Universal Journal of Public Health 11(1): 177-184, 2023 http://www.hrpub.org
DOI: 10.13189/ujph.2023.110119

Generalized Linear Model to Estimate Length of Stay in The Hospital due to Respiratory Diseases

Siti Wafiah Hanin Mohd Zulkifli, Humaida Banu Samsudin*, Noriza Majid

Department of Mathematical Science, Faculty of Science and Technology, National University of Malaysia, Malaysia

Received September 28, 2022; Revised November 30, 2022; Accepted December 15, 2022

Cite This Paper in the Following Citation Styles

***(a)**: [1] Siti Wafiah Hanin Mohd Zulkifli, Humaida Banu Samsudin, Noriza Majid, "Generalized Linear Model to Estimate Length of Stay in The Hospital due to Respiratory Diseases," Universal Journal of Public Health, Vol.11, No.1, pp. 177-184, 2023. DOI: 10.13189/ujph.2023.110119*

***(b)**: Siti Wafiah Hanin Mohd Zulkifli, Humaida Banu Samsudin, Noriza Majid, (2023). Generalized Linear Model to Estimate Length of Stay in The Hospital due to Respiratory Diseases. Universal Journal of Public Health, 11(1), 177-184. DOI: 10.13189/ujph.2023.110119*

Abstract Several studies reported various factors associated with length of stay in the hospital due to respiratory diseases. This study aims to select the best Generalized Linear Models (GLM) for predicting LOS and identify the effects of clinical and demographic factors on LOS. The study was carried out using data on registered admission and discharge at the government hospitals for central region states in Malaysia. A total of 526,511 cases were classified under diseases of respiratory systems coded J00 to J99 and factors such as gender, age group, ethnicity, marital status, discharge condition and diagnoses of the patients were used in this study. Two regression models were used to predict LOS: GLM Zero Truncated Poisson regression (ZTP) and GLM Zero Truncated Negative Binomial (ZTNB). The best count fit model was chosen according to Akaike's Information Criteria (AIC) and Bayesian Information Criterion (BIC). The median length of stay in the study was four days (Interquartile Range: three to six days). According to statistical comparisons, the best model for count data is GLM Zero Truncated Negative Binomial. The outcome of the model found that there were significant changes in log LOS for each predictor except for unknown marital status and patients diagnosed with other diseases of the respiratory system that were insignificant in ZTP but significant in ZTNB. ZTNB model helps to identify factors contributing to the log length of stay so that hospitals are well equipped with the facilities and prepared for the expenditures and resources.

Keywords Generalized Linear Model, Respiratory Diseases, Length of Stay, Zero Truncated Negative Binomial

1 Introduction

Respiratory diseases are often given attention as they commonly happen in our community and contribute to death. Respiratory diseases are defined as the type of diseases that affect the respiratory system and any disorder of the airways and the lung that affect human respiration [1]. Respiratory diseases include diseases of the upper respiratory tract, influenza, pneumonia, acute lower respiratory infection, lung diseases due to external agents, suppurative and necrotic conditions of the lower respiratory tract and pneumothorax. Upper respiratory tract diseases comprise acute nasopharyngitis, acute pharyngitis, acute tonsilitis, acute laryngitis and tracheitis, chronic sinusitis and chronic diseases of tonsils. Lower respiratory tract diseases involve chronic bronchitis, emphysema, chronic obstructive pulmonary diseases (COPD), asthma and bronchiectasis [2]. Respiratory diseases that are commonly highlighted are asthma, lung cancer, pneumonia and acute chronic pulmonary obstruction. According to the Department of Statistics Malaysia [3], the mortality rate per 100,000 population in Malaysia for total respiratory diseases is 37.20% with the highest contributor 25.7% is from pneumonia consisting of viral pneumonia and other types of pneumonia. Pneumonia and chronic lower respiratory diseases are listed as the two of five leading causes of death in Malaysia with 11.8% and 2.6% respectively.

Length of stay (LOS) is important to be estimated as it is the key indicator of hospital efficiency [4,5]. The efficiency of managing the hospital is measured by having good facility management and excellent patient quality of care while being admitted. Shorter hospital stays mean that more beds are available for more patients thus improving the bed turnover. Hence, patients with a shorter period of stay represent lower hospital consumption as well as reducing the expenditure of the hospital. Identifying LOS is associated with reduced risk of opportunist infections, side effects of medications, improved patient discharged conditions and reduced patients' mortality. Determining the factors contributing to longer or shorter lengths of stay may add an effort to control the resources and prepare the hospital for massive admissions for longer lengths of stay [6-8]. Hospital length of stay in the case of respiratory diseases can be caused by various factors such as clinical factors, comorbidity factors, and demographic factors [9-11]. Three forms of data can be used in LOS models according to Herwartz et al. [12]: i) data available after admission for example therapy, comorbidity, physical and mental problems, death status and insurance status ii) data collected when admitted to hospital such as age, sex, marital status, smoking habits, alcohol assumptions or health status during admission and iii) data from existing diseases, socioeconomic variables needed to identify the recurrence diseases. All these data are useful to maximize short-term resource utilization and prevent the recurrence of hospital admission in the long term. Therefore, to determine the factors associated with LOS, both demographic and clinical parameters are considered in this study.

Researchers explored LOS data with different objectives of the studies such as Freitas et al. [13] that used logistic regression with generalized estimating equation (GEE) to detect the factors influencing the high outliers in LOS and Nash et al. [14] used survey-method logistic regression to investigate prolonged LOS in Emergency Department by exploring patient-level characteristics. Many methods used to predict LOS and association factors are found in their studies such as Kulinskaya et al. [15] that used advanced robust method with truncated maximum likelihood to elucidate the factors influencing LOS, Verburg et al. [16] that used 8 regression models for predicting ICU LOS, Tsung et al. [17] having a prospective study by comparing univariate comparisons using non parametric Kruskal-Wallis test to identify the associated factors LOS among children and Dong et al. [18] that had similar approaches with Tsung et al. [17], with negative binomial regression model was fitted first before having Kruskal-Wallis test. The study conducted by Wen et al. [19] predicted LOS for Covid patients by using time-to-event modelling by having 6 survival models which are Cox linear, Deep Surv, Cox-CC, Random Survival Forest, Deep Hit and Multitask logistic regression. A study at Tamale hospital found that the best model is the Negative Binomial model in terms of influence on the expected length of stay in the hospital for malaria [20]. Previous study in Portugal used a multilevel model which includes the Poisson regression model, zero-inflated Poisson, negative binomial regression, zero-inflated negative binomial and random-effect model. A study found that zero-inflated negative binomials and the random-effect model showed the most appropriate model with a significant decrease in the length of stay. This is because, for the count data, using OLS and logistic regression can contribute to overdispersion hence creating a biased estimate and does not reflect the data [21]. Hence Poisson Regression Model is used for count data which highlights two main assumptions. Assumptions are mean and variance of this distribution must be equal, and the observation must be independent of each other. However, the equality of mean and variance can be violated as more variation in response compared to the model caused overdispersion. There are two ways to cater overdispersion whether by inflating standard error from the estimated dispersion factor or by using negative binomial regression [22]. Negative Binomial Regression model is used to encounter overdispersion since it has the benefit of introducing another parameter but with the same mean structure as Poisson. If overdispersion occurs, the confidence interval will get narrowed if compared to Poisson Model. Another method for handling count data is Zero Truncated Poisson and Zero Truncated Negative Binomial considering the dataset cannot be zero since the length of stay commence with the count of 1. Ordinary Negative Binomial will have problems with truncated data because it will include zero in the distribution. Meaning that, the mean of response contains no zero, but standard error and parameters might be biased because the model has tried to predict zero counts [21].

In this study, the length of stay in the hospital should be more than 0 day. Hence, GLM with Zero Truncated Models were used because of the count data for which zero value cannot occur and overdispersion exist. Zero Truncated Poisson and Zero Truncated Negative binomial are compared to find the best model estimating the length of stay due to respiratory diseases and investigate interaction effect to the length of stay in the hospital due to respiratory diseases. Investigating the factors that affect LOS is important in identifying how it affects LOS and determining how to optimize admission-related expenses, medical supplies, and facilities of the hospitals.

2 Materials and Methods

2.1 Data Collection

Data for this study are obtained from the Ministry of Health Malaysia, with approval from the Health Information Centre Malaysia. To protect the confidentiality, only anonymous datasets are included in the analyses described. Data on patients discharged due to respiratory diseases between 2014 to 2020 were collected focusing on the Central region of Malaysia, Selangor, Wilayah Persekutuan Kuala Lumpur and Wilayah Persekutuan Putrajaya. Data involved the patient's gender, marital status, ethnicity, discharged conditions and class of respiratory diseases. These data considered only patients hospitalized with respiratory diseases listed in Malaysia Health Indicators coded from J00 to J99. Data from this study only involved with the patients admitted in the government hospitals situated in Central region, Malaysia.

2.2 Outcome and Variables

The dependent variable of this study is the LOS (non-negative integer) which is defined as the number of days between a patient's admission and the day patients are discharged. Length of stay of hospital admission is recorded as count (n) and LOS data is truncated at 28 days which is based above 27 days at 99% percentile. Meanwhile, the determinant of this study includes the patient's gender, age, marital status, ethnicity, discharge conditions and diagnoses. Multivariables are used in this study by grouping, gender (0 = Male, 1 = Female) and age groups (0 -14 : young age, 15-64 : working age, 65 and above : old age). The age group classification follows the categorization of 3 main age groups in Malaysia listed in the Department of Statistics Malaysia. According to DOSM in the analysis of Labor Force Survey (LFC) in Malaysia, working age in Malaysia refers to the age structure of the economically active population, which is classified to be between the ages of 15 and 64. Other than that marital status (0 = Single, 1 = Married, 2 = Widowed, 3 = Divorced, 4 = Others) the definition of others in the marital status includes individuals with unknown relationships and does not fall under either of the marital categories listed. The fourth factor is ethnicity (0 = Malay,1 = Chinese, 2 = Indian, 3 = Others) with others that include the minority of the ethnics such as Iban, Bajau, Melanau, Murut, Bumiputera Sabah dan Sarawak, Orang Asli and others. The discharge conditions are categorized with (0 = Went Home, 1 = Transfer to other ward and facility, 2 = Discharge at patient's own risk and refuse treatment, 3 = Discharge without permission from the doctors, and 4 = Death). Diagnoses of respiratory diseases are divided into 9 categories which are (0 = Influenza, pneumonia and other acute lower respiratory infections, 1 = Diseases of the upper respiratory tract, 2 = Other diseases of upper respiratory tract, 3 = Chronic lower respiratory disease, 4 = Lung diseases due to external agent, 5 = Other respiratory diseases principally affecting the interstitium, 6 = Suppurative and necrotic conditions of lower respiratory tract, 7 = Other diseases of pleura, 8 = other diseases of the respiratory system). For all the categories, the reference level of the categorical variables was selected with the largest number of observations for each category to narrow the confidence interval and minimize the standard error.

2.3 Descriptive analysis and GLM regression

Descriptive analysis and univariate analysis were constructed. There are two regression models used to predict LOS for respiratory disease based on the count data which are GLM Zero Truncated Poisson regression and GLM Zero Truncated Negative Binomial. GLM is used to build a linear relationship that links the response variables and predictors to a linear model. Since the outcome variable of this study is count data hence Poisson regression is commonly used for the data type. However, Poisson distribution will try to predict zero counts even though there are no zero values and there is overdispersion by using Poisson Regression. Overdispersion in the count data is observed by looking at the dispersion value and alpha value. When the dispersion value is greater than one and the alpha value is greater than 0, overdispersion occurred. Even though Negative Binomial Regression attempted to overcome overdispersion in Poisson distribution, but it will also face difficulty with the count data as it will try to predict zero counts even though there are no zero values. Since LOS is recorded as a minimum of at least a day in the hospital and contained no zeros, hence Zero Truncated Poison and Zero Truncated Negative Binomial are used in this study. These models had been developed to overcome circumstances in which the origin of overdispersion is due to no zero values in the data.

Poisson distribution with parameter λ, the density is

$$Pr^{(P)}(Y=y)=\frac{\lambda^{y}e^{-\lambda}}{y!} \tag{1}$$

With the probability of zero is

$$Pr^{(P)}(Y=0)=e^{-\lambda} \tag{2}$$

Hence, the zero truncated Poisson distribution is considered positive values with conditional density

$$\begin{aligned}Pr^{(P)}(Y=y|Y>0)=&\frac{Pr^{(P)}(Y=y)}{1-Pr^{(P)}(Y=0)}\\=&\frac{\lambda^{y}e^{-\lambda}}{y!(1-e^{-\lambda})};y=1,2,3\end{aligned} \tag{3}$$

The log-likelihood function for ZTP is

$$L=\sum_{i=1}^{n}[y_{i}log(\lambda)-\lambda-log(1-e^{-\lambda})-log(y_{i})] \tag{4}$$

The general equation for the log of expected counts for LOS:

$$log(\lambda)=\beta_0+\beta_1x_1+\beta_2x_2+....+\beta_nx_n \tag{5}$$

Under assumption of heterogeneity, with parameter λ and α , the density for negative binomial is

$$Pr^{(NB)}(Y=y_i)=\frac{\gamma(y_i+\alpha^{-1})}{y!\gamma(\alpha^{-1})}(\frac{\alpha^{-1}}{\lambda_i+\alpha^{-1}})^{\alpha-1}(\frac{\lambda_i}{\lambda_i+\alpha^{-1}})^{y} \tag{6}$$

with probability of zero count in negative binomial is

$$Pr^{(NB)}(Y=0)=(1+\alpha\lambda)^{-1/\lambda} \tag{7}$$

Hence the probability of Zero Truncated Negative Binomial regression model is

$$Pr^{(NB)}(Y=y_i|y_i>0)=\frac{Pr^{(NB)}(Y=y)}{1-Pr^{(NB)}(Y=0)}=\frac{\gamma(y_i+\alpha^{-1})}{y!\gamma(\alpha^{-1})}\left(\frac{\alpha\gamma}{1+\alpha\gamma_i}\right)^{y_i}(1+\alpha\lambda_i)^{-\alpha^{-1}}(1-(1+\alpha\lambda_i)^{-\alpha^{-1}})^{(-1)} \tag{8}$$

The log likelihood for ZTNB is

$$log(L_{NB}) = \sum_{i=1}^{n}(\sum_{m=0}^{y_i} log(m+\alpha^{-1}) - logy_i! \\ + y_i log\alpha\gamma_i - (y_i+\alpha^{-1})log(1+\alpha\gamma_i) \\ - log[1-(1+\alpha\gamma_i)^{-\alpha^{-1}}]) \tag{9}$$

where:

$$\sum_{m=0}^{y_i} log(m+\alpha^{-1}) = log\gamma(y_i+\alpha^{-1}) - log\gamma(\alpha^{-1}) \tag{10}$$

The equation for ZTNB

$$ln(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_n x_{in} \tag{11}$$

2.4 Goodness of Fit for the Models

The best model was selected by analyzing the corresponding fit of different count models. The goodness of fit for each model was examined by comparing the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The lowest value for each criterion will indicate that the model is the most suitable to be used in this study. AIC considered the model to be closer to the truth by estimating the constant plus relative distance between the true and fitted likelihood functions [23,24]. BIC also considers the model to be a good model based on Bayesian that estimates the posterior probability of the model to be true [24,25]. All regression models and analyses were conducted using R programming version 3.6.3 with a few packages such as VGAM, ggplot2, foreign and boot. VGAM with factor is specialized software for all GLM regression [26].

3 Results and Discussion

A total of 526,511 patients were hospitalized from 2014 to 2020 with truncated LOS at 28 days. The mean of stay for the patient was 4.78±3.43 days with the median LOS for the Central region of Malaysia was 4 days. Table 1 summarizes the descriptive statistics of the variables considered in this study. Despite having the lowest percentage of admissions (17.16%) among the age group, the average LOS for elderly patients was longer if compared to young and working age groups. According to gender, males contributed more to respiratory disease than females by a difference of 8.74%. Being the dominant ethnic group in Malaysia, Malay patients have the highest incidence of respiratory diseases and all ethnicities have an average LOS of around 5 days. Apart from that, single patients were most likely exposed to respiratory diseases, but widowers and divorcees had lengthier LOS upon being hospitalized. Of all the patients in the Central region, 94.12% of the patients were discharged home and only 0.36% of them were discharged without permission from the doctors and contributing to the lowest mean of stay in the hospital. The average length of stay (LOS) in hospitals for patients who died was 7 days, which is longer than the average length of stay in hospitals generally. The largest contributors to admissions according to the diagnoses (48.01%), with a mean average LOS of 5.16 days, were patients with influenza, pneumonia, and other acute lower respiratory infections. Even though patients with lower respiratory tract, suppurative and necrotic diseases had the lowest admission rate with only 0.37%, but they have the longest average LOS of 11 days when they were admitted.

Figure 1 shows the frequency distribution of LOS at the central region of government hospitals from 2014 to 2020. The distribution of LOS skewed to the right with a median of 4, IQR of 3 days to 6 days. LOS from the histogram is shown to be heavily tailed with skewness and kurtosis that are 2.69 and 13.08, respectively.

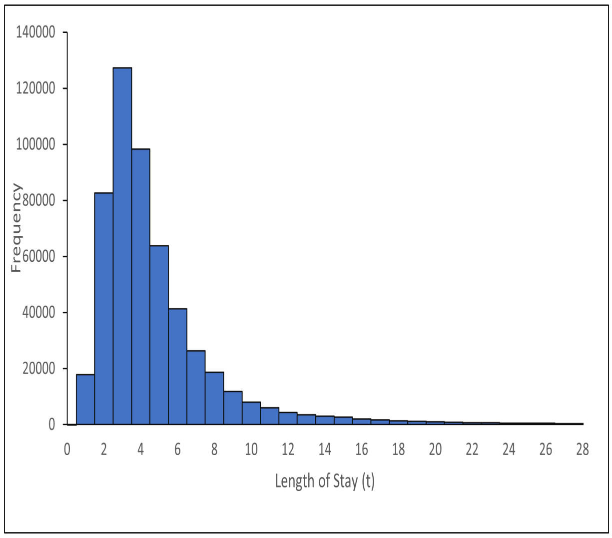

Figure 1. Frequency Distribution of Length of Stay at Central Region State

Table 2 presents the goodness of fit for each of the models that was also investigated by looking the Akaike Information Criterion AIC and the Bayesian information criterion (BIC), which estimators correspond to the best statistical models. Zero Truncated Negative Binomial has lower AIC and BIC values with 2386297 and 2386565 respectively, compared to Zero Truncated Poisson. Hence, the best model used to estimate the length of stay in the hospital due to respiratory diseases was GLM with Zero Truncated Negative Binomial. This study emphasized the day of admission must be greater or equal to 1 day and Zero Truncated Negative Binomial was the best model selected to counter the overdispersion and supported by having the lowest AIC and BIC.

Table 3 presents two GLMs that were employed to predict the length of stay for patients with respiratory illnesses in Malaysia's central region. The value coefficient for female patients in both models is -0.025, which indicates that they have 0.025 lower log count of stay than male patients. Patients with Chinese, Indian, and other ethnicities had longer log counts of stays than those of Malay ethnicity. Other ethnics contribute to the increment for log count of stay by 0.089 and 0.0964 than

Table 1. Descriptive statistics, frequency, and characteristics of the study sample (N = 526,511)

Variables	Number of Patients	% of Patients Admitted	Mean ± SD	P- value
Gender				$\leq$0.0001
Male	286,288	54.37	4.82±3.53	
Female	240,223	45.63	4.74±3.32	
Ethnicity				$\leq$0.0001
Malay	357,371	67.88	4.54±3.18	
Chinese	63,258	12.01	5.43±3.99	
Indian	87,659	16.65	5.21±3.77	
Others	18,223	3.46	5.25±3.89	
Age Group				$\leq$0.0001
Young Age (0 -14)	256,800	48.77	3.95±2.35	
Working Age (15-64)	179,379	34.07	5.15±3.79	
Old Age (65 and above)	90,332	17.16	6.39±4.43	
Marital Status				$\leq$0.0001
Single	330,981	62.86	4.22±2.82	
Married	188,043	35.71	5.71±4.1	
Widowed	2,269	0.43	6.25±4.29	
Divorced	3,640	0.69	6.44±4.41	
Others	1,578	0.30	5.20±3.89	
Discharge Conditions				$\leq$0.0001
Home	495,332	94.12	4.72±3.26	
Transfer to another ward or facility	6,233	1.18	5.4±5.05	
Refuse treatment / At Own Risk	8,956	1.70	4.15±3.94	
Without permission	1,882	0.36	3.95±2.94	
Die	13,908	2.64	7.07±6.25	
Diagnosis				$\leq$0.0001
Diseases of the upper respiratory tract	95,559	18.15	3.16±1.6	
Influenza, pneumonia, and other acute lower respiratory infections	252,775	48.01	5.16±3.5	
Other diseases of upper respiratory tract	14,610	2.77	3.73±2.45	
Chronic lower respiratory disease	4,735	0.90	4.24±2.23	
Lung diseases due to external agent	130,167	24.72	4.88±3.43	
Other respiratory diseases principally affecting the interstitium	11,999	2.28	6.51±4.48	
Suppurative and necrotic conditions of lower respiratory tract	1,961	0.37	11.39±7.09	
Other diseases of pleura	8,668	1.65	7.59±5.33	
Other diseases of the respiratory system	6,037	1.15	5.84±4.73	

Table 2. Results of AIC and BIC

Goodness of fit test	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	Loglikehood
Zero Truncated Poisson	2522901	2523158	-1261427
Zero Truncated Negative Binomial	2386297	2386565	-1193125

Malay ethnics for ZTP and ZTNB respectively. Other than that, both groups' working age and old age contribute to an increase in the log count of stay compared with the young age group. Older patients' log counts of stays are 0.03 higher than those of younger patients and the old age group has higher coefficient values than the working age group. Married, widowed and divorced were associated more with the rise in log count of LOS than single groups did. The married patient's log LOS was 0.04 greater than single patient. Additionally, widowers and divorcee patients contribute 0.08 to 0.09 more log LOS than single patients. In both models, patient's marital status stated as others was insignificantly correlated with log LOS. Patients who've been discharged to another ward or institution, those who refused treatment, and those who left without a doctor's approval had lower log count of stays are compared with those who were discharged home. In contrast, according to ZTP and ZTNB, respectively, the log count of patients who died in the hospital was 0.1366 and 0.1496 higher than that of patients who were discharged to their homes.

Patients with upper respiratory tract diseases, other respiratory tract diseases, chronic lower respiratory disease and lung diseases caused by external agents have shorter log LOS when being admitted than patients with influenza, pneumonia, and other acute lower respiratory illnesses. However, patients with other respiratory diseases principally affecting the interstitium, suppurative and necrotic conditions of the lower respiratory tract and other diseases of pleura diseases have longer log lengths of stay (0.07 for both models) than patients with influenza, pneumonia, and other acute lower respiratory conditions. Patients diagnosed with other diseases of the respiratory system were significant at $p \leq 0.05$ for ZTNB but insignificant for ZTP model. According to the first intercept values for ZTP and ZTNB, when all predictors are equal to zero, the log count of stay is 1.4847 and 1.4529, respectively. The second intercept in ZTNB has a value of 1.7543, which is the overdispersion parameter (alpha).

In this study, the median LOS for hospitals in central regions of Malaysia was four days which is comparable with findings reported in Prince Whale Hospital Hong Kong by Dong et al. [18], the median length of stay for bronchitis due to respiratory syncytial virus infection is 4 days and median length of stay for respiratory disease is 2 days, with IQR (1-3 days). For COPD, China's LOS for COPD at 10 days [19], followed by 13 European countries (7 days) [24], acute COPD exacerbation

Table 3. Regression on Length of Stay

	Zero Truncated Poisson	Zero Truncated Negative Binomial
Intercept: 1	1.4847 ***	1.4529 ***
Intercept: 2		1.7543 ***
Gender		
Male	Reference	Reference
Female	-0.0255 (0.0013) ***	-0.0257(0.0053) ***
Ethnicity		
Malay	Reference	Reference
Chinese	0.0387(0.002) ***	0.0434 (0.0028) ***
Indian	0.0556 (0.0018) ***	0.0572 (0.0024) ***
Others	0.089 (0.0034) ***	0.0964 (0.0048) ***
Age Group		
Young Age (0 -14)	Reference	Reference
Working Age (15-64)	0.2179 (0.0020) ***	0.2233 (0.0028) ***
Old Age (65 and above)	0.367 (0.0025) ***	0.3832 (0.0035) ***
Marital Status		
Single	Reference	Reference
Married	0.0390 (0.0019) ***	0.0446 (0.0027) ***
Widowed	0.0828 (0.0087) ***	0.0941 (0.0127) ***
Divorced	0.0856 (0.0069) ***	0.0941 (0.0101) ***
Others	0.0190 (0.0113)	0.01999 (0.0160)
Discharge Condition		
Home	Reference	Reference
Transfer to another ward or facility	-0.0528 (0.0056)***	-0.0556 (0.0080)***
Refuse treatment / At Own Risk	-0.2406 (0.0055) ***	-0.2576 (0.0074) ***
Without permission	-0.2687 (0.0122) ***	-0.2809 (0.0162) ***
Die	0.1366 (0.0034) ***	0.1496(0.0051) ***
Diagnosis		
Diseases of the upper respiratory tract	-0.4618 (0.0022) ***	-0.4859 (0.0028) ***
Influenza, pneumonia, and other acute lower respiratory infections	Reference	Reference
Other diseases of upper respiratory tract	-0.4025 (0.0047) ***	-0.4214 (0.0061)***
Chronic lower respiratory disease	-0.3201 (0.0074) ***	-0.3358 (0.001) ***
Lung diseases due to external agent	-0.1503 (0.0016)***	-0.1614 (0.0023)***
Other respiratory diseases principally affecting the interstitium	0.0714(0.0038)***	0.0783 (0.0056)***
Suppurative and necrotic conditions of lower respiratory tract	0.6914 (0.0068) ***	0.7157 (0.0117)***
Other diseases of pleura	0.2526 (0.0041)***	0.2662(0.0062)***
Other diseases of the respiratory system	-0.0111 (0.0055)	-0.0097 (0.0079) *

Coefficient (Standard error) significant ***0.001, *0.05

admitted at People's College of Medical Science & RC Bhopal found that the median length of stay (LOS) is 9 days [4]. Different environments, treatment strategies, or demographic samples across the countries can all be used to account for variations in LOS duration [27].

The study's findings suggested that Zero Truncated Negative Binomial, which had lower AIC and BIC values than Zero Truncated Poisson, was a better fit. A significant relationship between gender and log LOS was found in both models, even though a large number of sample sizes were used in this study. Similar results were discovered by Orooji et al. [28], who found that gender has significant effects on LOS for both univariate and regression models by evaluating the relative risk (RR) of the regressions. Baharlooci et al. [29] also discovered that multiple regression analysis demonstrated that both groups of gender were significant in predicting LOS. According to this study's findings for the age group, older patients had longer average stays than younger patients, which contributed to an increase in the log count of stays when they are admitted. Zeleke et al. [5] also discovered that the average LOS increased in the age categories of (60, 70), (70, 80), and 80+. The study conducted by Shanley et al.[30] similarly showed that extended LOS was linked to older ages (60 years).

The ethnicity of Malaysians was found to be another factor in this study, with Chinese, Indian, and other groups contributing to high log LOS by having average LOS values of more than 4 days. According to a study by Bruce et al. [31], Black Caribbean and African inpatients had longer lengths of stay (LOS) than their White British counterparts, and Black Caribbean patients continued to have the highest risk of extended lengths of stay when compared to other groups. Therefore, this study also found that different ethnicities have significant contributors to LOS. Other that that, this study found that patients who were discharged in a dead condition will have a higher mean LOS than those who were discharged alive. Similar research discovered by Orooji et al. [28] that the LOS of deceased patients was higher than that of other patients, which may have been associated with the severity of the patients.Patients who were discharged due to death were a factor that statistically lengthened the stay of elderly patients.

Factors that contributed to the increase in LOS will influence on hospital admission costs and hospitals must consider providing extra beds for patients with longer stay periods in the hospitals. Both demographics and clinical factors should be considered because they may have an impact on the length of stay (LOS) of patients. Due to study limitations, factors such as dietary status, tobacco use history, and physical activity could not be extensively studied in the current study. These factors were not fully recorded in the hospital information system. However, these factors listed in this study appear to be sensible decisions that identify the length of hospital stays.

4 Conclusion

In conclusion, the study's findings indicated that the GLM Zero Truncated Negative Binomial is the best model for LOS data. All predictors were significant when compared to their reference group for both models except for unknown marital status and patients diagnosed with other diseases of respiratory system but were significant for ZTNB. The demographic factors, such as age, marital status, and ethnicity contribute to increased log LOS when compared to their reference group. This resulted in a longer stay length being recorded, which resulted in increased hospital expenses for this group of patients. This implies that these groups will stay in the hospital for longer periods, hence they will incur higher healthcare expenses when being admitted. But female patients had a shorter log LOS when hospitalized, lower hospital costs were imposed compared to male patients. As for clinical factors, the deceased patients will have the higher length of stay if compared to the patients that were discharged home, patients diagnosed with other respiratory diseases principally affecting the interstitium and suppurative and necrotic conditions of the lower respiratory tract have greater log LOS if compared to patients with Influenza, pneumonia and other acute lower respiratory infections. These groups of patients will incur more costs as they will be contributing to longer stays in the hospital. As the GLM Zero Truncated model provided a better fit to the studied data hence it would be better for the health care field to use this model in predicting the log length of stay so that hospitals are well equipped with the facilities and prepare for the expenditures and resources.

Acknowledgements

The authors would like to thank the National University of Malaysia (UKM) that supported this research under Encouragement Research Grant (Geran Galakan Penyelidikan: GGP-2020-027). The sincere gratitude for Ministry of Higher Education Malaysia, Ministry of Health Malaysia, National Medical Research Register (NMRR) and Health Information Centre Malaysia for their cooperation in collecting data and analyzing the data.

REFERENCES

[1] R. Melinda. Respiratory System, Online available from https://www.webmd.com/lung/how-we-breathe

[2] Petunjuk kesihatan 2019, Online available from https://www.moh.gov.my/moh/resources/Penerbitan /Penerbitan%20Utama/HEALTH%20INDICATOR/ Petunjuk%20Kesihatan%202019%20(Web%20Version)/ mobile/index.html.

[3] Statistics on Causes of Death Malaysia, Online available from $https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat\&cat=401\&bul_id=QkxLckg3WjlzcEZyVzRIajllenBIQT09\&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09$

[4] B. Armocida, B. Formenti, S. Ussai, F. Palestra, E. Missoni. The Italian health system and the COVID-19 challenge, Lancet Public Health, No. 5, 2020.

[5] A.J. Zeleke, S. Moscato, R. Miglio, L. Chiari. Length of stay analysis of COVID-19 hospitalizations using a count regression model and quantile Regression: A study in Bologna, Italy, International Journal of Environmental Research and Public Health, Vol.19, No. 4, 2224, 2022.

[6] C. Terzano, V. Colamesta, B. Unim, S. Romani. A. Meneghini, G. Volpe, G. La Torre. Chronic obstructive pulmonary disease (COPD) exacerbation: impact of comorbidities on length and costs during hospitalization, Eur Rev Med Pharmacol Science, Vol. 21, No. 16, 3680-3689, 2017.

[7] F. De la Iglesia, P. Valino, S. Pita, V. Ramos, C. Pellicer, R. Nicolas, F. Diz-Lous. Factors predicting a hospital stay of over 3 days in patients with acute exacerbation of chronic obstructive pulmonary disease, Journal of Internal Medicine, Vol. 251, No. 6, 500-507, 2002.

[8] G. Agboado, J. Peters, L. Donkin. Factors influencing the length of hospital stay among patient's resident in Blackpool admittedwith COPD: A cross-sectional study, BMJ Open, Vol.2, No. 5, 1-10, 2012.

[9] S. Tandon, A. Budhraja, S.T. Nagdeote, K. Sharma, C.S. Dausage. Assessment of length of hospital stays of patients with acute exacerbation of chronic obstructive pulmonary disease, International Journal of Medical Research and Review, Vol. 2, No. 1, 97-103, 2016.

[10] R.K. Cydulka, E.R. McFadden Jr, C.L. Emerman, L.D. Sivinski, W. Pisanelli, A.A. Rimm. Patterns of hospitalization in elderly patients with asthma and chronic obstructive pulmonary disease, American journal of respiratory and critical care medicine, Vol. 156, No. 6, 1807 – 1812, 1997.

[11] I.M. Osman, D.J. Godden, J.A. Friend, J.S. Legge, J.G. Douglas. Quality of life and hospital re-admission in patients with chronic obstructive pulmonary disease, Thorax, Vol. 52, No. 1, 67-71, 1997.

[12] H. Herwartz, N. Klein, C. Strumann. Modelling hospital admission and length of stay by means of generalised count data models, Journal of Applied Econometrics, Vol. 31, No.6, 1159-1182, 2016.

[13] A. Freitas, T. Silva-Costa, F. Lopes, I. Garcia-Lema, A. Teixeira-Pinto, P. Brazdil, A. Costa-Pereira. Factors influencing hospital high length of stay outliers, BMC health services research, Vol. 2, No.1, 1-10, 2012.

[14] K.A. Nash, B.T. Zima, C. Rothenberg, J. Hoffmann, C. Moreno, M.S. Rosenthal, A. Venkatesh. Prolonged emergency department length of stay for US pediatric mental health visits (2005-2015). Pediatrics. Vol. 145, No. 5, 2021.

[15] E. Kulinskaya, D. Kornbrot, H. Gao. Length of stay as a performance indicator: robust statistical methodology, IMA Journal of Management Mathematics, Vol. 16. No.4, 369-381, 2005.

[16] I.W. Verburg, N.F. de Keizer, E. de Jonge, N. Peek. Comparison of regression methods for modeling intensive care length of stay, PloS one, Vol. 9. No. 10, 2014.

[17] L.Y. Tsung, K.C. Choi, E.A.S. Nelson, P.K.S. Chan, R.Y.T. Sung. Factors associated with length of hospital stay in children with respiratory disease, Hong Kong Med J, Vol.16, No. 6,440-6, 2010.

[18] F. Dong, K. Huang, X. Ren, S. Qumu, H. Niu, Y. Wang, Y. Li, M. Lu, X. Lin, T. Yang, J. Jiao. Factors associated with inpatient length of stay among hospitalised patients with chronic obstructive pulmonary disease, China, 2016–2017: a retrospective study, BMJ open, Vol.11, No. 2, 2021.0.

[19] Y. Wen, M.F. Rahman, Y. Zhuang, M. Pokojovy, H. Xu, P. McCaffrey, A. Vo, E. Walser, S. Moen, T.L.B. Tseng. Time-to-event modeling for hospital length prediction for Covid-19 patients, Machine Learning with Application, Vol. 9, 100365, 2022.

[20] A. Abdul Aziz, L. Munyakazi, N. Nsowah-Nuamah. Modelling length of stay in hospital using Generalized Linear Model. A case study at Tamale teaching hospital, American International Journal of Contemporary Research, Vol.3, No.1, 148-157, 2013.

[21] A.N. Shaaban, B. Peleteiro, M.R.O. Martins. Statistical models for analyzing count data: predictors of length of stay among HIV patients in Portugal using a multilevel model, BMC Health Services Research, Vol. 21, No. 1, 1-17, 2021.

[22] J.W. Hardin, J.M. Hilbe. Regression models for count data from truncated distributions, The Stata Journal, Vol. 12, No. 1, 226-246, 2015.

[23] H. Akaike. Factor analysis and AIC. Selected papers of hirotugu akaike, Springer, New York, NY, 371-386, 1987.

[24] S.I. Vrieze. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), Psychological methods, Vol. 17, No. 2, 228, 2012.

[25] G. Schwarz. Estimating the dimension of a model, The annals of statistics, 461-464, 1978.

[26] T.W. Yee, VGAM: Vector Generalized Linear and Additive Model R package version 1.1-7, Online available from https://CRAN.R-project.org/package=VGAM, 2022.

[27] M. Ruparel, J.L. López-Campos, A. Castro-Acosta, S. Hartl, F. Pozo-Rodriguez, C.M. Roberts. Understanding variation in length of hospital stay for COPD exacerbation: European COPD audit. ERJ Open Research, Vol. 2, No. 1, 2016.

[28] A. Orooji, E. Nazar, M. Sadeghi, A. Moradi, Z. Jafari, H. Esmaily. Factors associated with length of stay in hospital among the elderly patients using count regression models, Medical Journal of the Islamic Republic of Iran, Vol. 35, No.5, 2021.

[29] O. Baharlooei, M. Alavi, M. Adelmehraban. Psychosocial factors predicting length of hospitalization in elderly individuals with diabetes in selected hospitals of Isfahan University of Medical Sciences, Isfahan, Iran, in 2015, ARYA atherosclerosis, Vol. 13, No. 3, 103, 2017.

[30] L.A. Shanley, H. Lin, G. Flores. Factors associated with length of stay for pediatric asthma hospitalizations, Journal of Asthma, Vol. 52, No. 5, 471–477, 2015.

[31] M. Bruce, J. Smith. Length of stay among multi-ethnic psychiatric inpatients in the United Kingdom, Comprehensive Psychiatry, Vol. 102, 152201, 2020.