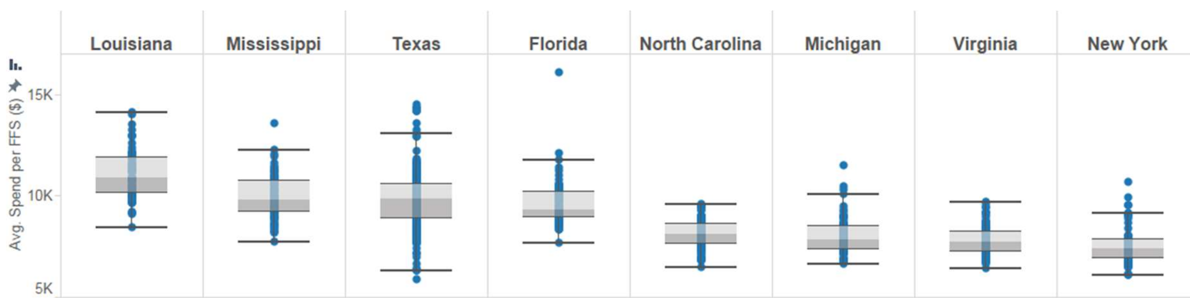

Figure 21: Average Cost per FFS Beneficiary by State

In Figure 20, each dot represents the average annual prevalence rate for a particular condition in a single county. Each of the 19 conditions are represented by a different color. This allowed us to quickly perform a visual comparison of the sickness profile of the Medicare beneficiaries in each state. The graphs confirmed that these states are composed of similar patient populations. In all eight of these states, patients experienced generally high levels of Hyperlipidemia, Hypertension, and Ischemic Heart Disease. Diabetes and Arthritis were also fairly common, while Autism, HIV/AIDS, and Hepatitis were rare. If patient health characteristics were the only driver of Medicare spending, we would have expected these eight states to demonstrate similar levels of spending. However, Figure 21 demonstrates that this was not the case. Each dot represents the average annual Medicare spending in a single county. We observed that Louisiana, Mississippi, Texas, and Florida all demonstrated higher average spending levels as well as wider ranges of spending levels. On the other hand, North Carolina, Michigan, Virginia, and New York demonstrated narrower ranges of spending as well as lower average spending.

The index was an extremely useful tool because it allowed us to identify patient populations with similar sickness characteristics. However, we observed that many states demonstrated a wide range of sickness indexes, indicating a variety of patient sickness profiles within a single state. In the next segment of research, we moved to a more granular level, examining data at the county level rather than the state level.

Clustering

The results of the regression analysis show that the cost does not differ too much across states. The principal component analysis shows that states with a similar sickness profile differ widely in cost. It has also been observed that the county-level costs have a much larger range as compared with the state-level costs. Thus it is essential to dissolve the state level boundaries and analyze the costs at the county level to get a deeper understanding of the cost drivers.

Looking at the sickness index and comparing states with similar sickness profiles, we were able to identify states where costs differed a lot. Drilling down into county-level, we wanted to achieve the same goal, i.e., to compare counties with similar sickness profiles and compare the costs. For this purpose, we first needed a way to be able to identify counties with similar sickness profiles. This was performed using the Clustering method.

Methodology

K-Means Clustering was used to cluster counties. The variable used for the distance measure was the Sickness Index. Clusters were created for each value of K between 2 and 30. The within-cluster sum of squares for each possible value of K was compared in the following scree plot.

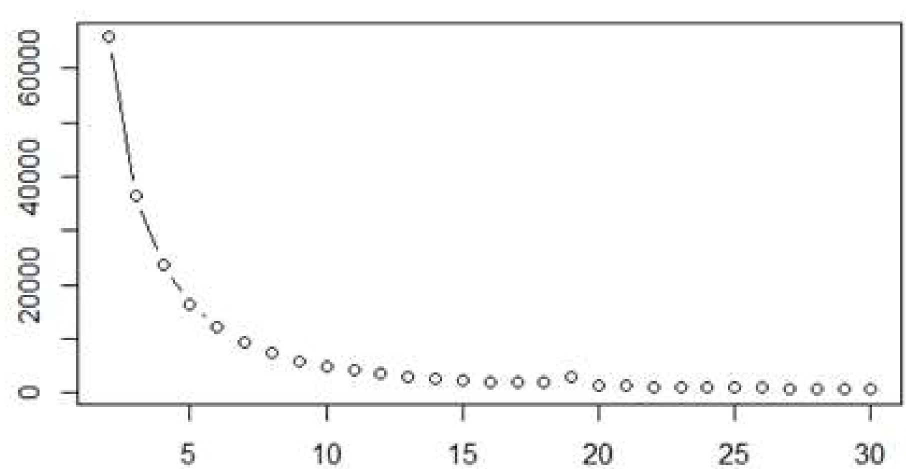

Figure 22: K-Means Clustering Scree Plot

The final number of clusters used was 7. At this configuration, the between-cluster sum of squares is 153,639.9. The within-cluster sum of squares for each cluster is given below.

Cluster ID	Within Cluster Sum of Squares
1	1255.35
2	1134.90
3	1187.68
4	1184.45
5	1164.57
6	1438.75
7	1958.24

Table 3: K-Means Clustering Within Cluster Sum of Squares

Thus a cluster number between 1 to 7 was assigned to each county for each year. It is possible that a county may be part of different clusters in different years if the prevalence rates changed over time and resulted in a change in Sickness Index.

The following map shows the results of the clustering analysis. Counties are colored based on the cluster they belong to in 2014. It can be seen that counties from various states group together into one cluster based on their sickness profiles. Thus this captures the granular information at county level which was being averaged out and lost in state-level analysis.

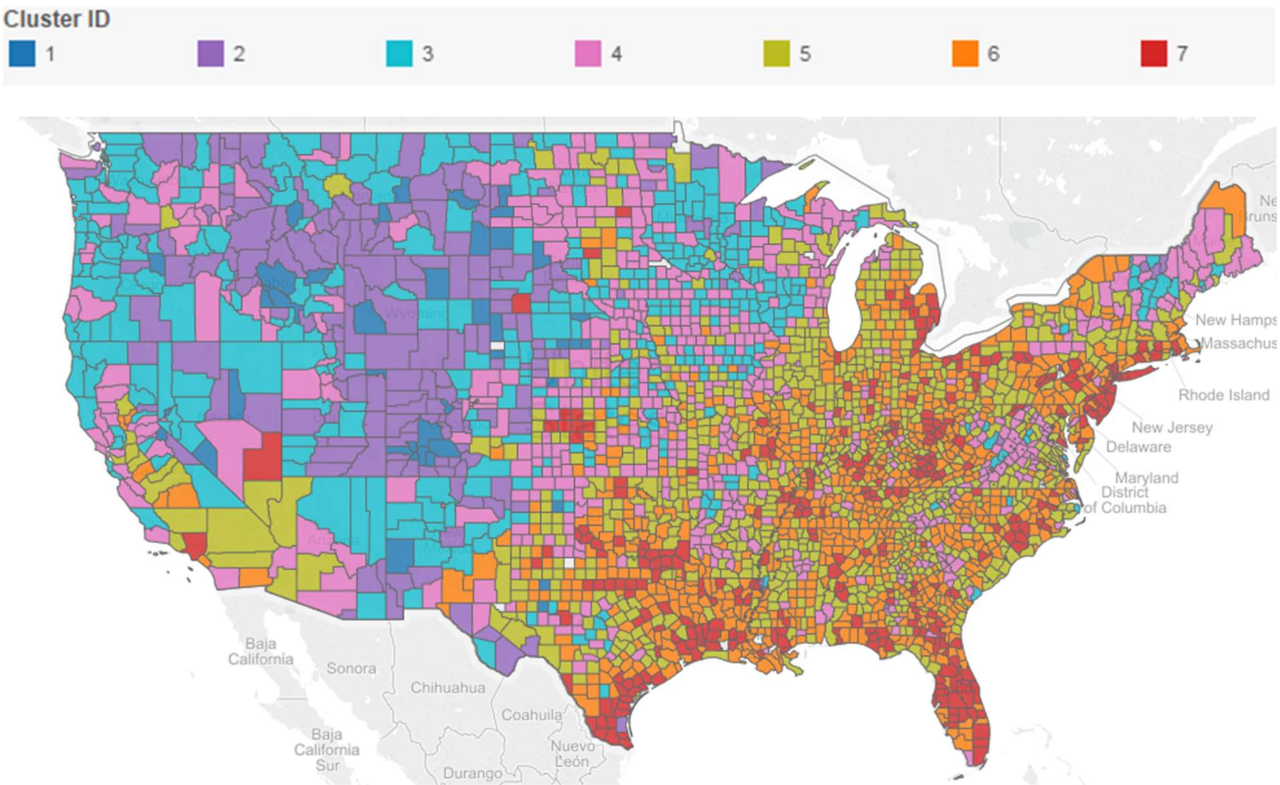

Figure 23: Counties Clustered by Sickness Index

Analysis

The following chart shows the result of the clustering analysis. The x-axis represents the Sickness Index with the level of sickness increasing from left to right. The y-axis shows the cost per FFS beneficiary. Each dot represents the cost per FFS beneficiary for a particular county in a particular year. The dots are colored based on the cluster ID. The counties with the lowest Sickness Index fall in cluster 1 and those with the highest Sickness Index fall in cluster 7.

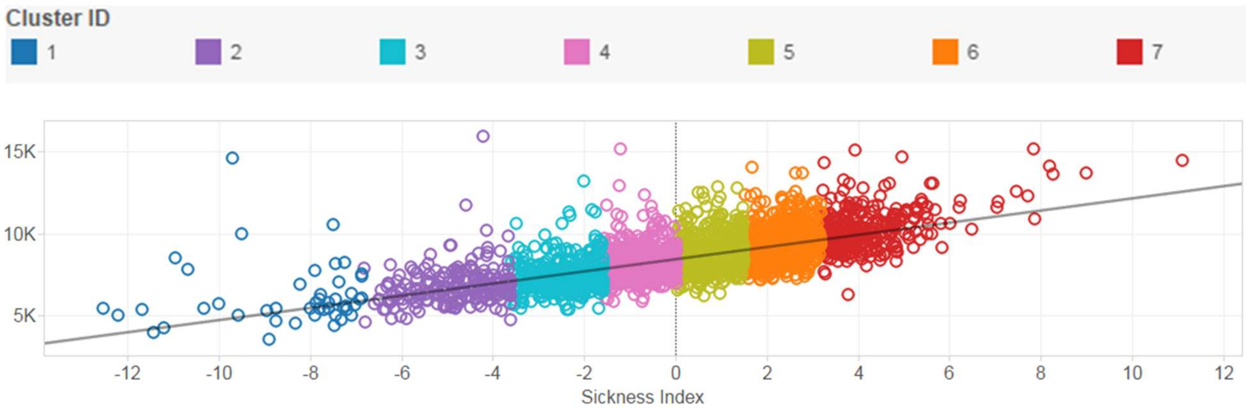

Figure 24: Cost by Sickness Index

The following chart shows the range of cost per FFS beneficiary for each cluster. Each dot represents the cost per FFS beneficiary for a particular county in a particular year. The color represents the cluster ID.

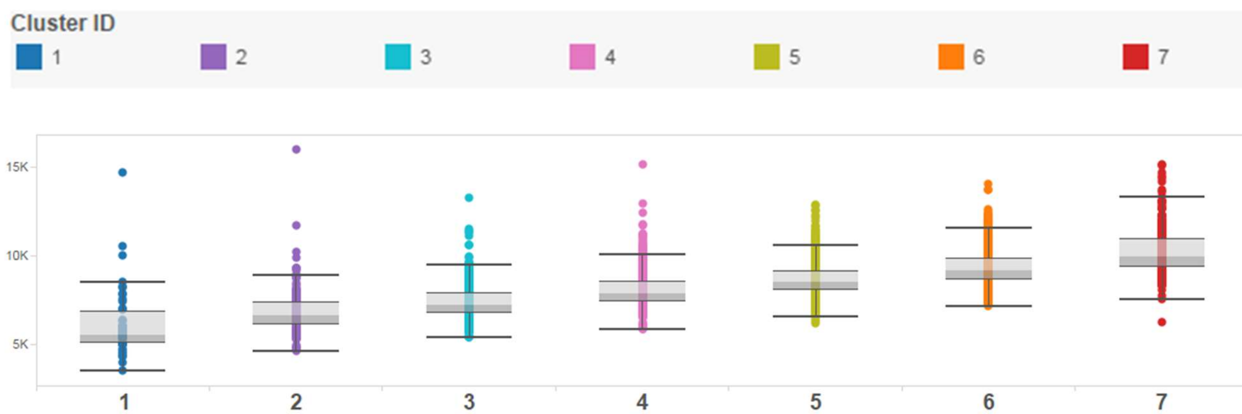

Figure 25: Cost by Cluster ID

These charts show a clear positive relationship between the sickness level of a county's population and the cost per FFS beneficiary for that county. On average, the cost per FFS beneficiary increases with an increase in sickness level of a population. However, there is cost variation within each cluster even though the sickness profiles are similar enough to get clustered together.

To investigate this in detail, we explored the sickness profiles of individual counties within a cluster and compared their costs. Following are some of the comparisons that are worth mentioning.

Outlier Counties

Some counties stand out on the chart for very high cost whereas some have very low costs. A detailed look at their sickness profiles shows that these counties have high prevalence rates of few chronic conditions whereas the prevalence rates for most other condition are zero. These counties have very small beneficiary populations (less than 100). The prevalence rate of zero may mean that there were fewer than 11 beneficiaries that were treated for a condition and so the data was suppressed.

In the chart below, the counties of Borden and Kenedy in Texas have an exceptionally high cost. Since we know that the beneficiary population here is small, it is possible that the counties are not well equipped for treatment and are extremely inefficient. But on the other hand, Clark, Idaho, despite being in a similar situation, is able to keep its costs very low. It may mean efficient treatment for the few conditions that they have to treat. It may also mean that they are not providing adequate care to the beneficiaries. It is definitely worth looking into these counties, especially the ones with high cost for possible insights for improvement. Although these insights would not result in significant savings in these particular counties because of their small populations, they may prove useful in other counties with larger population sizes.

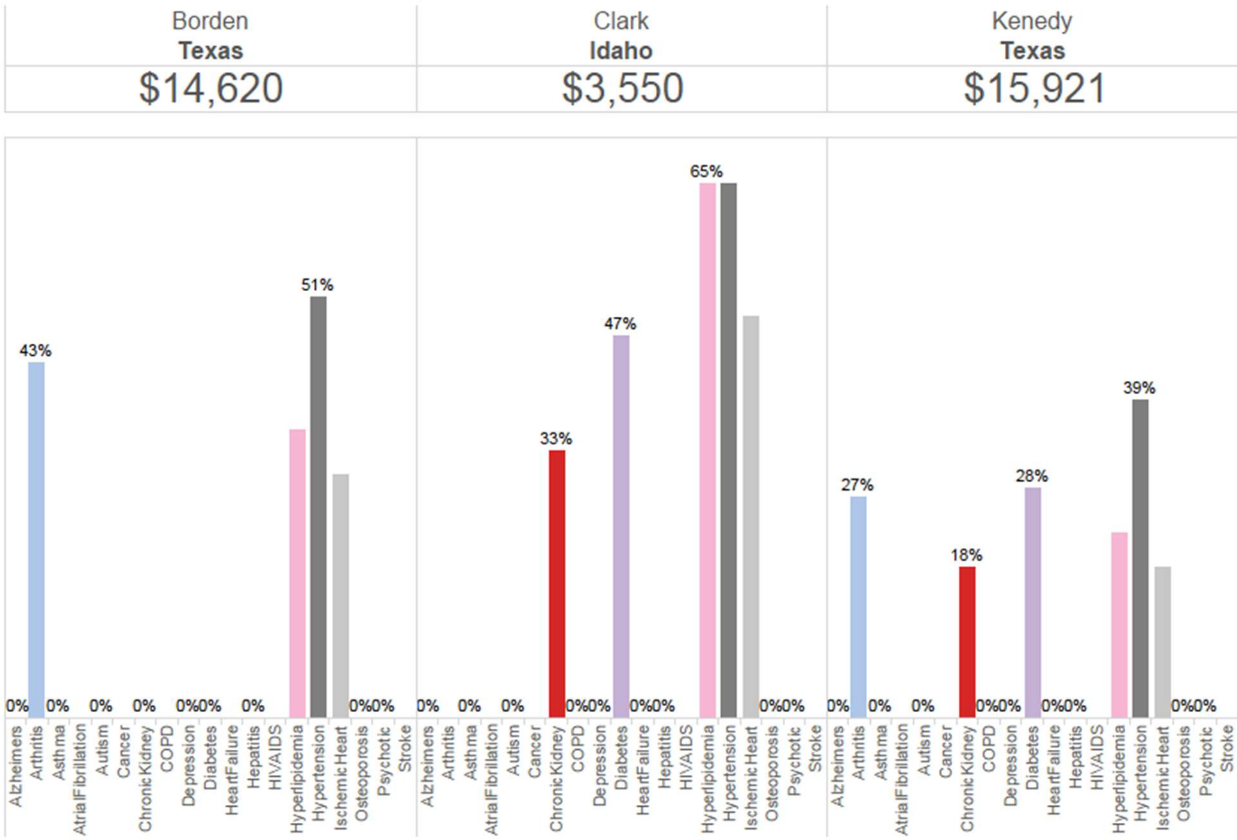

Figure 26: Sickness Profiles of Borden (Texas), Clark (Idaho), and Kenedy (Texas)

Kalkaska, Michigan vs. Montague, Texas

The bar chart below shows the sickness profiles of Kalkaska, Michigan and Montague, Texas by plotting the rates of prevalence for the 19 chronic conditions in these counties. The prevalence rates are very similar. However, the cost per FFS beneficiary (shown below the county names) differ by $2,595 per beneficiary. For a county in Texas to be spending this much more per beneficiary is alarming because the large population in this county will result in a high overall cost for the county. If Montague, Texas could save $2,595 per beneficiary, it would result in a total cost saving of $10.7 million per year.

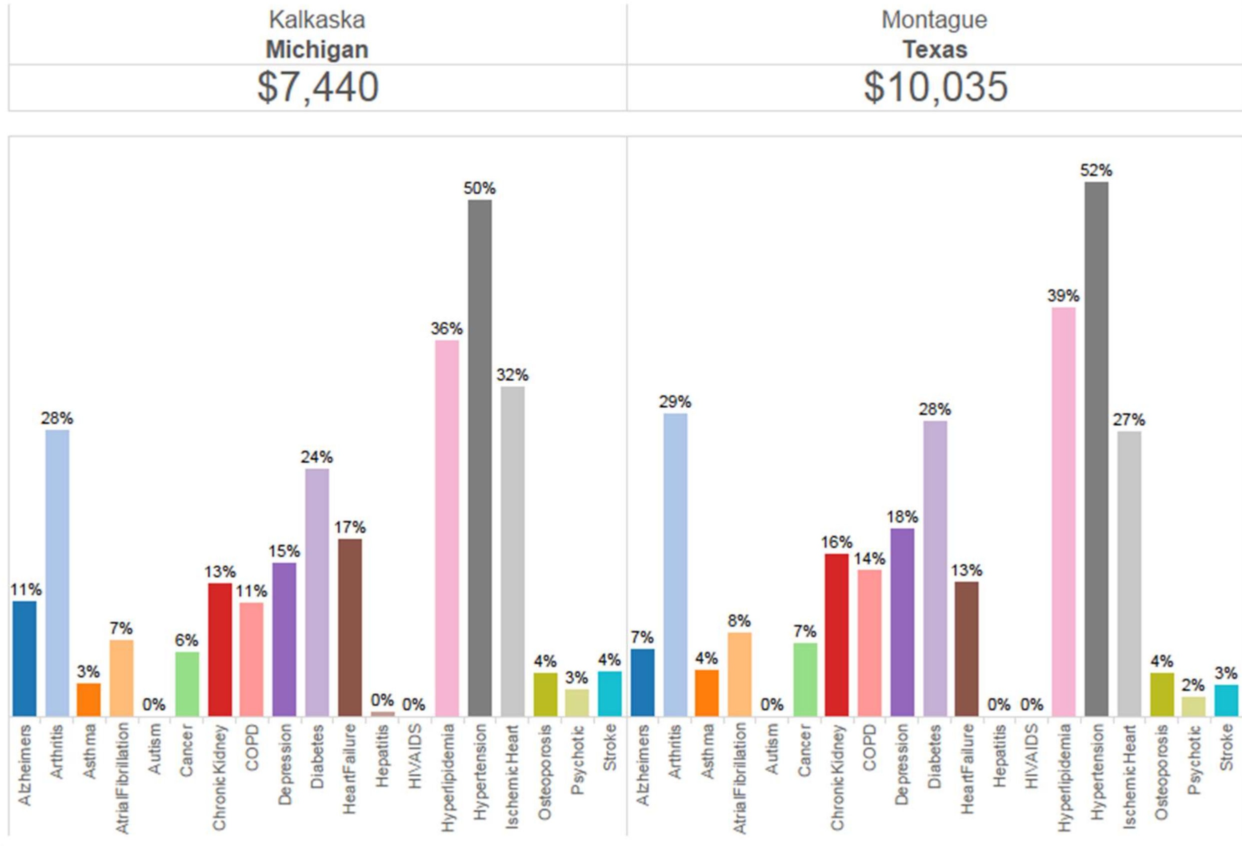

Figure 27: Sickness Profiles of Kalkaska (Michigan) and Montague (Texas)

Alger, Michigan vs. Clay, Mississippi

The bar chart below shows a similar comparison of the prevalence rates of the 19 chronic conditions for Alger, Michigan and Clay, Mississippi. The prevalence rates are very similar which shows that the counties are similar in sickness profile. However, the cost differs by $2,329. If Clay, Mississippi could close this difference, it would result in an annual cost saving of $8.01 million.

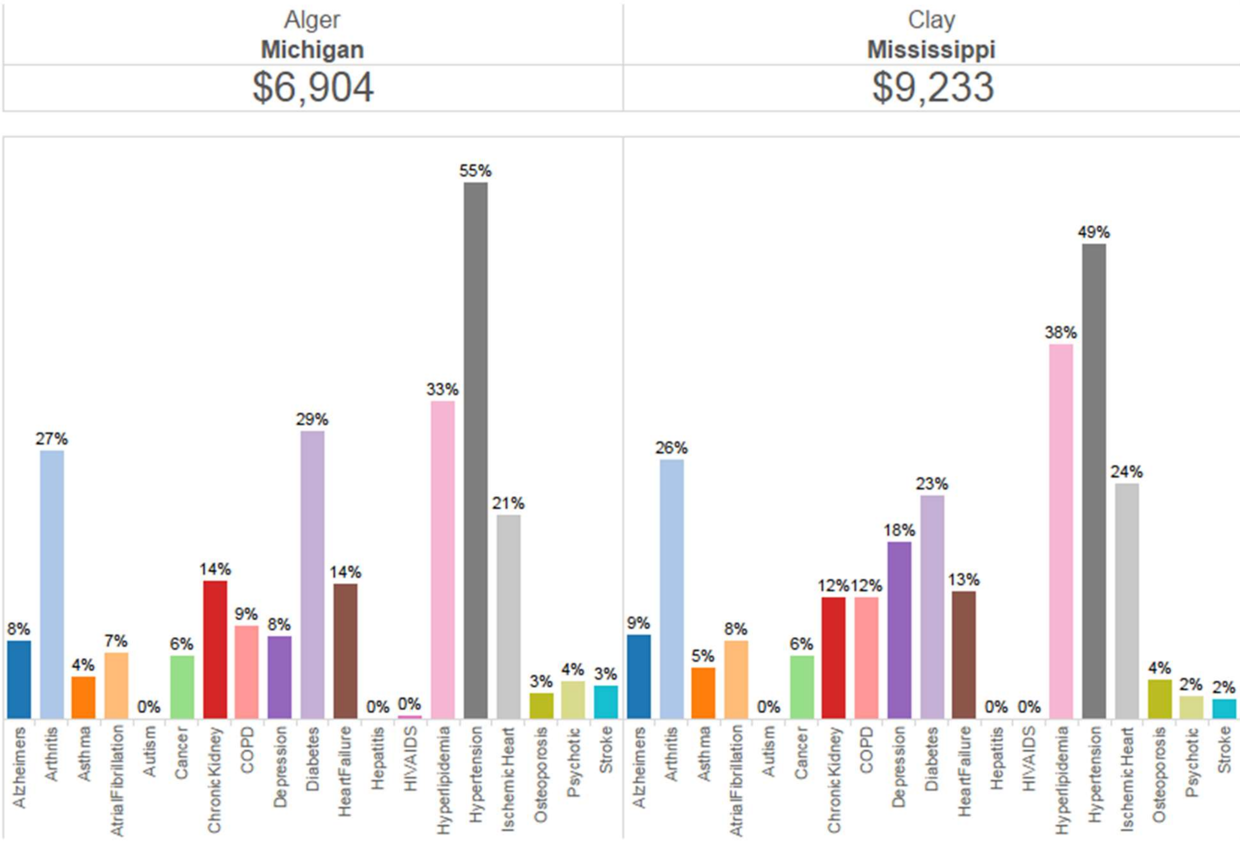

Figure 28: Sickness Profiles of Alger (Michigan) and Clay (Mississippi)

Tuscola, Michigan vs. Wise Texas

The following bar chart shows the comparison of the sickness profiles of Tuscola, Michigan and Wise, Texas. The counties have a similar sickness profiles but their costs per beneficiary differ by $1,782. Wise, Texas has a huge beneficiary population of 48,497. A reduction in the cost by $1,782 per beneficiary would result in a cost saving of $88.42 million per year.

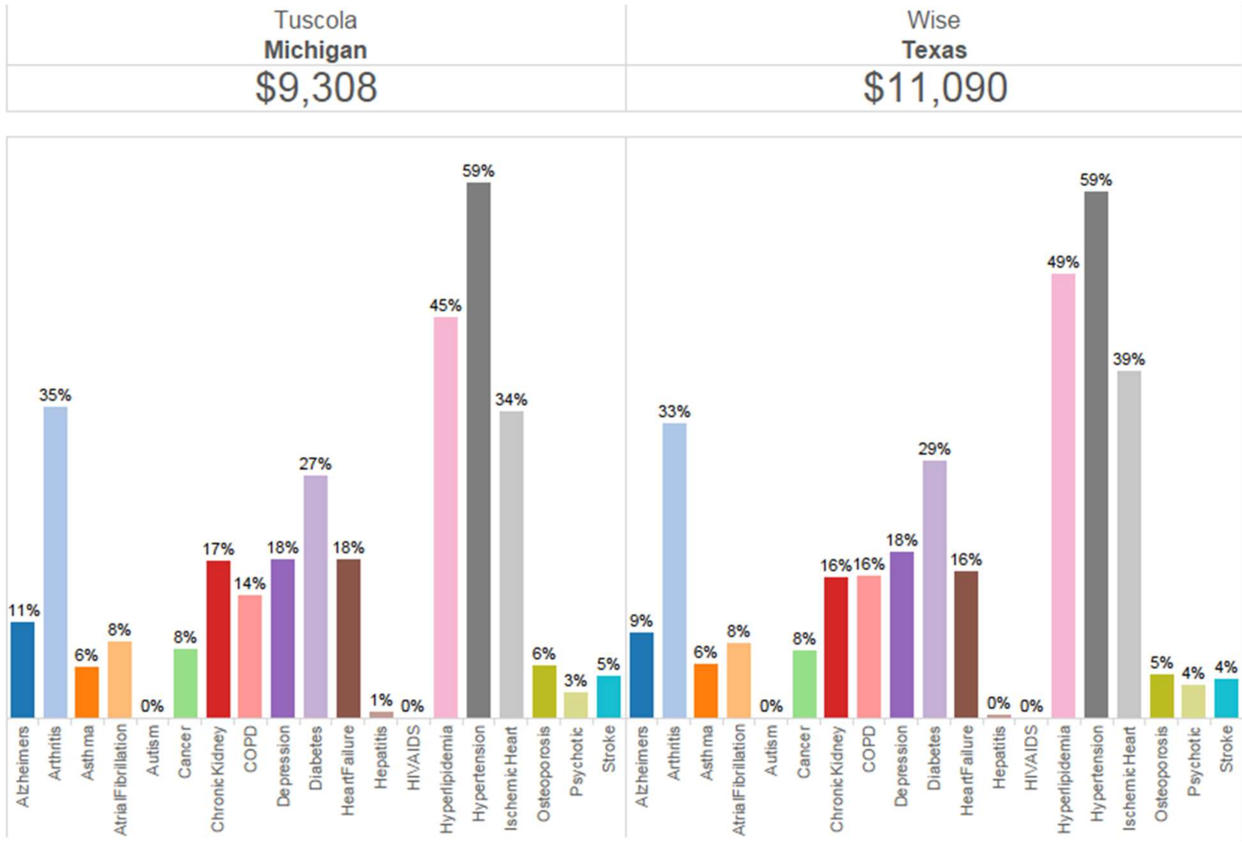

Figure 29: Sickness Profiles of Tuscola (Michigan) and Wise (Texas)

We previously identified that there is no one condition that is driving up the cost. We also determined that the sickness level of a county has a direct positive relationship with the cost per FFS beneficiary for that county. However, we also observe that counties of similar sickness profiles exhibit different spending per beneficiary. Thus it appears that even though the prevalence of chronic conditions is a definite factor contributing to cost, there are also other factors that are responsible for cost variations. Based on this analysis, we have been able to identify the starting point for further exploration by focusing on the counties that seem to be able to manage their costs better and possible apply the findings on the counties that have high cost.

Conclusions

The major findings of our analyses can be summarized as follows:

- In the data under consideration, Year did not stand out as a significant factor driving the Medicare costs. But the literature review, especially research into chronic conditions performed by Kaiser Family Foundation reveals that total Medicare cost is expected to rise sharply after 2018 due to the baby boomer generation reaching retirement. Therefore, cost optimization is needed before the anticipated increase in number of beneficiaries occur.
- The cost per FFS beneficiary does not vary greatly by state. None of the states stand out as exceptional cost drivers.
- The findings regarding conditions present somewhat of a paradox: expensive conditions are rare, and common conditions are inexpensive. Thus, none of the conditions alone has a large impact on cost.
- In general, cost per FFS beneficiary goes up as prevalence rates increase, and therefore, sicker counties spend more per beneficiary.
- Medicare spending differs for counties within the same state. Furthermore, many counties with similar sickness profiles nonetheless have very different costs.

Overall, this study has identified counties that should be targeted first in any further research performed into Medicare expenses. It has helped to narrow down the scope geographically and to focus attention on counties where major cost saving opportunities lie.

Future Research: Recommended Next Steps

This analysis identifies counties that have very high cost. Since other counties with a similar sickness profile have a lower cost, it shows that there is room for improvement in the high cost counties. This provides a starting point for future research by focusing on these problematic counties while gaining insights from other counties that are managing costs better. It is required to explore factors other than the sickness profile to identify the reason for these cost differences. We suggest looking into factors such as medical infrastructure, ambulatory costs, transportation cost, salaries of staff, specialists and doctors, costs of laboratory tests, and other administrative expenses. This exploration can help in bringing down the cost by identifying possible cost optimizations and cost saving measures.

HIV/AIDS and Autism are identified to be the most expensive conditions. They are very sensitive to change in prevalence. It is important to investigate why the treatment of these conditions is so expensive and if there is any way to reduce the cost. Also, in addition to optimizing treatment expenses, it is important to reduce their spread to keep overall costs down.

Conditions that have very high prevalence should also be examined further for possible cost optimizations to reduce overall cost. Hypertension can be targeted first because of its very high prevalence and high correlation with other conditions.

Our research has focused on the aggregate data available for the Medicare fee-for-service program (Parts A & B). However, we believe that a wealth of additional areas is available for further exploration:

- **Disaggregated data for individual conditions:** Rather than utilizing the total Standardized Cost per FFS Beneficiary measure, future researchers may wish to explore condition-level costs, such as typical annual costs associated with different illnesses. This may highlight inefficiencies in treatment of particular conditions.

- **Pharmaceutical expenses:** Projected Medicare cost growth is expected to be higher for Medicare Part D (which covers prescriptions) than for Parts A and B. Therefore, further analysis into Medicare practices regarding prescription and generic drugs may yield additional cost-savings insights.
- **Medicare Advantage:** The Medicare Advantage program allows private insurance companies who provide equivalent Parts A and B coverage to be reimbursed for medical claims provided to Medicare enrollees. This program accounts for 26% of all Medicare benefits payments in 2014, and enrollment in this program is expected to rise. As this program will become increasingly relevant in years to come, analysis into the cost structures of Medicare Advantage may provide cost-savings in years to come.

We hope that our research adds meaningful insights to the question about drivers of Medicare cost variation, and that it can be used by the medical community to define a beneficial cost saving strategy for Medicare in the years to come.

Data Sources

Centers for Medicare & Medicaid Services. (2016, January 07). CMS Beneficiary Enrollment and Characteristics File. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Medicare_Beneficiary_Characteristics.html

Centers for Medicare & Medicaid Services. (2016, January 07). CMS Chronic Conditions File. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main.html

Centers for Medicare & Medicaid Services. (2016, January 27). CMS Public Use Geographic Variation File. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html

Works Cited

Centers for Medicare & Medicaid Services. (2016, July 16). *NHE Fact Sheet*. Retrieved from https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html

Centers for Medicare & Medicaid Services. (2016, January). *Chronic Conditions among Medicare Beneficiaries: A Methodological Overview.* Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Methodology.html

Cubanski, J., Neuman, T., & White, C. (2015, October 09). *The Latest on Geographic Variation in Medicare Spending: A Demographic Divide Persists But Variation Has Narrowed.* Retrieved from http://kff.org/medicare/report/the-latest-on-geographic-variation-in-medicare-spending-a-demographic-divide-persists-but-variation-has-narrowed/

The Henry J. Kaiser Family Foundation. (2015, July 24). *The Facts on Medicare Spending and Financing.* Retrieved from http://kff.org/medicare/fact-sheet/medicare-spending-and-financing-fact-sheet/

Appendices

Appendix 1: Steps for Recreating Data

Steps for reproducing the data have been provided below. Researchers can replicate our data transformations from scratch using the original source files (option 1) or utilize the sanitized data provided in the csv files (option 2). Both methodologies are explained below.

Option 1: Reproducing the Data from Original Source Files

We have used Microsoft SQL Server 2014 to load the source files into staging tables, transform them and populate the final tables. However, any database tool should be able to perform these tasks.

All raw data files are available from the CMS website:

- Beneficiary Enrollment and Characteristics [ZIP, 6MB]
    - https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/Beneficiary_Enrollment_Tables.zip
        - County_Table_Bene_Enrollment_2007.xlsx
        - County_Table_Bene_Enrollment_2008.xlsx
        - County_Table_Bene_Enrollment_2009.xlsx
        - County_Table_Bene_Enrollment_2010.xlsx
        - County_Table_Bene_Enrollment_2011.xlsx
        - County_Table_Bene_Enrollment_2012.xlsx
        - County_Table_Bene_Enrollment_2013.xlsx
        - County_Table_Bene_Enrollment_2014.xlsx
- Prevalence State/County Level: All Beneficiaries by Age, 2007-2014 [ZIP, 32MB]
    - https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/CC_Prev_State_County_All.zip
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2007.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2008.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2009.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2010.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2011.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2012.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2013.xlsx
        - County_Table_Chronic_Conditions_Prevalence_by_Age_2014.xlsx
- State/County Table - All Beneficiaries [ZIP, 49MB]
    - https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/Downloads/State_County_Table_All.zip
        - County_All_Table.xlsx

Once all raw data were downloaded, we performed the following data transformations:

- Removed all rows in which the Total Standardized Cost per FFS Beneficiary was $0.
- Converted all asterisk-labelled fields to 0.
- Converted the Year variable to a categorical variable.
- Converted the 19 condition prevalence rate variables into a single categorical Condition variable with prevalence listed as the value in the next column for the regression analysis.

SQL Files	Purpose
Loading_StateAbbr.sql	Loads the state abbreviations (required for all of the following scripts)
Loading_Enrollment.sql	Loads the Beneficiary Enrollment and Characteristics files
Loading_Chronic.sql	Loads the Prevalence State/County Level: All Beneficiaries by Age, 2007-2014
Loading_Cost.sql	Loads the State/County Table - All Beneficiaries files
AlldataInsertion.sql	Combines all data sources into a single table

Table 4: Details of SQL files

Option 2: Using the Data from Pre-processed Files

The data obtained after all the above mentioned processing has been exported to the following csv files.

- AllData_Regression.csv
- AllData_PCA.csv

These files can be used directly for analysis.

Appendix 2: Files for Recreating the Analysis

The R files mentioned below take two input files and generate output files which can then be used to create Tableau visualizations:

Filename	Purpose	Data File
RegressionAnalysisCode.R	Implementation of Regression Analysis in R reads data from database or .csv file and performs regression after transforming the data.	AllData_Regression.csv
DataForTableau.R	County level analysis using PCA and Clustering. This file also exports data that has been used in Tableau.	AllData_PCA.csv

Table 5: R Files

Users seeking to reproduce our results are welcome to reference the steps in the SQL scripts and R files provided above to understand how we computed our results.

Users can download all visualizations used in this report from the Tableau Public dashboard, available at: http://tabsoft.co/2aPp9s3

Appendix 3: Regression Diagnostics

Regression Diagnostics

Measure	Value
R-Squared	0.5708
Adjusted R-Squared	0.5706
Residual Standard Error	954.4

Table 6: Regression Diagnostics

Below are the diagnostic plots:

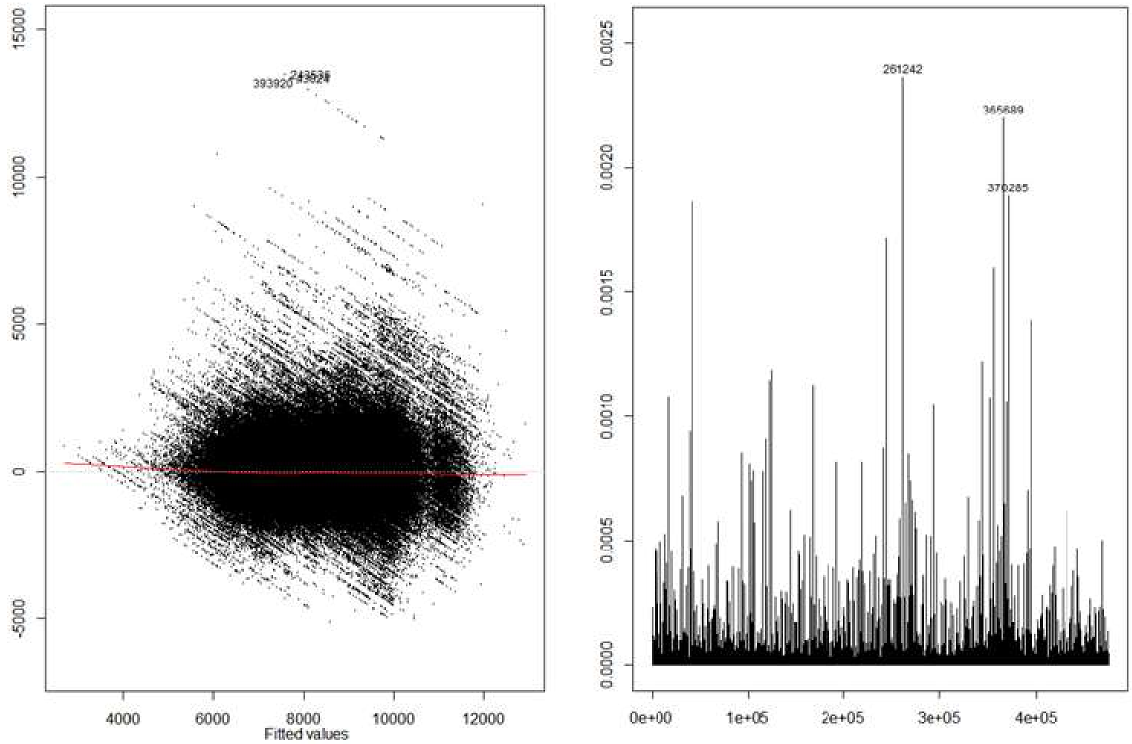

Figure 30: Residuals vs. Fitted and Cook's Distance Plot

Residuals are very well distributed around 0 except some points. Additionally, from Cook's Distance plot, there are no points having a high leverage on the model.

Appendix 4: Regression Coefficients

Year Coefficients

Year	Coefficient	p-value
2007	645.90	1.94E-09
2008	1007.39	7.98E-21
2009	1357.22	1.94E-36
2010	1571.03	3.26E-48
2011	1370.42	4.20E-37
2012	1396.68	1.80E-38
2013	1490.07	1.53E-43
2014	1547.75	7.59E-47

Table 7: Year coefficients with p-values

State Coefficients

State	Coefficient	p-value
Alabama	4310.19	0
Alaska	1447.20	1.29E-48
Arizona	2695.43	3.28E-160
Arkansas	3531.28	2.46E-290
California	2407.42	1.59E-135
Colorado	2256.45	1.23E-119
Connecticut	3734.49	1.38E-283
Delaware	3874.24	1.43E-254

District of Columbia	4272.65	2.37E-186
Florida	4842.58	0
Georgia	4051.89	0
Hawaii	845.32	8.87E-15
Idaho	2378.23	1.90E-131
Illinois	3560.50	1.94E-296
Indiana	3688.30	2.45E-317
Iowa	2850.56	8.07E-191
Kansas	4177.39	0
Kentucky	3938.38	0
Louisiana	6311.92	0
Maine	2630.25	1.47E-152
Maryland	4234.34	0
Massachusetts	3590.56	2.04E-278
Michigan	3281.81	1.93E-251
Minnesota	2601.91	5.29E-159
Mississippi	5321.27	0
Missouri	3597.78	6.34E-303
Montana	2216.79	4.10E-115
Nebraska	3399.71	2.28E-270
Nevada	2467.83	7.00E-136
New Hampshire	2544.41	5.54E-137
New Jersey	4417.65	0
New Mexico	1961.86	2.11E-89
New York	2795.56	4.29E-182
North Carolina	3431.98	1.23E-275
North Dakota	2842.19	9.81E-188
Ohio	4072.79	0
Oklahoma	4509.10	0
Oregon	1743.47	4.14E-71
Pennsylvania	3904.89	0
Puerto Rico	1751.42	1.94E-34
Rhode Island	3275.01	1.31E-202
South Carolina	4016.83	0
South Dakota	2824.07	2.97E-186
Tennessee	4361.12	0
Texas	5096.34	0
Utah	2641.93	2.30E-159
Vermont	1965.88	2.75E-85
Virginia	3131.40	1.49E-230
Virgin Islands	0.00	(Baseline)
Washington	2227.77	4.43E-115
West Virginia	3174.59	8.10E-234
Wisconsin	2510.47	9.75E-148
Wyoming	2128.22	1.11E-102

Table 8: State coefficients with p-values

Condition Coefficients

Condition	Coefficient	p-value
Alzheimers	1841.94	1.56E-243
Arthritis	1285.46	4.78E-114
Asthma	2562.47	0
Atrial Fibrillation	2909.04	0
Autism	3343.26	0
COPD	2335.85	0
Cancer	2742.31	0
Chronic Kidney	2040.10	5.21E-320
Depression	2434.43	0

Diabetes	1153.45	7.47E-96
HIVAIDS	3288.98	0
Heart Failure	1661.24	3.94E-205
Hepatitis	3268.52	0
Hyperlipidemia	2113.47	0
Hypertension	0	(Baseline)
Ischemic Heart	1428.71	1.94E-154
Osteoporosis	2718.81	0
Psychotic	2647.21	0
Stroke	2560.54	0

Table 9: Condition coefficients with p-values

State and Prevalence Interaction Coefficients

State * Prevalence	Coefficient	p-value
Alabama * Prevalence	-839.34	0.211634
Alaska * Prevalence	2165.00	0.001688
Arizona * Prevalence	713.73	0.298544
Arkansas * Prevalence	129.14	0.847673
California * Prevalence	742.97	0.269966
Colorado * Prevalence	2442.35	0.000296
Connecticut * Prevalence	-53.25	0.938728
Delaware * Prevalence	-882.13	0.217272
District of Columbia * Prevalence	824.51	0.344054
Florida * Prevalence	-229.81	0.732335
Georgia * Prevalence	-703.97	0.293772
Hawaii * Prevalence	-169.37	0.813697
Idaho * Prevalence	1378.34	0.041864
Illinois * Prevalence	-407.54	0.543740
Indiana * Prevalence	-212.52	0.751631
Iowa * Prevalence	273.61	0.683757
Kansas * Prevalence	168.32	0.802050
Kentucky * Prevalence	-645.97	0.335614
Louisiana * Prevalence	-641.12	0.340152
Maine * Prevalence	318.53	0.641663
Maryland * Prevalence	-576.78	0.393739
Massachusetts * Prevalence	620.57	0.365395
Michigan * Prevalence	-17.10	0.979692
Minnesota * Prevalence	1152.27	0.086857
Mississippi * Prevalence	-599.38	0.372216
Missouri * Prevalence	-106.33	0.874108
Montana * Prevalence	1557.07	0.021266
Nebraska * Prevalence	455.00	0.49834
Nevada * Prevalence	1842.49	0.007369
New Hampshire * Prevalence	499.19	0.471928
New Jersey * Prevalence	-540.22	0.424808
New Mexico * Prevalence	859.77	0.204361
New York * Prevalence	90.86	0.892523
North Carolina * Prevalence	-216.47	0.747087
North Dakota * Prevalence	82.83	0.902176
Ohio * Prevalence	-398.31	0.553051
Oklahoma * Prevalence	-286.66	0.669628
Oregon * Prevalence	1067.75	0.11553
Pennsylvania * Prevalence	-270.99	0.686836
Puerto Rico * Prevalence	-1409.95	0.078231
Rhode Island * Prevalence	-44.72	0.94958
South Carolina * Prevalence	-614.84	0.361071
South Dakota * Prevalence	707.24	0.293496
Tennessee * Prevalence	-387.91	0.56338

Texas * Prevalence	-266.47	0.690879
Utah * Prevalence	1355.56	0.046945
Vermont * Prevalence	829.85	0.229554
Virginia * Prevalence	-296.20	0.658817
Washington * Prevalence	1239.07	0.067322
West Virginia * Prevalence	-496.36	0.460558
Wisconsin * Prevalence	534.05	0.427366
Wyoming * Prevalence	2487.06	0.000316

Table 10: State and Prevalence Interaction Coefficients with p-values

Condition and Prevalence Interaction Coefficients

Condition * Prevalence	Coefficient	p-value
Hypertension * Prevalence	6323.77	9.49E-21
Alzheimer's * Prevalence	16134.23	1.53E-109
Arthritis * Prevalence	7565.21	1.32E-28
Asthma * Prevalence	20781.43	2.47E-144
Atrial Fibrillation * Prevalence	6310.55	9.49E-17
Autism * Prevalence	73638.22	1.05E-34
COPD * Prevalence	8715.01	1.40E-36
Cancer * Prevalence	8756.86	1.88E-28
Chronic Kidney * Prevalence	10234.70	2.41E-49
Depression * Prevalence	6742.99	1.66E-22
Diabetes * Prevalence	8686.48	2.93E-37
HIVAIDS * Prevalence	79161.64	8.53E-186
Heart Failure * Prevalence	11159.23	8.83E-59
Hepatitis * Prevalence	33967.77	2.88E-70
Hyperlipidemia * Prevalence	3215.81	1.87E-06
Ischemic Heart * Prevalence	6891.85	3.04E-24
Osteoporosis * Prevalence	12153.34	1.35E-59
Psychotic * Prevalence	22673.56	1.62E-170
Stroke * Prevalence	23778.74	2.02E-171

Table 11: Condition and Prevalence Interaction Coefficients with p-values