



Article Type: Research Article

Available online: www.ujmrst.twistingmemoirs.com

PATIENTS CHARACTERISTICS AND LENGTH OF HOSPITAL STAY

*Md Raihan¹, Md Rasel Mia², Md Abu Sufian³, Fatima-tuz-zohra⁴, Khondoker Monir Hossen⁵,
Md Delowar Hosain⁶, Fatima Amin⁷

^{1,2, 5,6,7}Bachelor of Public Health, German University Bangladesh, ³Department of Epidemiology,
National Institute of Preventive and Social Medicine, ⁴Department of Public Health, North South
University

Corresponding author: Md Raihan, Email: monoarmunna@yahoo.com

ABSTRACT

Data mining techniques are useful for analyzing data from many different dimensions and for identifying relationships. Association between patient characteristics and length of hospital stay using data mining technique was my ultimate goal. Data were collected from hospital records. Convenient sampling technique was used to collect data on the basis of inclusion and exclusion criteria and written consent was taken prior to interview. Descriptive as well as inferential statistics such as Chi-square were used to present data and show association. Nearly half (48%) of the patients were diagnosed as aortic sclerosis. Compressive chest pain, hypertension, intermittent chest pain, respiratory distress, central chest tightness, shortness of breath, sweating was major symptom. Moreover, mean hemoglobin was 12.25 gm/dl. In addition to this most of the patients stayed <2 days in hospital. No significant association was found between age group and length of hospital stay as well as sex. Statistically significant association was found between hemoglobin level and length of hospital stay. Every hospital should use their data-by-data mining technique to do better management of patients. If it could be done on a large scale the result would be more precise.

Keywords: Patients' characteristic, Length of hospital stay, Data mining technique

INTRODUCTION

The major reason that data mining has attracted a great deal of attention in information industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration. Data mining can be viewed as a result of the natural evolution of information technology. An evolutionary path has been

witnessed in the database industry in the development of the following functionalities: data collection and database creation, data management (including data storage and retrieval, and database transaction processing), and data analysis and understanding (involving data warehousing and data mining). For instance, the early development of data collection and database creation mechanisms served as a prerequisite for later development of effective mechanisms for data storage and retrieval, and query and transaction processing. With numerous database systems offering query and transaction processing as common practice, data analysis and understanding has naturally become the next target. Data mining in higher education is a new emerging field [1]. There are many works in this area pursued which proves its potentials to educational institutes. Romero and Ventura have surveyed on educational data mining research works during the period of 1995 and 2005 [2]. They conclude that educational Data mining is promising area of research and it has a specific requirements not presented in other domains. Merceron and Yacep provided a case study in where data mining is used to identify behaviour of failing students to warn students at risk before final exam. Followed by their work, Tripti Arjariya, Vijay K Chaudhari, Dinesh Varshney studied data mining in educational area to enhance educational process in higher educational system by improving its decision making process [3]. Finally, Waiyamai successfully explored data mining to assist in development of new curricula to help engineering students for selecting appropriate major courses in [4]. Time series data mining is an important problem and has many practical applications, limited research has been done in the past. The common practice is to use existing classification systems (e.g., decision trees and Naïve Bayes) to solve the problem. Bing Liu, Yiming Ma, and Ching Kian Wong and Philip S. Yu has shown in the scoring the data using association rules by their research [5] but it has no relation in time series. Comorbidity of two or more kinds of shock is the most important factor of length of ICU stay in the studied sample. Since there are differences of ICU patient characteristics between wards and hospitals, consideration of the data-mining technique should be given by the intensive visits as a length of ICU stay prediction tool [6]. Baseline information of the length of ICU stay can help to make appropriate decisions about the number of ICU beds; it can also result in better communications with patients and relatives because it is a key determinant of the costs of intensive care. Thus, the development of a framework to predicate the length of ICU stay is needed and will be useful in supporting clinical judgment in decision making for individual patients. In Bangladesh this type of study is very much needed to come an accurate clinical decision which is beneficial for both hospital authority as well as patients perspective. Strong data system as well as proper use of statistics will help hospital authority to develop their plan. The present study is an attempt to solve this problem.

METHODS

This was an observational cross-sectional study following the quantitative methods of data collection and analysis. This study was designed to grab more data in a short time, so that it can be used for assessing the association between patient's characteristics (age, sex, hemoglobin level, serum creatinine, erythrocytic sedimentation rate) and length of hospital stay using data mining technique. Data were collected from Ibrahim Cardiac Research Institute through secondary source (Hospital record). This study was conducted for a period of six month started from June 2014 to December 2014. Study population was all cardiac patients admitted for treatment in Ibrahim Cardiac Research Institute for treatment and diagnostic purpose with different clinical conditions. Non probability convenient sampling method was used to select sample population. Data were collected from hospital record. After data collection, data were sent to the researcher, which was sorted, scrutinized by the researcher herself by the selection criteria and then data were analyzed by personal computer by SPSS version 20.0 program. The open ended questions were grouped and categorized. Data were analyzed by descriptive statistics and inferential statistics. Descriptive statistics means mean, standard deviation (SD), minimum, maximum, frequency, percentage etc. Though my one of the research objective was

to show association between patient characteristics and length of hospital stay so I had to do Chi-square test. Chi-square is an inferential statistics to show association between two qualitative or categorical variables. In this study I have shown association between age group and length of hospital stay, sex and length of hospital stay, hemoglobin status and length of hospital stay, serum creatinine level and length of hospital stay, ESR and length of hospital stay. First of all it was continuous variable and then for analysis purpose it was transformed into categorical variables. Limitations of studies are very common in social work. One of the major difficulties was the time constraint. Data were obtained from secondary source. So there might be some discrepancy.

RESULTS

The cross-sectional study was conducted to assess association between patient's characteristics and length of hospital stay using data mining technique. The result of the study was presented in tabular and graphical form there by interpretation of the result under the following headings. The figure shows that 58%, 41% and 1% respondents came from 31-60 yrs, >60 yrs and <30 yrs age group respectively.

Age group of respondents

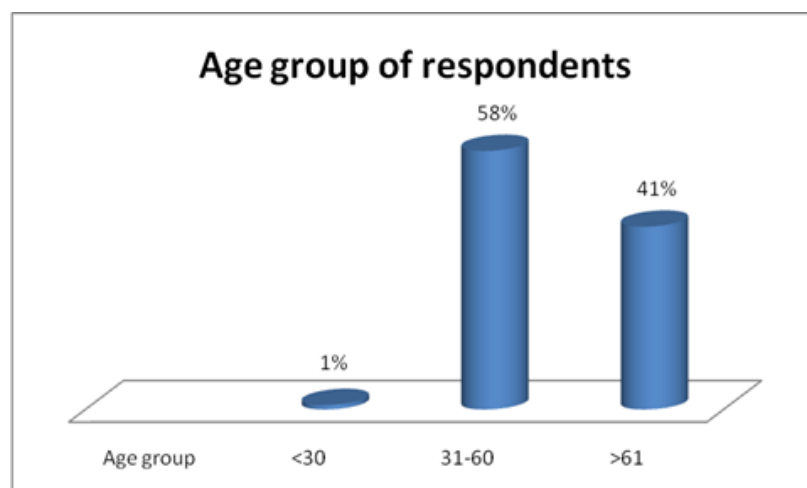


Figure 1. Distribution of respondents by age

Sex of respondents

Figure 2 shows that male and female distribution was 67% and 33%.

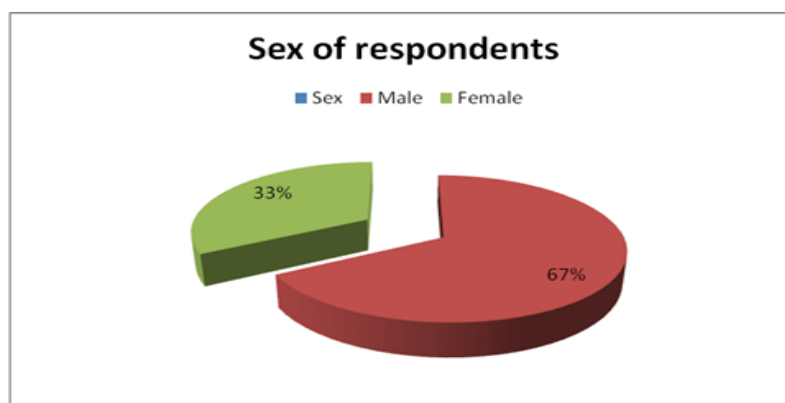


Figure 2. Distribution of respondents by sex

Table 1. Diagnosis of patients

Diagnosis	Frequency	Percentage
Aortic sclerosis	48	48
PCI	15	15
Hypothyroid	4	4
Br. asthma	4	4
CAD mild ISR	1	1
CAD non critical LAD	2	2
Adenocarcinoma of prostate	1	1
Sinus tachycardia	6	6
Respiratory distress	1	1
Stroke	1	1
IHD	11	11
Chronic kidney disease	2	2
RCA disease	2	2
Bradycardia	1	1
Hysterectomy	1	1

Table 1 shows that nearly half (48%) of the patients were diagnosed as aortic sclerosis. PCI, Ischemic heart disease (IHD), Sinus tachycardia, Br. Asthma was 15%, 11%, 6%, 4% consecutively.

Table 2: Chief complains of patients

Chief complain	Frequency	Percentage
Compressive chest pain	28	28
Hypertension	11	11
Intermittent chest discomfort	7	7
Respiratory distress	4	4
Uneasiness and feeling cool periphery of the body	1	1
Back pain	2	2
Palpitation and chocking sensation	4	4
Central chest tightness	10	10
Burning chest pain	4	4
Acute abdominal pain	1	1
Shortness of breath	13	13
Vertigo	3	3
Passing of black stool	1	1
Unconsciousness	1	1
Sweating	5	5
Lateral chest pain	1	1
Left leg pain	1	1
Orthopnea	2	2
Nausea & vomiting	1	1

Table 2 shows that compressive chest pain, hypertension, intermittent chest pain, respiratory distress, central chest tightness, shortness of breath, sweating was 28%, 11%, 7%, 4%, 10%, 13%, 5% respectively.

Table 3: Co-morbidity of patients

Co-morbidity	Frequency	Percentage
Hypertension	16	16
Diabetes	24	24
Hypertension+diabetes	33	33
Hypertension+dyslipidemia	14	14
Hypertension+diabetes+ dyslipidemia	10	10
Hyperthyroidism	3	3

Table 3 shows that 16%, 24%, 33%, 14%, 10%, 3% patients had hypertension, diabetes, hypertension+diabetes, hypertension+dyslipidemia, hypertension+diabetes+dyslipidemia, hyperthyroidism.

Table 4: Hemoglobin status of patients

Hb	gm/dl
Mean	12.25
Minimum	7.40
Maximum	15.80

Table 4 shows that mean hemoglobin was 12.25 gm/dl whereas minimum hemoglobin was 7.40 gm/dl and maximum was 15.80 gm/dl.

Table 5: ESR status of patients

ESR	mm
Mean	23
Minimum	5
Maximum	102

Table 5 shows that mean ESR was 23 mm whereas minimum ESR was 5 mm and maximum was 102 mm.

Figure 3: Length of hospital stay

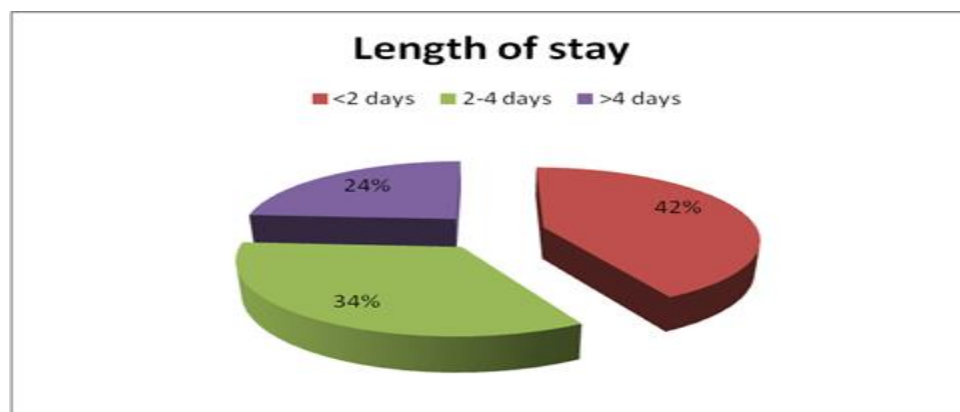


Figure 3 shows that 42%, 34% and 24% of patients stayed in hospital <2 days, 2-4 days and > 4 days respectively.

Table 6. Association between age group and length of hospital stay

Length of hospital stay in days	Age group in years			χ^2	p value
	<30	30-60	>60		
<2	1(1)	26(26)	15(15)	2.51	0.64
2-4	0(0)	20(20)	14(14)		
>4	0(0)	12(12)	12(12)		

Table 6 shows that no significant association was found between age group and length of hospital stay ($p=0.64>0.05$).

Table 7. Association between sex and length of hospital stay

Length of hospital stay in days	Sex		χ^2	p value
	Male	Female		
<2	25(25)	17(17)	1.87	0.39
2-4	25(25)	9(9)		
>4	17(17)	7(7)		

Table 7 shows that no significant association was found between sex and length of hospital stay ($p=0.39>0.05$).

Table 8. Association between hemoglobin level and length of hospital stay

Hemoglobin level gm/dl	Length of hospital stay			χ^2	p value
	<2	2-4	>4		
Anemic (<11.5)	19(19)	6(6)	5(5)	8.075	0.018
Normal (>11.5)	23(23)	28(28)	19(19)		

Table 8 shows that significant association was found between hemoglobin and length of hospital stay ($p=0.018<0.05$).

Table 9. Association between serum creatinine and length of hospital stay

Serum creatinine mg/dl	Length of hospital stay			χ^2	P value
	<2	2-4	>4		
Normal (0.6-1.2)	34(34)	24(24)	16(16)	1.932	0.381
Abnormal (>1.2)	8(8)	10(10)	8(8)		

Table 9 shows that no significant association was found between serum creatinine and length of hospital stay ($p=0.381>0.05$).

DISCUSSION

As computerization in healthcare services increase, the amount of available digital data is growing at an unprecedented rate and as a result healthcare organizations are much more able to store data than to extract knowledge from it. Today the major challenge is to transform these data into useful information and knowledge. It is important for healthcare organizations to use stored data to improve quality while reducing cost. The present study found that 42%, 34% and 24% of patients stayed in hospital <2 days, 2-4 days and > 4 days respectively. No significant association was found between age group and length of hospital stay. Moreover no significant association was found between sex and length of hospital stay. A study shows that LOS outliers decreased in the beginning of last decade but significantly increased after that. They also verified that readmission rates increased between 2000 and 2007 and started to decrease after that. These factors, the increase in LOS outliers and high readmission rates, can contribute to an important portion of hospital costs and therefore should be considered by hospital managers and health policy makers. They confirmed that emergent surgical admissions have significantly more outliers than planned surgical admissions. Moreover, patient age, the presence of specific co-morbidities, and the discharge status, visibly influence LOS outliers. Other important results

are those related to hospital type. Using different hospital grouping variables, and after adjustments for the patients' characteristics in the multivariate model, they only found statistically significant differences between teaching groups. All the three hospital related variables seem to have influence in LOS outliers but only large undergraduate teaching hospitals (in hospital teaching groups) have significantly more outliers (4.5%). In the other groups, they found that central hospitals (administrative groups) have more outliers than others hospitals, and that hospitals with higher technology, specialized and complex (Group I, economic groups) have also more outliers [7]. The proportion of LOS outliers is lower than that found in a study in Catalonia (Spain) for discharges in 1998 (4.5% outliers) [8]. In another study [9] the proportion of high outliers is even higher but, in that case, they used hospital real costs and not an approximation to cost through LOS. These differences cannot be easily explained given the possible differences in hospital case mixes of these studies, among other structural or health policy differences. The proportion of outliers does not seem to be related to their financial coverage: if anything, the proportion of high length of stay outliers increased in the years they were disregarded in public funding, from 3.9% in 2006 to 4.3% in 2009. The evolution of case mix indexes in the several economic groups may support this result: in all groups, case mix indexes are increasing over the years although in Group 1 this evolution is quite irregular and the index even diminishes in 2006, picking up in the following years. If hospitals had been able to easily control the volume of high LOS outliers, then it would be expectable to find some influence in the evolution of the case mix over the years. In fact, after 2006, case mix indexes consistently increased despite the lack of specific funding for inpatient days above the high trim point. Actually, hospital doctors normally try to avoid the long duration of the stays. They are typically not conscious of any resulting extra value on funding, but they are aware of the greater likelihood of inpatient complications, being the decrease of the average LOS one of their main concerns. The most important factor that influences the length of a hospital stay is the number of medical staff, such as medical laboratory scientists, physical therapists and occupational therapists. With regard to the services offered by a hospital, the physicians provide the direct health care services needed to improve patients' health, with medical staff providing supportive health care services. The number of medical staff employed is based on the organizational structure of a specific hospital, and reflects both the operating scale and services provided. For example, a small hospital with less than five physicians may only provide primary acute health care services, without a laboratory scientist or physical therapist. Therefore, patients with an expected length of stay over 30 days will prefer to seek health care services at medium or large hospitals, as these have more medical staff and better equipment to serve more health care services [10]. The present study found that 58%, 41% and 1% respondents came from 31-60 yrs, >60 yrs and <30 yrs age group respectively and male and female distribution was 67% and 33%. The online publication of information on the quality of medical care services offers patients a chance to reduce information asymmetries existed between patient and physician and to seek services. However, access to outside medical information is linked to a patient's socioeconomic status [11-12], in which can create a digital divide. Policy makers should thus target educational efforts to decrease or eliminate the persistent healthcare disparities among different socioeconomic groups. Educating and empowering e- health care consumers through online information enables them to become active participants in their own health care, thus potentially resulting in higher satisfaction.

CONCLUSION

Length of stay of patients in a hospital is important in providing them with better services and higher satisfaction, as well as helping the hospital management plan and managing hospital resources as meticulously as possible. Nearly half (48%) of the patients were diagnosed as aortic sclerosis. PCI, Ischemic heart disease, Sinus tachycardia, Br. Asthma was 15%, 11%, 6%, 4% consecutively. Besides compressive chest pain, hypertension, intermittent chest pain,

respiratory distress, central chest tightness, shortness of breath, sweating was 28%, 11%, 7%, 4%, 10%, 13%, 5% respectively. Moreover mean hemoglobin was 12.25 gm/dl whereas minimum hemoglobin was 7.40 gm/dl and maximum was 15.80 gm/dl. In addition to this 42%, 34% and 24% of patients stayed in hospital <2 days, 2-4 days and >4 days respectively. No significant association was found between age group and length of hospital stay as well as sex.

REFERENCES

1. Romero C and Ventura S. Educational Data mining: A survey from 1995 to 2005. *Expert systems With Application* 2007; (33): 135-146.
2. Waiyamai K. Improving Quality Graduate Student by Data Mining. Department of Computer engineering. Faculty of Engineering Kasetsart University, Bangkok Thailand, 2003.
3. Delavari N, and Beikzadeh MR and Amnuaisuk SP. Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. ITHET 6 th Annual International Conference. Juan Dolio, Dominican Republic. July 7 – 9, 2005.
4. Huan Wu, Zhigang Lu, Lin Pan, Rongsheng Xu. An Improved Apriori-based Algorithm for Association Rules Mining. Chinese Academy of Sciences Beijing 100049, China
5. Kumar KS, Pollachi R. A Survey on Association Rule Mining using Apriori Algorithm. *International Journal of Computer Applications* (0975 – 8887) Volume 45– No.5, May 2012
6. Xiao-Chun Zhang, Zhi-Dan Zhang, De-Sheng Huang. Prediction of Length of ICU Stay Using Data-mining Techniques: an Example of Old Critically Ill Postoperative Gastric Cancer Patients. *Asian Pacific J Cancer Prev*, 13, 97-101.
7. Freitas A et al. Factors influencing hospital high length of stay outliers. *BMC Health Services Research* 2012, 12:265.
8. Cots F, Mercadé L, Castells X, Salvador X: Relationship between hospital structural level and length of stay outliers: Implications for hospital payment systems. *Health Policy* 2004, 68:159–168.
9. Pirson M, Dramaix M, Leclercq P, Jackson T: Analysis of cost outliers within APR-DRGs in a Belgian general hospital: Two complementary approaches. *Health Policy* 2006, 76:13–25.
10. Colledge A, Car J, Donnelly A and Majeed A. Health information for patients: Time to look beyond patient information leaf lets. *Journal of the Royal Society of Medicine* 2008; 101: 447-453.
11. Lopez L, Green AR, Tan-McGrory A, King R and Betancourt JR. Bridging the digital divide in health care: The role of health information technology in addressing racial and ethnic disparities. *Joint Commission Journal on Quality and Patient Safety* 2011; 37: 437-445.
12. Appelros P. Prediction of length of stay for stroke patients. *Acta Neurol Scand* 2007; 116 (1):15-9.