

Medicare Cost Estimation for Heart Disease

Fariba Nowrouzi Kashan, University of Louisville, Louisville, Kentucky

ABSTRACT

The objective of this paper is to examine information about Medicare costs for the treatment of heart disease by using information from a Medpar database (the public-reporting billing forms used by Medicare). The database contains 4862 patient records related to treatment of the heart. The purpose is to examine the relationship between reimbursements, treatments, length of stay, and patient severity. Kernel density estimation demonstrates that the distributions are exponential or come from an exponential family. Because patient severity depends upon the uniformity of data entry, all assumptions of linear and logistic models are violated, putting into question the validity of hospital quality rankings. Reimbursements are also dependent upon the accuracy of data entry. The problems with validity will be demonstrated using SAS data mining techniques.

INTRODUCTION

The database contains 4862 patient records related to treatment of the heart and collected from two hospitals. The patients are classified in 32 different DRG groups. According to the DRG guidebook, “A DRG is one of 499 currently valid groups that classify patients into clinically cohesive groups that demonstrate similar consumption of hospital resources and length-of-stay patterns” (1). The numbers of observed data in some of the DRG groups are very small. In order to get significant results about the relationship between reimbursements, treatments, length of stay, and patient severity, we will omit some DRG groups with small numbers of data. It is difficult to define a measure of patient severity that does not depend upon patient outcomes. However, in order to compare outcomes, severity must be defined in another way. If patient severity rank is labeled for each patient according their feeling, it is hard to compare them as their outcomes. Therefore it would be better that have severity rank related to length of stay or mortality rate or other outcomes in order to define an objective standard.

Chart 1 illustrates that the distribution of patients in the two hospitals is not balanced; the ratio of the number of patients is almost 37 to 10 in the two hospitals.

Chart 1

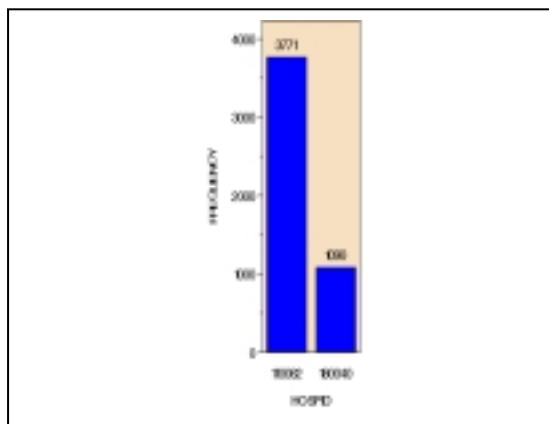


Chart 2

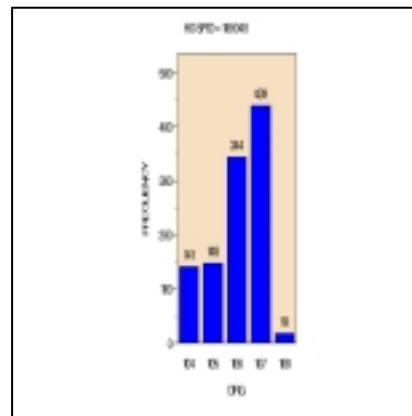
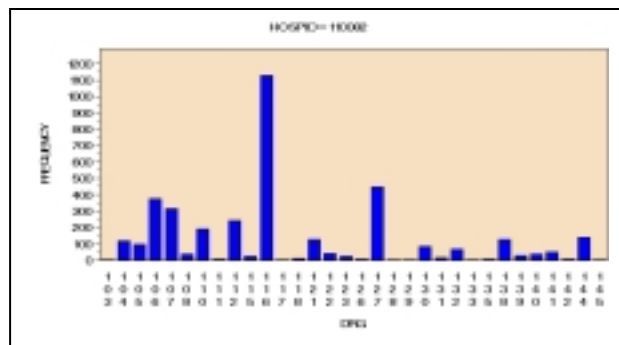


Chart 3
Charts 2 and 3
confirm that the
hospitals

ID = 180040, with 1090 patients record lists only five DRG groups, but the other hospital with 3771 patient records, gives treatment to all 32 DRG groups



Definition: Correlation between two variables illustrates the strength of the linear relationship between the variables. Pearson correlation between two variables x and y is based on the following formula

$$\rho = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the means of X and Y respectively. Calculating the pairwise comparisons of the Pearson correlations between the amount of reimbursement (AMTREIMB), total charge (TOTCHG), DRG, and length of stay (LOS) in both hospitals shows that there is no significant difference between the two hospitals. Therefore, we display Pearson correlation coefficients related to the whole database in table 1. The strongest correlation, 0.8539, is related to the amount of reimbursement and total charges. This means that there is a stronger linear relationship between these two variables compared to other variables. Graph 4 shows the correlation between reimbursement and total charge.

Table 1:

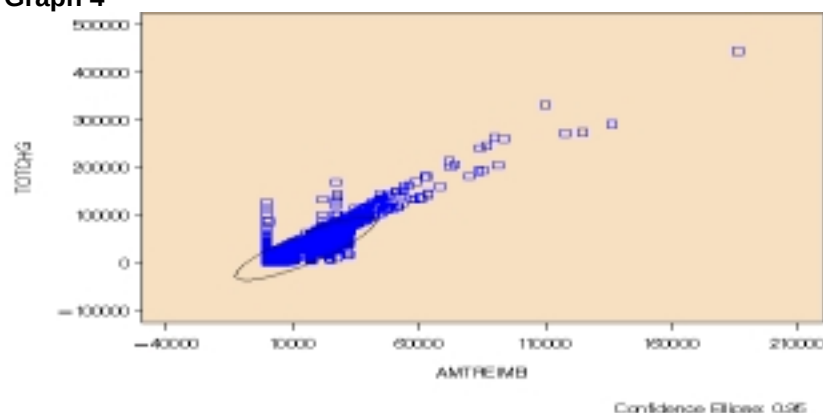
Pearson Correlation Coefficients, N = 4861

Prob > |r| under H0: Rho=0

	AMTREIMB	TOTCHG	LOS	DRG
AMTREIMB	1.00000	0.85391	0.65634	-0.70484
AMTREIMB		<.0001	<.0001	<.0001
TOTCHG	0.85391	1.00000	0.82072	-0.56360
TOTCHG	<.0001		<.0001	<.0001
LOS	0.65634	0.82072	1.00000	-0.29293
LOS	<.0001	<.0001		<.0001
DRG	-0.70484	-0.56360	-0.29293	1.00000
DRG	<.0001	<.0001	<.0001	

P values for all coefficients are less than .0001; that is, all the coefficients are highly statistically significant.

Graph 4

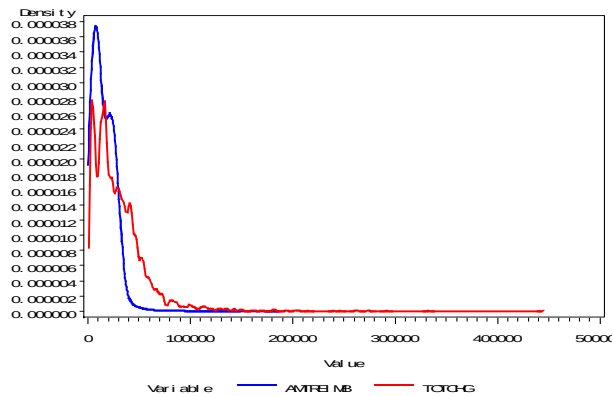


The graph 4 shows a strong positive association between reimbursement and total charge; that is, smaller values of reimbursement are associated with smaller values of total charge and larger values of reimbursement are associated with larger values of total charge.

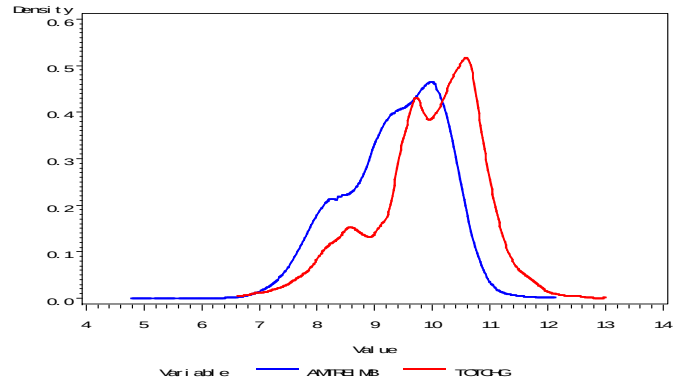
KERNEL DENSITY

A density function is the derivative of a cumulative distribution function. A kernel density is a non-parametric function; that is, the data are not assumed to be from a known parametric family distribution. The data determine the shape of the distribution. In order to compare the amount of reimbursement and total charge, we use the kernel density in graph 5. The graph clearly shows that the reimbursement and total charge distributions are not normal. By transforming the reimbursement and total charge using the natural logarithm function, we get a smoother kernel density function as shown in the graph 6.

Graph 5



Graph 6



Graph 6 illustrates that the two graphs are almost identical, but the reimbursement density is shifted to the left of the total charge with the difference of about 15% to 20%. This result indicates that hospitals are losing about that amount when charges are based upon average costs.

CLEANING THE DATA

The project database is associated with two hospitals. All data from one of the hospitals are related to only the five DRG codes, (104, 105, 106, 107, 108). Table 2 shows the DRG description (1). To find whether the different hospitals have different reimbursements, only the data related to the five common DRGs in the two hospitals will be considered. The filtering process has been done using SAS software. To define patient severity ranks without relying on the outcome variables requires the use of SAS Text Miner. The secondary diagnoses, called ICD-9 codes provide information as to the patients overall, general condition. ICD-9 codes typically are 5 digits in length with the first three digits providing a general category of problem. For example, the code 25000 represents diabetes. Although the codes are numeric, they represent nominal data, but data that are linked through the basic 3-digit “stem” of the code. To link all patient codes to the same patient, the codes are combined. SAS Text Miner is used to define the clusters that represent patient severity ranks. The ranks can be validated through comparisons with patient outcomes such as mortality and length of stay. Since the number of ranks is arbitrary, and defined through a clustering process, they are random effects.

Table 2

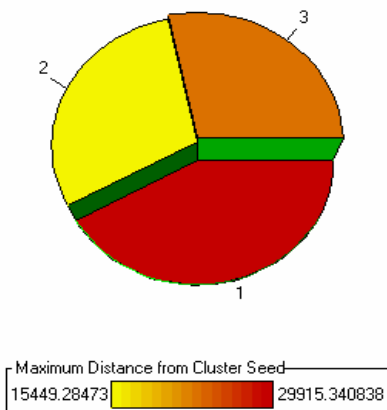
DRG	Description
104	Cardiac Valve Procedures and other Major Cardiothoracic Procedures with Cardiac Catheterization
105	Cardiac Valve Procedures and other Major Cardiothoracic Procedures without Cardiac Catheterization
106	coronary bypass with(PTCA
107	coronary bypass with cardiac cathetrization
108	other cardiothoracic procedures

To cluster the data, Enterprise Miner was used, and the relation of patient severity rank and some other variables has been studied. Data clustering has been performed using different clustering criteria. Results show that the common importance variables used for grouping the data are the natural logarithm of reimbursement (AMNTREIMB_log), total charge (TOTCHG), and DRGPRICE. The other variables that appear in different clustering category have a small importance level role in each cluster. Here we discuss the result of 3-clustering, where the least square criterion was used.

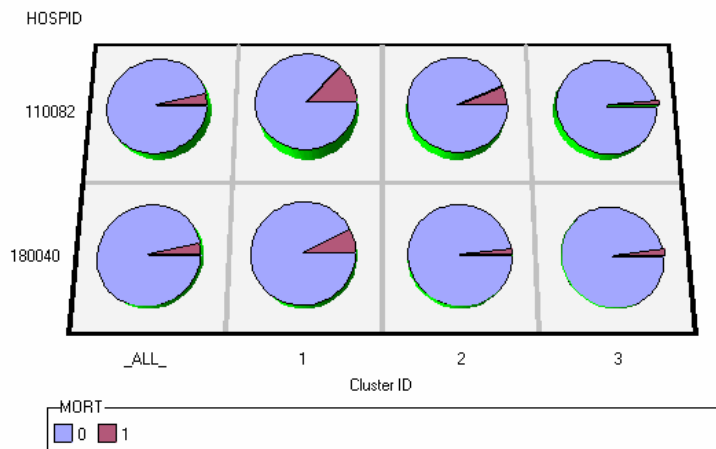
In graph (7), each Slice is standard deviation, Height is frequency, and Color is radius. The standard deviation is almost the same for all three clusters. The difference radii in the three clusters demonstrate that $\text{Max } |x_i - x_2| < \text{Max } |x_j - x_3| < \text{Max } |x_k - x_1|$ where x_i , x_j , and x_k are respectively any observation from groups 2, 3, and 1. Also, x_1, x_2, x_3 are the seeds of the groups 1, 2, and 3. We start with some profiles related to the three clusters to visualize the relation between some variables. Graph 8 shows the portion of mortality in each cluster, where height represents reimbursement.

Graph (7)

Clusters for SASUSER.MEDPAR_CLEARED_EDIT



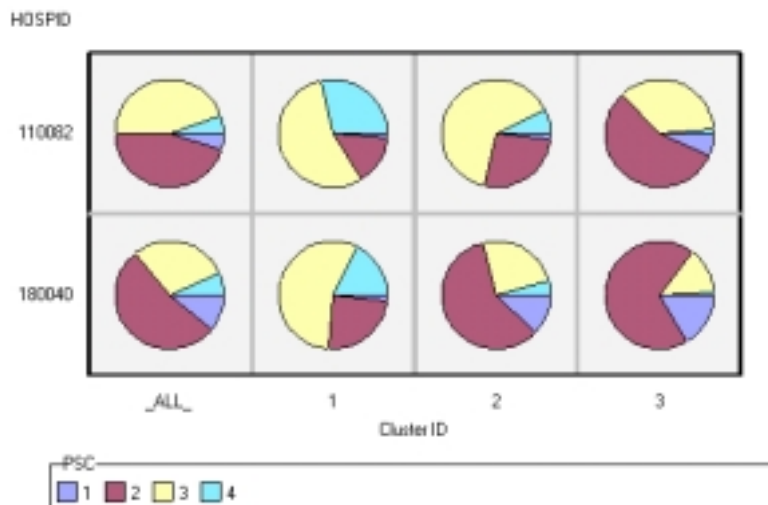
Graph 8



According to the first column of the graph (8), there are no significant differences between mortality rates in the two hospitals. In both hospitals, the highest mortality rate takes place in the first cluster. Also, comparing the heights in the rows demonstrates that the reimbursement amount in hospital with ID 110082 is higher than the other one.

The Graph (9) shows the relation of patient severity in the three clusters. Obviously, the highest patient severity, PSC =4, is related to the first cluster in both hospital.

Graph 9



To find the relation of patient mortality and severity with other variables, we need to consider the condition of each cluster. The table (3) displays the mean of some variables in each cluster.

Table (3) Mean value in each cluster

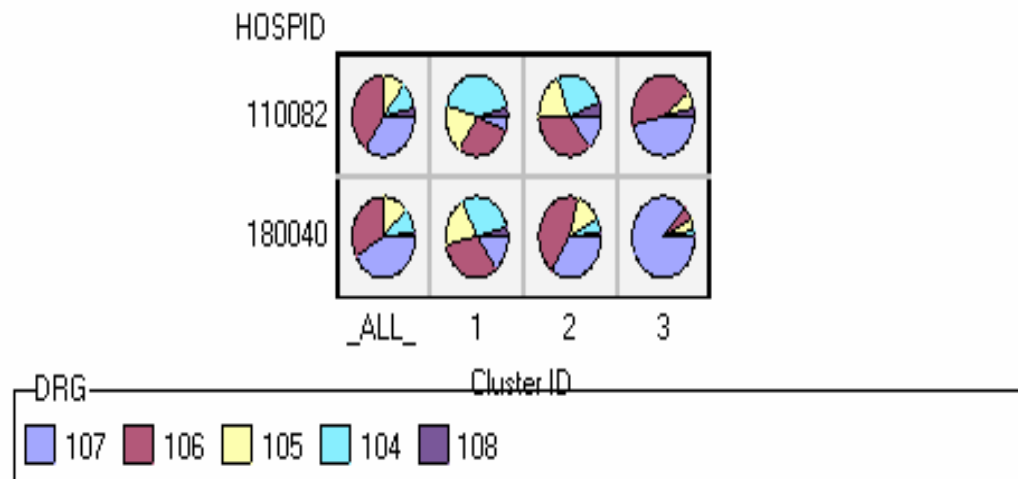
Cluster DRGPRICE	AGE	AMTREIMB_log	AMTREIMB	CARDCS	COVDAYS
1 26319.16667	71.47712	10.19640	27164.57190	1080.28105	13.36601
2 24055.46185	71.13922	10.05213	23595.36814	683.57430	8.25167
3 23033.97558	71.33822	10.01021	22596.83516	300.70452	6.06227
Cluster	LOS	TOTCHG			
1	13.37908	68958.38235			
2	8.25167	46291.69612			
3	6.06227	30673.07937			

Table (4)

Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation
1	306	1913.6	29915.3
2	747	1369.6	15449.3
3	819	1288.0	23091.0

Combining the summary of the clustering displayed in the table (3,) and information from the two graphs (8) and (9) shows that the number of observations in cluster 1, 306, is less than half of the numbers of the observations in clusters 2 and 3, 747 and 819 respectively. But the reimbursement mean (AMTREIMB), total charge mean (TOTCHG), the mortality rate, and the severity rate are higher in this cluster compared to other clusters. It implies that these variables are related to each other sufficiently. To find out the relation between the clusters and different DRG, we need to look at the graph (10).

Graph(10)



The first column in graph (10) makes it obvious that the portion of the different treatments (DRG) in both hospitals are almost the same. We next consider whether Reimbursement is a constant portion of total charge in all clusters. We compare the ratio of the reimbursement mean value and the total charge mean value for the three clusters, say R1, R2, R3. $R1 = 0.3939$, $R2 = 0.5097$, and $R3 = 0.7367$ so that the average of reimbursement is almost 39%, 51%, and 74% of total charge, respectively in clusters 1, 2, and 3. The reason for these differences is probability caused by the different DRGs. To find the effectiveness of DRG, DRGPRICE, total charge, patient severity, and length of stay to predict reimbursement, we examine different models.

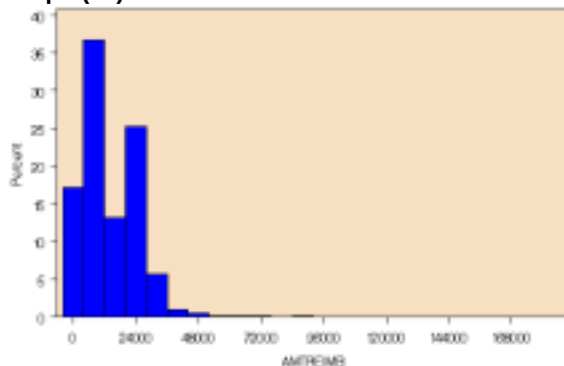
Estimating a model for reimbursement

We are interested in examining the variability of the reimbursement amount for patient treatment. Assuming reimbursement as an outcome variable and total charge, treatment (DRG), hospital, and patient severity (PSC) as predictors, we try to estimate a best model to fit the data. There are different kinds of models that can be applied according to the different assumptions. Some different methods are as follows:

- ANOVA or Analysis of Variance is a technique to analyze a linear model and to summarize how well the entire model fits the response. Required assumptions for ANOVA are independent observations, normally distributed data for each group, and equal variance from each group.
- PROC MIXED or Linear mixed model: Model assumptions are normality of the data distribution and the expected value of the dependent variable is linearly related to the independent variables. This is called mixed because of using both fixed and random effects in the model. It is accepted that variables that are blocked, or that are representative of a larger population are random. For instance, if data have been collected from different hospitals or clinics, the estimated parameter for that hospital or clinic can be considered as a random effect because they are selected from a bigger set, which contains more hospitals or clinics. But if data come from only two hospitals or clinics and the goal is to compare just those two hospitals, the estimated parameter can be considered as a fixed effect. In linear models, fixed effects and random effects are used to model the mean and variance-covariance structure of the dependent variable, respectively. The mixed model is $y = x\beta + z\gamma + \varepsilon$ where $E(y) = x\beta$, X is the matrix of fixed effects, and Z is the matrix for random effects. PROC GLIMMIX (generalized linear mixed model) generalizes PROC MIXED to model the data from non-normal distributions. If the expected value of the response variable, $E(y)$, is not a linear function of independent variables, a transformation function g , the link function, is defined, such that $g(E(y)) = x\beta + z\gamma$. This means to predict a dependent variable for given X and Z , fixed and random factors, the function $E(y) = g^{-1}(x\beta + z\gamma)$ will be used, where β and γ are fixed and random effects respectively

To begin a modeling procedure, the first step is to examine the distribution of the dependent variable, "reimbursement". A Histogram and density curve of observations is helpful to estimate a proper distribution model of data. Also, SAS Enterprise Guide provides three goodness-of-fit tests for different distributions (normal, lognormal, exponential, beta, gamma, kernel) based on the empirical distribution function (Anderson-Darling, Kolmogorov-Smirnov, and Cramer-von Mises). The analysis of reimbursement using whole data with 4861 observation shows that the data have Skewness = 2.4575300 and Kurtosis= 20.8460862. Skewness measures the symmetry of the data, and kurtosis measures the heaviness of the tails of the data distribution. In a normal distribution, both skewness and kurtosis should be near to zero.

Graph (11)



To get better comparison results of the reimbursement distribution in both hospitals, it is good to consider only the observations related to the common treatment (DRG) in two hospitals that have total charges less than 90,000 dollars. Table (5) shows the result of the statistical analysis.

Table5. Analysis Variable: AMTREIMB

HOSPID	N		Mean	Std Dev	N	Minimum	Maximum	Median
	Obs							
110082	890		25225.30	3579.22	890	16544.00	37694.00	25226.00
180040	982		22397.54	4656.62	982	11601.00	32813.00	23938.00

Analysis Variable : AMTREIMB

HOSPID	N		Skewness	Kurtosis
	Obs			
110082	890		0.2547002	-0.1276706
180040	982		0.4993490	-0.5353592

According to table (5), the values of skewness and kurtosis in both hospitals are near to zero; that is, the distribution in the two hospitals is close to normality. Each observation is related to different patients, so they can be assumed independent. Homogeneity of variances for the two hospitals and different treatments (DRG) is checked using Leven's test. Fitting a one-way ANOVA, where reimbursement is dependent and hospital is an independent variable, the following result is obtained.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
HOSPID	2	110082 180040
Number of Observations Read		1872
Number of Observations Used		1872

The ANOVA Procedure

Dependent Variable: AMTREIMB AMTREIMB

Source	DF	Squares	Sum of Mean Square	F Value	Pr > F
Model	1	3733197174	3733197174	213.74	<.0001
Error	1870	32660909168	17465727		
Corrected Total	1871	36394106342			
R-Square Coeff Var Root MSE AMTREIMB Mean					
		0.102577	17.60261	4179.202	23741.94

R^2 = (model sum of square) / (total sum of square), in this model R^2 is almost 0.1. It means that only 0.1 portion of variability of reimbursement can be explained by the hospitals' effect.

The ANOVA Procedure

Levene's Test for Homogeneity of AMTREIMB Variance ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
HOSPID	1	3.67E16	3.67E16	72.59	<.0001
Error	1870	9.453E17	5.055E14		

Levene's test null hypotheses that variances for hospitals are equal. Considering $\alpha = .05$ and P-value less than .0001, we do not have enough evidence to accept the null hypotheses. Therefore, the homogeneity of variance is not satisfied.

The ANOVA Procedure

Level of HOSPID	N	-----AMTREIMB-----	
		Mean	Std Dev
110082	890	25225.3034	3579.21566
180040	982	22397.5428	4656.62107

Dependent Variable	Source	Alpha	Power	Least Significant Number
AMTREIMB	HOSPID	0.05	0.999	37

Power= 0.999 is the probability of rejecting the null hypotheses when it is false.

Using One-Way ANOVA for reimbursement as response variable and DRG as an independent variable, the following results are found:

Source	DF	Anova SS	Mean Square	F Value	Pr > F
DRG	4	29332496593	7333124148	1938.78	<.0001
R-Square	Coeff Var	Root MSE	AMTREIMB Mean		
0.805968	8.191502	1944.821	23741.94		

Since the P-value is less than .0001, we reject the hypothesis that all DRG groups have the same mean of reimbursement.

The ANOVA Procedure

Levene's Test for Homogeneity of AMTREIMB Variance
ANOVA of Squared Deviations from Group Means

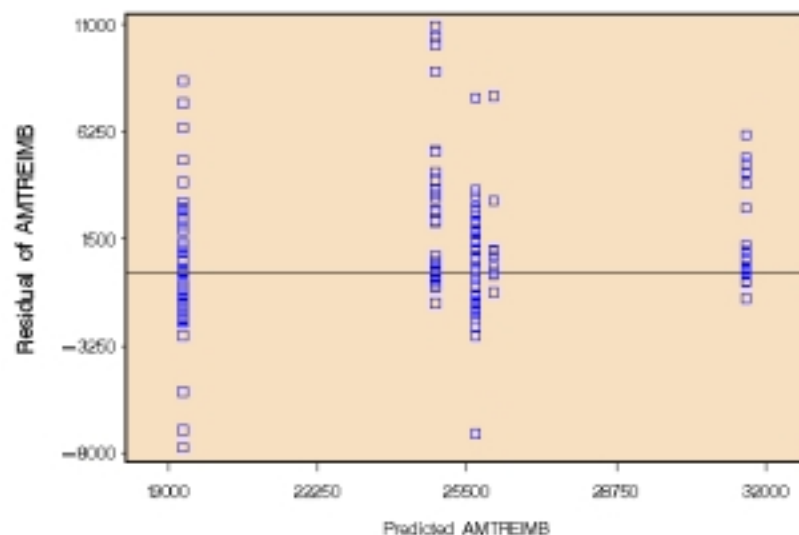
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
DRG	4	2.115E15	5.287E14	11.22	<.0001
Error	1867	8.797E16	4.712E13		

The Leven's test shows the p-value less than 0.0001. Therefore we reject the null hypotheses that all DRG groups have the same variances.

Since the assumption of ANOVA has been violated, the validity of the result is under question.

The ANOVA Procedure

Level of DRG	N	-----AMTREIMB-----	
		Mean	Std Dev
104	197	31572.3096	1205.79528
105	222	24815.0000	1874.97958
106	677	25694.5288	1871.76880
107	729	19334.8436	2203.43017
108	47	26083.7234	1450.68438



A special pattern of residuals in graph 12 confirms that it is not normally distributed. To examine the effectiveness of treatment (DRG), hospital (HOSPID), patient severity (PSC), and their interaction in reimbursement, Factorial ANOVA analysis has been used. The following result was obtained.

Graph 12.

The GLM Procedure

Class Level Information

Class	Levels	Values
DRG	5	104 105 106 107 108
HOSPID	2	110082 180040
PSC	4	1 2 3 4
Number of Observations Read		1872
Number of Observations Used		1872

The GLM Procedure

Dependent Variable: AMTREIMB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	37	33340225861	901087185	541.15	<.0001
Error	1834	3053880481	1665147		
Corrected Total	1871	36394106342			

The ANOVA test the hypotheses that the mean of reimbursement for all classes is the same. Since the p-Value is less than 0.0001, the null hypotheses should be rejected. Now the question is which factor(s) cause this difference.

R-Square	Coeff Var	Root MSE	AMTREIMB Mean
0.916089	5.435133	1290.406	23741.94

The R^2 of this model is almost 0.92, that is, 92 percent of variability of reimbursement can be explained by this model.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DRG	4	12045436179	3011359045	1808.46	<.0001
HOSPID	1	126501287	126501287	75.97	<.0001
PSC	3	53129375	17709792	10.64	<.0001
DRG*HOSPID	4	335487889	83871972	50.37	<.0001
DRG*PSC	12	32109725	2675810	1.61	0.0830
HOSPID*PSC	3	14690531	4896844	2.94	0.0320
DRG*HOSPID*PSC	10	160547403	16054740	9.64	<.0001

To test the null hypotheses that the effect of the factors in the model is insignificant; Type III sums of squares are used. The only p-value (0.083) related to DRG*PSC is bigger than 0.05; therefore there is not enough evidence to confirm that the interaction between DRG and PSC affects the variability of reimbursement. But we reject the null hypotheses strongly for the other factors with p-value less than 0.0001.

A Goodness-of-fit test was conducted to check the null hypothesis that the reimbursement residual has a normal distribution. Based upon the result of Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests in the following table and given p-values, the residual does not have a normal distribution. Therefore the validity of the models base on ANOVA is in question.

Goodness-of-Fit Tests for Normal Distribution
(residual reimbursement)

Test	---Statistic----		-----p Value-----	
Kolmogorov-Smirnov	D	0.219233	Pr > D	<0.010
Cramer-von Mises	W-Sq	24.947859	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	141.768324	Pr > A-Sq	<0.005

Enterprise Guide distribution tests were performed to determine a best distribution function that fits the response variable, reimbursement. No known and available significant distribution is found in a 90% confidence interval. According to the small value of skewness and kurtosis for the response variable (reimbursement), we may assume that reimbursement has a normal distribution. The set of the independent variables contains both fixed and random effects for reimbursement; therefore, both proc MIXED and GLIMMIX are used to examine a best model that can explain the variability of response variable. The code used for mixed is as follow:

```
proc mixed data=SASUSER.medpar_cleaned_edited_commondrg;
  class DRG PSC HOSPID;
  model AMTREIMB = TOTCHG LOS PSC HOSPID /SOLUTION;
  random DRG DRG*PSC / SOLUTION ;
run;
quit;
```

Solution for Fixed Effects

Effect	HOSPID	PSC	Estimate	Error	Standard DF	t Value	Pr > t	Alpha
Intercept			22252	1837.40	4.27	12.11	0.0002	0.05
TOTCHG			0.03974	0.004521	1860	8.79	<.0001	0.05
LOS			11.5882	12.0595	1860	0.96	0.3367	0.05

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
TOTCHG	1	1860	77.28	<.0001
LOS	1	1860	0.92	0.3367
PSC	3	9.08	0.65	0.6050
HOSPID	1	1854	1057.71	<.0001

1872 observations are used in this model. After 8 iterations, convergence criteria were met. According to given p-values, length of stay (LOS) and patient severity rank (PSC), effects are not significant where alpha is 0.05. Also, neither one of the random effects, DRG and DRG*PSC, are significant (the related p-value are bigger than 0.05). To examine another model, the following code is used.

```
proc mixed data=SASUSER.medpar_cleaned_edited_commondrg;
  class DRG PSC HOSPID;
  model AMTREIMB = TOTCHG DRGPRICE DRGPRICE*HOSPID /SOLUTION;

  random HOSPID DRG*PSC*HOSPID / SOLUTION ;
run;
quit;
```

Solution for Fixed Effects

Effect	HOSPID	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		1996.06	2778.86	1	0.72	0.6034
TOTCHG		0.06385	0.002195	1831	29.09	<.0001
DRGPRICE		0.8809	0.02379	1831	37.03	<.0001
DRGPRICE*HOSPID	110082	-0.1593	0.02966	1831	-5.37	<.0001

Using PROC GLIMMIX and running the same model gives the same result. But the ratio of Generalized Chi-Square and DE (Degrees of freedom), 894414.4, is far from 1. This implies that the variability in these data has not been properly modeled. Assuming different kinds of distributions and related link functions for the data, and different categories for independent variables, PROC GLIMMIX has been used to estimate a good model. However, since no significant result is found, the codes and results are not shown.

CONCLUSION

1872 observation from two hospitals related with five kinds of treatment or DRG have been used as a random sample. Clustering of the data does not confirm differences between mortality rates in the two hospitals. Also, it shows that DRG 104 may have higher mortality rate and severity rank compared to the other DRGs. According to some statistical tests, there is no significant evidence for a relationship between patient severity and reimbursement. The factor DRGPRICE explains variability of reimbursement more than total charge. According to the ANOVA and GLM (generalize linear model) result, the interaction effects of DRG*PSC and HOSPID*PSC are not significant in any models. That means the patient severity rank in different DRGs and hospitals are not significant in the variability of the dependent variable. Assuming the data normality using proc MIXED and proc GLIMMIX give the same result in fixed and random effects. But using GLIMMIX gives the ratio of Generalized Chi-Square and DE that is helpful to recognize how the variability of the response variable has been modeled properly. Changing some independent variable categories from fixed to random or vice versa, does not change the estimation parameter of other fixed effects. Unfortunately, we could not get any good model using GLIMMIX. Also assuming different distribution functions for response variable, PROC GLIMMIX does not give any proper model for the data. To get a better model that explains the variability of reimbursement, we may need to choose more hospitals randomly and consider more DRG groups. Also other information such as total cost, ICD-9 codes, may explain better the variability of reimbursement.

REFERENCE

- 1- DRG GUIDEBOOK, A comprehensive Reference to the DRG Classification System, 2001, seventeenth edition,

CONTACT INFORMATION

Author Name: Fariba Nowrouzi Kashan
 University of Louisville
 Department of Mathematics
 328 Natural Sciences BLDG.
 Louisville, Kentucky 40292
 Phone: (502) 852-6826
 Fax: (502) 852-7132
 E-mail: f0nowr01@louisvill.edu