Estimating Optimum Length of Stay in a Hospital to Control the Infection Spread during an Epidemic Using Left-Right Truncated Poisson Model

Alka Sabharwal[1], Babita Goyal[2,*] and Vinit Singh[3]

[1]*Department of Statistics, Kirori Mal College, University of Delhi, Delhi, India*

[2]*Department of Statistics, Ramjas College, University of Delhi, Delhi, India*

[3]*Department of Statistics, University of Delhi, Delhi, India*

Abstract: *Background*: The unprecedented havoc of COVID-19 pandemic stressed medical infrastructure of every affected country. The developing countries were more affected as their already inadequate medical resources were strained further.

Material and Methods: In order to estimate the time of onset of recovery through the period of hospitalisation stay, the retrospective data on the number of days that 83 COVID-19 patients stayed in a hospital in New Delhi, India was obtained. A Left-Right Truncated Poisson Distribution Model (LRTPD) was developed to estimate the average number of days that patients had to spend in the hospital before the onset of recovery and they were no longer infected. Left truncation is on the '*u*' left most classes of the random variable and right truncation is after '*v*' classes. The parametric estimates of the LOS were validated using the Monte-Carlo method.

Results and Conclusion: The models suggested that if appropriate truncation limits (both the data-specific and as per expert advice) are used in case of critical medical emergencies, approximately 90 percent of the patients will be able to get hospital admission, without over-burdening the hospital infrastructure. The median recovery onset time/ Length of stay (LOS) obtained using the Kaplan-Meier estimator was consistent with the results of the parametric modeling and simulation. However, the Kaplan-Meier method overestimated the mean LOS as compared to the parametric methods.

Keywords: COVID-19, Left-Right Truncated Poisson Model, Length of stay, Kaplan-Meier estimator, Monte-Carlo simulation.

INTRODUCTION

The unprecedented havoc of COVID-19 pandemic stressed medical infrastructure of every affected country. The developing countries were more affected as their already inadequate medical resources were strained further. In order to identify, map, and systemize existing knowledge about the relationships between COVID-19 and hospital infrastructure adaptation and capacity planning worldwide, Ndayishimiye C. *et al*. (2022) conducted a review of 102 studies. They suggested the importance of preparing and planning in the future for an outbreak affecting a hospital infrastructure [1]. A similar conclusion was drawn in the study by Filip R. *et al.* (2022) [2]. In order to investigate the effect of an intensive care unit (ICU) bed capacity optimization on the average length of stay and average cost of hospitalization, Zheng Q. *et al*. (2024) collected data on 5944 patients admitted by the outpatient and emergency access during the 2-month periods in China. They observed a significant positive impact of ICU bed optimization in mitigating the shortage of medical resources following an epidemic outbreak [3].

The truncated statistical distributions can be used in optimisation problems as these reduce the spread of the data by truncating less frequent events at the tails of the data. Depending upon the need of the data, a number of truncated models have been suggested. The theoretical truncated Poisson distribution, as introduced by Plackett (1953), arises when some specified values are not possible to record (in terms of process and not in terms of availability) either initially or at the end of a variate range [4]. The zero truncated Poisson distribution has been used under various conditions: (i) fertility trait phenotypes (Xu and Hu, 2011) (ii) word or species frequency count data (Ginebra and Puig, 2010), (iii) mental health services data (Elhai *et al*., 2008); and (iv) number of illegal immigrants in four large cities in the Netherlands (Van der Heijden *et al*., 2003), to name a few [5-8]. Fu, Liang, *et al*. (2014) used right truncated Poisson distribution to model the progression of asthma symptoms using data from Childhood Asthma Management Program. Their study established that the properties of asthma symptom severity progression distinctly differ from those of asthma prevalence [9]. Zhao, Shi, *et al*. (2021) developed a zero-truncated Negative Binomial distributed likelihood framework to estimate the individual heterogeneity in infectiousness. Their study demonstrated that the zero-truncated framework is recommended for less biased transmission hetero-geneity estimates than non-truncated version [10].

For the parameter estimation and validation of various distributions, simulation has been a preferred

*Address correspondence to this author at the Department of Statistics, Ramjas College, University of Delhi, Delhi, India;
E-mail: goyalbabita@gmail.com

E-ISSN: 1929-6029/24

tool. Alomair, Gadir, *et al*. (2024) assessed the performance of the two estimation methods viz. Maximum Likelihood Estimation (MLE) & Method of Moments while estimating the parameter of the Zero-Truncated Poisson-improved second-degree Lindley distribution through simulation. Their study established that both estimates were asymptotically unbiased and consistent [11]. A Abd El-Hady, H. *et al*. (2022) used simulation in order to evaluate the performance of the Endpoint-Inflated Double Truncated Poisson $(\varphi_0, \varphi_1, \lambda)$ distribution and the Endpoint-Inflated Double Truncated Poisson Regression model [12]. Aydin, D. (2018) designed and conducted a Monte-Carlo simulation study to evaluate the performances of the considered estimation methods *viz*. maximum likelihood, least squares and weighted least squares for Doubly-truncated exponentiated inverse Weibull (DTEIW) distribution. Their study concluded all the considered estimators of the parameters of DTEIW distribution are asymptotically unbiased and consistent [13].

If the underlying distribution of a data set is unknown; or if the data is censored, then in order to estimate the mean and median of the survival time, non-parametric methods, such as the Kaplan-Meier (K-M) method, are recommended. Zare, Ali, *et al*. (2014); Xie, J., Brayne, C., & Matthews, F. E. (2008); Kato, I., Severson, R. K., & Schwartz, A. G. (2001) employed the Kaplan-Meier method to estimate the median survival time of patients with gastric cancer, from date of onset of dementia, and breast carcinoma respectively [14-16]. Ashfaq *et al*. (2006) compared survival time of diabetic and non-diabetic groups to observe the effect of vein graft intervention [17]. Bruce, Sheppard and others (2004) compared survival times of three categories: no diabetes, diabetes without peripheral vascular disease and renal failure, and diabetes with peripheral vascular disease and/or renal failure [18]. Joss *et al*. (2002) work concludes that survival time of type-2 diabetic patients, once diabetic nephropathy has developed, becomes even worse after starting dialysis [19]. Rossing *et al*. (1995) study compared three levels of albuminuria in insulin dependent diabetic patients [20].

Objective and Novelty of the Present Study

For any disease, the period of hospitalisation is a direct function of onset of recovery. However, for a pandemic like COVID-19, standard operating procedures suggest an optimal period of hospitalisation as the patient, even if not recovered, is unable to spread the infection to others. The WHO guidelines suggest an optimal stay of 6-10 days before a COVID-19 patient is either discharged or is shifted to a general ward. The authors wished to estimate the mean time of onset of recovery through the period of hospitalisation stay. For this purpose, the retrospective data on the number of days that the COVID-19 patients stayed in a hospital in New Delhi, India was obtained. This was the only information available. We developed a Left-Right Truncated Poisson Distribution Model (LRTPD) in order to estimate the average number of days that patients had to spent in the hospital before onset of recovery (no death was reported for these patients) (i.e. they were no longer infected or were capable of spreading the infection) and to see if our results corroborated the directions of WHO on the desired duration of stay of the patients. Left truncation is on the '*u*' left most classes of the random variable and right truncation is after '*v*' classes. Justification of the left and right truncation has been discussed. The unknown model parameter λ has been estimated using

(i) Method of Maximum Likelihood Estimation

(ii) Method of Moments

(iii) Monte Carlo Method (λ_{ML_C})

The 95% confidence interval for the unknown parameter λ has been obtained. The mean has been calculated using distribution based methods viz. the Moment generating function and Method of Moments; and has been compared with Monte-Carlo method and the distribution free Kaplan-Meier Method. Further the median of the applied distribution is computed and compared with that obtained using the Kaplan-Meier Method.

The novelty of this study is optimisation of truncation limits in order to determine the length of hospital stay so as to optimise the medical facilities in order to accommodate the maximum number of patients. To the best of our knowledge, this is the first study which aims to achieve this objective through truncated models.

Besides introduction, this paper includes three more sections. In section 2 material and methods have been presented. Section 3 is about the results, section 4 is about the discussion of the model on real data. Section 5 is about conclusions.

MATERIAL AND METHODS

Material

The retrospective data used in this study is on the hospitalization times of 83 COVID-19 patients, admitted in a hospital in New Delhi, India, at the time of disease outbreak. No other information was available with respect to patients' age, gender, comorbidities (if any) and severity of disease etc.

Methods

Left Right Truncated Poisson (LRTPD) Model

Let X_{ij} $\left(i=u,...,v\ ;\ j=1,2,...,k\right)$ are independently distributed random variables following LRTPD $\left(\lambda\right)$. The probability mass function of the random variable X_{ij} for the j^{th} respondent with '*u*' classes from the left and classes after '*v*' on the right are truncated, is given by

$$P(X_{ij}=x_{ij})=\frac{\dfrac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{1-\sum_{x_{ij}=0}^{u-1}\dfrac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}-\sum_{x_{ij}=v+1}^{\infty}\dfrac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}\ ; \tag{1}$$

$x_{ij}=u,.....v$

Estimation of the Parameter λ of LRTPD

Method of Maximum Likelihood Estimation (MLE) has been used to estimate the unknown parameter λ. The likelihood function of a LRTPD with parameter λ is given by

$$L=\prod_{j=1}^{k}P(X_{ij}=x_{ij})=\prod_{j=1}^{k}\left(\frac{\dfrac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\dfrac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}\right)=\prod_{j=1}^{k}\left(\frac{\dfrac{\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \tag{2}$$

The log-likelihood of (2) is

$$\log\left(L\right)=\log\left(\lambda\right)\sum_{j=1}^{k}x_{ij}-\sum_{j=1}^{k}\log\left(x_{ij}!\right)-\sum_{j=1}^{k}\log\left(\sum_{x_{ij}=u}^{v}\left(\frac{\lambda^{x_{ij}}}{x_{ij}!}\right)\right)$$

Upon solving $\dfrac{\partial\log\left(L\right)}{\partial\lambda}=0$, we have

$$\frac{\sum_{j=1}^{k}x_{ij}}{\lambda}=\sum_{j=1}^{k}\left(\frac{\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}-1}}{(x_{ij}-1)!}}{\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \tag{3}$$

The maximum likelihood estimate $\hat{\lambda}$ of λ is obtained by solving (3) by the method of iteration.

The 95% Confidence Interval Based on MLE of λ

95% confidence interval (C.I.) of λ is

$$\left(\hat{\lambda}\mp1.96\sqrt{Var\left(\hat{\lambda}\right)}\right)=\left(\hat{\lambda}\mp1.96\left(1\Big/\sqrt{E\left(-\frac{\partial^2\log\left(L\right)}{\partial\lambda^2}\right)}\right)\right)$$

where $-E\left(\dfrac{\partial^2\log\left(L\right)}{\partial\lambda^2}\right)=I\left(\theta\right)$ is the Fisher's information matrix .

Again, differentiating equation (3) with respect to λ, we have

$$\frac{\partial^2\log(L)}{\partial\lambda^2}=-\frac{\sum_{j=1}^{k}x_{ij}}{\lambda^2}-\sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}-2}}{\left(x_{ij}-2\right)!}\right)}{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right)$$

$$+\sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}-1}}{\left(x_{ij}-1\right)!}\right)^2}{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right)$$

$$\Rightarrow E\left(-\frac{\partial^2\log(L)}{\partial\lambda^2}\right)=A+B+C;$$

where

$$A=\frac{\sum_{j=1}^{k}E\left(x_{ij}\right)}{\lambda^2}; \tag{4}$$

$$B=\sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}-2}}{\left(x_{ij}-2\right)!}\right)}{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right);\ \text{and}$$

$$C=\sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}-1}}{\left(x_{ij}-1\right)!}\right)^2}{\left(\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right)$$

and

$$E\left(X_{ij}\right)=\sum_{x_{ij}=u}^{v}\left(\frac{\dfrac{x_{ij}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\dfrac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \tag{5}$$

Hence

$$E\left(-\frac{\partial^2 \log(L)}{\partial \lambda^2}\right) = D + E + F$$

where

$$D = \frac{\sum_{j=1}^{k}\left(\sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right)\right)}{\lambda^2}$$

$$E = \sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}\right)\left(\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}-2}}{(x_{ij}-2)!}\right)}{\left(\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right); \text{ and} \quad (6)$$

$$F = \sum_{j=1}^{k}\left(\frac{\left(\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}-1}}{(x_{ij}-1)!}\right)^2}{\left(\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}\right)^2}\right)$$

Moment Estimator of λ

Here number of unknown parameters is 1. Therefore, we solve the equation given by

$$\mu_1' = m_1' \;\Rightarrow\; \mu_1' = \frac{1}{k}\sum_{j=1}^{k} x_{ij} \;\Rightarrow\; \mu_1' = \bar{x} \quad (7)$$

$\bar{x} = \mu_1' = E\left(X_{ij}\right)$ where $E\left(X_{ij}\right)$ is given by equation (5). On solving equation (7) for λ, the moment estimator is obtained.

Characteristics of LRTPD

Moment Generating Function

Moment generating function of LRTPD (λ) is given by

$$M_{x_{ij}}(t) = \sum_{x_{ij}=u}^{v} e^{tx_{ij}}\left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}\right) \quad (8)$$

The first four moments of LRTPD

Using the moment generating function

$$\textit{Mean} = E\left(X_{ij}\right) = \sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \quad (9)$$

and

$$Var\left(X_{ij}\right) = \sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}^2\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right) - \left(\sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right)\right)^2 \quad (10)$$

$$\mu_3' = \sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}^3\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \quad \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\left(\mu_1'\right)^3 \quad (11)$$

and

$$\mu_4' = \sum_{x_{ij}=u}^{v}\left(\frac{\frac{x_{ij}^4\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{\lambda^{x_{ij}}}{x_{ij}!}}\right) \quad \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\left(\mu_1'\right)^2 - 3\left(\mu_1'\right)^4 \quad (12)$$

The skewness and kurtosis can be calculated by $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{(\mu_2)^2}$.

Median and Mode of LRTPD

Median is obtained by solving the equation.

$$\sum_{x_{ij}=u}^{md}\left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{1-\sum_{x_{ij}=0}^{u-1}\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}-\sum_{x_{ij}=v+1}^{\infty}\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}\right) = \frac{1}{2} \quad (13)$$

The mode is that X_{ij} which satisfy

$$p(x_{ij}-1) \le p(x_{ij}) \;:\; p(x_{ij}+1) \le p(x_{ij}) \quad (14)$$

or

$$\left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}-1}}{(x_{ij}-1)!}}{\sum_{x_{ij}=u}^{v}\frac{e^{-\lambda}\lambda^{x_{ij}-1}}{(x_{ij}-1)!}}\right) \le \left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{e^{-\lambda}\lambda^{x_{ij}-1}}{x_{ij}!}}\right) : \quad (15)$$

$$\left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}+1}}{(x_{ij}+1)!}}{\sum_{x_{ij}=u}^{v}\frac{e^{-\lambda}\lambda^{x_{ij}+1}}{(x_{ij}+1)!}}\right) \le \left(\frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\sum_{x_{ij}=u}^{v}\frac{e^{-\lambda}\lambda^{x_{ij}-1}}{x_{ij}!}}\right)$$

Probabilities of Left and Right Truncated Classes

The tail probabilities of the truncated classes are given by

$$P(X_{ij} \geq v) = \sum_{x_{ij}=v+1}^{\infty} \frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\left(1 - \sum_{x_{ij}=0}^{u-1} \frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}\right)} \quad (16)$$

$$P(X_{ij} \leq u) = \sum_{x_{ij}=0}^{u-1} \frac{\frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}}{\left(1 - \sum_{x_{ij}=v+1}^{\infty} \frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!}\right)} \quad (17)$$

Hence the probability of X_{ij} to lie in the observation range is

$$P(u \leq X_{ij} \leq v) = \sum_{x_{ij}=u}^{v} \frac{e^{-\lambda}\lambda^{x_{ij}}}{x_{ij}!} \quad (18)$$

Mean and Median Survival Time using Kaplan-Meier Method

The mean survival time is approximated by summing the products of the observed survival times and the probability of surviving beyond each observed time point. Mathematically, it can be represented as:

$$\mu = \sum_{i=1}^{k-1} \hat{S}\left(t_i\right)\left(t_{i+1} - t_i\right) \quad (19)$$

where $\hat{S}\left(t_i\right)$ is the estimated survival probability at each time point and t_i and t_{i+1} are two consecutive survival times.

The median survival time is the time point at which the Kaplan-Meier survival curve crosses the 0.5 threshold. In other words, it is the time point where approximately 50% of the subjects have experienced the event of interest.

Monte Carlo Simulation

In this method the value of each output parameter is one particular value in the simulation run. Monte Carlo algorithm used in this study is:

(i) Each distinct value of the output parameter is used to run simulation.

(ii) Each run is performed 500 times.

(iii) Corresponding to each run parameter, mean of distribution (Poisson) is computed.

(iv) All the runs are used to compute the final value of the parameter.

RESULTS

The data for the study consisted of the hospitalisation period of 83 patients suffering from COVID-19, in a hospital in New Delhi, India. The hospital stay data was further truncated/ censored based on the criteria discussed below:

The Truncation Criteria for the Parametric Models:

Left Truncation

Patients were discharged either from OPD or after one day medical observation.

Right Truncation

Two different criteria were used for right truncation:

Criterion 1

Onset of recovery was not observed up to 8 days, so patients were either shifted to a different ward or were discharged under medical observation (due to the limited number of beds). This was Data_1.

Criterion 2

Onset of recovery was not observed up to 10 days, so patients were either shifted to different ward or were discharged under medical observation (due to the limited number of beds). This was Data_2.

Table 1: Day Wise Descriptive Statistics of the Respondents without and with Truncation

Descriptive Statistics	Complete data	Data_1*	Data_2**
Mean	6.662	5.122	5.727
Standard Deviation	4.693	1.918	2.356
Minimum	0	2	2
Maximum	25	8	10

*Data_1: Right truncation after 8 days, left truncated before 2 days.
**Data_2: Right truncation after 10 days, left truncated before 2 days.

Table 2: Frequency Table for the Number of Days an Individual Patient Stayed at Hospital

S. No.	Number of days stay at hospital (X_{ij})	Frequency (f_i)		
		Complete data	Data_1	Data_2
1	0	4	-	-
2	1	2	-	-
3	2	8	8	8
4	3	8	8	8
5	4	7	7	7
6	5	9	9	9
7	6	9	9	9
8	7	13	13	13
9	8	6	6	6
10	9	4	-	4
11	10	2	-	2
12	11	3	-	-
13	12	1	-	-
14	14	2	-	-
15	15	1	-	-
16	17	1	-	-
17	21	1	-	-
18	22	1	-	-
19	25	1	-	-

The Censoring Criteria for the Kaplan-Meier Method

Without assuming the form of the distribution, the following criteria have been used to obtain the mean and median of the survival time using Kaplan-Meier method:

Case 1

Observations are uncensored i.e. complete discharge time of each and every patient has been recorded.

Case 2

Observations are censored based on

(a) Onset of recovery was not observed for patients having stay time 0 or 1 days as expert did not feel the need of hospitalisation at that time (onset of recovery was unknown).

(b) Study was terminated after 8 (10) days.

The Table **1** above represents descriptive statistics of the 83 patients before and after truncation.

Table 3: Monte Carlo Simulation Result

Runs	1	2	3	4	5	6	7	8	9
Mean	11.8977	11.3262	10.4313	8.5195	6.9960	6.9061	5.9332	5.3595	5.1654
Runs	10	11	12	13	14	15	16	17	18
Mean	4.2983	4.0690	3.4221	3.0087	2.4548	1.9747	1.5832	0.9539	0.4731

Table 4: Estimates, Variances of the Estimates and the 95% Confidence Intervals for the 'Complete', 'Truncated' and 'Simulated Monte Carlo (MC)' Data

Model	$\hat{\lambda}$	$Var(\hat{\lambda})$	95% Confidence Interval	
			LCL	UCL
MLE (without Truncation)	6.6626	0.07841277	5.9594	7.0571
MLE (Data_1)	5.2404	0.008720993	5.0573	5.4234
MLE (Data_2)	5.98	0.0078587	5.8062	6.1537
MLE(MC)	5.2651	0.737442	5.0581	5.4601

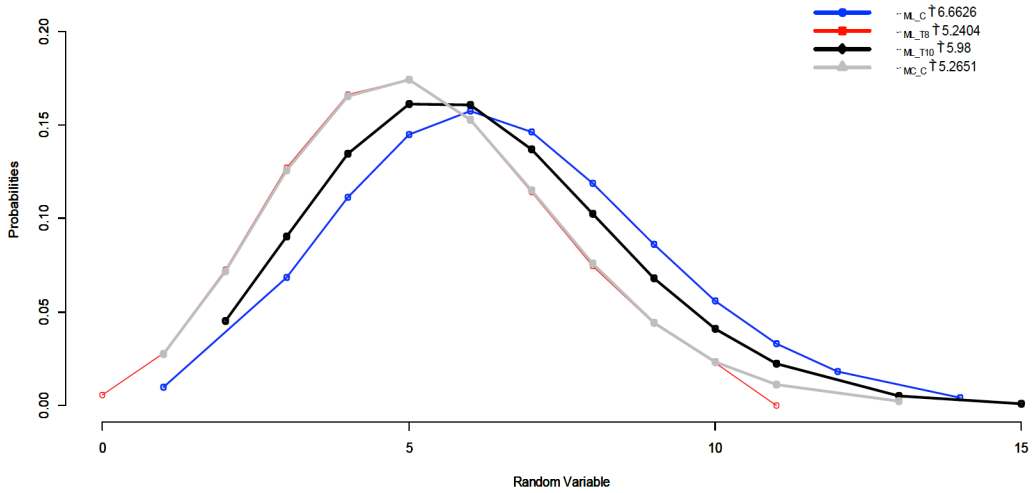

Figure 1: Probability plots of Poisson distribution and LRTPD for the estimated parameters.

The day wise frequency distribution of the number of patients is presented in Table **2** above.

As the sample size was small, we used Monte Carlo simulation to generate data for estimation of the unknown model parameter and validation of the proposed models results. Each distinct value of the number of days of hospital stay was taken as a parameter to for a single run of simulation. There were 18 distinct such values yielding 18 runs as indicated in Table **3** below:

The estimates of λ, $Var(\hat{\lambda})$ and 95% confidence intervals for λ obtained by the method of MLE (on non-truncated data and using both truncation criteria) and simulation are displayed in Table **4** below.

The probability plots of the model using the estimated values of the parameter λ in Table **4** are presented in Figure **1** below:

The Table **5** below shows first four moments about origin and mean for the number of days an individual patient stayed at hospital during COVID-19 pandemic (as identified by the values that the random variables X_{ij} assumes) for the 'Complete' and the 'Truncated' data respectively.

Mean and Median of the survival time for complete and censored data (based on specified criteria) are computed below in Table **6** using Kaplan-Meier method, where survival time *T* is the onset recovery time of the COVID patients. Survival probabilities have been

Table 5: First Four Moments about Origin and Mean for all the Four Models

Moments About	Complete data		Simulated data		Data_1		Data_2	
	Origin	Mean	Origin	Mean	Origin	Mean	Origin	Mean
1st	6.508	0	5.259	0	5.122	0	5.726	0
2nd	48.982	6.625	32.894	5.235	29.001	2.758	36.151	3.356
3rd	407.672	2.645	232.949	4.886	175.878	-0.957	245.401	-0.068
4th	3680.515	133.765	1821.155	84.517	1119.251	15.769	1764.483	30.091
Skewness		0.024		0.166		0.043		0.0001
Kurtosis		3.047		3.083		2.071		2.670
Median	6		6		4.5767		5.31816	
Mode	7		7		7		7	

Table 6: Means and Medians for Survival Times using Kaplan-Meier Method

Data	Means and Medians for Survival Time							
	Mean				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Complete	6.663	.515	5.653	7.672	6.000	0.413	5.191	6.809
Censored (8 days)	9.799	.974	7.891	11.707	6.000	0.399	5.218	6.782
Censored (10 days)	8.481	.807	6.898	10.063	6.000	0.399	5.218	6.782

calculated at using the censoring criteria: (i) at 8 days (left censoring before 2 days and right censoring after 8 days); and (ii) at 10 days (left censoring before 2 days and right censoring after 10 days).

DISCUSSION

Whenever a medical emergency (due to an epidemic, pandemic or natural disasters) arises, we need sufficient medical facilities to deal with it. However, a country is not always well equipped with adequate infrastructure. One of the dimensions of the medical infrastructure is the number of beds available for all the needy patients. This is directly linked to the stay time of the patients in the hospital. However the stay time of the patients is determined by the severity of the problem being faced by an individual patient.

With the outbreak of COVID-19 in 2020, the number of COVID cases rose sharply from January to June 2021 in many states in India. In order to provide adequate medical facility to all, it was imperative to predict hospital bed demand and to estimate the length of hospital stay (LOS) for any patient. This was crucial for decision-making and contingency planning. However, estimating LOS was not only crucial during the time mentioned above but is still a major problem as it is a fact that the risk COVID (or a similar pandemic) is not over globally. To plan disease prevention and control, researchers and policymakers need to make/suggest policies to handle these abnormal situations. With the objective of estimating hospital LOS, the retrospective data on the LOS of 83 COVID-19 patients was obtained from a Delhi-based hospital.

By selecting appropriate model(s), we extracted maximum possible information from the available data. Monte Carlo simulation was then applied to validate the models. This also helped us to overcome the limitation(s) of small sample. The models that we applied were (i) a parametric discrete Poisson distribution; and (ii) the distribution free Kaplan-Meier method to obtain the mean and median onset recovery days/LOS.

Under the first model, the range of LOS (n) was taken to be 0 - 28 days, which was based on the available data and probability of onset of recovery on a particular day was taken as 'p'. Then the parameter of the distribution was obtained as np (= λ) and hence the Poisson model was applied.

Next, we applied truncation to the parametric model. According to WHO, the risk of spreading the disease is over after 6-10 days (after confirmed diagnosis of the disease). So after this period, a patient may either be discharged or can be shifted to a general ward as s/he is safe from the risk of further spread. For our data, the mode of LOS for the complete data was 7 days. Hence the 8 days truncation was applied in the first case. After 10 days, the frequency of patients still requiring hospital stay reduced substantially. Hence 10 days truncation point was taken in the second case. These limits were data specific.

Left truncation is justified as either the patient was examined in OPD and was not found to be a fit case of admission. Secondly, the patient was found to be safe to be discharged after one day, under medical supervision. Hence 0 and 1 values were truncated.

The data based truncations limits were applied on the model(s) used in the study. The following results (Table **7**) were obtained:

Table 7: Probabilities of LOS using Different Truncation Criteria

Truncation	Probability
T < 2	0.0361308
T > 8	0.08782592
T > 10	0.04255144
2 <= T <=8	0.87604328
2 <= T <=10	0.92131776

When we applied both the left and right truncation of 8 days, it was found that 87% of the cases would be included in general (irrespective of data). If the right truncation is done at 10 days, the proportion of included cases would be 92%. According to this inference, the 10 days truncation on LOS is better justified as compared with 8 days truncation and it also agrees with the WHO upper limit recommendation (21).

While using the Kaplan-Meier estimator, the truncation limits were taken as the censoring values. For the patients whose discharge time was below 2 days, the recovery onset was unknown and hence censored. On the other hand, for the patients who were there in the hospital after 8 (10) days, the recovery time was not observed and hence censored. The median recovery onset time/ LOS obtained using the Kaplan-Meier estimator was consistent with the results of the parametric modeling and simulation. However, the Kaplan-Meier method overestimated the mean LOS as compared to the parametric methods.

CONCLUSION

The study establishes the role of truncation models for optimizing a scarce utility. The models suggested

that if appropriate truncation limits (both the data-specific and as per expert advice) are used in case of critical medical emergencies, approximately 90 percent of the patients will be able to get hospital admission, without over-burdening the hospital infrastructure. This should help the hospital management to optimize the use of their medical infrastructure.

REFERENCES

[1] Ndayishimiye C, Sowada C, Dyjach P, Stasiak A, Middleton J, Lopes H, Dubas-Jakóbczyk K. Associations between the COVID-19 pandemic and hospital infrastructure adaptation and planning—a scoping review. International Journal of Environmental Research and Public Health 2022; 19(13): 8195. https://doi.org/10.3390/ijerph19138195

[2] Filip R, Gheorghita Puscaselu R, Anchidin-Norocel L, Dimian M, Savage WK. Global challenges to public health care systems during the COVID-19 pandemic: a review of pandemic measures and problems. Journal of Personalized Medicine 2022; 12(8): 1295. https://doi.org/10.3390/jpm12081295

[3] Zheng Q, Zeng Z, Tang X, Ma L. Impact of an ICU bed capacity optimisation method on the average length of stay and average cost of hospitalisation following implementation of China's open policy with respect to COVID-19: a difference-in-differences analysis based on information management system data from a tertiary hospital in southwest China. BMJ open 2024; 14(4): e078069. https://doi.org/10.1136/bmjopen-2023-078069

[4] Ghosh SK, Sen S. Modeling pages left blank in university examination: A resolution in higher education process. arXiv preprint arXiv: 1910.04994 2019. https://doi.org/10.48550/arXiv.1910.04994

[5] Xu S, Hu Z. Mapping quantitative trait loci using the MCMC procedure in SAS. Heredity 2011; 106(2): 357-69. https://doi.org/10.1038/hdy.2010.77

[6] Ginebra J, Puig X. On the measure and the estimation of evenness and diversity. Computational Statistics & Data Analysis 2010; 54(9): 2187-201. https://doi.org/10.1016/j.csda.2010.04.001

[7] Elhai JD, Calhoun PS, Ford JD. Statistical procedures for analyzing mental health services data. Psychiatry Research 2008; 160(2): 129-36. https://doi.org/10.1016/j.psychres.2007.07.003

[8] Van Der Heijden PG, Bustami R, Cruyff MJ, Engbersen G, Van Houwelingen HC. Point and interval estimation of the population size using the truncated Poisson regression model. Statistical Modelling 2003; 3(4): 305-22. https://doi.org/10.1191/1471082X03st057oa

[9] Fu L, Freishtat RJ, Gordish-Dressman H, Teach SJ, Resca L, Hoffman EP, Wang Z. Natural progression of childhood asthma symptoms and strong influence of sex and puberty. Annals of the American Thoracic Society 2014; 11(6): 939-44. https://doi.org/10.1513/AnnalsATS.201402-084OC

[10] Zhao S, Shen M, Musa SS, Guo Z, Ran J, Peng Z, Zhao Y, Chong MK, He D, Wang MH. Inferencing superspreading potential using zero-truncated negative binomial model: exemplification with COVID-19. BMC Medical Research Methodology 2021; 21: 1-8. https://doi.org/10.1186/s12874-021-01225-w

[11] Alomair G, Tajuddin RR, Bakouch HS, Almohisen A. A Statistical Model for Count Data Analysis and Population Size Estimation: Introducing a Mixed Poisson–Lindley Distribution and Its Zero Truncation. Axioms 2024; 13(2): 125. https://doi.org/10.3390/axioms13020125

[12] A Abd El-Hady H, T El-Ghanam AT, E Abu-Hussien A. Endpoint-inflated double truncated poisson model. Scientific Journal of Faculties of Commerce Sector, Al-Azhar University 2022; 28(1): 109-33.

[13] Aydın D. The doubly-truncated exponentiated inverse Weibull distribution. Anadolu University Journal of Science and Technology B-Theoretical Sciences 2018; 6(1): 55-74. https://doi.org/10.20290/aubtdb.326782

[14] Zare A, Mahmoodi M, Mohammad K, Zeraati H, Hosseini M, Naieni KH. A comparison between Kaplan-Meier and weighted Kaplan-Meier methods of five-year survival estimation of patients with gastric cancer. Acta Medica Iranica 2014; 764-7.

[15] Xie J, Brayne C, Matthews FE. Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up. BMJ 2008; 336(7638): 258-62. https://doi.org/10.1136/bmj.39433.616678.25

[16] Kato I, Severson RK, Schwartz AG. Conditional median survival of patients with advanced carcinoma: surveillance, epidemiology, and end results data. Cancer 2001; 92(8): 2211-9. https://doi.org/10.1002/1097-0142(20011015)92:8%3C2211::AID-CNCR1565%3E3.0.CO;2-W

[17] Ashfaq S, Ghazzal Z, Douglas JS, Morris DC, Veledar E, Weintraub WS. Impact of diabetes on five-year outcomes after vein graft interventions performed prior to the drug-eluting stent era. The Journal of Invasive Cardiology 2006; 18(3): 100-5.

[18] Leavitt BJ, Sheppard L, Maloney C, Clough RA, Braxton JH, Charlesworth DC, Weintraub RM, Hernandez F, Olmstead EM, Nugent WC, O'Connor GT. Effect of diabetes and associated conditions on long-term survival after coronary artery bypass graft surgery. Circulation 2004; 110(11_suppl_1): II-41. https://doi.org/10.1161/01.CIR.0000138197.07051.e7

[19] Joss N, Paterson KR, Deighan CJ, Simpson K, Boulton-Jones JM. Survival Analysis Techniques for Censored and Truncated data. 2nd ed. Springer Science Business Media Inc. 2002.

[20] Rossing P, Tarnow L, Nielsen FS, Boelskifte S, Brenner BM, Parving HH. Short stature and diabetic nephropathy. BMJ 1995; 310(6975): 296-7. https://doi.org/10.1136/bmj.310.6975.296

[21] https://www.google.com/search?q=revise+discharge+policy+for+covid+19&oq=revise+discharge+policy+for+covid+&gs_lcrp=EgZjaHJvbWUqCQgBECEYChigATIJCAAQRRg5GJ8FMgkIARAhGAoYoAEyCQgCECEYChigATIHCAMQIRifBTIHCAQQIRifBdIBCTEzOTkxajBqN6gCCLACAQ&sourceid=chrome&ie=UTF-8.

Received on 18-08-2024 Accepted on 14-09-2024 Published on 10-10-2024

https://doi.org/10.6000/1929-6029.2024.13.20