Paper 107-2009

Using SAS Enterprise Guide 4.1 to Reduce the Cost of Diabetic Outpatients in Medicare

Xiao Wang, University of Louisville, Louisville, KY

ABSTRACT

The purpose of this paper is to evaluate the main causes for costs for diabetic outpatients in Medicare. The data sets of outpatient revenue and beneficiary summary used in this study are provided by the Chronic Condition Data Warehouse (CCW) from the Centers for Medicare and Medicaid. In this paper, we used Summary Statistics, One-Way Frequency, the General Linear Model, the Generalized Linear Model and Kernel Density Estimation. We utilized Filter and Query as well as Random Sampling to preprocess the dataset. We examined patient diagnosis information. The results of Summary Statistics show that the total charges are reduced considerably for payment. Kernel Density Estimation shows that for both male outpatients and female outpatients, the distributions of total charges follow a Gamma distribution. One-way frequency counts show the top 20 reasons for total charges and we used these 20 diagnoses to define 0-1 indicator functions. Both the outputs of the General Linear Model and the Generalized Linear Model show that diagnosis is significant to the total charges. After our study, we concluded that the top 20 treatments are related to the total charges and we could decrease the cost of the outpatient treatment for patients with diabetes through increased blood glucose monitoring. Race is also related; Caucasians should pay more attention to diabetes prevention than others.

INTRODUCTION

Diabetes is a chronic disease that can lead to many complications such as heart disease and kidney failure. The treatments and services for diabetic outpatients not only contain blood glucose monitoring and the blood count necessary for the diabetic patients, but also cover Electrocardiogram or routine EGG, unlisted dialysis procedure needed by persons with heart disease or end stage kidney disease. For Medicare, the dataset uses HCPCS codes to explain the claims for payments. There are two levels of HCPCS codes; Level I is comprised of CPT (Current Procedural Terminology) and Level II is a standardized coding system that is used primarily to identify products, supplies, and services not included in the CPT. Although Medicare covers inpatients and outpatients, we just consider the latter for this study.

Blood glucose monitoring is very important since checking the blood glucose level can help the patients take better care of their diabetes. The Diabetes Control and Complications Trial funded by the National Institutes of Health reported in 1993 that intensive glucose control prevents or delays the eye, nerve and kidney complications of type I diabetes (as cited in Glucose Control Cuts Risk of Heart Disease in Type 1 Diabetes, 2005); and DCCT/EDIC study (2005) illustrated that intensive glucose control lowers the risk of heart disease and stroke by about 50 percent in people with type I diabetes.

The aim of this paper is to examine ways to decrease payments from the Medicare program for treatment of diabetes; it is necessary to learn the treatments accounting for most of the total charges and their relationship to costs. After our study, we found that there exist linear relationships between the 20 most frequent treatments and total charges.

METHOD

The analysis used the datasets entitled outpatient_revenue_center as well as beneficiary _summary, both of which are from the Chronic Condition Data Warehouse. The revenue data contain 12,158,258 claim entries and 10,000 of them were randomly selected to perform the Generalized Linear Model. Beneficiary data cover 358,709 observations and only 10,000 of them were used in this study. We basically used the following variables:

1	REV_CNTR_RATE_AMT	Charges relating to unit cost associated with the revenue center code
2	REV_CNTR_CCINSRNC_WGE_ADJSTD_C	Revenue Center Coinsurance/Wage Adjusted Coinsurance Amount
3	REV_CNTR_PRVDR_PMT_AMT	the amount paid to the provider for the services reported
4	REV_CNTR_PMT_AMT_AMT	the line item Medicare payment amount for the specific revenue center.
5	REV_CNTR_TOT_CHRG_AMT	Revenue Center Total Charge Amount
6	HCPCS_CD	Revenue Center HCFA Common Procedure Coding System
7	BENE_SEX_IDENT_CD	Sex
8	BENE_RACE__CD	Beneficiary Race Code

Since the purpose of this paper is to examine ways to decrease the cost of outpatients with diabetes, the first step is to sort out all the columns related to cost in the revenue data set. We used Data->Filter and Query->Select buttons sequentially and generate a new data set. Figure1 shows the results as listed in Enterprise Guide 4.1.

Figure 1:Query_for_outpatient_revnue_cen (First 20 rows out of 12158258)

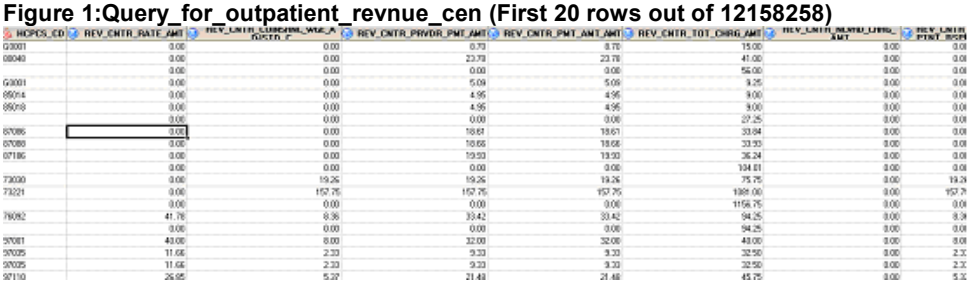


Table 1: Summary of REV Charges:

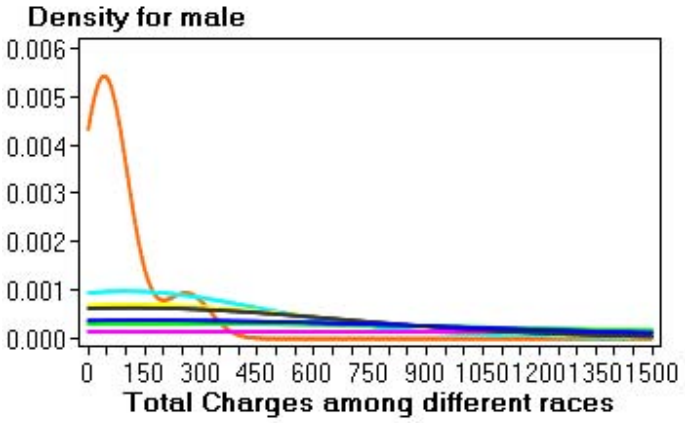
Variable	Mean	Std Dev	Minimum	Maximum	N	Median
Revenue Center Rate Amount	11.6548	46.9767	0	26624.00	12158258	3.00
Revenue Coinsurance/Wage Adjusted Coinsurance	13.8942	55.6397	0	23713.87	12158258	0
Revenue Center Provider Payment Amount	41.7001	200.8778	-2018.35	69700.60	12158258	8.32
Revenue Center Payment Amount Amount	42.1139	202.7057	0	69700.60	12158258	8.38
Revenue Center Total Charge Amount	399.9745	1767.31	0	237840.01	12158258	68.00

The results in Table 1 show that the average of total charges is almost 400 dollars, but the average payment is only 41.7 dollars. The differences are in the other payments such as coinsurance and deductions from the total charges. Therefore, the total charges are reduced considerably for Medicare reimbursement.

The following study is mainly concerned about the total charges. We need to preprocess the data before any analysis. We use Filter and Query->Add table to join columns from revenue and beneficiary data sets to generate another new data set containing beneficiary ID, HCPCS codes, total charges and so on. Next, we used Kernel Density Estimation for total charges. Here, we used the method of SNR (Simple normal reference) and set the upper bound at 1500 dollars and the lower bound to 0 dollars for the total charges. We grouped the total charges by race and sex. The SAS code is as follows:

```
proc sort data=sasuser.ransample out=sasuser.sransam;
by bene_race_cd  bene_sex_ident_cd ;
proc kde  data=sasuser.sransam;
univar rev_cntr_tot_chrg_amt /
gridl=0 gridu=1500
method=snr  out=sasuser.kdechar;
bY bene_race_cd bene_sex_ident_cd ;
run;
```

Figure2: KDE of Total Charges among Different Races:



Beneficiary race code:
0: Unknown 1: White 2: Black 3: Other 4: Asian 5: Hispanic
6: North American Native

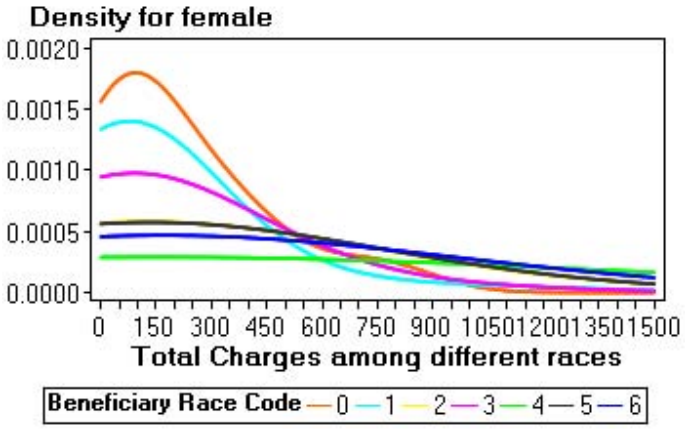


Figure 2 visualizes the distributions of the total charges. In figure 2, densities of total charges for males and females follow gamma distributions and all of them approach zero after 1500 dollars. The total charges of the whites are higher than the ones of the other known races. There exist differences between the males and the females; the densities arrive at the peak when the total charges reach 50 dollars for the males and 100 dollars for the females. Before 500 dollars, the densities for male are much higher than ones for female, but after 1000 dollars, the former ones become lower than the latter ones. Displays show that in most cases, the expenditures of the males with diabetes are less than 500 dollars, while the cost of the females with diabetes is between 0 and 1000 dollars.
After showing the distribution of the total charges, we wanted to find the top 20 reasons for the total charges; HCPCS codes just explain why the providers assign the charges. We used Describe->One-way frequencies and chose HCPCS_CD (HCPC codes) as the analysis variables and chose descending frequencies in the order of data output. We also chose to include frequencies and cumulative percentages to get the results in table 2:

Table 2 One-way Frequencies of HCPS codes: (Top 20 HCPCS codes)

Center HCFA Common Procedure Coding System				
HCPCS_CD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
82962	629284	6.78	629284	6.78
G0001	586056	6.31	1215340	13.09
97110	434128	4.68	1649468	17.77
85025	308946	3.33	1958414	21.10
85610	260886	2.81	2219300	23.91
80048	217654	2.34	2436954	26.25
80053	206596	2.23	2643550	28.48
83036	186689	2.01	2830239	30.49
90999	182848	1.97	3013087	32.46
97530	171227	1.84	3184314	34.30
Q4055	153939	1.66	3338253	35.96
80061	149025	1.61	3487278	37.57
97116	136731	1.47	3624009	39.04
97112	117931	1.27	3741940	40.31
99212	112990	1.22	3854930	41.53
93005	102165	1.10	3957095	42.63
99213	101873	1.10	4058968	43.72
A4657	91004	0.98	4149972	44.70
84443	87193	0.94	4237165	45.64
71020	82548	0.89	4319713	46.53

Table 3: Explanation for HCPCS codes:

1	82962	Glucose, blood by glucose monitoring device(s) cleared by the FDA specifically for home use
2	G0001	Routine venipuncture for collection of specimen(s)
3	97110	Therapeutic procedure，therapeutic exercises to develop strength and endurance, range of motion and flexibility
4	85025	Blood count; complete (CBC), automated (Hgb, Hct, RBC, WBC and platelet count) and automated differential WBC count
5	85610	Prothrombin time;
6	80048	Basic metabolic panel
7	80053	Comprehensive metabolic panel
8	83036	Hemoglobin; glycosylated (A1C)
9	90999	Unlisted dialysis procedure, inpatient or outpatient
10	97530	Therapeutic activities, direct (one-on-one) patient contact by the provider (use of dynamic activities to improve functional performance),
11	Q4055	Injection,
12	80061	Lipid panel
13	97116	Therapeutic procedures ;gait training (includes stair climbing)
14	97112	Therapeutic procedure,； neuromuscular reeducation of movement, balance, coordination, kinesthetic sense, posture, and/or proprioception for sitting and/or standing activities
15	99212	Office or other outpatient visit for the evaluation and management of an established patient; Physicians typically spend 10 minutes face-to-face with the patient and/or family.
16	93005	Electrocardiogram, routine ECG with at least 12 leads; tracing only, without interpretation and report
17	99213	Office or other outpatient visit for the evaluation and management of an established patient physicians typically spend 15 minutes face-to-face with the patient and/or family.
18	A4657	Syringe, with or without needle, each
19	84443	Thyroid stimulating hormone (TSH)
20	71020	Radiologic examination, chest, two views, frontal and lateral;

The extracted 20 HCPCS codes are recorded as binary 0-1 indicator functions in 20 columns with SAS coding. We created 0-1 indicator functions using the following code:

```
data sasuser.sortedcodes;
set Sasuser.sorteddatawithcode;
if (HCPCS_CD   EQ: '82962')
then  CODE82962=1;
else CODE82962=0;
if(HCPCS_CD   EQ: 'G0001')
then CODEG0001=1;
else CODEG0001=0;
...
```

We then used Analyze->ANOVA->Linear Model with total charges (REV_CNTR_TOT_CHRG_AMT) as the dependent variables; the newly-generated 20 indicator variables were used as the class variables in the model. While identifying the model, we defined the 20 class variables as the main effects and we chose Model III because the result is invariant of the order of input variables. Since we used more than one classification variable, we included least squares tests in post-hoc tests. The results are given below:

Table4：Overall model results:

R-	Coeff	Root MSE	REV_CNTR_TOT_CHRG_AMT Mean
0.033	434.381	1737.414	399.9746

Source	DF	Sum of	Mean Square	F Value	Pr > F
Model	20	1.2740058E12	63700290153	21102.5	<.0001
Error	1.22E7	3.6700937E13	3018606.8466		
Corrected	1.22E7	3.7974943E13			

The r-square value is 3.3%, which means that 3% of the variability of the total charges can be explained by the above 20 factors. Although this is a small number, it gives us an idea of what is significant.

Table5: Type III Sum of Squares:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Blood glucose monitoring	1	144951012755	144951012755	48019.2	<.000
Routine venipuncture	1	135540500264	135540500264	44901.7	<.000
Therapeutic exercises to develop strength and	1	76340503128	76340503128	25290.0	<.000
Blood count	1	60850533348	60850533348	20158.5	<.000
Prothrombin time	1	57263486670	57263486670	18970.2	<.000
Basic metabolic panel	1	37054519078	37054519078	12275.4	<.000
Comprehensive metabolic panel	1	30419948211	30419948211	10077.5	<.000
Hemoglobin; glycosylated	1	37500526373	37500526373	12423.1	<.000
Unlisted dialysis procedure,	1	417296629150	417296629150	138241	<.000
Therapeutic activities, direct patient contact by the	1	34114806835	34114806835	11301.5	<.000
Injection	1	169829929901	169829929901	56261.0	<.000
Lipid panel	1	26676907909	26676907909	8837.49	<.000
Therapeutic procedure, gait training	1	29359815777	29359815777	9726.28	<.000
Neuromuscular reeducation of movement, balance	1	24582913341	24582913341	8143.79	<.000
office or other outpatient visit, typically 10 minutes	1	20034618172	20034618172	6637.04	<.000
Electrocardiogram, (EGG)	1	13816375074	13816375074	4577.07	<.000
Office or other outpatient visit, typically15minutes	1	17996085232	17996085232	5961.72	<.000
Syringe,	1	20096703187	20096703187	6657.61	<.000
Thyroid stimulating hormone	1	15385539003	15385539003	5096.90	<.000
Radiologic examination	1	8767439243.1	8767439243.1	2904.47	<.000

Table 5 shows that the 20 variables are significant to the total charges. However, since the distributions of the total charges are gamma distributions, we should also consider the generalized linear model with a gamma link function. The advantage of the model is that it can extend the analysis to predict the mean of variables that are not reasonably assumed to be normally distributed, but it takes considerable time to get the results. Therefore, before we did regression, we used Random Sample to reduce the sample size to 10,000. Finally, we used Analyze->Regression->Generalized linear model and we chose the total charges as the dependent variables, and the newly-generated variables as the classification variables. Then we chose the Gamma function with the log link function; we choose Type I analysis. The results are given in the table6.

Table 6: Criteria for assessing goodness of fit:

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	9964	22374.1763	2.2455
Scaled Deviance	9964	12474.8191	1.2520
Pearson Chi-Square	9964	130667.2143	13.1139
Scaled Pearson X2	9964	72854.0725	7.3117
Log Likelihood		-63303.4243	

In table 6, the value of deviance divided by degree is 2.25 .The value shows the adequacy of this model, which means that the generalized linear model with a gamma distribution function fits the data reasonably well. In table 7, all the p values are smaller than 0.0001, indicating that all 20 variables are statistically significant. The chi-square value of 1070.13 for blood glucose monitoring represents twice the difference in log likelihoods between fitting a model with only an intercept and a model with an intercept and blood glucose monitoring. Similarly, every chi-square value for each variable represents the differences in log likelihoods between successive models. The output shows that the chi-square values for venipuncture and blood glucose monitoring are the highest among all the chi-square values; therefore, we could conclude that blood glucose monitoring and venipuncture have more important effects on the total charges than any other treatments do.

Table 7: Type I analysis:

LR Statistics For Type 1 Analysis				
Source	2*LogLikelihood	DF	Chi-Square	Pr > ChiSq
Intercept	-131538.88			
Blood glucose monitoring	-130468.75	1	1070.13	<.0001
Routine venipuncture	-129441.88	1	1026.87	<.0001
Therapeutic exercises	-129197.95	1	243.93	<.0001
Blood count	-128915.08	1	282.87	<.0001
Prothrombin time	-128568.73	1	346.35	<.0001
Basic metabolic panel	-128424.85	1	143.88	<.0001
Comprehensive metabolic panel	-128315.64	1	109.21	<.0001
Hemoglobin; glycosylated	-128081.87	1	233.78	<.0001
Unlisted dialysis procedure	-127823.38	1	258.49	<.0001
Therapeutic activities, direct patients contact by the	-127633.20	1	190.18	<.0001
Injection	-127565.57	1	67.63	<.0001
Lipid panel	-127459.23	1	106.34	<.0001
Therapeutic procedure, gait training	-127263.69	1	195.55	<.0001
Neuromuscular reeducation of movement, balance	-127112.20	1	151.48	<.0001
Office or other outpatient visit, typically 10 minutes	-127004.16	1	108.04	<.0001
Electrocardiogram, EGG	-126952.54	1	51.62	<.0001
Office or other outpatient visit, typically 15 minutes	-126855.49	1	97.05	<.0001
Syringe	-126711.41	1	144.09	<.0001
Thyroid stimulating hormone	-126640.03	1	71.37	<.0001
Radiologic examination	-126606.85	1	33.18	<.0001

Table 8: Pearson Correlation:

Pearson Correlation Coefficients, N = 12158258	
	Rev total charges
Blood glucose monitoring	-0.05110<.0001
Routinevenipunctureforcollectionspeci	-0.04919<.0001

In Table 8, the negative correlation coefficients between the total charges and blood glucose monitoring, routine venipuncture indicate that if we increase the monitoring and routine venipuncture, we can decrease the total charges.

CONCLUSION

After the study, we could draw the conclusion that in order to decrease the cost of outpatients with diabetes in Medicare, we should increase blood glucose monitoring and routine venipuncture of the outpatients; and we also should emphasize the diabetes prevention of the whites.

REFERENCES

1. Cerrito, P. (2007) *Exploratory Data Analysis: An Introduction to Data Analysis Using SAS*. Lulu.com.
2. Cerrito, P. (2008) *Student Papers in Introductory Statistics for Mathematics Majors*. Lulu.com.
3.The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications (DCCT/EDIC) Study Research Group.
(Dec.22, 2005).” Intensive Diabetes Treatment and Cardiovascular Disease in Patients with Type 1 Diabete”, *The New England Journal of Medicine*, Volume 353:2643-2653. http://content.nejm.org/cgi/content/full/353/25/2643
.Dec.12th, 2008
4.University of Iowa Health Science Relations. Dec. 22, 2005, “Glucose Control Cuts Risk of Heart Disease in Type 1 Diabetes”, http://www.news-releases.uiowa.edu/2005/december/122205glucose_control.html.

ACKNOWLEDGMENTS

We appreciate the use of the CCW Data Warehouse from the Centers for Disease Control. Thanks to Dr. Patricia Cerrito, for aiding in SAS code and interpretation of some results in this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Xiao Wang

Ph.D. student in Applied Mathematics

Department of Mathematics

University of Louisville

Louisville, KY 40292

Work Phone: 502-852-6022

E-mail: x0wang16@louisville.edu.