Statistical Inference on Medicare Payment for Stent Insertion Procedure in US Midwest

Based on Referral Network and Physician Information

Zaiwei Liu, Xiao Hu, Hao Xin, Chenxi He, Rong Ma

University of Wisconsin-Madison

December, 2015

Abstract

In this report, a statistical inference on Medicare payment for stent insertion procedure in the Midwest is performed based on the referral network data and medicare data. The result shows that the Medicare payment amount is influenced by geographical pattern, network characteristics and physician/beneficiary information. In a word, Medicare payments for stent insertion procedure are significantly different among states; physicians in urban areas have relatively higher payments than those in rural areas. In terms of network analysis, the connectedness and referred counts of a physician is not significant; it's the hierarchical centrality that matters. For cardiologists, the age, total counts and the disease patterns of his/her beneficiaries will significantly make a difference in the Medicare payment.

Keywords: Medicare Payment, Stent Insertion Procedure, US Midwest, Statistical Inference, Network Analysis

Contents

1. Introduction

1.1 Background

In the United States, Medicare is a national social insurance program, administered by the U.S. federal government since 1966, currently using about 30 private insurance companies across the United States.[1] This program is aimed for people who are 65 or older, certain younger people with disabilities, End-Stage Renal Disease[2] or amyotrophic lateral sclerosis[3]. Medicare serves a large population of elderly and disabled individuals.

Medicare is composed of four parts. Part A covers inpatient hospital stays such as food, test and semiprivate room. Part B is medical insurance. It helps pay for certain services and products not covered by Part A, which is generally on an outpatient basis. Part C is the Medicare advantage plan. Original Medicare beneficiaries who choose to enroll in Part C will receive the same standard benefits and an annual out of pocket (OOP) limit, which is not included in Original Medicare. Besides, for almost all Part C plans, the beneficiary is required to have a primary care physician; that is not a requirement of Original Medicare. Part D is the prescription drug plan. Anyone with part A or B is eligible for Part D. It's a federal government program to subsidize the costs of prescription drugs and prescription drug insurance premiums for Medicare beneficiaries.

Medicare has several sources of financing. Part A is funded by revenue from a 2.9% payroll tax levied on employers and workers (each pay 1.45%). Part B and D are partially funded by premiums paid by Medicare enrollees and general fund revenue. Several surtaxes are added to fund both part in the following years. In 2011, Medicare spending accounted for about 15% of the Federal budget. This share is projected to exceed 17% by 2020.[4] However, over the long-term, Medicare faces significant financial challenges because of rising overall healthcare costs, increasing enrollment as the population ages, and a decreasing ratio of workers to enrollees.

Obviously it's a complex system. Here we focus on the Part B of Medicare. Part B coverage begins once a patient meets his or her deductible ($166 in 2016), then typically Medicare covers 80% of approved services, while the remaining 20% is paid by the patient.The enrollees must then cover the remaining approved charges either with supplemental insurance or with another form of out-of-pocket coverage. Out-of-pocket costs can vary depending on the amount of health care a Medicare enrollee needs. They might include uncovered services—such as long-term, dental, hearing, and vision care—and the supplemental insurance. However, the patient may have to pay more depending on certain type of doctor. There are three types of Original Medicare doctors: Participating doctors, Non-participating doctors and Opt-out doctors. Participating doctors are required to submit a

---

[1] From Wikipedia "medicare" page. https://en.wikipedia.org/wiki/Medicare_(United_States)
[2] Permanent kidney failure requiring dialysis or a transplant, sometimes called ESRD
[3] Also known as Lou Gehrig's disease and Charcot disease, is a specific disorder that involves the death of neurons
[4] A Primer on Medicare Financing | The Henry J. Kaiser Family Foundation. Kff.org (January 31, 2011). Retrieved on 2013-07-17.

bill (medical claim) to Medicare for care the patients receive. Medicare will process the bill and pay the doctor directly. For Non-participating doctors the patient needs to pay the full cost and Medicare will reimburse the patient directly. But Opt-out doctors can charge their Medicare patients whatever they want. These doctors do not submit any bill (medical claims) to Medicare and are not subject to the Medicare law that limits the amount doctors may charge patients.

1.2 Problem of Interest and Thesis

Our objective here is to find the potential factors that would affect the Medicare payment amount. In details, we're going to analyze the relationship between geographical patterns, referral network characteristics, physician information and Medicare payment amount. We focus on the average Medicare payment of stent insertion procedure for 1129 cardiologists who performed stent surgery in Midwest. The cardiologists belong to three of the specialities: Cardiology, Cardiac Surgery, Cardiac Electrophysiology.

There are several factors that may have effects on the amount a physician is going to charge and how much reimbursement the Medicare would pay. For instance, a physician's location might be influential. For network characteristics, the payments for a highly skilled or more centralized physician, whose degree and centrality measure are large in the physician referral network, might be higher than others. And we want to see whether the "hierarchical structures" of network would make some difference on the Medicare payment, which means that we're going to compute the network features in different types of networks. Besides, there are other potential factors such as economy and population.

To be more specific, we'll focus on the payments of Stent Insertion Procedure. Such a procedure is quite important an operation in cardiology and relatively expensive. For patients with coronary heart disease, stents are used to open narrowed arteries and help reduce symptoms or to help treat a heart attack. We also narrow down the range of physicians to be within the U.S. Midwest area, which consists of 12 states in the north central United States.

1.3 Data Source

Taking those factors into consideration, several datasets are included in our analysis, such as Medicare data, physician referral network data, economy data and so on. Here's some basic information about the datasets.

1.3.1 The Medicare Dataset

The medicare data used here is the Medicare Provider Utilization and Payment Data from the CMS website[5]. It includes several features of physicians such as NPI, address, Medicare payment and some basic information of his or her patients. Physicians are divided into 84

[5]From:https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html

different provider types. Here we'll focus on the Cardiology, Cardiac Surgery and Cardiac Electrophysiology types, which account for 2.86% of all types.

Generally, patients who are referred to Cardiology are from Primary Care. The Primary Care providers includes Family Practice, Internal Medicine, Obstetricians, Gynecologists and Pediatricians, which accounts for 14.10% of all types. In the detailed datasets, data on specific procedures of physicians are included. The HCPCS code of the procedure we're interested in is 92928, which is the procedure for stent insertion, namely the "Catheter insertion of stent in major coronary artery or branch".[6]

1.3.2 Other Datasets Involved

Besides, we also use other several datasets such as the referral network dataset and population census data.The census data corresponds with zip-code and can be linked to each physician. The referral network dataset includes physician's' NPI on both sending and receiving ends, with exact number of patients and visitors.

1.4 Outlines

Firstly in Section 1, the basic information about the Medicare is introduced. Then, according to the problem of interest, the objective we're going to study is well-defined, and related datasets are selected for further analysis. Section 2 provides the main result of our report. It's basically consisted of three parts: geograph, network features and regression analysis. Section 3 includes more detailed analysis from geographical exploration, network analysis and regression models. Taking physician information and referral network into consideration, we finally come up with the final inference model. Lastly, discussions of the problems we run into during this project are made in Section 4.

2. Main Results

Main results given by exploratory analysis and statistical modelling in our reports are summarized as below. For Midwest physicians within stent insertion procedure, we have the following results:

Geographically, large cities such as Chicago, Detroit, and St. Louis have more physicians than small cities or rural area. Meanwhile, physicians from large cities have relatively higher medicare payment amount, whereas in rural areas or small cities they have lower medicare payment amount. Also, there's no significant correspondence between the city size and the average submitted charge amount, but the physicians in Wisconsin actually tend to submit more charge amount to Medicare than other Midwest states. Finally, there is no significant correlation between submitted charge amount and the actual Medicare payment amount.

As for the network features, the degree (connectedness) and referred counts (three types) of a physician are not significant factors to the payment amount in the six networks. The

[6]From:https://www.cms.gov/apps/ama/license.asp?file=http://download.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Downloads/Medicare_Provider_Util_Payment_PUF_a_CY2013.zip

eigenvector centrality scores based on hierarchical structures are important factors: more centralized cardiologists tend to get higher Medicare payment amounts. When checking the relationship between the network characteristics and the payments, we can see that the eigenvector centrality measures for all types of physicians in Midwest area and within each state are significant; and the centrality for specific types of physicians in Midwest are also significant. When considering other control variables and dividing the data by Midwest centrality, the centrality measures for nationwide and Midwest networks matter.

Statistically, the degree and referred counts of a physician are not significant. It's the hierarchical centrality that matters: more centralized cardiologists tend to get higher Medicare payments. Disease pattern affects payment amount for physicians with higher centrality. Also, the amount a physician get payed is highly depend on which state he/she is in. Combined result from geographical distribution of medicare payment amount, urban physicians have higher payment amounts than physicians in rural area.

3. Statistical Analysis

In this section, we'll provide detailed statistical analysis and evidences supporting our thesis presented in section 2. We'll firstly take a look at the general numerical and geographical distribution of Medicare payment amount, then use proper statistical models to make statistical inference of our observational findings. We'll also explore the physician referral network and network based modeling, and examine some other potential factors such as economic, physician and beneficiaries information.

Basically, the following questions are going to be answered in this section: Are there any interesting patterns within geographical representation of the Medicare payments for stent insertion procedure? Do such medicare payments have any connections to the size of the city (in both economic and geographical senses)? How is the submitted charge amount correlated to the actual medicare payment amount? These questions are of great importance to our following study.

3.1 A General View of Medicare Payment Amount

A general view of Medicare payment amount can be obtained in various ways. In our analysis, we'll use geographical representations as our main visualization to the data. We will pay much attention to the zipcode, city and state, of those Midwest physicians who performed stent insertion procedure in their medical practice.

We first look at the the average medicare payment amount for Midwest physicians who performed stent insertion procedure in cardiology related provider types. From the histogram, we can see the distribution is roughly normal, without being highly skewed. So in later studies, we won't perform any transformations of the average payment amount.

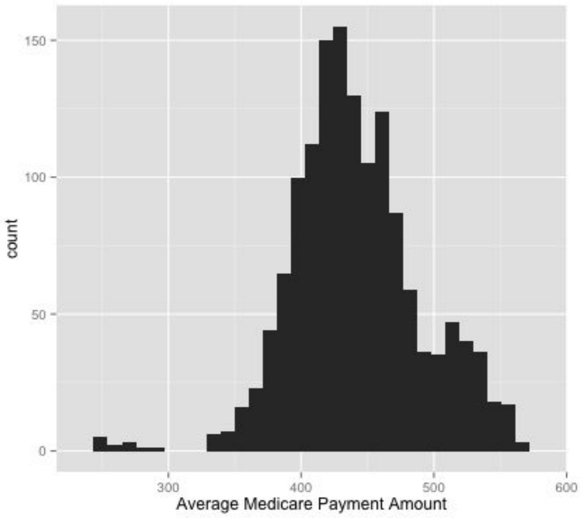


Figure 1: Histogram of average medicare payment amount

Our next step is to show what it looks like when each physician is located on the map with the colors indicating their average payment amount. Intuitively, for the same medical procedure, physicians from larger cities or economically better developed areas may obtain higher medicare payment, whereas physicians from smaller towns or rural areas may be assigned with relatively lower medicare payment.

Figure 2 shows such a distribution of physician's medicare payment amounts on a map. The color of the points indicates the value of the medicare payment amount; red means high payment amount, yellow means low payment amount.

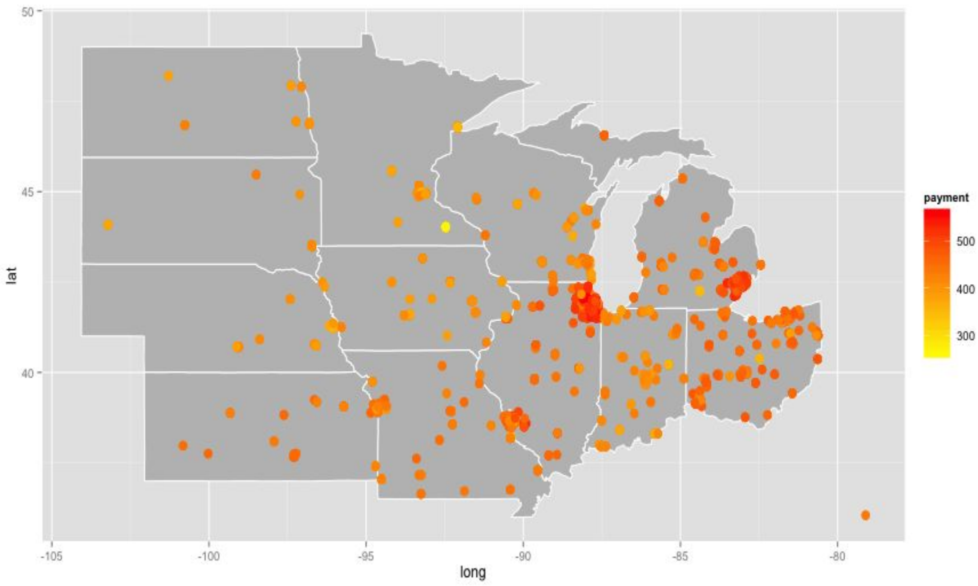


Figure 2: Plot showing the distribution of each physician's medicare payment amount on a map.

The following observations can be obtained from Figure 2. Firstly, large cities such as Chicago, Detroit, and St. Louis have more physicians. Secondly, physicians from large cities have relatively higher medicare payment amount, whereas in rural areas or small cities the color of the points are lighter, which indicates lower medicare payment amount.

Here we have some interesting reflections from this plot: (1) The density of physicians in certain area might correlate to the density of actual population of that area. (2) Large cities are usually economically better developed and the cost for living in those places is usually much higher, so physicians dealing with the same medical procedure must earn more money to support their living in big cities (3) Medicare payment only constitutes a part of the charge, so is by no means a potential indicator of the income of those physicians. (4) Physicians in large cities where population density is usually higher may have more patients than physicians from small cities, and that might be a reason for earning more money.

Nevertheless, merely from Figure 2, we’ve already seen the tendency that physicians from large cities obtain higher medicare payment amounts, although we should conduct further statistical inference to justify our observations.

3.2 Inference via Statistical Modeling with Categorical Variable "City"

To statistically justify our observations and the previous intuitive results, in this subsection, we will conduct statistical inference via certain modeling technique to obtain some further results.

Our main thesis from the previous subsection was that "physicians from large cities have relatively higher medicare payment amount". In order to make some inference from the data, we will use generalized linear models as the basic analytical tool in our study. Since our main focus is on "city", we will set city names as a categorical variable, and use this variable as predictor to the average medicare payment amount. Ideally, if the regression model fits well, we can make inference from the fitted model such that, without allowing intercept term, a larger coefficient would imply a higher average medicare payment amount in the corresponding city, whereas a smaller coefficient would imply a lower average medicare payment amount in that city.

In short, our regression model have the following form:

$$p = \sum_i c_i A_i + \varepsilon$$

where A_i's are indicators of cities, c_i's are coefficients for dummy variables, p is the medicare payment amount, ε is the error term, and the summation sign is over all the cities.

The following table gives the goodness-of-fit quantities of the fitted model (Table 1). We can see the model fits well. Also, by looking at the model diagnostic plots (due to limited space we won't present here), we can conclude that the model fits well and can provide valid inference.

Table 1: Model Information

Residual standard error (df = 1117)	23.1
Multiple R-squared	0.8214
Adjusted R-squared	0.772
p-value of F statistic	<2.2e-16

In Figure 3, we mark the 13 cities with the smallest coefficients on the map. The color of the points indicate the value of the corresponding coefficients. The darker the point, the greater the coefficients. In sum, except for Minneapolis, which is rather a large city, all the other cities are small cities. So the figure also supports our conclusion that we made in previous section.

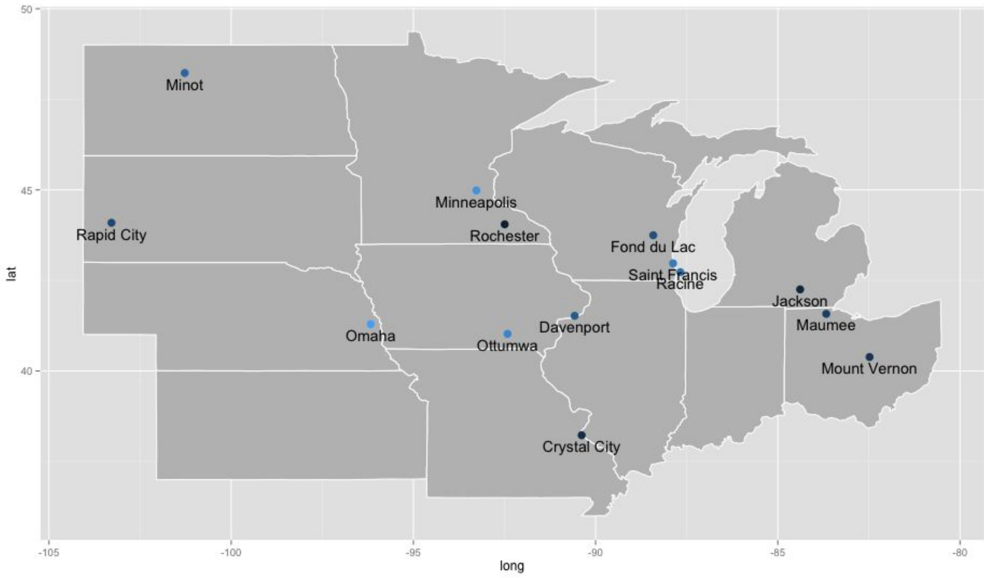


Figure 3: The plot showing the 15 cities whose regression coefficients are the smallest ones.

3.3 Relation to Submitted Charge Amount

From exploratory analysis to statistical modeling, we have already seen the tendency that physicians from large cities obtain higher medicare payment amount. Next, we will turn to submitted charge amount of those Midwest physicians on the same Medicare procedure.

3.3.1 Exploratory Visualization

In this subsection, we're going to make a plot showing the average submitted charge amount on the map (Figure 4). Similar to Figure 2, we still use colors from red to yellow to indicate the level of a quantity. The red points correspond to higher submitted charge amount and yellow points correspond to lower amount.

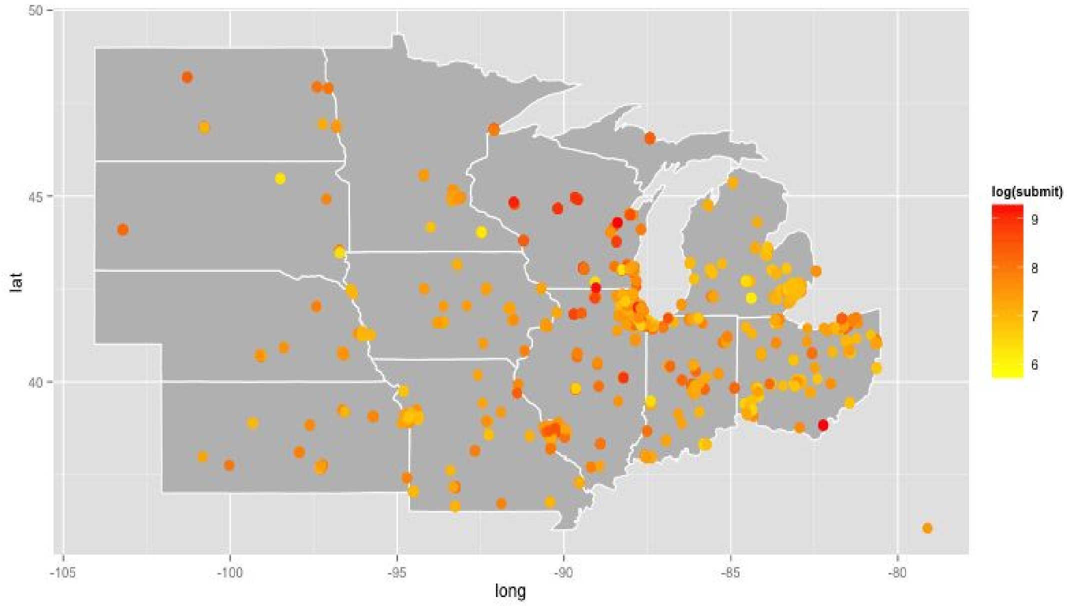


Figure 4: The plot showing the average submitted charge amount of physicians from Midwest who performed stent insertion procedure.

From the plot above, we can see some reddish points in Wisconsin and some in Illinois, while in general most of the places, including those large cities such as Chicago and Detroit, are filled with points with certain orange to yellow color. In other places where the density of physicians is relatively low, such as North Dakota and South Dakota, some physicians are submitting requiring relatively higher charge amount. Therefore, at least only by looking at the plot, we cannot conclude the same result as in previous part.

3.3.2 Statistical Inference via Categorical Variable "State"

From the previous analysis, we've seen that the submitted charge amount in Wisconsin seems higher than other states in Midwest. To give a justification through statistical modeling and inference, we still use GLM with "state" as the categorical predictor while average submitted charge amount as the response variable. The regression model informations are given in the following table. Note that our regression model doesn't include intercept, so the model coefficients are actually the mean values of submitted charge amount within in each state.

Table 2: Model Information

Residual standard error (df = 1415)	1236
Multiple R-squared	0.7533
Adjusted R-squared	0.7532
p-value of F statistic	<2.2e-16

The result shows the model fits well. The following plot shows the mean values of submitted charge amount within each state, arranged in decreasing order.

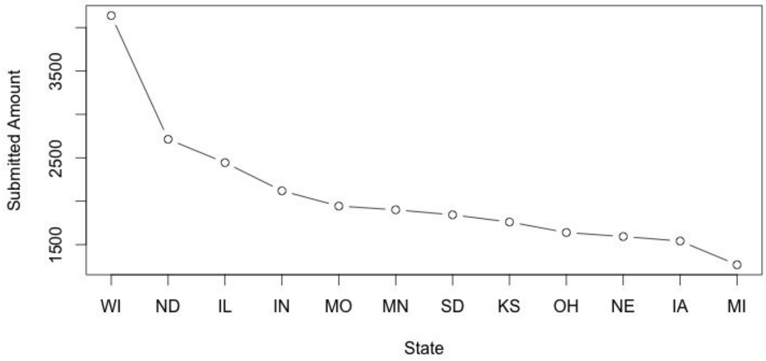


Figure 5: Ranking of the mean submitted charge amount for each state.

We can see that the average submitted charge amount in Wisconsin is approximately twice as the state with the second largest coefficient, North Dakota. However, to explain the underlying reason for this phenomena, we need some other information about the subject matter, though we won't make such explanation here in our paper.

To make a comparison, Figure 6 shows the Medicare payment amount in Wisconsin is only in the middle among all the Midwest states. Actually it can be seen that there’s no significant correlation between submitted charge amount and the actual Medicare payment amount. Note that Minnesota has the smallest coefficient among all the Midwest states. This result agrees with the previous regression model (payment against city) output, that even Minneapolis is a large city, its coefficient is still one of the smallest among all the cities Midwest.

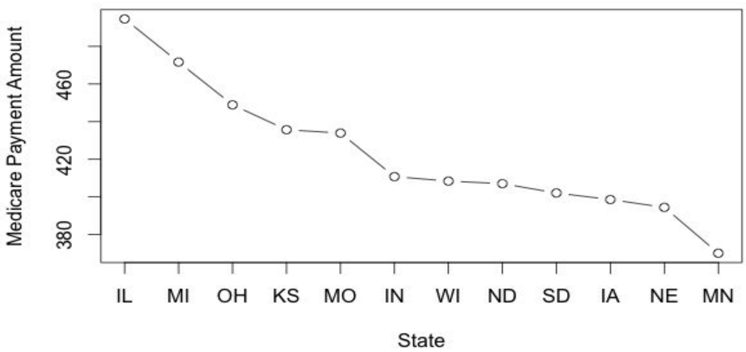


Figure 6: Ranking of the mean Medicare payment amount for each state.

3.3.3 Insights through Correlation Matrix

Now let's look at the correlation among the three quantities, namely, average medicare payment amount, average submitted charge amount and the difference between these two amounts.

Table 3: Correlation among average medicare payment amount (payment), average submitted charge amount (submitted) and their difference.

	Payment	Submitted	Difference
Payment	1.000	-0.067	-0.101
Submitted	-0.067	1.000	0.999
Difference	-0.101	0.999	1.000

From Table 3, we can see that the correlation between the average submitted charge amount and the difference rate is 0.999 (nearly 1), whereas the correlation between Medicare payment amount and submitted charge amount is -0.067 (almost 0). It seems that physicians who submitted higher charge amount will also face with higher difference rate between their submitted charge amount and the actual medicare payment amount. And Medicare may not take the submitted charge amount into consideration when they decide how much they should pay for the physicians.

To briefly conclude our findings in this section, we can summarize our points as following:

- Large cities such as Chicago, Detroit, and St. Louis have more physicians than small cities or rural area;
- Physicians from large cities have relatively higher medicare payment amount, whereas in rural areas or small cities they have lower medicare payment amount;
- There is no significant correspondence between the city size and the average submitted charge amount;
- The physicians in Wisconsin tend to submit more charge amount to Medicare than other Midwest states;
- There is no significant correlation between submitted charge amount and the actual Medicare payment amount.

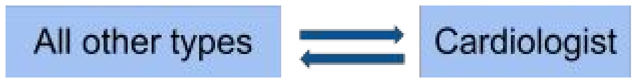
3.4 Network Analysis

3.4.1 Hypothesis

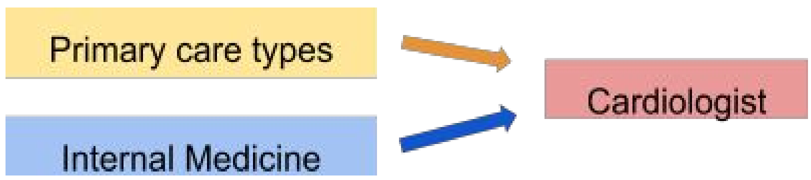
The hypothesis is that the average Medicare payment amount of stent insertion procedure for a physician in Midwest has some certain relationship with the physician referral network characteristics. We also want to see if the hierarchical structures of network would have different influences on the output. Thus, we'll capture different types of network structures including 2 kinds of ***speciality structures*** for referral networks and 3 kinds of ***regional structures*** for referral networks.

3.4.2 Network Structures and Graphs

For the speciality structures of network, the first one is the referral network between other provider types physicians and cardiologist:



We get the second network by restricting the specialities and make the network directed. It is the referral network from primary care types (Family Practice, Nurse Practitioner, Physician Assistant) and Internal Medicine physicians to cardiologists. This network is more sensible, since Primary Care and Internal Medicine are sorts of general physicians, and most referrals happening here are between these general physicians and other specialized physicians. As we know, the Primary Care is the day-to-day health care given by a health care provider. Typically, such a provider acts as the first contact and principal point of continuing care for patients within a healthcare system, and coordinates other specialist care that the patient may need. The referral structure is as below:



We can plot this kind of network in Wisconsin as an example (Figure 7):

Referral Network from Primary Care and Internal Medicine to Cardiologist in Wisconsin

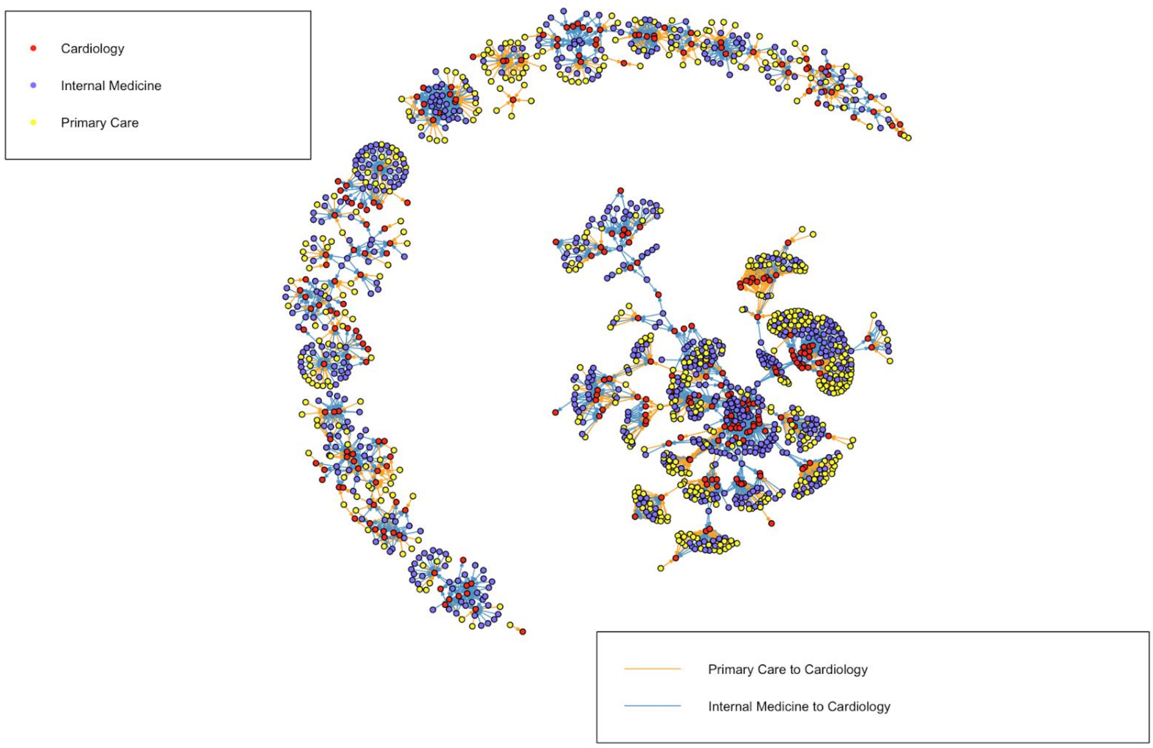


Figure 7: This directed graph represents the referral network from Primary Care and Internal Medicine physicians to Cardiology physicians in Wisconsin. The vertices in red are the Cardiologist, the purples are Internal Medicine physicians, and the yellows are Primary Care physicians. The blue edges with arrows represent the referrals from Internal Medicine physicians to Cardiologists, and the orange edges with arrows represent the referrals from Primary Care types physicians to Cardiologists.

For the regional structures of network, we try to learn three levels of network structures. To interpret it clearly, for each of the regional level network, we make a toy network on the map and one related example for the real network to illustrate the type of the network.

1. The nationwide physician referral network. This is the network in the nationwide level, where Midwest Cardiologists can send and receive patients in and out of Midwest area. Since the nationwide network is rather big and hard to plot, we just make a sketch plot for some of the Midwest Cardiologists who have performed stent surgery.

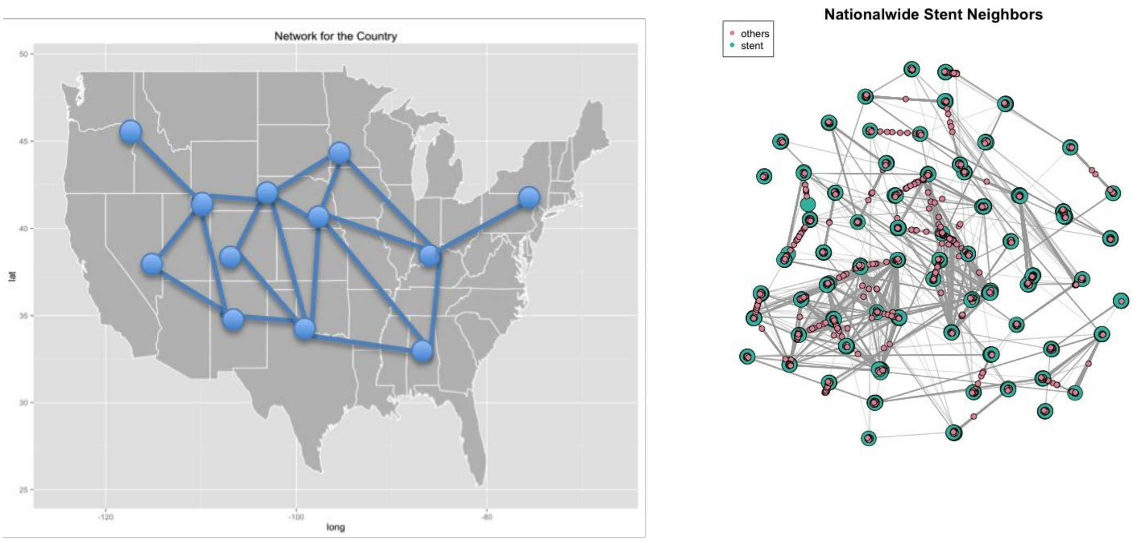


Figure 8: The left panel is the toy network to illustrate the nationwide level network. The right panel is the network showing the Midwest Cardiologists who'd performed stent surgery (the green vertices) and their neighbors. To simplify the plot, only the 100 Cardiologists with the highest degrees are plotted. The red vertices are physicians with other types.

2. The Midwest physician referral network. In this level, we narrow down the range of the network to Midwest. We plot the Midwest network and emphasize the cardiologist who'd performed stent surgery in big green vertices (Figure 9, right panel).

Note that the network graph is composed of several notable smaller components, which are obvious communities. The different colors of edges represent different relationships from the green vertices and the red vertices. The referrals between cardiologists who'd performed stent surgeries should be in black edges; however, they occur so rarely and could hardly be seen in the graph. Thus, the network within cardiologists who performed stent surgeries should be so sparse that is useless to investigate in. In fact, the cardiologist are morely likely to be connected with other types physicians than cardiologists themselves.

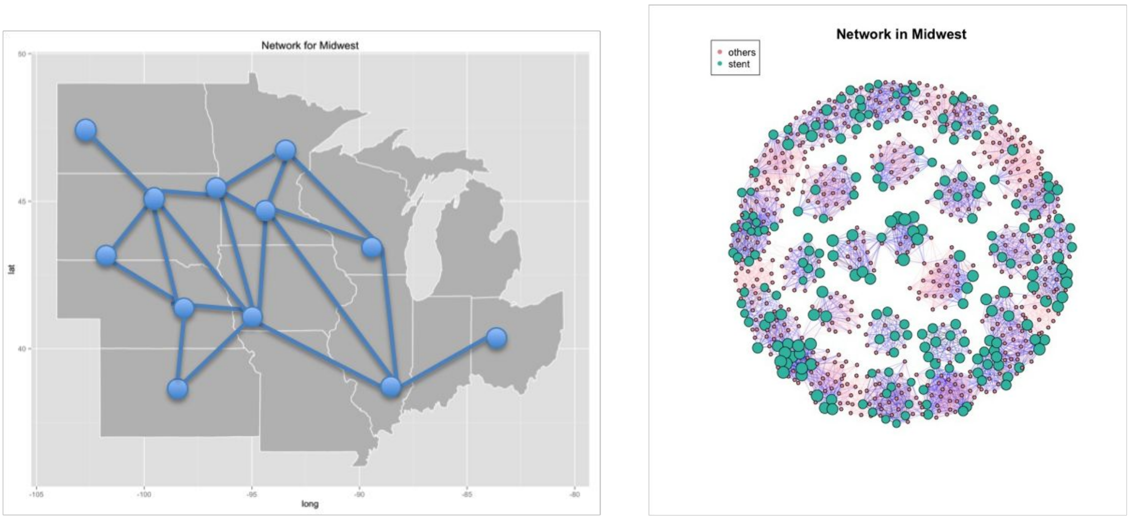


Figure 9: The left panel is the toy network to illustrate the Midwest level network. The right panel is the network between the Midwest physicians. The cardiologists who did stent surgery are the big green vertices, and the Midwest physicians with other types are small red vertices. To simplify the plot, we set the coreness to be greater than 12. The blue edges represent the referral relationship between the cardiologists and the other types physicians, while the pink edges represent the referral relationship between other types physicians.

3. The within-state physician referral network. This is the state-level network, where the referrals between physicians happen within a state in the Midwest area, as the left panel of Figure 10 describes. We also take Wisconsin network as an example to make a graph.

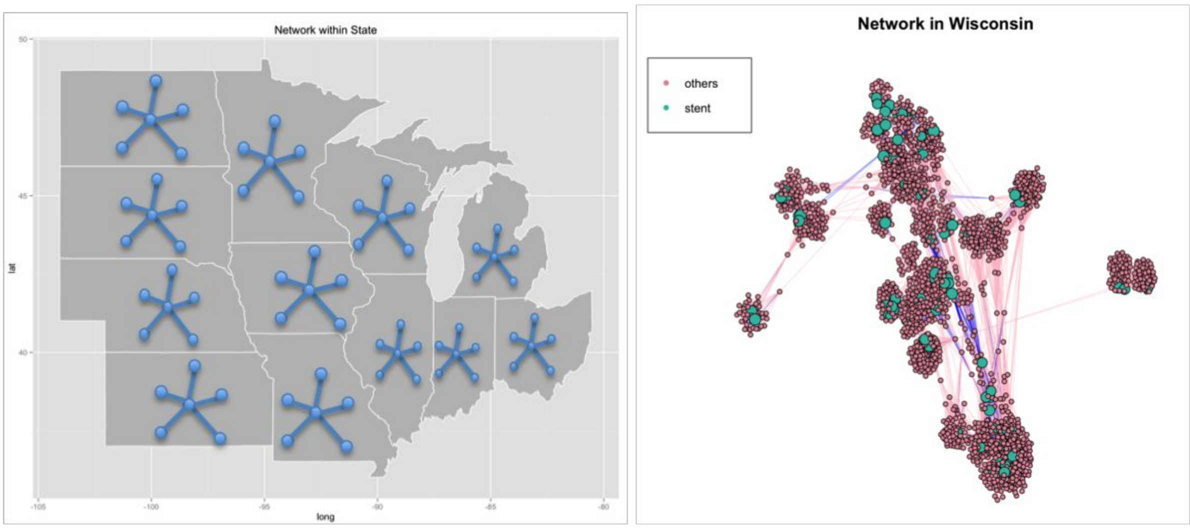


Figure 10: The left panel is the toy network to illustrate the state level network. The right panel is an example network between the physicians in Wisconsin. The cardiologists who did stent surgery are the big green vertices, and the physicians with other types are small red vertices. To simplify the plot, we set the coreness to be greater than 25. Again, the blue edges represent the referral relationship between the cardiologists and the other types physicians, while the pink edges represent the referral relationship between other types physicians.

3.4.3 Graph Characteristics and Variables Related to Networks

To study the relationship between the with the physician referral network and the average Medicare payment amount, several network characteristics are selected as predictors.

1. Degree d_v of a node v

Degree counts the number of edges in E incident upon v in a network graph G = (V,E). Degree captures connectedness of the physicians.

2. Eigenvector Centrality Score

To describe the importance of the physician in the network, we use eigenvector centrality score. Centrality describes node's position, representing how important or influential it is. The eigenvector-based centrality measure captures importance of nodes' friends. In our networks, it suggests that the physician's importance comes from being connected to other important physicians.

The mathematical definition of eigenvector centrality score is as below:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

where $\sum_{t \in M(v)} x_t$ is the neighbourhood of x_v, $\sum_{t \in G} a_{v,t} x_t$ is the adjacency matrix of the graph.

3. Referred Counts

In order to measure the referral counts from referring physicians to referred physicians, we also construct some measures for each physician, including the number of referrals, the number of unique beneficiaries, and the number of referrals within a same day.

3.4.4 Summary

In our model, there are six types of network included, and five network variables are calculated within each of the network. Thus, totally 30 network variables will be generated and analysed. The variable structures are as below:

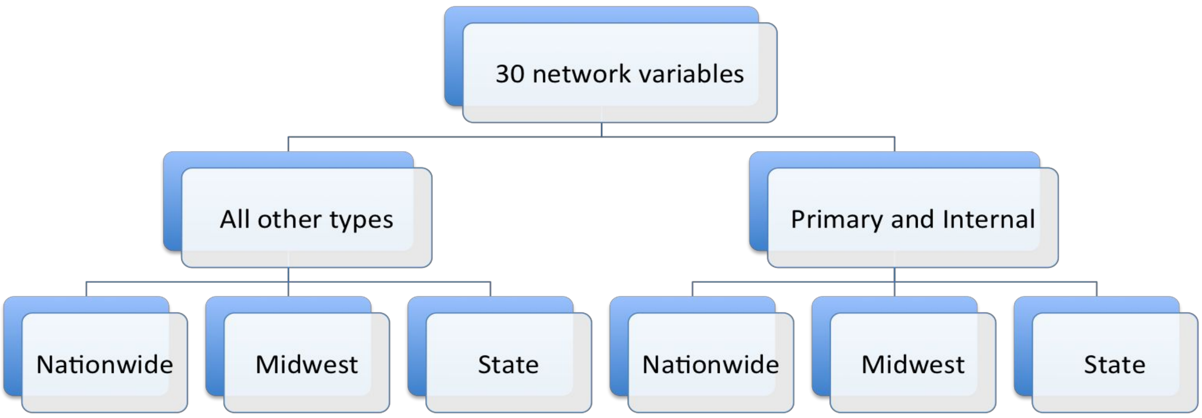


Figure 11: The frame represents the structure of the networks. It shows that there are 30 variables based on two hierarchies, where one is the speciality structures of network, and the other is the regional structures of network.

3.4.5 Inference via Statistical Modelling on Network Variables

First, we scale all of the graph variables, then regress the payment amount on all of the network variables, and perform stepwise model selection by BIC to build up the model.

Table 4: Regression results of linear model for network variables

Variable	Estimate	Pr(>|t|)	Significance
Intercept	445.499	< 2e-16	***
m_central_score	10.921	3.17e-15	***
state_central_score	-11.106	4.10e-13	***
sum_vis	-18.613	0.000277	***
sum_pat	51.741	< 2e-16	***
sum_vis_day	-14.104	0.000103	***
m_small_central_score	5.911	1.70e-05	***
small_sum_pat	-45.994	< 2e-16	***
small_sum_vis_day	13.043	0.000520	***
s_small_sum_vis	18.417	8.50e-05	***

Residual standard error: 40.92 on 1119 degrees of freedom
Multiple R-squared: 0.2043, Adjusted R-squared: 0.1979
F-statistic: 31.92 on 9 and 1119 DF, p-value: < 2.2e-16

The results can be summarized as follows:

- The eigenvector centrality measures are very significant for three kinds of networks: the one among all types physicians in Midwest area, the one within each state, and the one among specific types physicians in Midwest
- Referral Counts among three kinds of networks are all significant;
- No degree variables are significant in all of the networks.

However, since other variables are not controlled here, it is not solid to make conclusions till now. We’re going to consider other physician information in the next step, and include them as control variables in the final model.

3.5 Physician Information Factors

Here in this part, we're going to investigate the Medicare Provider Utilization and Payment Data[7] to see if the background information of a physician have an impact on the medicare payment of stent insertion.

3.5.1 Descriptions

In general, the 52 variables in the dataset can be categorized into three categories:

- Physician background, i.e. gender, number of services...
- Geographical information, i.e. the physician's state, or zipcode-linked information...
- Beneficiary characteristics, i.e. the age, sex, or disease of the patients…

In details, the variables related to Physician Background include:
Physician gender, number of HCPCS, number of services, number of unique beneficiaries (total / stent Procedure), number of female/male beneficiaries, number of beneficiaries with medicare only entitlement.

The variables related to Geographical information include:
Total number of establishment, population, zip-code median income.

The variables related to Beneficiary characteristics include:
Percents of beneficiaries identified with 16 kinds of disease, beneficiary age distribution, sex ratio, race ratio, average HCC risk score.

3.5.2 Linear Regression

To see if the physician information factors are significant, an OLS regression is performed, then selected by stepwise BIC. The detailed outputs are described as below:

First of all, the number of unique stent insertion beneficiaries for a physician and the number of HCPCS procedures a physician has performed are negatively related to the average Medicare payment amount for the stent insertion procedure, while the number of total services is positively related to the payment amount. Intuitively, within the stent procedure, the more a physician is specialized and personalized, the more he/she will get paid; while in general, the more active a physician is, the more payment he/she will get.

The percents of beneficiaries identified with certain kinds of diseases can also make a huge difference in the stent insertion payment. It's difficult for us to interpret the mechanism here due to the lack of medical knowledge, but it's reasonable to claim that the existence of some diseases may cause complications in heart disease and make it harder to perform a stent

[7] https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html

insertion surgery. Physicians dealing with harder surgeries are natural to get better Medicare payments.

As for the zip code-linked variables, the numbers of establishments is significantly positively related to the Medicare payment amount. Intuitively, the number of establishments measures the economic condition of a zipcode area. The more establishments an area has, the more “developed” the area would be, and hence the physicians there will tend to be highly paid by the Medicare.

Finally, it’s shown that the age distribution and the race ratio of the beneficiaries are significant, while the genders of the beneficiaries as well as the physicians are not important in this issue.

The adjusted R-squared of the stepwise regression is 0.4542, which suggests that these physician information variables are powerful in explaining the variation of Medicare payments, and hence should be valid to be used as control variables in the final model.

3.6 Final Modelling and Inference

In the final part, with all of the variables discussed above, we’re going to build up the final model and use it to do statistical inference for the Medicare payment of Stent Insertion procedure in the Midwest. The idea is basically as below:

- Multicollinearity is detected when we explored the datasets in the above sections and we’ll deal with it first.
- Based on the result of decision tree algorithm, we’re going to split the dataset into two parts to improve the performance of our model.
- Proper transformations would be made and the data will be standardized.
- Perform stepwise BIC to build up the final model.

Note that the shrinkage methods and nonparametric methods can not significantly reduce the mean squared error (MSE) as we’ve tried. More importantly, most of those methods are hard to interpret compared to linear models.

3.6.1 Deal with the Multicollinearity

In statistical inference, multicollinearity is an unexpected phenomenon in linear regressions which will make it hard to interpret the result. Our initial modelling suggests that the “raw” model with simple linear regression on all predicting variables has severe collinearity, hence model selection is necessary in this case.

There are several approaches for model selection. One can perform model selection algorithms such as stepwise regression, LASSO or PLS, or he can directly check the correlations between variables and manually remove some of the highly correlated variables.

Since interpretability is an important issue here, shrinkage methods such as LASSO and PLS are not preferable in this case. Instead, stepwise regression is performed here for its simpleness in interpretation. To make the stepwise method more stable, some of the highly correlated predictors (which can be deemed as redundants) are deleted before modelling.

3.6.2 Split the dataset Based on Physician's Centrality

After remedying the collinearity, we split the dataset into two parts, according to the value of a physician's eigenvector centrality score for the Midwest referral network. The splitting point is $centrality\ score = 2.6049 * 10^{-7}$, which is detected and determined by the regression tree algorithm and is almost the median of this variable. Note that the centrality score is an important characteristic of the referral network; the result actually verifies the effectiveness of the network analyses performed in our report.

The statistical reason why we split the dataset is that fitting linear models separately on the splitted datasets can significantly improve the prediction accuracy. From table 5, we can see that the LOOCV MSE for different models can generally be reduced by 5-10% after splitting.

Table 5: LOOCV MSE for different models

Model	Without Stepwise	Stepwise BIC
Full	717.55	664.18
Deleted1	729.40	703.35
Deleted2	729.16	718.26
Full + Split	707.47	611.61
Deleted1 + Split	669.19	634.88
Deleted2 + Split	662.61	650.76
Deleted2 + Split + Trans/Scale	643.84	**626.65**

Intuitively, there should be some essential difference between the behaviors of high-centrality physicians and low-centrality physicians, and that's the reason why the two-part model can improve our results by far. Further details of the models will be discussed in the next subsection.

3.6.3 Stepwise BIC, After transformation and Standardization

Before applying the stepwise regression, the predictors are properly transformed and standardized to make the coefficients more reasonable. Note that we didn't transform or scale our response variable, since it's already normally distributed and has a reasonable range of values.

Then the stepwise regression is performed based on the Bayesian Information Criteria (BIC). Statistically, BIC has better properties than AIC, while practically it's more conservative than AIC and would introduce less variables into the model. In a word, stepwise BIC gives us a simple but powerful model in this case, which is a huge advantage for making an inference.

The summary information of our final model is given in the following table:

Table 6: Summary of the Final Model

	Lower Midwest Centrality		Higher Midwest Centrality	
	Coefficient	P-Value	Coefficient	P-Value
Intercept	402.04	0.000	421.97	0.000
IL	55.81	0.000	87.38	0.000
KS	34.74	0.000		
MO	31.20	0.000		
OH	47.71	0.000	34.67	0.000
MI			43.16	0.000
Bene_Count	-6.35	0.000	-3.25	0.007
Bene_65-74_Ratio	2.72	0.004		
%Alzheimer/Dementia			8.43	0.000
%Depression			-11.35	0.000
%Diabetes			6.17	0.000
%Osteoporosis			6.18	0.000
Midwest Centrality			10.49	0.000
National Centrality			4.87	0.000

It suggests that for Medicare payment, situations are different among different states. It can also be seen that the Medicare payment is negatively related to the number of beneficiaries, which suggests that less-paid physicians are generally more popular.

The beneficiaries' age ratio is significant for the low-centrality physicians, while such a statement is not true for the high-centrality physicians. Note that the disease patterns make difference on the payment amount for the high-centrality physicians. The reason might be that high-centrality physicians tend to deal with more complicated stent insertion surgeries where the physical conditions of patients matter a lot.

As for the referral network characteristics, the variables *Midwest Centrality* and *National Centrality* are selected to be in the high-centrality model. For this part of model, more centralized physicians tend to get higher Medicare payments for the stent insertion procedure.

4. Discussions

In the last section, we'll make a discussion on the limitations of our study and the problems we ran into. It would be interesting for further researchers to investigate in these issues.

In terms of the referral network, the network characteristics we chose may not cover all of features we would like to study. Analysis on the multilevel networks are somehow demanding, since there are six types of network, some of the features would be redundant. The degree and referral counts in different networks are somehow correlated, which made the interpretation of the modeling results a bit difficult. We've tried several dimension reduction methods such as PCA for these variables but the result was not so satisfying. The question remains for the further research to come up with better statistics to grasp the network pattern.

Another potential extension of our research is to include more regional data into analysis. One suggestion is to make some adjustments to the Medicare payment amount with the regional price index, so as to control the different standards of living among places.

Reference

Agresti, A., & Kateri, M. (2011). *Categorical Data Analysis.* Springer, Berlin Heidelberg.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning.* Springer, Berlin: Springer series in statistics.

ED Kolaczyk, G Csrdi. (2014) *Statistical Analysis of Network Data with R*, Springer

Uddin, S., Hossain, L., Hamra, J., & Alam, A. (2013). A study of physician collaborations through social network and exponential random graph. *BMC health services research,* 13(1), 234.

Hollingsworth, J. M., Funk, R. J., Garrison, S. A., Owen-Smith, J., Kaufman, S. R., Landon, B. E., & Birkmeyer, J. D. (2015). Differences between physician social networks for cardiac surgery serving communities with high versus low proportions of black residents. *Medical care*, *53*(2), 160-167.

Mandl, K. D., Olson, K. L., Mines, D., Liu, C., & Tian, F. (2014). Provider collaboration: cohesion, constellations, and shared patients. *Journal of general internal medicine*, *29*(11), 1499-1505.

Barnett, M. L., Christakis, N. A., O'Malley, A. J., Onnela, J. P., Keating, N. L., & Landon, B. E. (2012). Physician patient-sharing networks and the cost and intensity of care in US hospitals. *Medical care*, *50*(2), 152.

Jackson, M. O. (2008). *Social and economic networks* (Vol. 3). Princeton: Princeton University Press.

Acknowledgement

This material is a coursework requested by the Social Network Analysis course at Department of Statistics, UW-Madison in 2015 Fall. We warmly appreciate Professor Karl Rohe for his instruction during the semester.