

Comparative Regional Analysis of Bacterial Pneumonia Readmission Patients in Medicare category

Heramb Joshi, Oklahoma State University; Aditya Sharma, NXP Semiconductors;
Dr. William Paiva, CHSI, Tulsa; Dr. Goutam Chakraborty, Oklahoma State University

ABSTRACT

Recently hospital re-admissions have become a major growing concern in the healthcare industry. It causes excessive utilization of medical resources, time, monetary aspects and above all patient dissatisfaction. Excessive re-admissions also result in penalties incurred by hospitals. Hospital re-admission is defined as more than one inpatient visit made by an individual for the same diagnosis in a short span of time. It is identified as 30 days by Medicare in United States. This paper aims at the analysis of hospital re-admissions to identify diagnostic procedural anomalies and recognize preventive measures to benefit patients, hospitals, government and insurance companies.

INTRODUCTION

Some of the major diseases such as Chronic Obstructive Pulmonary Diseases (COPD), Congestive heart Failure (CHF) and Bacterial Pneumonia [1] have recorded high re-admission rates reflecting discrepancies in medical procedures. There is a growing need to improve diagnostic procedures at hospitals to reduce hospital re-admissions. In case of readmissions the hospital bill is generated more than once and adds to the overall cost. This excessive cost, in case of Medicare patients, is covered by the government. This analysis aims to analyze the Cerner Database to identify ICD-9 medical procedures contributing significantly towards hospital readmissions of patients with Bacterial Induced pneumonia. The focus will be on Medicare patients, aged 65 or above, in the Northeast and South region of United States.

The analysis and suggestions provided by the study will benefit hospitals, patients and the government in number of ways. The hospitals can device remedies in order to better handle medical procedures which contribute towards re-admissions. It will help improve patient health and reduce expenditure. Reduced readmissions will avoid hospital penalties on base medical reimbursements by the government reducing the paperwork. It will also improve hospitals' credentials in the healthcare industry.

DATA

We received data from Cerner's database provided by Center for Health Systems Innovation, OSU Tulsa. The dataset for Medicare patients (Payer_code = "Medicare") of North Eastern region was based on ICD-9 482 codes (Bacterial Pneumonia) as primary diagnosis. The dataset covered time period from 2003 till 2014. We have considered only procedures and not medication variables because the medication data involved attributes such as brands, dosage, dosage quantities, costs etc. which were not part of the procedural analysis. Following are some of the attributes present in the raw data.

Variable Name	Measurement Level	Data Type
ENCOUNTER_ID	Nominal	Varchar2(20)
HOSPITAL_ID	Nominal	number
PATIENT_ID	Nominal	number
PATIENT_TYPE_ID	Nominal	number
GENDER	Nominal	
AGE_IN_YEARS	Interval	number
MARITAL_STATUS	Nominal	
RACE	Nominal	
DISCHARGE_CARESETTING_ID	Nominal	
ADMITTED_DT_ID	Nominal	number

Variable Name	Measurement Level	Data Type
DISCHARGE_DT_ID	Nominal	number
PAYER_ID	Nominal	number
ADMISSION_TYPE_ID	Nominal	number
ADMISSION_SOURCE_ID	Nominal	number
DISCHARGE_CARESETTING_ID	Nominal	number
URBAN_RURAL_STATUS	Nominal	CHAR
DISCHARGE_DISP_CODE	Nominal	number
BED_SIZE_RANGE	Nominal	VARCHAR
CATH_LAB_FULL_IND	Binary	number
CATH_LAB_DIAGNOSTIC_IND	Binary	number
CENSUS_DIVISION	Nominal	number
PROCEDURE_CODE	Nominal	CHAR
PROCEDURE_DESCRIPTION	Nominal	VARCHAR

Table 1 List of variable names, measurement level, data types

Health Insurance Portability and Accountability Act (HIPAA) requires that any personal information which would identify a patient must be kept confidential. The dataset used for this research was HIPAA compliant. The dataset used for this analysis incorporated “The International Statistical Classification of Diseases and Related Health Problems”, usually called by the short-form name International Classification of Diseases (ICD) codes [2]. The ICD is maintained by the World Health Organization (WHO). The ICD is designed as a health care classification system, providing a system of diagnostic codes for classifying diseases. The dataset incorporated ICD-9 codes ranging from ICD-9001 to ICD-9999

DATA PREPARATION

The original dataset comprised of around 55,000 observations including Northeast and South region of the United States with around 55 attributes. We had to filter the data based on project's scope and requirements by removing undesired columns. After removing redundant columns and all the duplicate rows associated with it, we were left with 7,962 encounters for 6,417 patients of Northeast region and around 8,615 encounters for around 7,478 patients of South region. The admitted date and discharged date were computed based on the corresponding date time variables.

Census Region	Unique Encounters	Unique Patients
Northeast	7,962	6,417
South	8,615	7,478

Table 2 Patients and Encounters statistics for different census regions

READMISSION COMPUTATION

For our analysis, we needed an indicator to represent the readmissions. Consider following procedure for the detailed algorithm of readmission computation [2].

1. A subset of data is created comprising of only "encounter_id", "patient_sk", "admitted_dt" and "discharged_dt".
2. Then for a given patient, the encounters were sequenced in descending order of their admitted date. This ensures that the latest encounter will occur first.
3. Use LAG function in SAS to find the admitted date of the previous encounter for a patient. The difference of this date with the discharged date of the current encounter is the number days in between the two admissions.
4. If this day count is less than 30, the "readmit30" flag is set to 1.

After computing the readmissions, this dataset is joined with original encounters data making the newly created readmit indicators available with rest of the attributes. The number of days in between two readmissions for few patients was observed to be negative. This was because error in recording admitted date and discharge date in the original dataset. The encounters of such patients is removed. The final count of readmissions is as follows.

Census Region	Unique Encounters	Readmissions
Northeast	5,852	788
South	5,968	335

Table 3 Total Readmissions

The final dataset for analysis has 32 attributes displayed below.

Role	Measurement Level	Frequency Count
ID	NOMINAL	1
INPUT	BINARY	3
INPUT	INTERVAL	1
INPUT	NOMINAL	11
REJECTED	INTERVAL	6
REJECTED	NOMINAL	9
TARGET	BINARY	1

Figure 1 Variable Summary

Encounter_ID is the ID variable which is unique identification for the encounters. The three indicator variables such as Teaching_facility_ind, Cath_lab_full_ind and cath_lab_diagnostic_id are the hospital facility indicators used in the model building. The dataset has only one interval variable, age, which had no missing values. Among the class variables, only "urban_rural_status" had around 28 missing values which are very less compared to total observations. Such observations will be ignored by regression models. The nominal variable 'Discharge disposition' had 28 levels. These levels are grouped into 6 levels in a logical manner using SAS code. The hospital ID's are also clustered separately for Northeast and South regions. The target variable "readmit30" is a binary variable with level '0' representing no readmission and '1' representing the readmission.

DESCRIPTIVE STATISTICS

PATIENT DEMOGRAPHICS

The Gender distribution is observed to be consistent in both the regions. There are fairly equal number of male and female patients observed in both the regions.

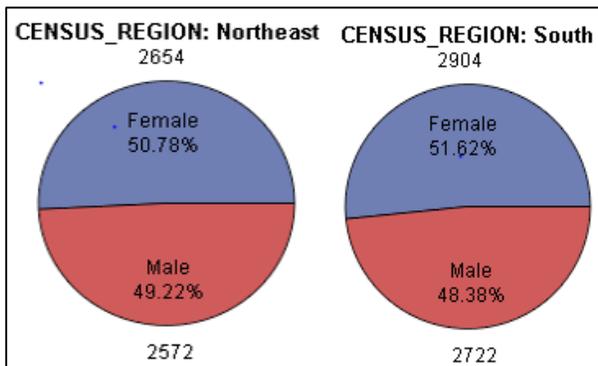


Figure 2 Gender Distribution

The race distribution is also observed to be similar in both the regions. Almost 85% of the patients are “Caucasians” followed by around 10% “African Americans”.

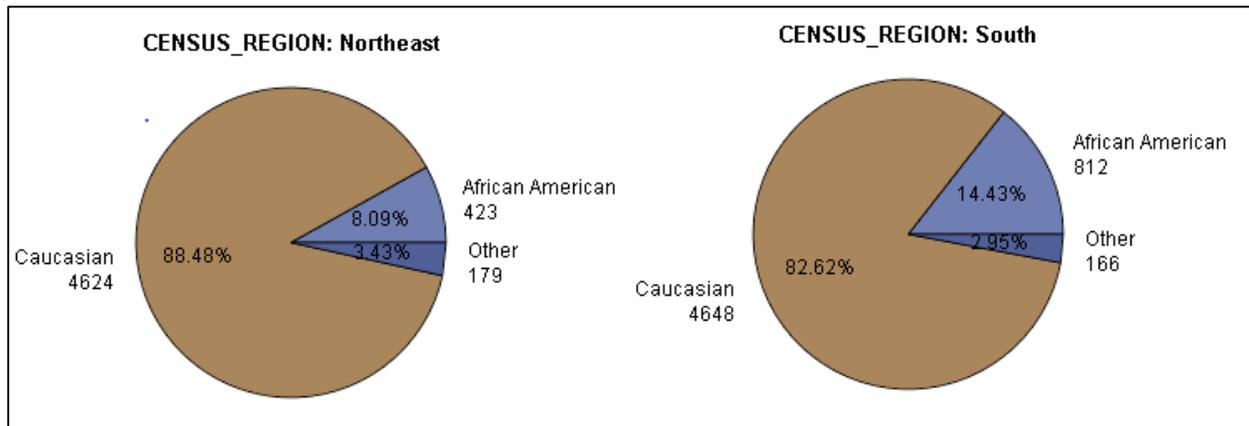


Figure 3 Race Distribution

HOSPITAL DEMOGRAPHICS

The dataset consists of demographic details of 42 hospitals in Northeast region and 57 hospitals in South region. The hospitals having catheterization laboratory are indicated as follows. Catheterization laboratory also known as Cath Lab is the facility in the hospitals or clinics with diagnostic imaging equipment which is used to visualize and monitor arteries of the heart and treat any abnormalities found [4]. There are more number of hospitals in Northeast region having Cath lab facility compared to South.

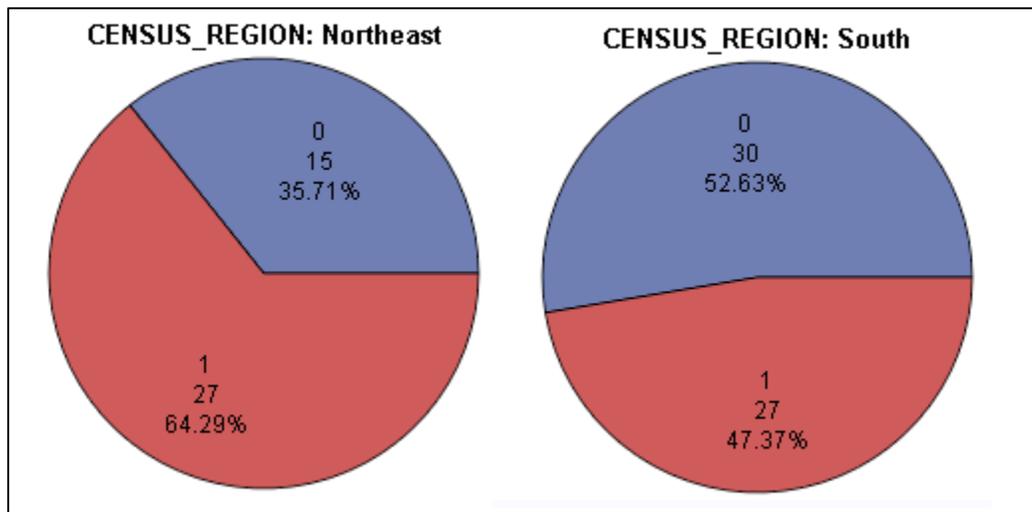


Figure 4 Cath Lab facility distribution

ENCOUNTER DEMOGRAPHICS

Discharge disposition or discharge status is the patient’s anticipated location after discharged from the hospital or medical facility. It is observed that around 35% of the patients are discharged to home followed by around 22% are transferred to Skilled Nursing Facility in both the regions.

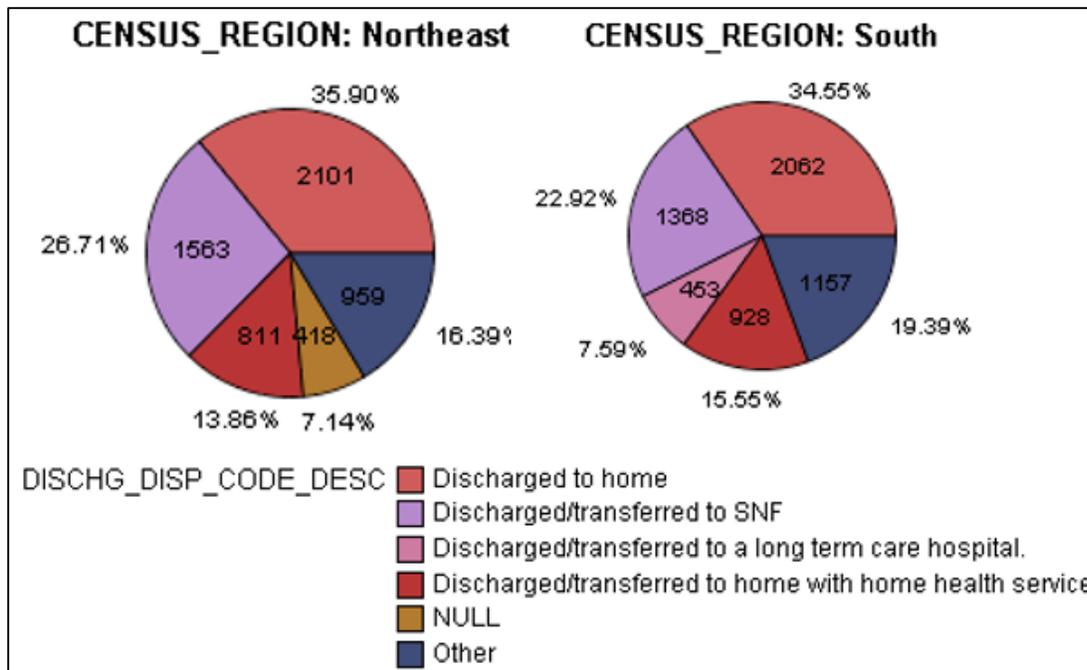


Figure 5 Discharge disposition distribution

The patient type determines the type of encounter visit for a given patient. For example, Inpatient which means the patients whose condition requires admission to the hospital. Outpatient patients are the patients receiving medical treatment without being admitted to the hospital. For example, patients visiting the hospital for diagnosis or treatment. It is observed that most of the encounters are inpatient encounters in both the regions.

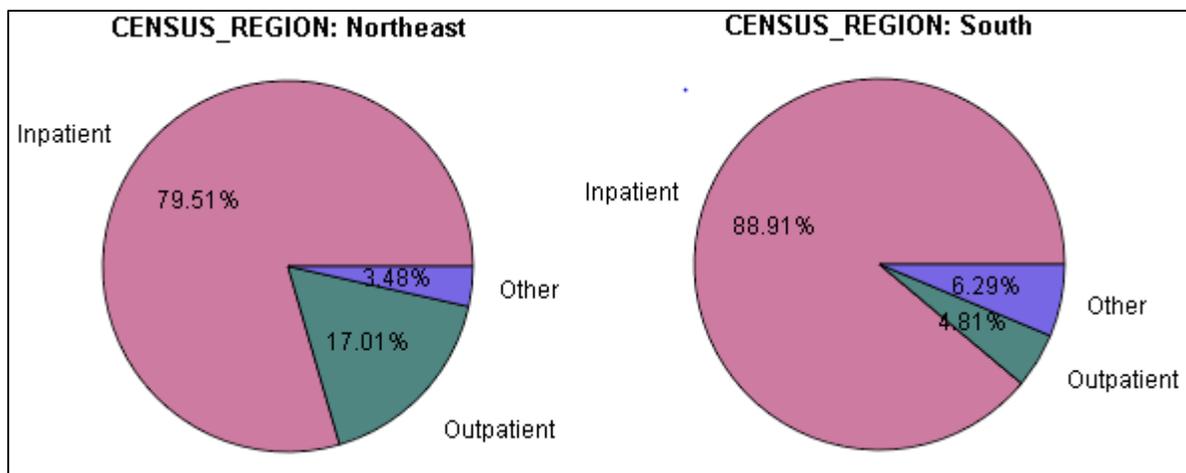


Figure 6 Patient Type Distribution

MODEL BUILDING

As there are very few readmissions, this is a case of rare events. Thus oversampling [5] is used for addressing the case of rare events. The prior probabilities of both the dataset (Northeast and South) are determined and used in the target profile. The target profile setting for analyzing Northeast region is mentioned as follows.

Level	Count	Prior
1	788	0.1347
0	5064	0.8653

Figure 7 Prior Probabilities setting for Oversampling

The data is partitioned into 70% training and 30% Validation for honest assessment. The dataset is oversampled (50% event and 50% non-event) as discussed above for handling the rare events. The parameter estimates will be adjusted based on the prior probabilities mentioned above. Various models such as Decision Trees, Logistic regression with forward selection, stepwise selection and backward elimination are built for predicting the readmission. The Modeling diagram is as follows.

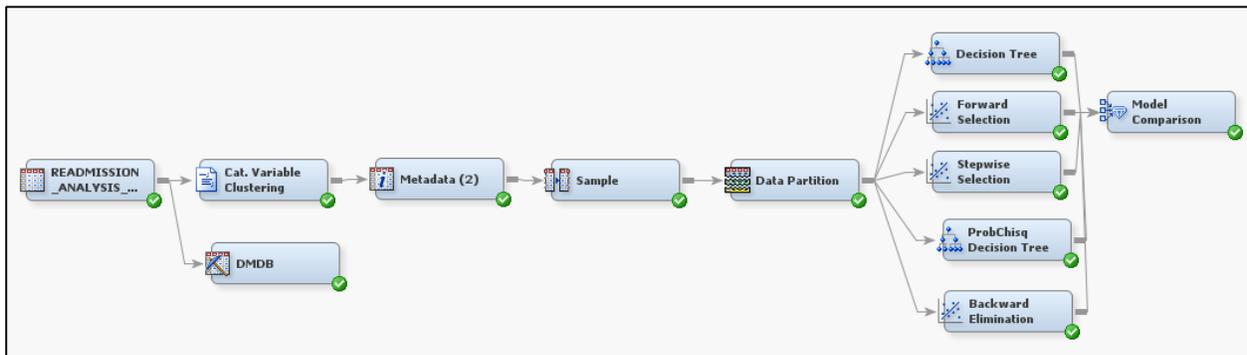


Figure 8 Modeling diagram for readmission analysis

MODEL FOR NORTHEAST REGION

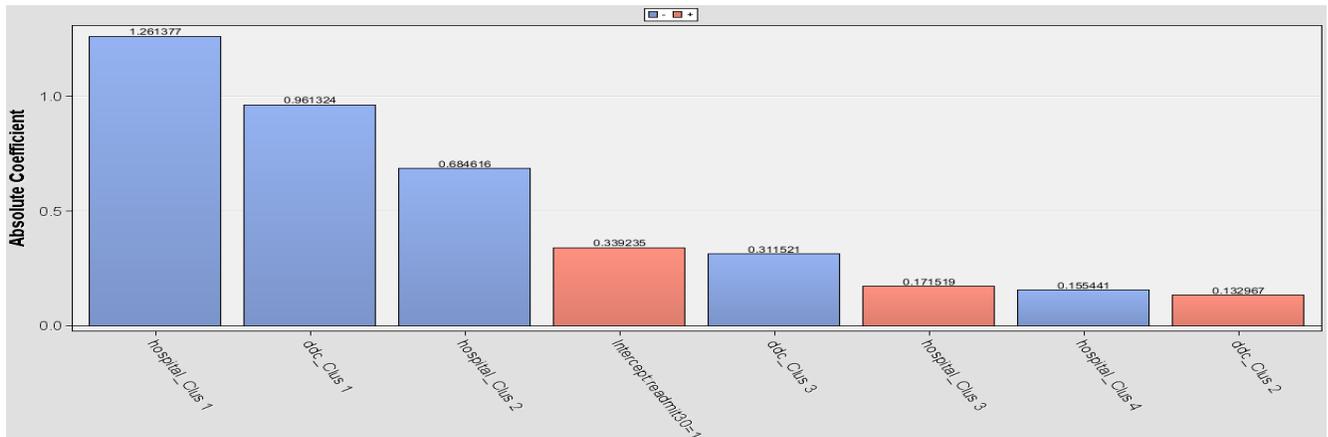
Logistic regression with stepwise selection is chosen to be the best model based on least Validation Misclassification rate (28.9%).

Selected Model	Predecessor Node	Model Description	Target Variable	Selection Criterion: Valid: Misclassification Rate
Y	Reg	Stepwise Selection	readmit30	0.289641
	Reg2	Forward Selection	readmit30	0.289641
	Reg3	Backward Selection	readmit30	0.291755
	Tree	Decision Tree	readmit30	0.501057
	Tree2	ProbChisq Decision Tree	readmit30	0.501057

Output 1 Fit Statistics of Model for Northeast region

Explaining the best model (Logistic Regression with Stepwise selection)

Stepwise selection criteria is used for variable selection in the logistic regression. This method resulted in the model with the least validation misclassification rate.



Output 2 Parameter Estimates of Logistic Regression with Stepwise selection

The effects plot of the important variables is as shown above. The variable 'hospital_clus1' has the highest parameter estimate of 1.2614 (-1.2614 is the parameter estimate). The important predictors are as follows.

- Discharge disposition code cluster 1, which consists of discharge dispositions "Discharged/Transferred to SNF" has lower chances of readmission compared to discharge disposition cluster 2, which consists of "transferred to another short term hospital". (Odds ratio 0.122)
- Also discharge disposition cluster 3, which consists of patients "Discharged to home" has lower chances of readmission.
- The hospital clusters 1 and 2 which consists of the hospitals which does not have Cath Lab facilities have lower chances of readmissions compared to cluster 3 which consists of hospitals with such facilities.

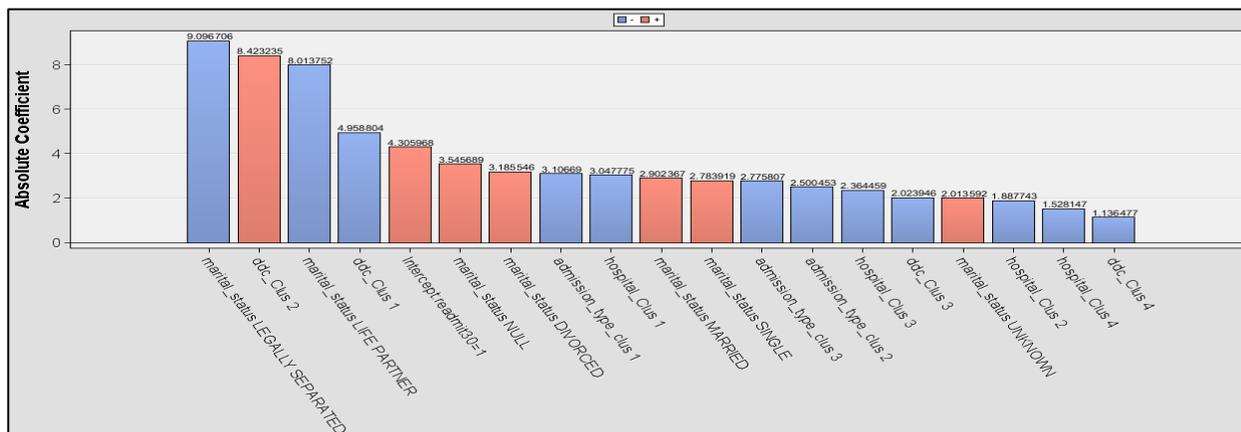
MODEL FOR SOUTH REGION

The Logistic regression model with backward elimination method is chosen to be the best model based on lowest validation misclassification rate (34%).

Fit Statistics				
Selected Model	Predecessor Node	Model Description	Target Variable	Selection Criterion: Valid: Misclassification Rate
Y	Reg8	Backward Elimination	readmit30	0.343195
	Reg4	Stepwise Selection	readmit30	0.390533
	Reg5	Forward Selection	readmit30	0.390533
	Tree3	Decision Tree	readmit30	0.502959
	Tree4	ProbChisq Decision Tree	readmit30	0.502959

Output 3 Fit Statistics of Model for South region

Explaining the best model (Logistic Regression with Backward Elimination)



Output 4 Parameter Estimates of Logistic Regression with Backward Elimination

The effects plot for the important variables is as shown. Marital Status “Legally_separated” has lower chances of readmission compared to married patients. The important predictors are as follows.

- Discharge disposition code cluster 1, which consists of discharge dispositions “Discharged/Transferred to SNF” has lower chances of readmission compared to discharge disposition cluster 2, which consists of “transferred to another short term hospital”. (Odds ratio 0.10)
- Also discharge disposition cluster 3, which consists of patients “Discharged to home” has lower chances of readmission.
- The patients with marital status “Married” has around 25% higher chances of readmission compared to “Widowed”. (Odds ratio 1.25)

Association analysis of procedures

Through both the models, we observe that the hospitals with Cath Lab facilities are having more readmission likelihood. Thus analyzing procedures performed at such hospitals is necessary in order to better understand the readmission scenario. Following settings were used for association analysis.

Association	
Maximum Items	2
Minimum Confidence Level	1
Support Type	Count
Support Count	20
Support Percentage	5.0
Sequence	
Chain Count	3
Consolidate Time	0.0
Maximum Transaction Duration	0.0
Support Type	Percent
Support Count	.
Support Percentage	2.0
Rules	
Number to Keep	200
Sort Criterion	Default
Number to Transpose	200
Export Rule by ID	Yes
Recommendation	No

Figure 9 Association Node setting

The association rules are as follows.

Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule
22.73	72.73	13.99	3.20	40	Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours ==> Insertion of Endotracheal Tube
19.23	61.54	13.99	3.20	40	Insertion of Endotracheal Tube ==> Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours
22.73	28.35	12.59	1.25	36	Venous Catheterization, Not Elsewhere Classified ==> Insertion of Endotracheal Tube
44.41	55.38	12.59	1.25	36	Insertion of Endotracheal Tube ==> Venous Catheterization, Not Elsewhere Classified
44.41	58.18	11.19	1.31	32	Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours ==> Venous Catheterization, Not Elsewhere Classified
19.23	25.20	11.19	1.31	32	Venous Catheterization, Not Elsewhere Classified ==> Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours
11.19	36.92	8.39	3.30	24	Insertion of Endotracheal Tube ==> Continuous Invasive Mechanical Ventilation for Less Than 96 Consecut
22.73	75.00	8.39	3.30	24	Continuous Invasive Mechanical Ventilation for Less Than 96 Consecut ==> Insertion of Endotracheal Tube
8.39	38.18	7.34	4.55	21	Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours ==> Temporary Tracheostomy
19.23	87.50	7.34	4.55	21	Temporary Tracheostomy ==> Continuous Invasive Mechanical Ventilation for 96 Consecutive Hours

Output 5 Association Rules for Procedural Analysis

We see that these procedures are performed together causing the readmission.

- “Insertion of endotracheal Tube or endotracheal intubation” which involves inserting tube through the mouth down into the trachea, can result the aspiration of stomach which can in turn cause pneumonia. The patients can potentially get affected within 8 days by pneumonia [6].
- Also the patients who underwent cardiac surgery (venous cauterization procedure is the major surgical procedure carried out in open heart surgery) experienced infection caused by methicillin-resistant Staphylococcus Aureus. Mediastinal and other infections caused by methicillin-resistant Staphylococcus aureus have a significant morbidity in cardiac surgical patient [7].

CONCLUSION

We found that the patients discharged to SNF are less prone to readmissions. This insinuated the possible role of procedures carried out in these hospitals towards the readmission. We also found that the cardiovascular procedures are the probable cause of bacterial pneumonia readmissions within 2-3 weeks for Medicare patients.

The most common procedure performed was found to be Venous Catheterization which comes under cardiovascular procedure groups. The patients operated by these procedures are diagnosed with methicillin susceptible pneumonia (diagnosis code 482.41). Staphylococcus is a common bacterium causing infections. Its infection is caused through a cut in the skin (cuts carried out in surgical operations). Also it was found by research that one of the most common conditions (50%) associated with MRSA [6] are found in Pneumonia patients aged 65 years i.e. Medicare patients. This corroborates with our common medical knowledge.

Majority of the readmission cases were reported from the hospitals having full catheterization facility. This signifies the patients are infected with bacterial pneumonia when they are operated with cardiovascular surgical procedures at these hospitals.

RECOMMENDATIONS

- Hospitals with catheterization facility should be extra cautious in operating Medicare patients while performing any open cuts or other factors that may cause pneumonia.
- Extending the length of stay to monitor the development of any of the bacterial pneumonia symptoms can be costlier many times so the alternative can be that these patients may possibly be discharged to the home with healthcare facility. This alternative is supported by the data analysis as the patients who are discharged to home under medical supervision have significantly lower chances of readmission.
- In cases where the medical facility could not be provided at home, timely follow up is recommended for discharged patients who underwent procedures under cardiovascular procedure groups. Monitoring patients operated with venous catheterization and endotracheal intubation may possibly limit the readmissions due to bacterial pneumonia.

REFERENCES

1. Medical News Today. 2011. "Pneumonia Biggest Problem Following Heart Surgery." November 15, 2011. <http://www.medicalnewstoday.com/articles/237714.php>
2. Wikipedia "International Statistical Classification of Diseases and Related Health Problems". Available at https://en.wikipedia.org/wiki/International_Statistical_Classification_of_Diseases_and_Related_Health_Problems
3. Fan, Weifeng and Sarfarazi, Maryam 2014. "SAS®Solutions to Identifying Hospital Readmissions" *Proceedings of the SAS Global Forum 2014 Conference*, 16622-2014, Washington, DC. :SAS Institute Inc. Available at <http://support.sas.com/resources/papers/proceedings14/1622-2014.pdf>
4. Wikipedia "Cath Lab" available at https://en.wikipedia.org/wiki/Cath_lab
5. "How to Implement Oversampling in SAS® Enterprise Miner" discussions on SAS community. Available at <https://communities.sas.com/message/155166>
6. Rello J, Soñora R, Jubert P, Artigas A, Ruém, Vallés J "Pneumonia in intubated patients: role of respiratory airway care." Available at <http://www.uptodate.com/contents/endotracheal-tube-management-and-complications/abstract/14>
7. Carrier M1, Marchand R, Auger P, Hébert Y, Pellerin M, Perrault LP, Cartier R, Bouchard D, Poirier N, Pagé P. , January 2002, "Methicillin-resistant *Staphylococcus aureus* infection in a cardiac surgical unit", NCBI, Available at <http://www.ncbi.nlm.nih.gov/pubmed/11782754>

DISCLAIMER

The authors do not possess medical expertise. The results are based purely on the data and are not intended to serve as a medical advice.

ACKNOWLEDGMENTS

We thank the Center for Health Systems Innovation for providing us the Cerner Data for analysis of Bacterial pneumonia patients.

CONTACT INFORMATION

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

Dr. Goutam Chakraborty
Oklahoma State University
Stillwater, OK, 74078
goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is Ralph A. and Peggy A. Brenneman professor of marketing and founder of SAS® and OSU data mining certificate and SAS® and OSU marketing analytics certificate at Oklahoma State University. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired 2007 data mining conference. He has over 25 years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.

Dr. William Paiva
Center of Health Systems Innovation
Tulsa, OK, 74107
wpaiva@okstate.edu

Dr. William Paiva is Executive Director of Oklahoma State University's Center for Health Systems Innovation (CHSI). He is a health and biotech venture capitalist who has served as a board of director for the CHSI since its establishment in 2012. He has than 25 years of venture-capital, biomedical-industry and management-consulting experience. He specializes in identifying early-stage venture opportunities at the intersection of traditional information technology sectors and health care.

Heramb Joshi
Oklahoma State University
Stillwater, OK, 74075
heramb.joshi@okstate.edu

Heramb Joshi is a graduating Master's student in Management Information Systems from Oklahoma State University. He has a bachelor's degree in Electronics and Telecommunication engineering from Mumbai University. He worked as a graduate research assistant at CHSI Tulsa during June – August 2015. Earlier he worked as an ETL developer for 3 years managing data warehousing projects at L&T Infotech Ltd. Mumbai, India. He has two years' experience in using SAS® tools for Database Marketing and Data Mining. He is a Base SAS® 9 certified professional, SAS Certified Advanced Programmer for SAS 9 Credential, SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling Credential, and holds the SAS and OSU Data Mining certification.

Aditya Sharma
Oklahoma State University
Stillwater, OK, 74075
aditya.sharma@okstate.edu

Aditya Sharma is a Master's graduate in Management Information Systems from Oklahoma State University. He has a bachelor's degree in Information Technology from Mumbai University. He has two years' experience in using SAS® tools for Database Marketing and Data Mining. He is a Base SAS® 9 certified professional, SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling Credential, and holds the SAS and OSU Data Mining certification. He is currently working as an analyst at NXP Semiconductors, Austin.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.