A Platform based on Multiple Regression to Estimate the Effect of in-Hospital Events on Total Charges

Dimitrios Zikos
Department of Health Administration
Central Michigan University
Mt. Pleasant, MI, United States
zikos1d@cmich.edu

Dhanashri Ostwal
Computer Science and Engineering Department
University of Texas at Arlington
Arlington, TX, United States
dhanashrivilas.ostwal@mavs.uta.edu

Abstract—**Recently hospitals struggle to control the cost of care while maintaining optimal outcomes. To respond to this challenge, we developed an interactive web platform which utilizes a multiple linear regression model. The user can create and furthermore alter a clinical scenario, during a patient hospitalization to see the dynamic prediction of total charges, via interactive sessions. The R^2 value of our model is 0.655 and the standard error of the estimate is $38,732. Predictors with high coefficient scores include the cardioverter implantation, mechanical ventilation, implant of pulsation balloon and hospital-acquired conditions such as staphylococcus aureus septicemia. Our findings indicate that (a) integration of predictive models into clinical decision support systems is feasible and use of regression methods provide direct feedback on the effect of any clinical practice to the in-hospital charges (b) medical claims data can provide a useful estimation of the in-hospital charges (c) hospital acquired conditions have significant impact on the in-hospital charges.**

Keywords—*total charges; multiple linear regression; prediction; decision making*

I. INTRODUCTION

Hospitals in the United States are in a constant effort to provide high-quality services without undergoing unneeded procedures. There is a need of maintaining a balance between optimal health outcomes and the cost of the provided care. Novel practices and therapeutic methods are being introduced into the clinical practice, hospitals purchase new equipment and capacity to provide modern services, often with important amortization considerations to be made during budgeting. Increased health care costs have not necessarily led to improved outcomes. According to the American Hospital Association, overdiagnosis and overuse of treatments have increased health care costs with barely any improvement in health outcomes [1]. While there is a lot of research associating nursing and quality of care, very little has been done on the impact that clinical and nursing practices have, to the cost of care and the total charges of an in-hospital stay [2].

At the same time, we recognize an unmet need for services that provide dynamic, individualized estimations of the effect of clinical interventions to in-hospital charges, during the clinical practice. Such a dynamic estimation would not only provide an insight on the projected financial burden of the hospital stay, but it could also be used to drive decisions via the interaction of therapists with clinical decision support systems which integrate the aforementioned functionality.

It is not uncommon for multiple regression techniques to model the cost as a function of covariates that are observed in the patients. The estimated beta coefficients have been reported to provide an estimation of the total cost for each admission case [3]. Generalized Linear Models have also been used in the past for the cost of care estimation [4]. In another case, researchers utilized data from stroke patients and used DRGs, and other hospital variables in order to construct a regression model which explained a 61% of the cost of care variance [5]. Cost prediction models are often driven by limitations related with the non-availability of features which would help explain a higher percentage of the variability. In [6], researchers used hospital admission information and their model could explain a rather small ratio of the total charges variance, which was no higher than 34%.

In the case of the Intensive Care Unit (ICU) cost estimation, Moran et al. [7] used a combination of ICU activity indices and severity scores for cost prediction. In a similar work by Ramianira et al. [8] researchers estimated the costs and then used a standard linear regression model to correlate cost units and their predictors. The study identified as importation predictors, the patient gender and age, the admission type (urgency/elective), ICU admission, blood transfusion, the admission outcome (death/no death), the complexity of medical procedures, and a risk-adjustment index. Researchers from MIT presented an algorithmic approach to predict the cost of care [9] by utilizing classification trees and clustering algorithms on claims data from more than 800,000 patients. The authors of this study stressed the limitations of using the R^2 value as the primary evaluator of the prediction accuracy.

The majority of the aforementioned studies have used regression methods to predict the cost of care and have approached the problem in a conventional statistical manner. There are no research examples in the literature, though, of efforts to integrate predictive models into decision support systems which can be used by the hospital administration and clinicians in an interactive manner, during the course of the healthcare provision. To respond to this unmet need, we first developed and evaluated a multiple linear regression model and then we integrated the model into an interactive web interface, which provides direct feedback to the hospital administration and to clinicians. During the clinical care, users are presented with an estimation of the total charges based on the selection of their preferred attributes of clinical care. Subsequently, they can

alter any attribute value to see the effect of such a change to the cost of care, and overview a comparison of consecutive runs. There are many levels of interest and a variety of possible use case scenarios; The hospital administration would be provided with a realistic snapshot of the total charges per patient as well as per unit. Clinicians, being members of the hospital team, would extend to the strategic goals of the hospitals since they would have available tools assisting them to make important cost-benefit considerations, during the clinical practice.

The contribution and importance of our study is the introduction of an online platform, which is built around a reasonably performing regression model, rendering the system easy to use without any prior in-depth understanding of statistics, and providing direct meaningful feedback hospital administrators and clinicians.

The paper is organized as follows: Section II describes the data that we used for the development of our platform and the preprocessing. Section III provides detailed information on the training and the performance of the predictive model. Finally, Section IV presents the architecture and functionality of the web platform and an example use-case scenario.

II. DATA SELECTION AND PREPARATION

A. Description of the Data

Our platform utilizes a comprehensive Medicare in-hospital claims file which contains records of Medicare beneficiaries who used hospital inpatient services in Texas, the United States during the year 2013 [10]. The dataset is de-identified and includes more than one million tuples, each representing a hospital admission. The attributes can be classified into the following categories: (i) admission information and demographics (ii) discharge information (iii) clinical outcomes (iv) hospital procedures (v) diagnoses (vi) cost of care and diagnosis related groups. Table I presents some important descriptive statistics of our dataset.

Medicare is an enormous U.S social insurance program and provides health insurance for Americans aged 65 and older who have worked and paid into the system, as well as to younger people with disabilities, end-stage renal disease and amyotrophic lateral sclerosis [11]. The total number of Medicare beneficiaries for the year 2015 exceeded 49 million of patients, while Medicare is the primary payer for the 47.2 percent of total aggregate inpatient hospital costs in the United States [12].

TABLE I.   DESCRIPTIVE STATISTICS OF THE TARGET DATASET

Indicator	Descriptive Statistics
% admissions of female patients	54.0%
Mode Age group	65-69 years (16.9% of total)
In-hospital mortality ratio (%)	3.1%
Length of stay (days)	Mean=6.38 (sd=7.69)
Total charges (U.S Dollars)	Mean=49,548 (sd=64,719)
ICU Use (%)	31.4%
Admitted from home (%)	73.7%
Type of Admission (%)	Emergency: 53.9% Elective: 27.4% Urgent: 17.8%

B. Data Preprocessing

To facilitate the estimation of the cost of care with the use multiple linear regression (MLR), we transformed the dataset to a sparse data file by computing multiple binary attributes for the unique values of the original dataset. The categories for all non-ordinary nominal attributes were transformed to new binary attributes, essentially describing the existence (value=1) or non-existence (value=0) of a diagnosis, a medical procedure, or a hospital-acquired condition (HAC), acting like a switch. These binary attributes are going to be used as our features to predict the total charges (dependent variable). The user will know the impact of a change of an attribute value to the total charges, by observing the beta coefficient of the attribute. In a linear regression equation, the beta coefficient of any attribute is equal to the units of change to the dependent variable (in our case the total charges) when the value of that attribute increases by one unit.

We removed from the dataset any attributes that would normally be unavailable at the point of the decision, in a real hospital context. The point of the decision can any time after the admission and during the patient hospitalization. The attributes we removed include the Diagnosis Related Groups (DRG) price, the discharge destination, the discharge status, and all cost-related attributes. We opted for the inclusion of the HACs since this information is acquired at any temporal point during the patient hospitalization.

C. The Multiple Linear Regression Model

We calculated a multiple linear regression model using SPSS version 22 [13] to predict the dependent variable "total charges". We utilized 391 variables as predictors of the total charges in our model. We generated a selection of dummy attributes for the diagnosis and procedure variables codes with the highest frequency (>1,000 cases) in the dataset, instead of generating thousands of dummy variables, equal to the icd-9 size (14,000 codes). We observed that the cost of adding all these dummy variables was a substantial increase to size of the data file (~100 GB), significantly longer training time during our experiments, but only a negligible improvement to the model performance.

The variables include information about: the type of hospital admission, source of admission, admitting diagnosis, the day of admission, age group, sex, discharge diagnosis, hospital acquired conditions, intensive care unit stay, the length of stay, surgery indicator and primary diagnosis. We added all the independent variables into the analysis simultaneously, using the enter method.

The R^2 value shows how close the data are to the fitted regression line and was found to be equal to 0.655, indicating that 65.5% of the variability in the response is explained by the explanatory variables. The standard error of the estimate was equal to $32,237.17 (Table II).

TABLE II. R^2 VALUE AND STANDARD ERROR OF ESTIMATE

R	R Square	Std. Error of the Estimate
0.809	0.655	32237.166

We wanted to validate that there exists a significant linear regression relationship between the response variable (total

charges) and the predictor variables and for this reason we conducted an Analysis of Variance (ANOVA) test. A significant regression equation was found (f=505.47, p<0.01), indicating that there is a significant effect of the amount of the independent variables on the total charges at the p<0.01 level (Table III).

TABLE III.   ANALYSIS OF VARIANCE

Source	df	Sum of Squares	Mean Square	F	Sig.
Between Groups	385	269840547401083.6	700884538704.1	505.47	0.00
Within Groups	102574	142229779619092.03	1386606543.8		
Total	102959	412070327020175.6			

The top 10 attributes whose coefficients were calculated with the highest precision (highest t-score) were the in-hospital length of stay, five clinical procedures (continuous mechanical ventilation, cardioverter implantation, extracorporeal circulation auxiliary to open heart surgery, (re)fusion of 4-8 vertebrae, implant of pulsation balloon), three ICU stay types (Surgical, General, Intermediate) and the Diagnosis 'Acute Kidney Failure with Lesion of Tubular Necrosis' (Table IV).

TABLE IV.   ATTRIBUTES WITH THE HIGHEST T-STATISTIC

Attribute	Beta coefficient	Std. Error	t*
Length of Stay	4080.8	19.1	213.5
Procedure 9672: Continuous invasive mechanical ventilation for > 96h	74141.7	1233.9	60.1
Procedure 3794: Implantation/ replacement of automatic cardioverter /defibrillator, total	98486.1	2846	34.6
Surgical ICU stay	39551.9	1145.9	34.5
Procedure 3961: Extracorporeal circulation auxiliary to open heart surgery	55548.6	1799.5	30.7
General ICU stay	12895.5	420.9	30.6
Intermediate ICU stay	9340.5	340.9	27.4
Procedure 8163: (Re)fusion of 4-8 vertebrae	81254.4	3167.6	25.6
Procedure 3761: Implant pulsation balloon	65649.5	2778.4	23.6
Diagnosis 5845: Acute Kidney Failure with Lesion of Tubular Necrosis	22671.8	1061.3	21.4

*The t-statistic for all attributes was significant at the 1% significance level

The predictors of the total charges with the highest coefficient scores were found to be the pediatric intensive care unit stay, four clinical procedures (cardioverter implantation, (re)fusion of 4-8 vertebrae, continuous mechanical ventilation, implant of pulsation balloon) and, not surprisingly, five hospital acquired conditions, including the displacement of lumbar intervertebral disc, Methicillin Resistant Staphylococcus Aureus Septicemia and Pneumonia, Complications of Transplanted Kidney and Intestinal Or Peritoneal Adhesions With Obstruction (Table V).

TABLE V.   PREDICTORS WITH THE HIGHEST COEFFICIENT VALUES

Attribute	Beta coefficient	Std. Error	t*
Procedure 3794: Implantation/ replacement of automatic cardioverter /defibrillator, total	98486.1	2846	34.6
Pediatric ICU stay	97490.2	11804.8	8.3
Hospital Acquired Condition 72210: Displacement of Lumbar Intervertebral Disc w/o Myelopathy	97342.9	26405	3.7
Procedure 8163: (Re)fusion of 4-8 vertebrae	81254.4	3167.6	25.7
Hospital Acquired Condition 03812: Methicillin Resistant Staphylococcus Aureus Septicemia	80512.9	7263.5	11.1
Procedure 9672: Continuous invasive mechanical ventilation for > 96h	74141.7	1233.9	60.1
Hospital Acquired Condition 48242: Methicillin Resistant Pneumonia from Staphylococcus Aureus	70460.9	4721.9	14.9
Hospital Acquired Condition 99681: Complications of Transplanted Kidney	66905.5	7638.5	8.8
Procedure 3761: Implant of pulsation balloon	65649.5	2778.4	23.6
Hospital Acquired Condition 56081: Intestinal or Peritoneal Adhesions with Obstruction	60503.6	15287.2	3.9

*The t-statistic for all attributes was significant at the 1% significance level

D. Testing of the Model using Binarized classes

We used the median of total charges (50% percentile) as a cutoff point to generate a "low charges" and "high charges" class with equal number of observations. The median total charges were $31,228. With this experiment we want to provide an alternative means to evaluate the performance of our method, being aware of the reported limitations of using the R^2 value as the primary evaluator of the prediction accuracy [9]. We grouped the observed and predicted total charges into the "low charges" or "high charges" class. The overall accuracy of the classification was 80.6%. The recall for the "low charges" class was equal to 74.9% and the precision 83.9%. The recall for the "high charges" was found to be 86.1% and the precision 77.9% (Fig. 1).
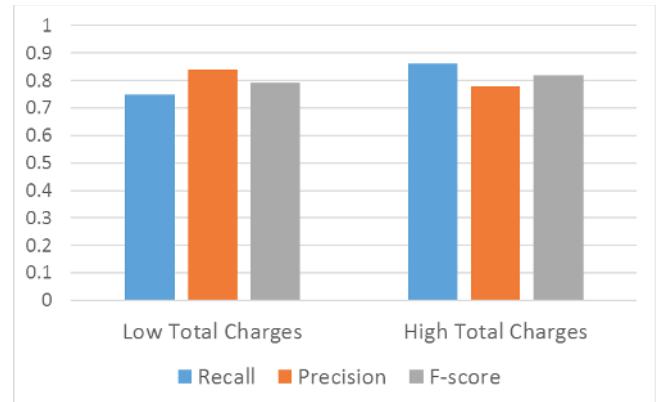

Fig. 1.   Classification performance of binarized total charges class

We performed a similar experiment, this time by generating five total charges categories with a range of $60,000 each. This cut-off point would simply serve as an example, to allow us to

explore the performance when the cost estimation problem becomes a multiclass one. The overall accuracy of the classification was found to be 80.3%. The precision and recall for the class '$0-$60,000' were found to be 89.4% and 93.1% respectively. For the class '$60,000-$120,000', the precision fell to 57.6% and the recall to 48.1%. There was further decline to the precision and recall for the next two "total charges" classes while the performance slightly improved for the very expensive (>$240,000) class (recall=47.5% and precision=71.3%). The linear model does not properly fit hospital stays with total charges lying across the middle range.

Finally, we wanted to compare the performance of the binarized grouping with the performance of classifiers which handle classes of discrete nature. For these experiments, we used the binarized attribute as a class. We explored the performance Naïve Bayes, as a baseline and found that the overall accuracy was equal to 73.4%. The recall was 78.3% for the "low charges" class and 68.4% for the "high charges" class. The precision was 71.3% and 75.9% for the "low charges" and "high charges" class, respectively. The classification performance was significantly better in the case of the logistic regression, with an overall accuracy equal to 83.5%. The recall was found to be equal to 86.5% for the "low charges" and 79.6% for the "high charges" class, whereas the precision was 80.9% and 85.5% for the "low charges" and the "high charges" class, respectively. Finally, the AdaBoost meta-classifier showcased performance comparable to our method.

IV. THE PLATFORM

A. Human-Computer Interaction

With our interactive web platform, the user can create a clinical scenario, overview the total charges prediction and consequently make any changes to the clinical scenario to see the effect of those changes to the total charges.

The system is session-based. As soon as a new session is initiated, the user can enter data for the attributes of care. This view allows the user to input information such as the patient age and the expected length of stay and select all the existing medical procedures, diagnoses, and hospital acquired conditions for that patient. The aforementioned is the input to the multiple regression function, which will output the predicted total charges in US dollars. The predicted value is stored as a temporary variable. Within the same session, the user has the choice to add or remove any binary clinical care attribute or change the value for a continuous variable (i.e. length of stay) to see an updated prediction of the total charges. The user can continue trying out additional case scenarios during one session and all runs are shown in tabular format and via a chart, as shown in Fig. 3. A clear button allows the user to clear the previous runs and start a new session.

During a session, a table displays details for all the previous runs of that session. This table can be sorted by clicking on the table headers. A histogram is also shown that displays the total costs for the previous runs along with the time stamps. This representation provides to the user a quick view of how cost has changed over time and for all the different parameters selected during the session.

B. System Architecture

The front end consists of a simple HTTP server that runs on a CherryPy web framework. This framework was used primarily because of its compatibility with the python programming language and since it provides a reliable, built-in HTTP-compliant, web server gateway interface (WSGI) thread-pooled server [14]. This made it possible to incorporate a web application that can be accessed via HTTP-compliant web browsers (Fig. 2).
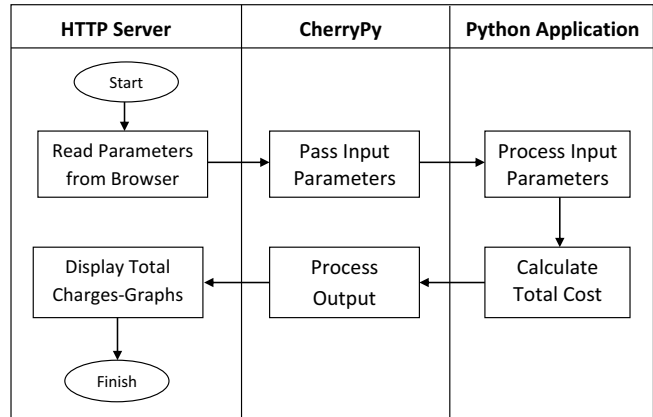

Fig. 2.   System Architecture

C. Use case scenario

In our scenario (Fig.3), a patient has been admitted to the hospital to undergo a total knee replacement (icd-9 code: 815.4). The patient belongs to the age group 6 (75-79 years old) and the doctor in charge believes that the in-hospital length of stay is expected to be around five days. The given information would output a predicted total charge amount equal to $48,934.
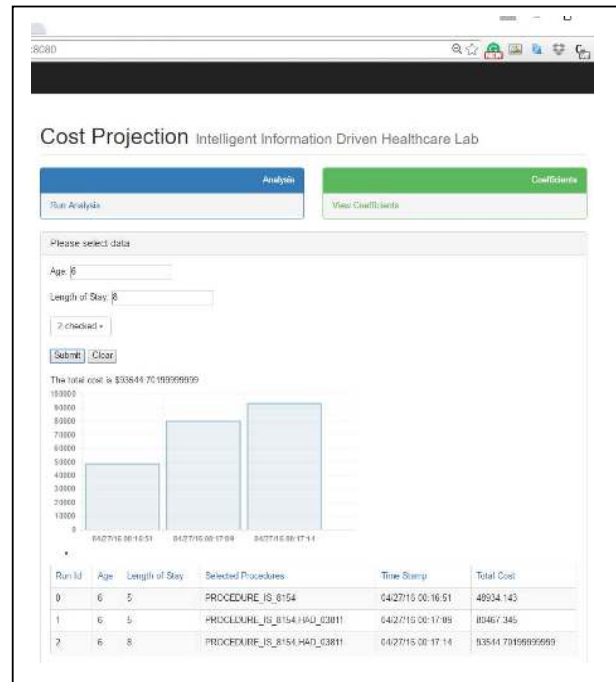

Fig. 3.   An example session of the web platform

During the course of the hospital stay, the patient was found to have a hospital-acquired condition, such as Methicillin susceptible Staphylococcus aureus septicemia (ICD-9 code: 038.11). After this addition, a second run of this session, outputs a new estimation for the total charges, which is significantly higher and equal to $80,467. Given this complication, the doctor in charge decided that the patient would need to prolong his stay for another three days (length of stay= 8). This change (run 3) would change the input to the regression function and the new estimate of the total charges would rise even more, up to $93,544.

V. DISCUSSION

The results of our study indicate that the integration of predictive models into clinical and administrative decision support systems is feasible since all data that we used as predictors in our models are readily available in Electronic Medical Records. Use of regression models in such systems can provide direct feedback to the hospital administration, on the effect of any clinical practice to the total charges during a hospital stay. We also strongly believe that the use of medical claims datasets provides a useful resource for research. Medicare datasets have been used in many studies for research purposes, in secondary data analysis, although not specifically for hospital charges or cost estimation. Examples that can be found in the literature include the identification of clinical events [15], evaluation of the effectiveness of medical devices [16] and the study of rare conditions [17].

While there is a plethora of cost estimation studies in the literature, in various hospital contexts and different patient groups, we are not aware of any research that specifically uses medical claims data to estimate hospital charges. As a consequence, direct performance comparisons would not generate easy to interpret conclusions. In a comparable approach though [5], the R^2 value was found to be slightly lower when compared to the model fit we estimated in our study. In a more recent study, Loginov et al. wanted to determine future health care costs from prior costs, demographics, and diagnoses, using ordinary linear regression and reported adjusted R^2 results between 0.37 and 0.4 [18], while in the case of community case psychiatric, Amaddeo et al. [19] used the ordinary least-squares regression method, which explained between 20% and 69% of the cost variation for new coming patients. There are also few examples in the literature on the prediction of hospital charges that use non regression methods, such as Artificial Neural Networks and decision trees [20].

Linear models are most useful when the variability across the whole spectrum of the dependent variable is same (there is minimal heteroscedasticity). When predicting in-hospital total charges, the nature of the medical claims data is such that, the variability is low when the total charges are either low or very high, but the variability appears to be higher when total charges lie across the middle range.

The majority of studies found in the literature, have been designed and implemented with a traditional statistical mindset, without further considering how the results can be directly utilized for the prediction of hospital charges, dynamically, during the provision of in-hospital health care services.

Therefore, in those studies, one would expect that no consideration was taken regarding:

(i) the inclusion/exclusion of attributes based on the data availability in the real context

(ii) the integration of results into interfaces which not only output the regression equation but also present, via a user friendly interface, the effect that any change of the clinical practice would have on the total cost

With our study, we addressed these two limitations, and this summarizes the contribution and importance of our methodology and implementation.

Since the primary use of our system is to quantify the individual effect of a change of a clinical practice to the total charges, we were equally interested in (i) the classification performance, and (ii) the quantified estimation of the effect of each variable to the total charges. We, therefore, did not consider to integrate classification methods which cannot quantify the effect of each individual variable and, naturally, we excluded methods which can only handle categorical classes. We wanted to know, though, how the performance of our method would compare to modern meta-classifiers such as Adaboost, to probabilistic methods, such as Naïve Bayes and to other traditional regression methods which can only handle discrete classes, such as logistic regression. The results of our experiments showed that when we binarized the total charges variable, only the logistic regression outperformed the performance of our model, by no more than 3%. Naïve Bayes, on the other hand, demonstrated poor performance and Adaboost showcased similar performance when compared to our method.

It is evident from the results of our study that many hospital-acquired conditions drastically contribute towards a substantial increase in the total charges during a hospital stay. Hospital-acquired conditions that are often preventable, such as the displacement of lumbar intervertebral disc, methicillin-resistant staphylococcus aureus septicemia, were found to be significant predictors of the total charges sharing some of the highest coefficient scores. With the use of our system, hospitals will know, prospectively or retrospectively, the quantified contribution of those conditions to the projected charges. This is of great importance, considering that insurance companies do not pay for expenses generated during the treatment of hospital-acquired conditions.

As a conclusion, we believe that our interactive platform can provide an impactful insight to the hospital administration and to health care professionals, by quantifying the contribution of the clinical practice dynamics to the expected hospital charges. This is especially important, considering that unwanted overtreatment practices keep increasing health care costs substantially and our system provides invaluable evidence against such practices.

REFERENCES

[1] American Hospital Association. Appropriate use of medical resources. Available from: http://www.aha.org/content/13/appropusewhiteppr.pdf.

[2] J. Needleman and S. Hassmiller, "The role of nurses in improving hospital quality and efficiency: real-world results," Health Affairs vol 28, No 4, 2009, pp. 625-633.

[3] A.R. Willan and B.J. O'Brien, "Cost prediction models for the comparison of two groups," Health Econ. Vol 10, No 4, June 2001, pp. 363-6.

[4] J. L. Moran, P.J. Solomon, A.R. Peisachand and J. Martin, "New models for old questions: generalized linear models for cost prediction," J Eval Clin Pract Vol 13, No 3, Apr 2007, pp.381-9.

[5] S. Evers, G. Voss, F. Nieman, A. Ament, T. Groot, J. Lodder, A. Boreas, and G. Blaauw "Predicting the cost of hospital stay for stroke patients: the use of diagnosis related groups". Health Policy Vol 61. No 1, Jul. 2002, pp. 21-42.

[6] W. M. Tierney, J. F. Fitzgerald, M. E. Miller, M. K. James and C. J. McDonald "Predicting inpatient costs with admitting clinical data" . Med Care Vol 33, No 1. Jan 1995, pp.1-14.

[7] J. L. Moran, A.R. Peisach, P.J. Solomon and J. Martin, "Cost calculation and prediction in adult intensive care: a ground-up utilization study". Anaesth Intensive Care Vol 32, No 6, Dec 2004, pp. 787-97.

[8] R. Ramiarina, R. M. Almeida and W. C. Pereira, "Hospital costs estimation and prediction as a function of patient and admission characteristics". Int J Health Plann Manage vol 23, No 4. Oct-Dec 2008, pp. 345-55.

[9] D. Bertsimas, M. Bjarnadóttir, M. Kane, C. Kryder, R. Pandey and G.Wang, "Algorithmic Prediction of Health-Care Costs" Operations Research. Vol. 56, No. 6, 2008, pp. 1382–1392

[10] Medicare Provider Analysis and Review (MEDPAR) available from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/MEDPAR.html tember_2014_Issue/336QQML_Journal_2014_Johnston_Sept_619-626.pdf, last accessed 10/26/2015

[11] D. Altman and W. H. Frist, "Medicare and Medicaid at 50 Years: Perspectives of Beneficiaries", Health Care Professionals and Institutions, and Policy Makers". JAMA vol 314 No 4, Jul 2015, pp. 384–395.

[12] C. M. Torio and R. M. Andrews, "National Inpatient Hospital Costs: The Most Expensive Conditions by Payer", 2011. HCUP Statistical Brief #160. Agency for Healthcare Research and Quality, Rockville, MD. August 2013.

[13] IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

[14] Cherrypy. Available from: http://www.cherrypy.org/

[15] A. M. Kucharska-Newton, G.Heiss, H. Ni, S. C. Stearns, N. Puccinelli-Ortega, L. M. Wruck and L. Chambless, "Identification of Heart Failure Events in Medicare Claims: The Atherosclerosis Risk in Communities" (ARIC) Study. Journal of cardiac failure, Vol 22, No 1, 2016, pp. 48-55.

[16] A. P. Shah, E. M. Retzer, S. Nathan, J. D. Paul, J. Friant, K. E. Dill and J. L. Thomas, "Clinical and economic effectiveness of percutaneous ventricular assist devices for high-risk patients undergoing percutaneous coronary intervention", The Journal of invasive cardiology Vol 27, No 3, 2015, pp. 148-154.

[17] M. Menis, R. A. Forshee, S. Kumar, S. McKean, R. Warnock, H.S. Izurieta, et. al, "Babesiosis Occurrence among the Elderly in the United States, as Recorded in Large Medicare Databases during 2006–2013." PloS one Vol 10, No 10, 2015, e0140332.

[18] M. Loginov, E. Marlow and V. Potruch, "Predictive Modeling in Healthcare Costs Using Regression Techniques", Arch 2013.1 Proceedings, 2012.

[19] F. Amaddeo, J. Beecham, P. Bonizzato, A. Fenyo, M. Tansella and M. Knapp, "The costs of community-based psychiatric care for first-ever patients. A case register study", Psychol Med Vol 28, No 1, 1998, pp. 173-83.

[20] J. Wang, M. Li, YT. Hu and Y. Zhu, "Comparison of hospital charge prediction models for gastric cancer patients: neural network vs. decision tree models", BMC Health Serv Res Vol 9, Sept 2009, pp. 161.