**ESTIMATION OF SURVIVAL TIMES OF COVID-19 PATIENTS USING SOME LIFETIME DISTRIBUTIONS**

***Caner TANIŞ1,[[1]](#footnote-1)\****

*1 Science Faculty, Statistics, Çankırı Karatekin University, Çankırı, Turkey*

|  |
| --- |
| **Abstract** In this study, the survival time (in days) of Covid-19 patients from hospitalization and death were modeled with some known lifetime distributions such as Weibull, transmuted Weibull, exponentiated Weibull, and generalized Lindley distributions. The maximum likelihood method is considered for point estimation. We present four data sets on Covid-19 patients. The goodness of the fitted distribution is evaluated via some selection criteria such as Akaike information criterion, the Bayesian information criterion, the Kolmogorov-Smirnov test statistic, the Anderson Darling statistic, the Cramér von Mises statistic, and the p-value criteria Also, the estimated probabilities of the survival times of Covid-19 patients were calculated via the invariance property of maximum likelihood estimation. In dying Covid-19 patients, the average survival time is estimated approximately 15 days. |
| Keywords: Covid-19, Lifetime distribution, Maximum likelihood, Point estimation |

1. **Introduction**

The new coronavirus, also known as the Covid-19 virus, appeared in Wuhan, China in December 2019 and rapidly exceeded its borders and affected the whole world. The WHO officially reported that Covid-19 is a global pandemic on March 11, 2020 [1]. As of March 27, 2023, it is globally announced that there have been 761,402,282 confirmed Covid-19 cases with 6,887,000 deaths by the World Health Organization (WHO) [2]. Tian et al. [3] analyzed the findings of many studies on mortality in hospitalized Covid-19 patients. They reported that patients with damaged organs such as the heart, liver, kidney, had a high risk of death due to Covid-19. Covino et al. [4] investigated the clinical properties and prognostic indicators in Covid-19 patients aged more than 80 years. The findings of Covino et al. [4] show that the risk of death could be not age dependent in Covid-19 patients aged more than 80 years while severe dementia emerged as a risk factor in this group. Cheng et al. [5] investigated the death periods in hospitalized Covid-19 patients. In the literature, many statistical distributions are used to model data obtained in many fields such as biology, chemistry, engineering, and medical sciences. Some of the popularly used lifetime distributions can be shown as Weibull, Lindley, and various modified versions of these distributions. In this study, different from other studies in the literature, the estimates will be provided about the survival times of Covid-19 patients using some known lifetime distributions including Weibull, transmuted Weibull, and generalized Lindley distribution. The rest of this paper is organized as follows: In Section 2, we describe the emphasized lifetime distributions. In Section 3, the point estimation of the examined distributions is given. In Section 4, Model evaluation is given. In Section 5, the results and discussion are presented. Finally, the conclusions are given in Section 6.

1. **Materials and Methods**
	1. **Weibull Distribution**

Weibull distribution is one of the most popular distributions used to model lifetime data. The cumulative distribution function (CDF) and probability density function (PDF) and first moment of the Weibull distribution are given by

, (1)

 (2)

and

 (3)

respectively, where, is a shape parameter,  is a scale parameter, and .

* 1. Exponentiated Weibull Distribution

The EW distribution suggested by Pal et al. [6] is a generalization of Weibull distribution. The CDF and PDF and first moment of the EW distribution are given by

, (4)

 (5)

and

, (6)

where,  denote natural number. and  are shape parameters, is a scale parameter, and [7]. The EW distribution is reduced Weibull distribution for.

* 1. Transmuted Weibull Distribution

Transmuted Weibull (TW) distribution proposed by Aryal and Tsokos [8] is another generalization of Weibull distribution. The CDF, PDF and first moment of TW distribution are

, (7)

 (8)

and

 (9)

respectively, where, and. The TW distribution is reduced Weibull distribution for.

* 1. Generalized Lindley Distribution

Lindley distribution introduced by [9] can be used in various fields such as biology, engineering, and medical sciences is another popular lifetime distribution. Ghitany et al. [10] mentioned that Lindley distribution is particularly useful for modeling mortality data.

The GL distribution proposed by [11] is a generalization of Lindley distribution. The CDF, PDF and first moment of the GL distribution are

 (10)

 (11)

and

 (12)

where, , , and[11].

1. **Point Estimation**

 denote a random sample from the Weibull  distribution and indicate the observed value of for . For the Weibull  distribution, the log-likelihood function is

 (13)

For the Weibull  distribution, the maximum likelihood estimators of the and  parameters maximize the function in Eq. (13).

 denote a random sample from EW  distribution and  indicate the observed value  for. The log-likelihood function is

 (14)

For EW  distribution, the maximum likelihood estimators of the, and  parameters maximize the function in Eq. (14).

 denote a random sample from TW  distribution and  indicate the observed value  for. The log-likelihood function is

 (15)

For the TW  distribution, the maximum likelihood estimators of the , and  parameters maximize the function in Eq. (15).

 denote a random sample from GL  distribution and indicate the observed value  for. The log-likelihood function is

 (16)

where 

For GL  distribution, the maximum likelihood estimators of the  and  parameters maximize the function in Eq. (16). The optim function in the R program is used to maximize the log-likelihood functions in Eqs. (13, 14, 15, 16).

1. **Model Evaluation**

We consider some selection criteria including Akaike information criterion (AIC), Bayesian information criterion (BIC), Anderson-Darling statistics (A\*), Cramér-von Mises statistics (W\*), Kolmogorov-Smirnov statistics (K-S), and p-values (A\*, W\*, KS) for Covid-19 patients data analysis.

* 1. **Data Description**

We consider two Covid-19 patients data sets (the survival time of 75-84 aged Covid-19 patients (100 observations), the survival time of 85+ Covid-19 patients (100 observations)). The survival time (in days) refers to the time between deaths and hospitalizations of Covid-19 patients. The survival time of 75-84 aged Covid-19 patients data consist of 100 individuals of 1047 patients with certain characteristics (Age: 74-85, Source of Covid-19: Contact with confirmed) were selected by simple random sampling method. The survival time of 85+ Covid-19 patients data consist of 100 individuals of 1323 patients with certain characteristics (Age: 85+, Source of Covid-19: Contact with confirmed) were selected by simple random sampling method. The survival time of 75-84 aged Covid-19 patients and the survival time of 85+ Covid-19 patients data sets were collected between June 15, 2020 and January 20, 2021 and released to the public by the Israel Ministry of Health on 20 January 2021 (anonymous and open-access data, [12]). The descriptive statistics of the data sets are given Table 1.

**Table 1.** Descriptive statistics of Covid-19 data sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data set | Minimum | Maximum | Mean | Median |
| Data set 1 | 1 | 58 | 15.20 | 13 |
| Data set 2 | 1 | 41 | 14.48 | 13 |

1. **Results and Discussion**

In this subsection, we present the results of data analysis for described data sets in Section 4. Table 6 shows the maximum likelihood estimates (MLEs) of fitted models for four data sets. In Table 7, the selection criteria statistics to assess the fits of models to data sets are given. From Table 7, the EW distribution is the best-fitted model according to the KS statistic and its p-value. From Table 7, it is seen that the GL distribution is the best-fitted model according to all criteria except A \*, W \* statistics, and their p values. The best model was determined as TW according to A \*, W \*, and their p values for the first data set. In the second data analysis, the GL distribution is the best-fitted model according to AIC, BIC, A \*, and its p-value while the EW distribution is the best-fitted model according to KS and its p-value. The W \* statistics of the TW and EW distributions are very close to each other. According to this criterion, TW and EW distributions are the two best-fit models due to the higher number of the selection.

**Table 2.** MLEs and standard errors (SE) of parameters of all models for data sets

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Data set | Model |   |   |   |   |   |   |
| Data set 1 | Weibull | 1.461815 | 16.84369 | - | 0.109226 | 1.216951 | - |
|  | TW | 1.074265 | 10.79041 | -0.94037 | 0.14124 | 1.649272 | 0.211835 |
|  | EW | 0.088046 | 1.038861 | 2.003125 | 0.083591 | 0.240331 | 0.935173 |
|  | GL | 0.136562 | 1.206187 | - | 0.013607 | 0.183067 | - |
| Data set 2 | Weibull | 1.730377 | 16.28494 | - | 0.131714 | 0.993283 | - |
|  | TW | 1.861176 | 19.98046 | 0.596923 | 0.142753 | 2.965782 | 0.370521 |
|  | EW | 0.036983 | 1.322898 | 1.677508 | 0.040665 | 0.295992 | 0.726261 |
|  | GL | 0.162691 | 1.599501 | - | 0.015149 | 0.253754 | - |

**Table 3.** Selection criteria of fitted models for data sets

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data | Model | AIC | BIC | KS | A\* | W\* | p-value (KS) | p-value (A\*) | p-value (W\*) |
| Data set 3 | Weibull | 727.3092 | 732.5196 | 0.0758 | 0.6560 | 0.0923 | 0.6131 | 0.5963 | 0.6247 |
|  | TW | 726.7592 | 734.5747 | 0.0593 | **0.3627** | **0.0431** | 0.8737 | **0.8845** | **0.9175** |
|  | EW | 726.8427 | 734.6582 | 0.0596 | 0.3820 | 0.0426 | 0.8701 | 0.8661 | 0.9201 |
|  | GL | **724.8983** | **730.1087** | **0.0562** | 0.3711 | 0.0450 | **0.9102** | 0.8766 | 0.9071 |
| Data set 4 | Weibull | 699.3657 | 704.5761 | 0.0800 | 0.4140 | 0.0678 | 0.5437 | 0.8343 | 0.7663 |
|  | TW | 700.1762 | 707.9917 | 0.0689 | 0.2932 | **0.0469** | 0.7288 | 0.9430 | **0.8957** |
|  | EW | 700.2063 | 708.0218 | **0.0640** | 0.2896 | **0.0469** | **0.8080** | 0.9456 | **0.8957** |
|  | GL | **698.4261** | **703.6365** | 0.0666 | **0.2878** | 0.0472 | 0.7673 | **0.9469** | 0.8938 |

The invariance property of the maximum likelihood estimator is used to estimate the mean survival time of Covid-19 patients and the probabilities of dying for data sets. We compute the estimated probabilities of survival times in Covid-19 patients for data sets using the following procedure.

Let define  be a random variable from one of the examined distributions (Weibull, TW, EW, GL) with estimated parameters given in Table 6, and  denotes the survival times of Covid-19 patients. In this regard, we can calculate the probability via the following equation.

 (17)

where  denotes the PDF of the relevant distribution (Weibull, TW, EW, GL). Thus, we provide the estimated probabilities given in Tables 4-5 for data sets. Further, we estimate the average survival times of dead Covid-19 patients for the first and second data sets (Table 6).

**Table 4.** Estimated probabilites of survival times in Covid-19 patients for data set 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Survival time* | *Weibull* | *TW* | *EW* | *GL* |
| <7 days | 0.241981 | 0.232418 | 0.235266 | 0.226725 |
| <8 days | 0.285923 | 0.280867 | 0.284638 | 0.274363 |
| <9 days | 0.329706 | 0.329238 | 0.333658 | 0.322227 |
| <10 days | 0.3729 | 0.376798 | 0.381606 | 0.369567 |
| <11 days | 0.415158 | 0.422988 | 0.427953 | 0.415795 |
| <12 days | 0.456198 | 0.4674 | 0.472325 | 0.460462 |
| <13 days | 0.495802 | 0.509744 | 0.514471 | 0.503238 |
| <14 days | 0.533799 | 0.549833 | 0.55424 | 0.543894 |
| <15 days | 0.570064 | 0.587558 | 0.591556 | 0.582283 |
| <16 days | 0.604511 | 0.622875 | 0.626405 | 0.618327 |
| <17 days | 0.637088 | 0.655789 | 0.658818 | 0.652 |
| <18 days | 0.667769 | 0.686344 | 0.688856 | 0.683321 |
| <19 days | 0.696554 | 0.71461 | 0.71661 | 0.71234 |
| <20 days | 0.723462 | 0.74068 | 0.742182 | 0.739133 |
| <21 days | 0.748531 | 0.764659 | 0.765688 | 0.763793 |
| <4 weeks | 0.8778 | 0.883842 | 0.882602 | 0.886159 |
| <5 weeks | 0.945679 | 0.944485 | 0.942626 | 0.947507 |
| <6 weeks | 0.977684 | 0.973993 | 0.972355 | 0.976499 |
| <7 weeks | 0.991465 | 0.987981 | 0.986797 | 0.989696 |
| <8 weeks | 0.996943 | 0.994502 | 0.993732 | 0.995553 |

**Table 5.** Estimated probabilites of survival times in Covid-19 patients for data set 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Survival time* | *Weibull* | *TW* | *EW* | *GL* |
| <7 days | 0.207055 | 0.20091 | 0.201186 | 0.200513 |
| <8 days | 0.253461 | 0.249238 | 0.251857 | 0.252803 |
| <9 days | 0.301197 | 0.299307 | 0.303951 | 0.306571 |
| <10 days | 0.349531 | 0.350188 | 0.356379 | 0.360542 |
| <11 days | 0.397804 | 0.40104 | 0.408227 | 0.413675 |
| <12 days | 0.44544 | 0.451126 | 0.458754 | 0.465156 |
| <13 days | 0.491945 | 0.499816 | 0.507381 | 0.514376 |
| <14 days | 0.536902 | 0.546596 | 0.553674 | 0.560908 |
| <15 days | 0.57997 | 0.591061 | 0.597329 | 0.604479 |
| <16 days | 0.620886 | 0.632915 | 0.638151 | 0.644943 |
| <17 days | 0.65945 | 0.671958 | 0.676041 | 0.682253 |
| <18 days | 0.695529 | 0.708081 | 0.710972 | 0.716441 |
| <19 days | 0.729046 | 0.74125 | 0.742981 | 0.747596 |
| <20 days | 0.759973 | 0.771495 | 0.772149 | 0.775851 |
| <21 days | 0.788327 | 0.798899 | 0.798596 | 0.801365 |
| <4 weeks | 0.922258 | 0.924054 | 0.92083 | 0.918609 |
| <5 weeks | 0.976673 | 0.974378 | 0.971797 | 0.968443 |
| <6 weeks | 0.994213 | 0.992304 | 0.990696 | 0.988173 |
| <7 weeks | 0.998802 | 0.997993 | 0.997122 | 0.995668 |
| <8 weeks | 0.999792 | 0.999554 | 0.999159 | 0.998439 |

**Table 6.** Theestimated average survival times of Covid-19 patients for data sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Data Set* | *Weibull* | *TW* | *EW* | *GL* |
| *Data set 1: The estimated average of the survival times (in days)* | 15.2556 | 15.1888 | 15.1972 | 15.1887 |
| *Data set 2: The estimated average of the survival times (in days)* | 14.5134 | 14.4496 | 14.4774 | 14.4896 |

According to Tables 4-5, it was observed that the patients died with a probability of almost 90% within 4 weeks. We estimated that patients over 85 years of age died within 14.5 days and patients in the 74-85 age group died within approximately 15.2 days. (Table 6). The purpose of this paper is to estimate the survival times (in days) of Covid-19 patients by modeling some lifetime distributions. In a previous research, Cheng et al. [5] emphasized that the median of days from illness onset to admission was 10 days, and the median of stay in hospital was 5 days for dead Covid-19 patients. This means that a patient dies approximately 15 days after being infected [5]. Our results support previous studies, the survival time of an infected patient has been estimated at about 15 days. It can be said that this result is similar to Cheng’s [5] findings. Another subject of discussion is that there are other factors affecting survival times in Covid-19 patients such as the presence of chronic disease and treatment method. We have provided a new study on the survival times of Covid-19 patients only in terms of age and gender.

1. **Conclusion**

In conclusion, we also have shown that some statistical distributions (Weibull, TW, EW, GL) have the potential to be used in modelling the Covid-19 data. We recommend using GL distribution and EW distributions to model and estimate the survival time of Covid-19 patients in the data sets due to the high number of select according to selection criteria in Section 4. The study has the limitations. Firstly, the Covid-19 data sets are anonymous and open access URL-1, (2021). The datasets on Israeli Covid-19 patients are consisted by filtering variables (age and gender). Secondly, the sample size is limited and an accurate assessment of the survival time of Covid-19 patients is not accurate. The survival time may vary depending on age, ethnicity, country, and other global factors. The study is designed to show the survival time can be estimated by lifetime distributions. Extensive research should be conducted on the survival time of Covid-19 patients.

**References**

1. World Health Organization, (2020) WHO Director-General's Opening Remarks at the Media Briefing on Covid 19 - 11 March 2020. World Health Organization, Geneva.
2. World Health Organization, (2023) Available at: https://Covid19.who.int/, Retrieved March 27, 2023.
3. Tian, W., Jiang, W., Yao, J., Nicholson, C. J., Li, R. H., Sigurslid, H. H., ... , *&* Malhotra, R. (2020) Predictors of mortality in hospitalized COVID‐19 patients: A systematic review and meta‐analysis. *Journal of medical virology*, 92(10), 1875-1883.
4. Covino, M., De Matteis, G., Santoro, M., Sabia, L., Simeoni, B., Candelli, M., ... , *&* Franceschi, F. (2020) Clinical characteristics and prognostic factors in COVID‐19 patients aged≥ 80 years. *Geriatrics & gerontology international*, 20(7), 704-708.
5. Cheng, Y., Luo, R., Wang, K., Zhang, M., Wang, Z., Dong, L., ... , *&* Xu, G. (2020) Kidney impairment is associated with in-hospital death of COVID-19 patients. *MedRxiv*. <https://doi.org/10.1101/2020.02.18.20023242>.
6. Pal, M., Ali, M. M., *&* Woo, J.. (2006) Exponentiated weibull distribution. *Statistica*, 66(2) (2006) 139-147.
7. Gradshteyn, I. S., *&* Ryzhik I. M. (2014) Table of integrals, series, and products. London: Academic press.
8. Aryal, G. R., *&* Tsokos, C. P. (2011) Transmuted Weibull distribution: A generalization of the Weibull probability distribution. *European Journal of pure and applied mathematics*, 4(2), 89-102.
9. Lindley, D.V. (1958) Fiducial distributions and Bayes theorem. *Journal of the Royal Statistical Society B*, 20, 102–107.
10. Ghitany M. E., Alqallaf F., Al-Mutairi D. K., *&* Husain H. A. (2011) A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulation*, (81) 6, 1190-1201.
11. Nadarajah, S., Bakouch, H. S., *&* Tahmasbi, R. (2011) A generalized Lindley distribution*. Sankhya B*, 73(2), 331-359.
12. Israel Ministry of Health. (2021) Available at: https://data.gov.il/dataset/COVİD-19, Retrieved January 25, 2021.
1. \* Corresponding author. *e-mail address: canertanis@karatekin.edu.tr* [↑](#footnote-ref-1)