DOI: 10.14744/ejmo.2021.90258
EJMO 2021;5(2):132–143

Research Article

Evaluation of Parametric Method Performance for Left-Censored Data and Recommendation of Using for Covid-19 Data Analysis

Mustafa Agah Tekindal,[1] Harun Yonar,[2] Saadet Kader[3]

[1]Department of Biostatistics, Izmir Katip Celebi University Faculty of Medicine, Izmir, Turkey
[2]Deparment of Biostatistics, Selcuk University Faculty of Veterinary Medicine, Konya, Turkey
[3]Deparment of Biochemistry, Karapinar State Hospital, Konya, Turkey

Abstract

Objectives: Left-censored data, which is commonly seen in clinical studies, are frequently encountered in the literature, especially in the fields of food, environment, microbiology, and biochemistry. In this study, the most appropriate distribution between the negatively skewed distributions for left-censored data in Parametric Inverse Hazard Models was tried to be determined.

Methods: Within the scope of the study, firstly, the data were produced uncensored according to different parameters of each distribution. Then, simulation studies were carried out in different censorship rates (15%, 25% and 35%) and various sample sizes (1000, 2000 and 3000) in order to determine the most appropriate distribution. AIC, AICC, HQIC, and CAIC information criteria were employed to compare the distribution performances. Since it was not possible to study simulations of all possible scenarios, scenarios similar to each other were generally preferred over others.

Results: In the simulation results, the most appropriate distributions to be used for left-censored data in Parametric Inverse Hazard Models were found as Generalized Inverse Weibull as well as Log-Logistic, Log-Normal, Inverse Normal and Gamma distributions. It was also detected that the Marshal-Olkin distribution revealed a superior performance compared to the Modified Weibull, Generalized Gamma, Gamma, and Flexible Weibull distributions. Log logistics distribution gave the most appropriate result among the analyzed distributions in the examination made with real data application.

Conclusion: The use of censored data analysis in evaluations in terms of Covid-19 is quite additive, considering that more statistical evaluation will be needed in the next period of the epidemic. Improved estimates can be obtained with this approach, especially in Covid-19 data analysis.

Keywords: Covid-19, left-censored, limit of detection, Parametric Inverse Hazard Models

Cite This Article: *Tekindal MA, Yonar H, Kader S. Evaluation of Parametric Method Performance for Left-Censored Data and Recommendation of Using for Covid-19 Data Analysis. EJMO 2021;5(2):132–143.*

In survival analysis, censoring is neglecting data that can not be observed and known precisely for any reason due to some limitations in the lifetime data of the unit or individuals of interest, such as time, cost, and values below the determination limit. In cases when it is not certain what the data will be during the process or after the process, that is when the start time is certain but the end time is not clear; if the unit or individual leaves the experiment before the defined event takes place, the data is right-censored.

Prior to starting to study the individual, if some of the data have some exclusion criteria or if there are data below the measurement limit, this type of data are left-censored. In-

Address for correspondence: Mustafa Agah Tekindal, Ph. D. Izmir Katip Celebi Universitesi Tip Fakultesi, Biyoistatistik Anabilim Dali, Izmir, Turkey
Phone: +90 555 399 05 99 **E-mail:** matekindal@gmail.com
Submitted Date: February 11, 2021 **Accepted Date:** March 18, 2021 **Available Online Date:** June 10, 2021

terval censorship and double censorship obtained using right and left censorings are also known as generalized censorship types.[1]

In censoring, there is a certain measurement limit for the units and the selection of the censoring limit is determined by the researcher. This point is the Limit of Detection (LOD) value which has slightly exceeded the +3 standard deviation value around 0 or the Limit of Quantification (LOQ) value which is known as the +10 standard deviation value around 0. While it represents the smallest precision measurement that can be employed in detection limit validation, Quantification limit represents the highest reliable limit value of concentration (a measurable variable) that can be reported.[2]

Let L_i be censorship time and T_i be the lifetime of the individual; when $T_i < L_i$, the individual's survival time is smaller than the observed survival time and thus this individual's lifetime is said to be left-censored.

$$\delta_i = \begin{cases} 0 & T_i \le L_i \\ 1 & T_i > L_i \end{cases} \qquad i = 1, 2, 3, \ldots, n$$

If $\delta_i = 0\ (T_i \le L_i)$, the individual is censored and if $\delta_i = 1\ (T_i > L_i)$, the individual is observed.[1,3]

In the analysis of left-censored data, biases arise due to the fact that observation values cannot be determined for various reasons. Statistical methods used to eliminate these biases caused by censoring are divided into four general categories; **substitution methods**;[4,5] **semi-parametric methods** (Regression on Order Statistics (ROS));[6–8] **parametric methods** (Maximum Likelihood Estimation (MLE))[9] and **non-parametric methods** (Kaplan Meier estimator (KM)).[10,11] Hewett and Ganser[12] conducted a review on the comparison of different approaches to the processing of censored data.

The Maximum Likelihood Estimator (MLE) is employed in the parametric method, and the logarithmic form of the likelihood function of the distribution in which censored and uncensored data are included together is as follows:

$$l = \sum_{i=1}^{n} \{\delta_i \ln h(t_i) + \ln S(t_i)\} = \{\delta_i \ln h(t_i) - H(t_i)\}$$

In the literature, there are performance comparisons with different sample sizes and various censoring rates for non-parametric, parametric, semi-parametric and substitution methods. However, studies that can guide the relevance of left-censored data to distributions are quite insufficient.[13] As it is broadly employed in survival analysis and approximately fits the data, there are evaluations made for left-censored data by acting entirely upon experimental assumptions conducted under the assumption of Log-Normal, even Normal distribution, especially Weibull distribution. In Table 1, studies about the distributions employed in the evaluation of censored data are presented.

Table 1. Literature research for censored data analysis

Authors	Context
(Odell, Anderson, & D'Agostino, 1992)[14]	They obtained MLE estimates for interval-censored data by using Weibull Based Accelerated Failure Time Model.
(Luczynska, Sterne, Bond, Azima, & Burney, 1998)[15]	Concerning left-censored observations, they assumed that the distribution was Normal for values below a certain level that could not be measured, without any confirmation.
(R. D. Gupta & Kundu, 1999)[16]	They introduced a new family of distributions, a generalized exponential distribution very similar to the corresponding shape of the 3-parameter Gamma or Weibull distribution. These distributions can often be used as an alternative to censored data in survival analysis, and have even proved to be much more suited to the data structure in survival analysis.
(Thompson, Voit, & Scott, 2000)[17]	They compared the suggested (Log-Normal and Weibull) distributions using probability graphics (P-P and Q-Q) to identify the most suitable distribution for left-censored data.
(Pajek, Kubala-Kukuś, Banaś, Braziewicz, & Majewska, 2004)[18]	They compared non-parametric Kaplan-Meier and Nelson-Aalen estimators for data generated from Log-Normal distribution inspired by previous experimental studies in terms of left-censored data.
(Isingo et al., 2007)[19]	They tried to estimate survival times with Weibull models in left truncated and right-censored data.
(Annan, Liu, & Zhang, 2009)[20]	In cases where they assumed that the distribution is assumed exponential, they compared the Kaplan-Meier, MLE and ROS estimators for left-censored data in their simulation studies with numbers generated from the exponential distribution. In their study, they examined the distributions from the exponential family, which are frequently encountered in censored data such as Normal, Log-Normal, and Gamma distribution.
(Kremer, Weißbach, & Liese, 2014)[21]	In the conducted Finance implementation study, it was identified that the rating histories of left-censored observations were rarely viewed. They investigated the use of the Parametric Inverse Hazard Model in the prediction of the rating transition hazard of the credit source.
(Pesonen, Pesonen, & Nevalainen, 2015)[22]	They conducted simulation studies using an ML-based covariance matrix estimator, which also allows the left censored values to be detected without being substituted with some constants or without being completely ignored.
(Islam, 2016)[13]	They studied Parametric Reversed Hazards (PRH) Model for a variety of distributions which may be appropriate for left censored data.
(Achcar, Coelho-Barros, Cuevas, & Mazucheli, 2018)[23]	They modeled left-censored longitudinal data by using asymmetric distributions (Lèvy distribution). They concentrated on the implementation of classical and Bayesian inference methods to analyze the data produced for various variables.
(Fusek, Michálek, Buňková, & Buňka, 2020)[24]	They tested and compared the number of biogenic amines which are showed compliance for Weibull and exponential distributions with Left-censored data.

There are very few studies on Covid-19 on censored data in the literature, and there is no left-censored analysis among these studies.

Omar et al. (2020)[25] conducted research on interval censorship in evaluating the course of the disease at home or in the hospital in case of severe acute life with the onset of symptoms of Covid-19 patients.

Sreedevi and Sankaran (2020)[26] used the Proportional hazards mix treatment model to estimate the recovery rate of Covid-19 patients in India, including the effect of age and gender.

Abdel-Salam and Mollazehi (2020)[27] used the censored data from Singaporean Covid-19 patients and compared recovery times with non-Singaporean patients using Several parametric models.

Since there was no clear suggestion about the most suitable model and distribution in the analysis of left-censored data in the literature, the study finding were derived from negatively skewed distributions in order to guide the researchers on the use of the most suitable distribution for left-censored data. The most suitable Parametric Inverse Hazard Models were tried to be identified with the help of information criteria for the derived data. The study findings were produced from left-handed distributions in order to guide the researchers on the use of the most appropriate distribution for left-censored data. The most suitable Parametric Inverse Hazard models were tried to be determined with the help of information criteria for the derived data. Since it was not possible to study simulations of all possible scenarios, scenarios similar to each other were generally preferred over others. The main motivation of the study was the reliability of the estimates made by using the Parametric Inverse Hazard Models without using the substitution method, the semi-parametric method, and the non-parametric methods. The distribution and sample sizes that were most suitable for these estimates were tried to be determined. In addition, it was observed that the real data application and the simulation findings were compatible.

Parametric Inverse Hazard Model

In survival studies, the completed and incomplete data can be seen together and the goal is to analyze the failure structure related to the completed variable with the help of Hazard functions. Parametric models are based on a certain distribution assumption for the Hazard rate. The shape of the Hazard functions varies depending on the type of distribution, and an accurate determination of the Hazard function of survival times is necessary for unbiased estimation of parameters. The Hazard function, which is the inverse of the survival function, is the risk of occurring in a small time interval, such as t and $t+\Delta t$, under the condition that an event of interest is not yet present at time t (continuing to survive) and it is defined as follows:[28]

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t+\Delta t \,|\, T \ge t)}{\Delta t}$$

The Hazard function used here to explain when the data are censored from the right.[29]

In survival studies, when lifetime data are left-censored, Inverse Hazard Rate (IHR) models are preferred since Hazard rate estimators are tending to be unstable.[30]

IHR is defined as follows for an event of interest:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t-\Delta t \le T \,|\, T \le t)}{\Delta t}$$

The function of $\lambda(t)$ suggested by Barlow, Marshall, and Proschan (1963)[31] was employed in the estimation of the left censored distribution function;[32] in defining a new stochastic scheme;[33] in identifying the characterization of lifetime distributions;[34] in investigating the aging behavior;[35,36] in developing new repair and maintenance strategies;[37,38] in mixed proportional hazard models[4] and in stress hybrid hazard models.[29,39]

Let T be non-negative random variable with the distribution function $F(t)$, which represents an individual's lifetime, and probability density function f (t); IHR can be written as follows:

$$\lambda(t) = \frac{f(t)}{F(t)}$$

Let X be a (p×1) vector of covariates; in case of any factor affecting survival times in survival analyzes, Hazard function is modeled as Parametric Inverse Hazard (PIH) model defined by

$$\lambda(t|X) = \lambda_0(t) g(\beta; X)$$

Here $\lambda_0(t)$ is the base Hazard function; $g(.)$ is a non-negative function of X; β is a $(px1)$ regression parameters.[29,40]

The distribution function of the PIH model, which is a fully parametric model based on IHR, is as follows:

$$F(t|X) = F_0(t)^{g(\beta;X)}$$

where $F(t|X)$ is the distribution function of the lifetime T given X, $F_0(t)$ is the part representing the basic hazard function in the absence of covariates.

Let assume that the lifetime random variable T is randomly left-censored with random Z. In practice, $I(.)$ being the indicator function for observed vectors (Y,δ,X) with $Y=\max(T,Z)$ and $\delta = I(T \le Y)$; then the likelihood function can be written as follows:

$$L(\beta; y) = \prod_{i=1}^{n} f(y_i | x_i)^{\delta_i} F(y_i | x_i)^{1-\delta_i}$$

Here δ_i representing censor situation, when $\delta_i = 0$ it represents censored observation; when $\delta_i = 1$ it represents uncensored observation.[3,13,29,41] Since this method contains assumptions for left censored data, it also brings certain limitations; estimates can be calculated for parameters in this model using MLE method.[29,42]

In this study, PRH models have been formed for Exponential, Log-Normal, Inverse Normal, Gamma, Generalized Gamma, Inverse Gamma, Log-Logistic, Weibull, Inverse Weibull, Generalized Inverse, Flexible Weibull, Marshal-Olkin, Strong Generalized Weibull, Modified Weibull and Gompertz distributions that may be suitable for left-censored data. These distributions have been obtained as follows:[29]

- Inverse Weibull Distribution

$$l(\beta,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta - \frac{1}{\gamma}\sum_{i=1}^{n} \delta_i t_i - \sum_{i=1}^{n} \delta_i \ln\gamma + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left(1 - e^{-t_i/\gamma}\right)$$

- Exponential Distribution

$$l(\beta,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta - \frac{1}{\gamma}\sum_{i=1}^{n} \delta_i t_i - \sum_{i=1}^{n} \delta_i \ln\gamma + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left(1 - e^{-t_i/\gamma}\right)$$

- Log-Normal Distribution

$$l(\mu,\sigma,t) = \sum_{i=1}^{n} \delta_i x_i \beta - \sum_{i=1}^{n} \delta_i \ln(\sqrt{2\pi}\sigma t_i) + \sum_{i=1}^{n} \delta_i \frac{[\ln(t) - \mu]^2}{2\sigma^2} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left[\Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right]$$

- Inverse Normal Distribution

$$l(\alpha,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \frac{1}{2}\sum_{i=1}^{n} \delta_i \ln\left(\frac{\gamma}{2\pi {t_i}^3}\right) - \sum_{i=1}^{n} \left[\frac{\delta_i \gamma (t_i - a)^2}{2a^2 t_i}\right] + \sum_{i=1}^{n} [e^{x_i\beta} - \delta_i] \ln\left\{\Phi\left[-\sqrt{\frac{\gamma}{t_i}}\left(\frac{t_i}{a} + 1\right)\right]\right\} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left[1 - \exp\left(\frac{2\gamma}{a}\right)\right]$$

- Log-Logistic Distribution

$$l(\alpha,\omega,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\left(\frac{\omega}{t_i}\right) - \sum_{i=1}^{n} \delta_i \ln\left[1 + \left(\frac{t_i}{a}\right)^{\omega}\right] + \sum_{i=1}^{n} e^{x_i\beta} \ln\left\{1 + \left(\frac{t_i}{a}\right)^{-\omega}\right\}$$

- Gompertz-Makeham Distribution

$$l(\alpha,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln a + \sum_{i=1}^{n} \delta_i \gamma t_i - \sum_{i=1}^{n} \frac{\delta_i a}{\gamma}(e^{\gamma t_i} - 1) + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left\{1 - \exp\left[-\frac{a}{\gamma}(e^{\gamma t_i} - 1)\right]\right\}$$

- Gamma Distribution

$$l(\alpha,\omega,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i a \ln\omega + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i \omega t_i + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln[\gamma(a, \omega t_i)] - \sum_{i=1}^{n} e^{x_i\beta} \ln[\Gamma(a)]$$

- Generalized Gamma Distribution

$$l(\alpha,\omega,\lambda,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i a \ln(\omega\lambda) + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i (\lambda t_i)^{\omega} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left[\gamma\left[\frac{a}{\omega}, (\lambda t_i)^{\omega}\right]\right] - \sum_{i=1}^{n} e^{x_i\beta} \ln\left[\Gamma\left(\frac{a}{\omega}\right)\right]$$

- Inverse Gamma Distribution

$$l(\alpha,\omega,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i a \ln\omega - (a+1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \frac{\delta_i \omega}{t_i} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln[\gamma(a, t_i)] - \sum_{i=1}^{n} e^{x_i\beta} \ln[\Gamma(a)]$$

- Weibull Distribution

$$l(\alpha,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln a - \sum_{i=1}^{n} \delta_i a \ln\gamma + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i \left(\frac{t_i}{\gamma}\right)^{a} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left[1 - e^{-\left(\frac{t_i}{\gamma}\right)^{a}}\right]$$

- Generalized Inverse Weibull Distribution

$$l(\alpha,\gamma,\lambda,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln a + \sum_{i=1}^{n} \delta_i \ln\gamma + \sum_{i=1}^{n} \delta_i a \ln\lambda - (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \gamma\left(\frac{\lambda}{t_i}\right)^{a} e^{x_i\beta}$$

- Modified Weibull Distribution

$$l(\alpha,\gamma,\lambda,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\gamma - \sum_{i=1}^{n} \delta_i \ln(a + \lambda t_i) + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i \lambda t_i - \sum_{i=1}^{n} \delta_i \gamma {t_i}^{a} e^{\lambda t_i} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln[1 - \exp(-\gamma {t_i}^{a} e^{\lambda t_i})]$$

- Flexible Weibull Distribution

$$l(\alpha,\gamma,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\left(a + \frac{\gamma}{{t_i}^2}\right) + \sum_{i=1}^{n} \delta_i \left(a t_i - \frac{\gamma}{t_i}\right) - \sum_{i=1}^{n} \delta_i e^{a t_i - \frac{\gamma}{t_i}} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left[1 - \exp\left(-e^{a t_i - \frac{\gamma}{t_i}}\right)\right]$$

- Strong Extended Weibull Distribution

$$l(\alpha,\gamma,\lambda,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln a + \sum_{i=1}^{n} \delta_i \ln\gamma - \sum_{i=1}^{n} \delta_i a \ln\lambda + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i + \left(\frac{1}{\gamma} - 1\right)\sum_{i=1}^{n} \delta_i \ln\left[1 + \left(\frac{t_i}{\lambda}\right)^{a}\right] + \sum_{i=1}^{n} \delta_i \left[1 - \left(1 + \left(\frac{t_i}{\lambda}\right)^{a}\right)^{\frac{1}{\gamma}}\right] + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln\left\{1 - \exp\left[1 - \left(1 + \left(\frac{t_i}{\lambda}\right)^{a}\right)^{\frac{1}{\gamma}}\right]\right\}$$

- Marshal-Olkin Distribution

$$l(\alpha,\gamma,\lambda,t) = \sum_{i=1}^{n} \delta_i x_i \beta + \sum_{i=1}^{n} \delta_i \ln\gamma + \sum_{i=1}^{n} \delta_i \ln a - \sum_{i=1}^{n} \delta_i a \ln\lambda + (a-1)\sum_{i=1}^{n} \delta_i \ln t_i - \sum_{i=1}^{n} \delta_i (\lambda t_i)^{a} + \sum_{i=1}^{n} (e^{x_i\beta} - \delta_i) \ln[1 - e^{-(\lambda t_i)^{a}}] - \sum_{i=1}^{n} (e^{x_i\beta} + \delta_i) \ln[1 - (1-\gamma) e^{-(\lambda t_i)^{a}}]$$

This study aim, determine Parametric, non-parametric and semi-parametric Models for various distributions that may be suitable for left-censored data. The performances of these derived models in analyzing HIV viral load data were compared using extensive simulations and a guide was created in which the distribution was optimal. Each simulation created varies with the sample size and censorship rate to find a consistently high distribution of performance. In our study; Karapınar State Hospital data were reviewed retrospectively, and a bootstrap study aimed to provide more information on the appropriateness of distributions in the analysis of HIV viral load data. Such data are frequently encountered in the literature, especially in biological studies. In this study, the advantages and disadvantages of the methods used for left-hand censored (Under the Limit of Detection) data commonly encountered in clinical studies will be examined. According to this data structure, it will be investigated which statistical method will be more appropriate to use and the reasons for this will be discussed in the study. In addition, another aim of our study is to compare different parameter estimation methods (nonparametric, parametric and semi-parametric) and determine the strength of these methods in left-censored data sets.

Methods

In earlier studies, performances between nonparametric, parametric, semi-parametric and substitution methods with different sample sizes and various censoring rates were compared and road maps were determined. However, in the parametric approach, no performance comparison has been executed between the distributions suitable for left-censored data. Therefore, this study was designed to determine a road map by conducting performance evaluations of 15 different distributions, with different sample sizes (1000-2000-3000) and different censoring rates (15%, 25% and 35%).

- Simulation studies were conducted at 15%-25% and 35% left-censored rates. While forming the censoring rate, 0 represented a censored observation and otherwise, the censoring indicator was 1. Then, censoring was determined by assigning the indicator to random observation rate. AIC, AICC, HQIC and CAIC information criteria were used to evaluate which distribution model was most suitable.
- After censoring, censored data were estimated with Parametric method (MLE) for 15 distributions with the help of the package of NADA (Nondetects And Data Analysis: Statistics for Censored Environmental Data) written by Lopaka Lee and updated on July 2, 2014 in the R program (Version i386 3.0.2). The repetition number for the simulation was determined as 5000. In addition, the evaluation was done by increasing the sample sizes by 1000-2000-3000. In order for the variations of the distributions to be equal, the data were derived from the following distributions: ; *Inv. Weibull* (1.5;1), *Exponential* (3.6;0.25), *Log-Normal* (0.5;2), *Inv. Gaussian* (3;1), *Log-Logistic* (3;0.3), *Gompertz-Makeham* (0.2;0.05;0.85), *Gamma* (3;2), *Gen. Gamma* (2;1.5;1), *Inv.Gamma* (2;1), *Weibull* (1.8;1), *Gen.Inv.Weibull* (1.8;1;1), *Modified Weibull* (2.1;0;1.5;1;1), *Flexible Weibull*(1.5;1), *Power Generalized Weibull* (1;1.5), *Marshal-Olkin* (3;2;1). The parameters of all distributions were different, but the variation coefficient was 0.5. The likelihood density functions of these distributions are given in Figure 1.

Results

The results of the simulation study are summarized in Table 2-4. Table 2 shows the results of simulated data with censoring rate of 15%, Table 3 for data with censoring rate of 25%, and Table 4 for data with censoring rate of 35%. The brief results of the information criteria for different Censorship rates of distributions are given in Figure 2. According to these results, the Generalized Inverse Weibull distribution has the lowest mean AIC, AICC, HQIC and CAIC values. Then, there are Log-Logistic, Log-Normal, Inverse Gaussian and Gamma distributions, respectively. This situation is consistent across all censoring rates and sample sizes. The

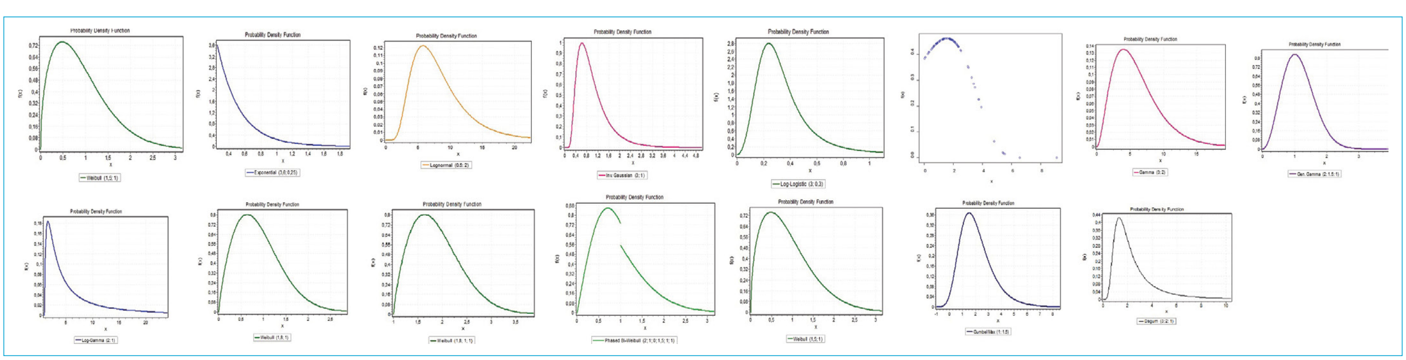

Figure 1. Likelihood density functions of the distributions.

distributions that consistently perform as the worst are as follows respectively: Modified Weibull, Inverse Weibull, Inverse Gamma, Strong Generalized Weibull and Exponential distributions.

Besides, as a result of the conducted simulation studies, the best distributions to be used for Parametric Inverse Hazard Models in the left-censored data, are the Log-Logistic, Log-Normal, Inverse Normal and Gamma distributions, respectively along with the Generalized Inverse Weibull distribution. We can also say that the Marshal-Olkin distribution shows superior performance compared to Modified Weibull, Generalized Gamma, Gamma, and Flexible Weibull distributions.

Real Data Application

In this section, it is aimed to determine the optimal distribution for various censoring rates (15%-25%-35%) in left-censored data for Parametric Inverse Hazard Models. Between 2016 and 2019, it has collected at the Konya Provincial Directorate Kararapınar State Hospital. The study was designed as a retrospective. Karapinar State Hospital data were screened retrospectively. As a bootstrap study, the analysis of HIV viral load data is intended to provide more information on the appropriateness of deployments. Anti-HBs, Anti Hcv, Anti-HIV and HBsAg were 30%, 11%, 2.3% and 51% respectively. Approval was obtained from the Ethics Committee of Bilecik Public Health Directorate with the code 959222041-449 for the study.

26557 individuals diagnosed with Anti-HBs, Anti Hcv, Anti-HIV and HBsAg have been summarized with frequencies and percentages in Table 5.

Firstly, the number of left-censored data was determined for each diagnosis, and censored data numbers have been given in Table 6.

The censored data for each diagnosis was determined and various descriptive statistics of the uncensored data are given in Table 7.

The number of diagnosis and censored data and also descriptive statistics of uncensored data of diagnosis have been given in Table 5-7, respectively.

After this descriptive step, the assessment of fit for distributions was made for each diagnosis.

Firstly, 65 distributions were tested for the Anti-HBS diagnosis, and Exponential, Gamma, Log-Logistic and Logistic distributions were fitted. Then fitted tests have been performanced for Anti HCV diagnosis. According to test results, Gamma, Inv. Gaussian, Log-Logic, Logic, Lognormal and Weibull distributions have been fitted. The same examination was made for ANTİ-HIV and HBsAg diagnoses

Table 2. Brief results of information criteria obtained according to distributions conducted with 5000 repetitions of which their censor rate was 15%, and sample sizes were 1000-2000-3000

Sample size	**1000**				**2000**				**3000**			
Distribution	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**
1. Exponential	8752,701	8751,463	8753,305	8757,737	17503,31	17502,23	17503,57	17510,25	26255,19	26253,08	26256,96	26259,53
2. Log-Normal	6161,563	6161,492	6164,413	6172,64	12321,95	12324,1	12325,76	12335,6	18484,83	18484,81	18489,03	18497,06
3. Inverse Normal	6161,491	6162,98	6166,177	6175,346	12324,17	12324,16	12329,02	12338,48	18487,94	18489,03	18492,83	18504,12
4. Gamma	6198,17	6197,826	6200,949	6209,82	12396,35	12397,33	12401,51	12407,22	18594,91	18596,69	18600,02	18610,63
5. Generalized Gamma	6207,292	6205,223	6210,888	6222,996	12408,71	12408,84	12416,18	12429,48	18613,75	18615,5	18620,56	18634,08
6. Inverse Gamma	10620,34	10620,66	10625,85	10632,32	21238,89	21239,88	21242,63	21251,21	31856,42	31857,7	31860,67	31871,36
7. Log-Logistics	6113,914	6115,107	6116,691	6124,624	12229,36	12227,44	12232,84	12243,41	18345,63	18344,06	18349,75	18357,4
8. Weibull	6500,858	6500,61	6505,212	6513,266	13006,5	13006,66	13010,78	13019,86	19512,24	19514,34	19518,93	19527,81
9. Inverse Weibull	13988,18	13991,78	13994,87	14003,31	27980,79	27980,69	27982,42	27993,47	41969,08	41969,2	41972,86	41983,3
10. Generalized Inverse Weibull	5958,398	5957,893	5962,879	5977,059	11914,01	11913,96	11922,91	11934	17871,92	17871,88	17879,83	17893,25
11. Flexible Weibull	7766,727	7767,223	7770,535	7777,048	15485,55	15487,57	15491,88	15500,93	23309,28	23311,69	23316,83	23324,4
12. Marshal-Olkin	7530,959	7530,511	7536,762	7549,605	15058,41	15058,38	15063,65	15078,76	22585,66	22585,45	22592,23	22605,92
13. Power Generalized Weibull	10490,43	10491,43	10496,58	10503,69	20979,16	20978,21	20984,28	20992,55	31467,66	31468,08	31472,88	31481,81
14. Modified Weibull	2,97E+63	2,97E+63	2,97E+63	2,97E+63	1,27E+68	1,27E+68	1,27E+68	1,27E+68	1E+68	1E+68	1E+68	1E+68
15. Gompertz	8176,282	8176,16	8180,808	8186,574	16176,17	16178,11	16180,96	16189,17	24097,66	24098,09	24103,62	24112,41

Table 3. Brief results of information criteria obtained according to distributions conducted with 5000 repetitions of which their censor rate was 25%, and sample sizes were 1000-2000-3000

Sample size	1000				2000				3000			
Distribution	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**
1. Exponential	7751,092	7749,683	7754,429	7756,775	15501,75	15500,47	15502,53	15508,73	23253,39	23251,57	23255,75	23260,94
2. Log-Normal	5568,799	5570,782	5572,471	5580,358	11139,29	11139,66	11141,96	11153,63	16710,21	16710,76	16713,71	16724,78
3. Inverse Normal	5571,307	5570,532	5574,913	5583,953	11141,23	11143,97	11146,92	11158,14	16713,27	16714,45	16717,16	16727,38
4. Gamma	5602,585	5602,036	5606,052	5613,745	11206,01	11204,33	11209,72	11217,36	16808,47	16806,55	16811,65	16821,12
5. Generalized Gamma	5610,511	5610,754	5616,96	5625,553	11216,55	11217,61	11225,11	11239,57	16825,97	16825,28	16832,15	16846,46
6. Inverse Gamma	9728,889	9729,646	9733,385	9741,775	19457,25	19457,08	19462,3	19470,88	29183,65	29183,7	29189,63	29197,92
7. Log-Logistics	5519,84	5519,301	5523,517	5533,149	11040,81	11040,49	11044,33	11054,14	16561,16	16560,37	16565,02	16575,84
8. Weibull	5847,881	5846,966	5850,535	5858,704	11698,79	11700,61	11702,95	11713,15	17550,17	17548,89	17554,79	17564,34
9. Inverse Weibull	12428,21	12429,13	12434,67	12441,15	24856,58	24856,74	24858,48	24868,74	37283,94	37281,76	37287,32	37298,27
10. Generalized Inverse Weibull	5399,195	5399,099	5406,7	5417,052	10796,93	10797,16	10802,57	10816,21	16196,99	16195,08	16202,12	16216,63
11. Flexible Weibull	6841,362	6842,428	6846,43	6854,086	13626,13	13627,76	13630,93	13638,42	20402,14	20402,07	20404,51	20415,68
12. Marshal-Olkin	7115,576	7118,48	7122,583	7134,742	14228,55	14229,84	14236,35	14249,44	21341,57	21343,95	21347,82	21363,44
13. Power Generalized Weibull	9251,347	9250,307	9254,689	9260,674	18497,21	18497,74	18502,02	18511,34	27744,59	27746,31	27749,97	27760,13
14. Modified Weibull	6,15E+63	6,15E+63	6,15E+63	6,15E+63	1,95E+63	1,95E+63	1,95E+63	1,95E+63	5,2E+64	5,2E+64	5,2E+64	5,2E+64
15. Gompertz	7116,079	7114,564	7119,731	7126,904	14421,55	14424,49	14425,78	14436,94	21421,04	21421,08	21426,13	21434,06

Table 4. Brief results of information criteria obtained according to distributions conducted with 5000 repetitions of which their censor rate was 35%, and sample sizes were 1000-2000-3000

Sample size	1000				2000				3000			
Distribution	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**	**AIC**	**AICC**	**HQIC**	**CAIC**
1. Exponential	6737,475	6739,796	6742,002	6745,848	13479,96	13477,43	13479,25	13484,69	20220,76	20218,51	20219,21	20226,02
2. Log-Normal	4951,469	4950,285	4957,528	4963,767	9904,529	9902,312	9908,291	9918,099	14859,49	14859,61	14862,78	14875,52
3. Inverse Normal	4952,628	4952,528	4955,593	4963,524	9907,178	9908,463	9911,179	9920,755	14863,39	14863,75	14869,86	14879,3
4. Gamma	4979,935	4981,464	4983,989	4993,026	9961,035	9965,01	9966,726	9977,424	14945,89	14947,29	14949,7	14961,52
5. Generalized Gamma	4986,936	4984,293	4991,783	5002,674	9969,986	9970,296	9974,619	9992,468	14959,87	14960,62	14967,24	14981,58
6. Inverse Gamma	8647,607	8649,49	8652,795	8660,086	17297,73	17297,08	17302,65	17310,5	25943,16	25944,53	25950,2	25957,97
7. Log-Logistics	4904,27	4903,875	4907,82	4916,772	9806,907	9807,704	9813,026	9822,888	14715,58	14716,51	14719,89	14727,63
8. Weibull	5175,644	5170,906	5176,507	5183,791	10346,45	10345,96	10350,93	10360	15526,38	15526,41	15529,78	15539,81
9. Inverse Weibull	10837,63	10841,3	10843,28	10851,16	21676,71	21675,92	21681,34	21687,95	32512,8	32512,31	32517,49	32528,91
10. Generalized Inverse Weibull	4812,903	4813,966	4819,703	4830,538	9622,236	9623,351	9630,946	9645,391	14437,89	14438,83	14445,05	14461,07
11. Flexible Weibull	5924,483	5924,522	5931,111	5936,027	11892,07	11889,09	11895,29	11903	17875,54	17876,87	17882,2	17889,62
12. Marshal-Olkin	6702,979	6702,162	6708,416	6720,46	13401,13	13401,25	13407,57	13423,23	20100,12	20100,65	20108,25	20121,29
13. Power Generalized Weibull	8001,563	8001,103	8006,717	8014,057	16001,25	16003,06	16005,16	16014,21	24001,95	24001,49	24003,88	24015,63
14. Modified Weibull	4,6E+63	4,6E+63	4,6E+63	4,6E+63	2,09E+62	2,09E+62	2,09E+62	2,09E+62	9,34E+65	9,34E+65	9,34E+65	9,34E+65
15. Gompertz	6218,138	6216,796	6223,172	6228,448	12432,6	12434,07	12437,87	12448,56	18725,48	18723,95	18730,27	18739,52

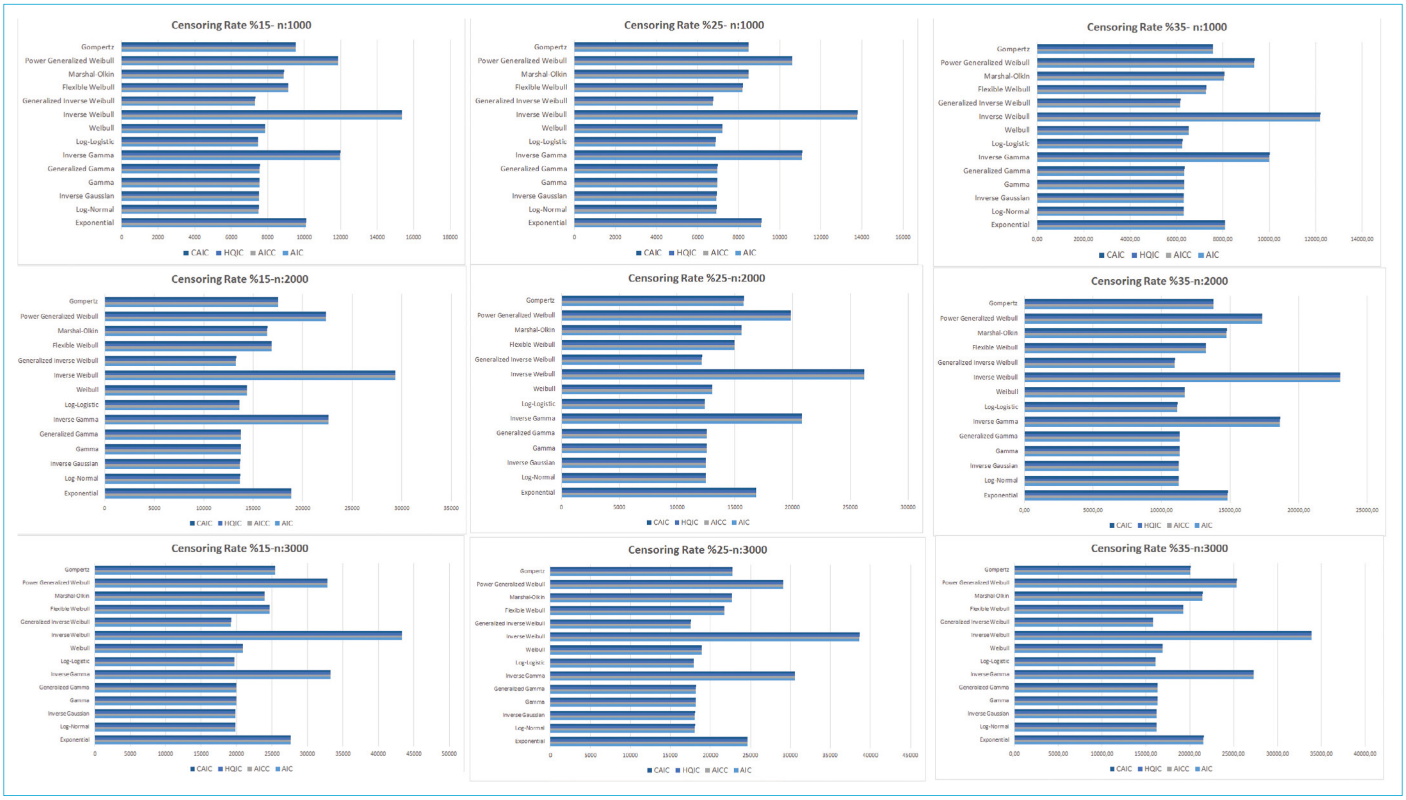

Figure 2. Brief results of information criteria obtained according to distributions conducted with 5000 repetitions of which their censor rate was 15-25-35%, and sample sizes were 1000-2000-3000.

Table 5. The diagnosis numbers of individuals

	n	%
Anti-HBs	6376	24
Anti Hcv	6353	23.9
Anti-HIV	6058	22.8
HBsAg	7770	29.3
Total	26557	100

Table 6. The values and frequencies of various censored data for each diagnosis

Diagnosis	n	%
Anti-HBs		
(<3.10)	2017	31.6
(>1000,00)	709	11.1
Anti HCV		
(<0.02)	708	11.1
(<0.10)	1	0
(>11.00)	19	0.3
(<0.001)	1	0
Anti-HIV		
(<0.05)	142	2.3
(<0.050)	1611	26.6
(<0.001)	1	0
HBsAg		
(<0.10)	3969	51.1
(>1000,0)	202	2.6

Table 7. The descriptive statistics of the uncensored data

	n	$\bar{x}$	Standart Deviation	Median	Min	Max
Anti-HBs	3650	169.945	235.104	56.74	3.1	999.99
Anti HCV	5624	0.1382	0.2889	0.11	0.01	10.08
Anti-HIV	4301	0.1958	0.18507	0.168	0.05	7.47
HBsAg	3599	3.018	36.9432	0.24	0.1	841.67
Total	17174	36.8451	129.667	0.19	0.01	999.99

Min: Minimum; Max: Maximum.

seperately, and appropriate distributions and parameters estimations' results were summarized in Table 8.

After the parameter estimation, left-censored data was generated for the distributions which are Log-Logic, Exponential, Lognormal and Gamma distributions. Generated data entegrated the main data and the descriptive statistics of the each diagnosis are given in Table 9.

Finally, the mean, standard deviation and median differences were calculated for each diagnosis, and the results are given in Table 10.

The biggest difference is Gamma, Log normal, and Exponential, respectively; The least difference was obtained for the Log logistic distribution.

Table 8. Fitting results of the Anti-HBs, Anti HCV, Anti-HIV, HBsAg diagnosis with parameters' estimations

Diagnosis	Distribution	Parameters
Anti-HBs	Exponential	λ=0.00588
	Gamma	∝=0.52251 β=325.25
	Log-Logistic	∝=0.99376 β=52.014
	Logistic	σ=129.62 μ=169.94
Anti HCV	Gamma	∝=0.22893 β=0.60381
	Inv. Gaussian	λ=0.03165 μ=0.13823
	Log-Logistic	∝=2.5347 β=0.10395
	Logistic	σ=0.15928 μ=0.13823
	Lognormal	σ=0.70316 μ=-2.2631
	Weibull	∝=1.7642 β=0.14413
Anti-HIV	Cauchy	σ=0.05344 μ=0.16036
	Gamma	∝=1.1236 β=0.17468
	Gen. Gamma	k=1.3293 ∝=1.3912 β=0.17468
	Log-Logistic	∝=3.2453 β=0.16553
HBsAg	Cauchy	σ=0.08885 μ=0.2142
	Gen. Gamma	k=0.58736 ∝=0.24281 β=452.16
	Inv. Gaussian	λ=0.23482 μ=3.0102
	Log-Logistic	∝=1.6918 β=0.26708

Discussion

New and effective treatments for Covid-19 are only possible by analyzing and evaluating case data. Although the analyzes, models, simulations and evaluations of Covid-19 are in the literature, they are not at a sufficient level in terms of solution. In this sense, it is thought to contribute to the literature and guide researchers by highlighting the censored data analysis approach in the investigation of the Covid 19 outbreak.

Conclusion

Many studies have presumed that there are certain fundamental distributions based on guidelines for the distribution functions of censored data or analyzed data. The distributions were identified as negatively skewed, on the grounds that all of the common distributions associated with left/right-censored data such as the Normal, Log-Normal and Gamma distributions belong to the exponential family.

Table 9. The descriptive results of uncencored and left-censored data for diagnosis

	n	$\bar{x}$	Standart Deviation	Median	Minimum	Maximum
Uncencored						
Anti-HBs	3650	169.9445	235.1041	56.74	3.1	999.99
ANTİ HCV	5624	0.1382	0.2889	0.11	0.01	10.08
ANTİ HIV	4301	0.1958	0.18507	0.168	0.05	7.47
HBsAg	3599	3.018	36.94318	0.24	0.1	841.67
Total	17174	36.8451	129.6673	0.19	0.01	999.99
Log-Logistic						
Anti-HBs	5667	168.6555	234.1395	55.739121	3.1	999.99
ANTİ HCV	6334	0.1093	0.276246	0.0159	0.01	10.08
ANTİ HIV	6058	0.1814	0.12867	0.102516	0.05	7.47
HBsAg	7568	3.0095	36.91428	0.22433	0.1	841.67
Total	25627	36.7572	129.5693	0.1745352	0.01	999.99
Exponential						
Anti-HBs	5667	167.3665	232.8505	54.450121	3.1	999.99
ANTİ HCV	6334	0.0804	0.247346	0.015882	0.01	10.08
ANTİ HIV	6058	0.1525	0.09977	0.073616	0.05	7.47
HBsAg	7568	2.9806	36.88538	0.19543	0.1	841.67
Total	25627	36.7283	129.5404	0.1456352	0.01	999.99
Lognormal						
Anti-HBs	5667	166.0775	231.5615	53.161121	3.1	999.99
ANTİ HCV	6334	0.0515	0.218446	0.0719	0.01	10.08
ANTİ HIV	6058	0.1236	0.07087	0.084716	0.05	7.47
HBsAg	7568	2.9517	36.85648	0.16653	0.1	841.67
Total	25627	36.6994	129.5115	0.1167352	0.01	999.99
Gamma						
Anti-HBs	5667	164.7885	230.2725	51.872121	3.1	999.99
ANTİ HCV	6334	0.0226	0.189546	0.0708	0.01	10.08
ANTİ HIV	6058	0.0947	0.04197	0.075816	0.05	7.47
HBsAg	7568	2.9228	36.82758	0.13763	0.1	841.67
Total	25627	36.6705	129.4826	0.0878352	0.01	999.99

Table 10. The differances of mean, standart deviation and median for uncencored and and left-censored data

	n	Mean difference	Standart deviation difference	Median difference	Minimum	Maximum
Log Logistics						
Anti-HBs	5667	1.289	0.9646	1.0008789	3.1	999.99
ANTİ HCV	6334	0.0289	0.012654	0.0941	0.01	10.08
ANTİ HIV	6058	0.0144	0.0564	0.065484	0.05	7.47
HBsAg	7568	0.0085	0.0289	0.01567	0.1	841.67
Total	25627	0.0879	0.098	0.0154648	0.01	999.99
Exponential						
Anti-HBs	5667	2.578	2.2536	2.2898789	3.1	999.99
ANTİ HCV	6334	0.0578	0.041554	0.094118	0.01	10.08
ANTİ HIV	6058	0.0433	0.0853	0.094384	0.05	7.47
HBsAg	7568	0.0374	0.0578	0.04457	0.1	841.67
Total	25627	0.1168	0.1269	0.0443648	0.01	999.99
Log Normal						
Anti-HBs	5667	3.867	3.5426	3.5788789	3.1	999.99
ANTİ HCV	6334	0.0867	0.070454	0.0381	0.01	10.08
ANTİ HIV	6058	0.0722	0.1142	0.083284	0.05	7.47
HBsAg	7568	0.0663	0.0867	0.07347	0.1	841.67
Total	25627	0.1457	0.1558	0.0732648	0.01	999.99
Gamma						
Anti-HBs	5667	5.156	4.8316	4.8678789	3.1	999.99
ANTİ HCV	6334	0.1156	0.099354	0.0392	0.01	10.08
ANTİ HIV	6058	0.1011	0.1431	0.092184	0.05	7.47
HBsAg	7568	0.0952	0.1156	0.10237	0.1	841.67
Total	25627	0.1746	0.1847	0.1021648	0.01	999.99

This study is aimed to perform simulation studies with different sample size (1000-2000-3000) and various censoring rates (15%-25%-35%) for evaluating the performance of derived Parametric Inverse Hazard models and to compare these models for identifying the most suitable distribution for the left-censored data. Since it was not possible to study simulations of all possible scenarios, scenarios similar to each other were generally preferred over others. In the simulation studies of the given scenarios, it was concluded that the most appropriate distribution to be used for Parametric Inverse Hazard models for left-censored data is the Log-Logistic distribution when the censoring rates were 15%-25% and 35%.

Simulation studies revealed that the best distributions to be used under the Parametric Inverse Hazard models were (in order of performance) Log-Logistic, Log-Normal, Inverse Gaussian and Gamma distribution, and Generalized Inverse Weibull distribution, respectively. As a result of the simulation study, it was revealed that the "Marshal-Olkin" distribution followed by "Modified Weibull", "Generalized Gamma", "Gamma" and "Flexible Weibull" distributions demonstrated conservative performance, respectively. In the results made with real data application, it has been determined that the best result is Log logistics distribution. Simulation results and real data application results support each other.

Since there is no clear suggestion about the most suitable model and distribution in the analysis of left-censored data in the literature today, the findings of this may guide researchers in terms of the use of more accurate models for left censored data.

Covid 19 is a pandemic that has affected the whole world, and the science world predicts that the effects and consequences of the pandemic will affect the world for many years.

In this process, technological inadequacies, people's attitudes towards the pandemic, insufficient infrastructure in the health sector, and the governments' wrong strategies with economic concerns are the most important problems and affect the functioning negatively. The investigations made by data analysts working on epidemic statistics have revealed the course of the epidemic, and updating the data, changing it or not sharing it properly caused mistakes.

The data insufficiency detected in the epidemic data of the countries causes serious differences in the results of the researchers involved in predicting the course of the epidemic.

This situation may cause inadequacies in the measures to be taken by the countries and incorrect evaluations in health services. Since the inadequacies, inaccuracies and changes experienced in the data may affect the consistency of the estimates developed by the researchers, it has led to the idea that censorship can be made in such cases.

Disclosures

Ethics Committee Approval: The ethics committee of Bilecik Public Health Directorate provided the ethics committee approval for this study (959222041-449).

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Authorship Contributions: Concept – M.A.T.; Design – M.A.T., H.Y.; Supervision – M.A.T.; Materials – S.K.; Data collection &/or processing – M.A.T., H.Y.; Analysis and/or interpretation – M.A.T., H.Y.; Literature search – H.Y.; Writing – M.A.T., H.Y.; Critical review – M.A.T., H.Y.

References

1. Nelson WB. Applied life data analysis. John Wiley and Sons; 1982.
2. Huston C, Juarez-Colung E. Guidelines for computing summary statistics for data-sets containing non-detects. Smithers, Canada: Bulkley Valley Research Center; 2009. Available at: https://bvcentre.ca/files/research_reports/08-03GuidanceDocument.pdf. Accessed Apr 20, 2021.
3. Lawless JF. Statistical models and methods for lifetime data. New York: John Wiley and Sons; 2003.
4. Horny G. Inference in mixed proportional hazard models with K random effects. Stat Pap 2009;50:481–99.
5. Glass DC, Gray CN. Estimating mean exposures from censored data: exposure to benzene in the Australian petroleum industry. Ann Ocp Hyg 2001;45:275–82.
6. Mulhausen J, Damiano J. A strategy for assessing and managing occupational exposures. 2nd ed. Fairview, VA: American Industrial Hygiene Association; 1998. p. 349.
7. Hawkins NC, Norwood SK, Rock JC. A strategy for occupational exposure assessment. Fairview, VA: American Industrial Hygiene Association; 1991.
8. Lee L, Helsel DR. Statistical analysis of water-quality data containing multiple detection limits: S-language software for regression on order statistics. Comput and Geosci 2005;31:1241–8.
9. Finkelstein MM, Verma DK. Exposure estimation in the presence of nondetectable values: another look. AIHAJ 2001;62:195–8.
10. She N. Analyzing censored water quality data using a non-parametric approach. JAWRA 1997;33:615–24.
11. Shumway RH, Azari RS, Kayhanian M. Statistical approaches to estimating mean water quality concentrations with detection limits. Environ Sci Technol 2002;36:3345–53.
12. Hewett P, Ganser GH. A comparison of several methods for analyzing censored data. Ann Occup Hyg 2007;51:611–32.
13. Islam F. Parametric reversed hazards model for left censored data with application to HIV. Master's thesis. Carolina, US: University of South Carolina; 2016. Available at: https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=4894&context=etd. Accessed Apr 20, 2021.
14. Odell PM, Anderson KM, D'Agostino RB. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. Biometrics 1992;48:951–9.
15. Luczynska C, Sterne J, Bond J, Azima H, Burney P. Indoor factors associated with concentrations of house dust mite allergen, Der p 1, in a random sample of houses in Norwich, UK. Clin Exp Allergy 1998;28:1201–9.
16. Gupta RD, Kundu D. Theory & methods: Generalized exponential distributions. Aust N Z J Stat 2002;41:173–88.
17. Thompson RE, Volt EO, Scott GI. Statistical modeling of sediment and oyster PAH contamination data collected at a South Carolina estuary (complete and left-censored samples). Environmetrics 200;11:99–119.
18. Pajek M, Kubala-Kukus A, Banas D, Braziewicz J, Majewska U. Random left-censoring: a statistical approach accounting for detection limits in x-ray fluorescence analysis. XRSPAX 2004;33:306–11.
19. Isingo R, Zaba B, Marston M, Ndege M, Mngara J, Mwita W, et al. Survival after HIV infection in the pre-antiretroviral therapy era in a rural Tanzanian cohort. AIDS 2007;21:S5–S13.
20. Annan SY, Liu P, Zhang Y. Comparison of the Kaplan-Meier, maximum likelihood, and ROS estimators for left-censored data using simulation studies; 2009. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.211.1833&rep=rep1&type=pdf. Accessed Apr 20, 2021.
21. Kremer A, Weißbach R, Liese F. Maximum likelihood estimation for left-censored survival times in an additive hazard model. J Stat Plan 2014;149:33–45.
22. Pesonen M, Pesonen H, Nevalainen J. Covariance matrix estimation for left-cencored data. Comput Stat Data Anal 2015;92:13–25.
23. Achar JA, Coelho-Barros EA, Cuevas JRT, Mazucheli J. Use of Lèvy distribution to analyze longitudinal data with asymmetric distribution and presence of left censored data. CSAM 2018;25:43–60.
24. Fusek M, Michálek J, Buňková L, Buňka F. Modelling biogenic amines in fish meat in Central Europe using censored distributions. Chemosphere 2020;251:126390.
25. Omar S, Bartz C, Becker S, Basenach S, Pfeifer S, Trapp C, et al; Palatina Public Health Study Group. Duration of SARS-CoV-2 RNA detection in COVID-19 patients in home isolation, Rhineland-Palatinate, Germany, 2020 - an interval-censored survival analysis. Euro Surveill 2020;25:2001292.
26. Sreedaevi EP, Sankaran PG. Statistical methods for estimating cure fraction of COVID-19 patients in India. Medrxiv. 2020 Jun

30. Doi: https://doi.org/10.1101/2020.05.30.20117804. [Epub ahead of print].
27. Mollazehi M, Mollazehi M, Abdel-Salam GAS. Modeling survival time to recovery from COVID-19: a case study on Singapore. Research Square. 2020 Mar 30. Doi: 10.21203/rs.3.rs-18600/v2. [Epub ahead of print].
28. Duchateau L, Janssen P. The frailty model. 1st ed. New York: Springer-Verlag; 2008.
29. Variyath AM, Sankaran PG. Parametric regression models using reversed hazard rates. J Probab Stat 2014;2014:645719.
30. Hossain S. Multivariate granger causality between economic growth, electricity consumption, exports and remittance for the panel of three SAARC countries. Eur Sci J 2012;8:347–76.
31. Barlow RE, Marshall AW, Proschan F. Properties of probability distributions with monotone hazard rate. Ann Math Statist 1963;34:375–89.
32. Lwless JF. Statistical models and methods for lifetime data. New York: John Wiley and Sons; 2011.
33. Keilson J, Sumita U. Uniform stochastic ordering and related inequalities. Can J Stat 1982;10:181–98.
34. Block HW, Savits TH, Singh H. The reversed hazard rate function. Probab Eng Inf Sci 1998;12:69–90.
35. Gupta RC, Gupta PL, Gupta RD. Modeling failure time data by lehman alternatives. Commun Stat-Theor M 1998;27:887–904.
36. Lai CD, Xie M. Stochastic ageing and dependence for reliability. 1st ed. New York: Springer-Verlag; 2006.
37. Li X, Xu M. Reversed hazard rate order of equilibrium distributions and a related aging notion. Stat Pap 2008;49:749–67.
38. Marshall AW, Olkin I. Life distributions. 1st ed. New York: Springer-Verlag; 2007.
39. Tojeiro CA, Louzada F. A general threshold stress hybrid hazard model for lifetime data. Stat Pap 2012;53:833–48.
40. Hosmer Jr DW, Lameshow S, May S. Applied Survival Analysis: Regression modeling of time-to-event data. 2nd ed. New York: Wiley; 2008.
41. Gijbels I. Censored data. Wiley Interdiscip Rev Comput Stat 2010;2:178–88.
42. Tekindal MA, Erdoğan BD, Yavuz Y. Evaluating left-censored data through substitution, parametric, semi-parametric, and nonparametric methods: a simulation study. Interdiscip Sci 2017;9:153–72.