American Journal of Theoretical and Applied Statistics
2023; 12(5): 103-109
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20231205.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

SciencePG
Science Publishing Group

Survival Analysis of COVID-19 Patients: A Case Study of AIC Kijabe Hospital

Roseline Achieng Oburu*, Joseph Eyang'an Esekon, Martin Mutwiri Kithinji

Department of Pure and Applied Sciences, School of Pure and Applied Sciences, Kirinyaga University, Kerugoya, Kenya

Email address:
oburu："roseline@gmail.com (Roseline Achieng Oburu)
*Corresponding author

To cite this article:
Roseline Achieng Oburu, Joseph Eyang'an Esekon, Martin Mutwiri Kithinji. Survival Analysis of COVID-19 Patients: A Case Study of AIC Kijabe Hospital. *American Journal of Theoretical and Applied Statistics*. Vol. 12, No. 5, 2023, pp. 103-109.
doi: 10.11648/j.ajtas.20231205.11

Received: August 4, 2023; **Accepted:** August 28, 2023; **Published:** September 12, 2023

Abstract: COVID-19 is an infectious disease caused by the novel coronavirus: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. The disease quickly spread, resulting in an epidemic in China and a number of cases in other countries around the world. This led to inconsistent health conditions and caused unceasing loss of human lives. Globally, the current COVID-19 pandemic posed a significant and imminent threat to healthcare systems, including Kenya, in terms of patient triage and allocation of limited resources. In this study we analyze time to recovery of the COVID-19 patients at AIC Kijabe hospital. A retrospective cohort study was used to review the existing medical records of 66 patients who tested positive for COVID-19 at AIC Kijabe Hospital. Kaplan-Meier curves were used in determining the probability of recovery. For statistical comparison of the survival curves, Log-rank test statistic was used while Cox proportional hazards model was used to investigate the relation between time to recovery and the predictor variables. Results showed that female patients recovered faster than male patients while there was no significant difference between the survival curves for gender and marital status among the COVID-19 patients. In the Cox proportional hazards model, only age was significant with a p - value (0.0463) and therefore affected the time to recovery of the COVID-19 patients while the rest of the variables, gender and marital status were not significant. In conclusion, age was the only variable that had an effect on the time to recovery of COVID-19 patients.

Keywords: COVID-19, Kaplan-Meier, Log-rank Test, Cox Proportional Hazards Model, Survival Analysis

1. Introduction

At the end of 2019, a cluster of Pneumonia cases (COVID-19) in Wuhan, Hubei Province, China was traced to a new coronavirus [15]. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease quickly spread, resulting in an epidemic in China and a number of cases in other countries around the world [5]. This led to inconsistent health conditions and caused unceasing loss of human lives. The pandemic had claimed 6.95 million lives and resulted in over 768 million cases as of 2nd August 2023, making it one of the life-threatening diseases in human history [14]. Globally, the COVID-19 epidemic had a significant influence on public mental health of general population. The pandemic was rapidly spreading throughout Sub-Saharan Africa, including Kenya. World Health Organization declared COVID-19 as a pandemic in 11th March 2020 [12]. The first confirmed case in Kenya occurred in 12th March 2020, with initial reports emanating from Nairobi [8]. As of 2nd August 2023, Kenya had $343,918$ confirmed cases and $5,689$ deaths [14]. As a result of the effects of COVID-19 pandemic, Kenya joined the fight against the pandemic by vaccinating its adult population in March 2021 and adolescent population in November 2021 [10]. The world prioritized health care provision to all through the SDG's 2030 which was put under a great test by the COVID-19 pandemic. Studying the differences in survival of COVID-19 patients in relation to their age, gender and marital status helps in fighting of the disease and preparedness for any other unforeseen outbreak of another health pandemic in order to protect the health of the people as discussed in SDG number 3. Previous studies have identified factors contributing to low survival rate among COVID-19 patients. Factors such as age was found to be significant explanatory factor which should be considered during the study since most risk factors and risk of disease change with age [1]. Models in survival analysis are used to study time to an event of interest. Further, models in survival are also used in investigating the association between time to occurrence of a certain event but not predominantly time to recovery. This study aims at determining the time to recovery of COVID-19 patients using the Kaplan Meier curves and to determine the factors affecting time to recovery of COVID-19 patients using the Log-rank test. Cox Proportional Hazards model is used to determine the association between the factors and time to recovery.

2. Literature Review

A research on the effects of COVID-19 on the Mexican population was carried out in [9]. Kaplan-Meier curves showed that men and women had different mortality rate. Men were more likely to die from COVID-19. Cox Proportional hazards model showed that old age, chronic kidney disease and hospitalization were independent risk variables that increased the likelihood that both men and women would die from COVID-19.

A survival analysis study of symptomatic COVID-19 in Phuentsholing municipality of Bhutan was done with the objective of identifying risk factors for patients who tested positive for COVID-19 to experiencing symptoms [4]. Survival curves showed a statistically significant variation between the categories of absolute lymphocyte count, risk factors present and case origin. The curves also revealed that there was no discernible gender difference in the likelihood of developing COVID-19 (with symptoms). Using the hazard model, the study found out that patients who received COVID-19 vaccine were 77% less likely to experience COVID-19 symptoms than those who did not receive the vaccine, with an adjusted hazard ratio of 0.23. However, there was no significant variation in the chance of developing symptoms of the COVID-19 infection between the various age groups, occupations and sexes.

In [7] survival analysis was used to determine survival probability and predictors of mortality among patients hospitalized for COVID-19 in Ethiopia. Log-rank test statistic showed that difference in cumulative probability of survival of the various factors were statistically significant. Multivariate Cox regression model revealed that patients with increased age had a higher risk of dying. The median time from the onset of symptoms till death was found to be 9 days, which was similar to the 11 days reported in China [2] but different from the 19 days reported in Mexico as shown in [13].

3. Methodology

3.1. Research Design

The study involved a retrospective cohort and data collected from the existing medical records of confirmed COVID-19 patients who attended AIC Kijabe Hospital during the study period through extraction of electronic administrative data. The study started from 17th May 2020 to 9th March 2023. The inclusion criteria were positively diagnosed COVID-19 patients with an admission date, age, gender, marital status, date of discharge and COVID-19 status. The starting point of the selected cohort was at the date of admission, and the end point was the date of discharge, or death or the closing date of the study which was 9th March 2023. Those who did not experience the event which was recovery or study ended before they experienced the event were considered as censored patients. We identified all patients that fulfilled the criteria and among the 316 COVID-19 patients, only 66 of them were eligible for the study.

3.2. Survival Data Analysis

Survival analysis allows for determination of time taken to recover from COVID-19, to compare time to recovery of the patients using the variables, and to determine the association between time to recovery of COVID-19 patients and the predictor variables. Survival analysis models were used to estimate time to recovery of COVID-19 patients. The event of interest in this study was recovery from COVID-19. Kaplan Meier curves were used to show the general pattern of time to recovery of COVID-19, Log-rank test to compare the time to recovery of COVID-19 patients using different variables and Cox Proportional Hazards model to investigate the relation between time to recovery and the predictor variables.

3.3. Kaplan Meier Model

This is a survival analysis method used to analyze time-to-event data. It is one of the often employed techniques in measuring the fraction of subjects who live for a certain amount of time after an event of interest has occurred [6]. Kaplan Meier estimator was used to fit the survival function $S(t)$ used to determine the time to recovery for different groups of COVID-19 patients. The median time to recovery of COVID-19 patients was obtained using the Kaplan Meier curves. The survival function $S(t)$ is calculated as follows

$$S(t) = \prod_{t_j \leq (t)} \frac{n_j - d_j}{n_j} \tag{1}$$

where: n_j is the number of COVID-19 patients at risk and d_j is the number of COVID-19 patients experiencing the risk.

3.4. Assumptions of Kaplan Meier Estimate

1. After the last duration at which an event is observed, the estimate of the survival function remains constant.
2. Only those at risk at the observed lifetime, t_j, contribute to the estimate.

3.5. Log-rank Test Statistic

The Log-rank test statistic is used in comparing the Kaplan-Meier survival curves for different samples. The null hypothesis tested is that the survival curves for the groups do not differ. The study compared different groups that are in age (youthful, middle-aged and aged), gender (female and male) and marital status (single, married and widowed). Log-rank test makes use of observed values versus the expected cell counts for several result categories. The formula is given as in (2).

$$\text{LR} = \frac{(O_1 - E_1)^2}{\text{var}(O_1 - E_1)} \tag{2}$$

where O_1 are the observed values and E_1 are the expected values.

3.6. Cox Proportional Hazards Model

Proportional hazards model estimates the durational impact on the hazard function. The Cox model is used to determine the influence of different variables on the time to recovery since the subject in the group may have additional characteristics that may affect their outcome [3]. The Cox Proportional Hazards model is designed in such a way that it can assess the effects of each predictor on the shape of the survival curve and show the estimated regression coefficient together the p-value for each coefficient. The model is of the form:

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k) \tag{3}$$

where: $h_0(t)$ is the baseline hazard function, $X_1, X_2, ?, X_k$ are the covariates and $\beta_1, \beta_2, ..., \beta_k$ are the model parameters being estimated describing the effect of the predictors on the overall hazard. The factors associated with COVID-19 were age, gender and marital status. This study determined these factors likely to influence the time taken to recover from COVID-19. In particular, the study aimed at the degree of influence of these variables on time to recover from COVID-19.

3.6.1. Assumptions for the Cox Proportional Hazards Model

1. Hazard rate remains constant across the study period and same for each time interval.
2. The baseline $h_0(t)$ is unspecified.

3.6.2. Testing Proportional Hazards Assumption

To ascertain the proportional hazards assumption, a Schoenfeld residuals [11] test is done for each predictor variable in the study. A variable that has a significant p - value suggests a violation of the proportional hazards assumption.
The Cox Regression was done in two sections:

3.6.3. Univariate Cox Regression

This was used to describe the time to recovery of COVID-19 patients according to one variable under investigation, but ignore the impact of any other variable. The models were as follows

$$h(t, X_1) = h_0(t)\exp(\beta_1 X_1) \tag{4}$$

$$h(t, X_2) = h_0(t)\exp(\beta_2 X_2) \tag{5}$$

$$h(t, X_3) = h_0(t)\exp(\beta_3 X_3) \tag{6}$$

where: X_1, X_2 and X_3 are gender, age and marital status variables.

3.6.4. Multivariate Cox Regression

This model was fitted to describe how the factors jointly impacted the time to recovery of COVID-19 patients. The model is of the form as in (7).

$$h(t, X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) \tag{7}$$

3.6.5. Estimation of Parameters of the Cox Model

The regression coefficients are estimated by maximizing the Cox partial likelihood. The partial likelihood function is given as:

$$\text{PL}(\beta) = \prod_{i=1}^{k} \frac{\exp(\beta X_{(t_i)})}{\sum_{j \epsilon Z_i} \exp(\beta X_j)} \tag{8}$$

where, k is the number of distinct outcome events, $X_{(t_i)}$ is the covariate vector at time t_i for the individual who has the event at t_i and Z_i is a group of people whose observed censoring or failure time is higher than or equal to t_i.

3.6.6. Evaluation of Cox Proportional Hazards Model

Hazard ratio is one of the components used in evaluation of the Cox Proportional Hazards model. It is obtained by exponentiating the coefficients of the predictor variables. Hazard ratio is expressed as;

$$\text{HR} = \exp[\beta_i(X_i^* - X_i)] \tag{9}$$

Where X_i^* is the set of predictors for one individual and X_i is the set of predictors for the other individual.
A p-value is also obtained from the Cox model where any variable whose p-value is significant is considered to form the best Cox proportional hazards model.

4. Results and Discussion

4.1. Descriptive Statistics

A total of 66 confirmed COVID-19 patients were admitted to AIC Kijabe Hospital from 17th May 2020 to 9th March 2023. The median age of the COVID-19 patients was 64 years with the minimum and maximum ages being 25 and 90 years respectively. Among the 66 patients, 50 experienced the event which was recovery and had a median time to recovery of 9 days. More than half (62.12%) were females with a median time to recovery of 8 days. Majority of the COVID-19 patients (83.33%) were aged with (10.61%) patients being single. Table 1 shows descriptive statistics of the COVID-19 patients which is gender, age, marital status and event.

Table 1. *Characteristics of Sample.*

Characteristics	Number of Patients	Percent
Gender		
Female	41	62.12
Male	25	37.88
Age group (Years)		
0-34	6	9.09
35-45	5	7.58
46+	55	83.33
Marital Status		
Single	7	10.61
Married	55	83.33
Widowed	4	6.06
Event		
Recovered	50	75.76
Censored	16	24.24

A comparison of the patients according to gender was done by plotting Kaplan Meier curve (Figure 1) where ‘gender=0’ represented female patients while ‘gender=1’ represented male patients. The median time to recovery of female patients was 8 days as opposed to 11 days of male patients. This showed that on average, female patients recovered faster than male patients.

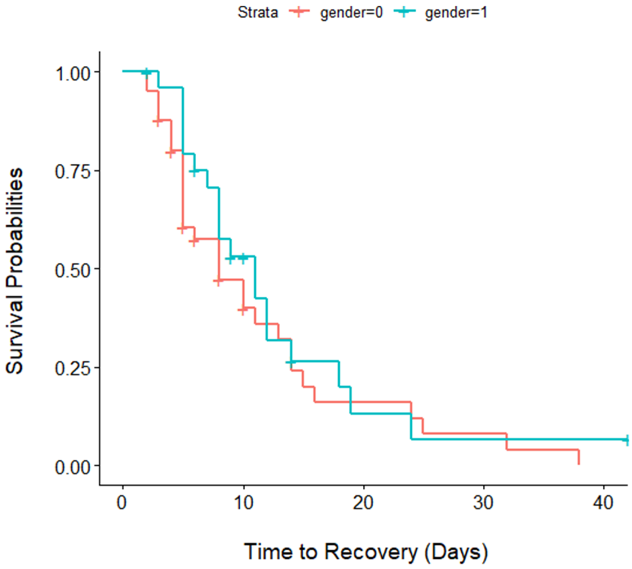

Figure 1. *KM curve for gender.*

Since age was a continuous variable, there was need to categorize for easier comparison. Therefore, age was categorized into three groups namely: $0 - 34$ years classified as youthful, $35 - 45$ years classified as middle-aged and $46+$ years classified as aged. There were 6 youthful patients with 4 of them experiencing the event. A Kaplan Meier curve was used to show the pattern for the three categories where ‘age=0’ are youthful patients,‘age=1’ are middle-aged patients and ‘age=2’ are the aged patients as shown in Figure 2.

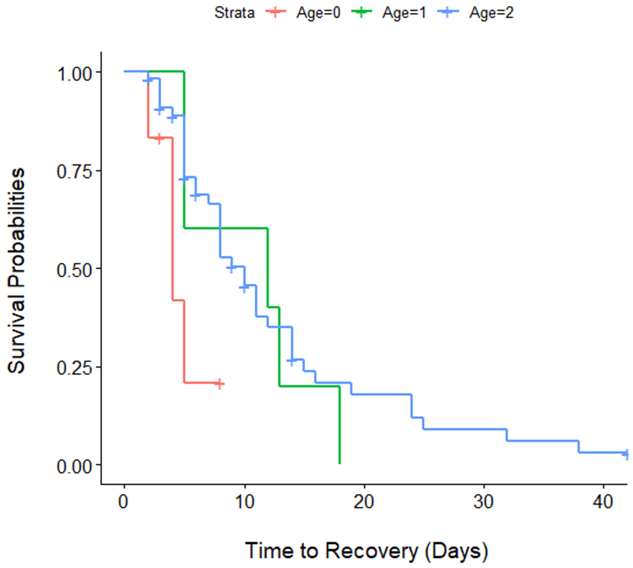

Figure 2. *KM curve for age.*

Among the 66 COVID-19 patients, using their marital status, 7 were single, 55 married and the rest widowed. From the Kaplan Meier curve, the median time to recovery was 10 days for both single and married patients while the widowed had a median time to recovery of 5 days. This implied that both the single and married patients recovered at a slower rate than the widowed patients as shown in Figure 3 where ‘maritalstatus=0’ represent single patients, ‘maritalstatus=1’ represent married patients and ‘maritalstatus=2’ represent widowed patients.

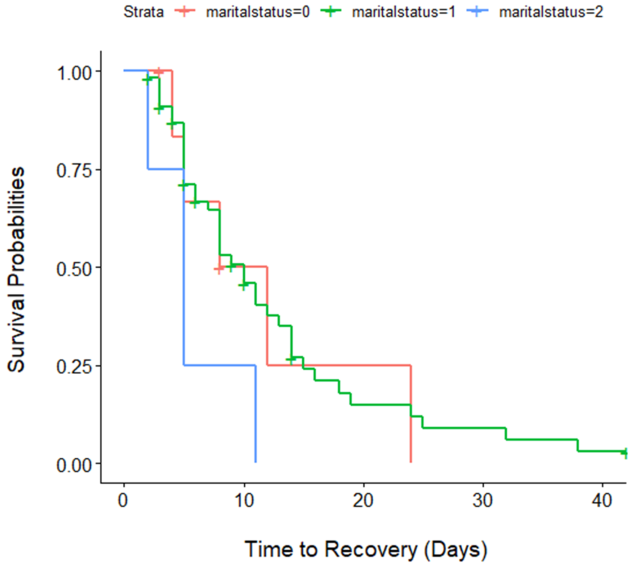

Figure 3. *KM curve for marital status.*

4.2. Testing Proportional Hazards Assumption

To evaluate validity of the proportional hazards assumption, a Schoenfeld residuals test was used as shown in Table 2. The results showed that none of the variables were significant (p - values > 0.05) therefore, the proportional hazards assumption was not violated which led to the use of Log-rank test.

Table 2. *Testing the PH Assumption.*

	chisquare	degrees of freedom	p-value
Gender	0.500	1	0.48
Age	0.211	1	0.65
Marital Status	1.755	1	0.19
global	2.644	3	0.45

4.3. Log-rank Test

A confirmatory test ascertaining the significance of the variables gender, age and marital status to the time to recovery of COVID-19 patients was done using the Log-rank test. Log-rank test was used to test the hypothesis that:

H_0 : There is no significant difference between the survival curves

H_1 : There is significant difference between the survival curves

From Figure 1, female patients recovered faster than male patients. However, after the Log-rank test was done the results showed that there was no significant difference between the time to recovery of male and female patients (p - value = 0.4) as shown in Table 3. For the variable age, the p-value (0.05) is equal to the level of significance (0.05), the null hypothesis is rejected and conclusion made that there is a significant difference in the time to recovery among the youthful, middle-aged and aged patients as shown in Table 4. Finally for marital status variable, (Figure 3) showed that both single and married patients recovered at a slower rate than the widowed patients. However, after Log-rank test was done, it was confirmed that there was no significant difference between the single, married and widowed patients (p - value = 0.1) as shown in Table 5.

Table 3. *Log-Rank Output For Gender.*

	N	Observed	Expected	$(O-E)^2$/E	$(O-E)^2$/V
Female	41	31	28.1	0.290	0.743
Male	25	19	21.9	0.373	0.743
chisq 0.7	on 1	degrees of freedom,	p = 0.4		

Table 4. *Log-Rank Output For Age.*

	N	Observed	Expected	$(O-E)^2$/E	$(O-E)^2$/V
Youthful	6	4	1.40	4.8250	5.6937
Middle-aged	5	5	4.44	0.0711	0.0885
Aged	55	41	44.16	0.2263	2.2031
chisq 5.9	on 2	degrees of freedom,	p = 0.05		

Table 5. *Log-Rank Output For Marital Status.*

	N	Observed	Expected	$(O-E)^2$/E	$(O-E)^2$/V
Single	7	5	4.96	0.00027	0.00034
Married	55	41	43.39	0.1321	1.1365
Widowed	4	4	1.64	3.3828	4.0265
chisq 4.0	on 2	degrees of freedom,	p = 0.1		

4.4. Describing Effects of Variables on Survival Time

Effects of the variable on the time to recovery was done using the Cox Proportional Hazards model. We began by computing univariate Cox model for each variable.

4.4.1. Univariate Cox Regression for Gender

The gender variable is not statistically significant (p - value = 0.401), therefore, gender of COVID-19 patients did not have an effect on the time to recovery of the patients as shown in Table 6.

Table 6. *Univariate Cox Output for Gender.*

	Coef	exp(Coef)	Se(Coef)	Z	P
Gender	-0.2459	0.7820	0.2930	-0.839	0.401
n = 66,	events = 50				

The model is as follows

$$h(t, X_1) = h_0(t)\exp(-0.2459X_1). \tag{10}$$

4.4.2. Univariate Cox Regression for Age

It is evident that age is statistically significant (p - value = 0.0463). This implies that age of COVID-19 patients affected the time to recovery of the patients as shown in Table 7.

Table 7. *Univariate Cox Output for Age.*

	Coef	exp(Coef)	Se(Coef)	Z	P
Age	-0.5137	0.5983	0.2579	-1.992	0.0463
n = 66,	events = 50				

The model is as follows

$$h(t, X_2) = h_0(t)\exp(-0.5137X_2). \tag{11}$$

4.4.3. Univariate Cox Regression for Marital Status

Results from marital status variable show that the variable is not statistically significant (p - value = 0.32) to the time to recovery of COVID-19 patients as shown in Table 8.

Table 8. *Univariate Cox Output for Marital Status.*

	Coef	exp(Coef)	Se(Coef)	Z	P
Marital Status	0.4257	1.5306	0.4282	0.994	0.32
n = 66,	events = 50				

The model is as follows

$$h(t, X_3) = h_0(t)\exp(0.4257X_3). \tag{12}$$

4.4.4. Multivariate Cox Regression Analysis

A multivariate Cox model is fitted to describe how the factors jointly impacted the time to recovery of COVID-19 patients. Table 9 shows the output of the results. From the results, among the three variables, only age is significant (p - value = 0.0105). This implies that only age affected the time to recovery of the COVID-19 patients. The hazard ratios for variables gender (0.8302) and age (0.4837) show a negative association with the event probability hence positively associated with the time to recovery while hazard ratio for marital status (2.1826) increased the hazard therefore decreasing the time to recovery of COVID-19 patients.

Table 9. *Multivariate Cox Regression.*

	Coef	exp(Coef)	Se(Coef)	Z	P
Gender	-0.1861	0.8302	0.2953	-0.630	0.5286
Age	-0.7264	0.4837	0.2837	-2.560	0.0105
Marital Status	0.7805	2.1826	0.4397	1.775	0.0759
n = 66,	events = 50				

The model is as follows

$$h(t, X) = h_0(t)\exp(-0.1861X_1 - 0.7264X_2 + 0.7805X_3). \tag{13}$$

5. Conclusion

The main purpose of the study was to model time to recovery of COVID-19 patients using survival analysis models. Variables included in the study were gender, age and marital status of COVID-19 patients. Findings showed that among the 66 COVID-19 patients, 50 experienced the event which was recovering from the disease. Female COVID-19 patients recovered faster than the male patients. Both single and married patients recovered at a slower rate than the widowed patients. Youthful patients recovered faster than both the middle-aged and aged patients. Using the Log-rank test statistic, there was no significant difference between the survival curves for male and female COVID-19 patients (p - value = 0.4). The proportional hazards assumption was not violated for all the three variables since none of them was significant (p - value > 0.05). In the Cox proportional hazards model, age was the only variable that was significant with a p-value of 0.0463 and therefore affected the time to recovery of COVID-19 patients while gender and marital status were not significant at 5% level of significance hence did not affect the time to recovery of the COVID-19 patients. In conclusion, the study recommends that future work could involve adding more variables such as race and place of residence of the COVID-19 patients.

Acknowledgements

The authors would like to express their gratitude to the administration and staff at AIC Kijabe Hospital.

References

[1] Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: survival analysis. Critical care, 8, 1-6.

[2] Chen, R., Liang, W., Jiang, M., Guan, W., Zhan, C., Wang, T., ... & for COVID, M. T. E. G. (2020). Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China. Chest, 158 (1), 97-105.

[3] Cox, D. R. (1972). Regression models and life tables. Journal of the Royal Statistical Society: Series B (Methodological), 34 (2), 187-202.

[4] Gyeltshen, K., Tsheten, T., Dorji, S., Pelzang, T., & Wangdi, K. (2021). Survival analysis of symptomatic COVID-19 in phuentsholing municipality, Bhutan. International Journal of Environmental Research and Public Health, 18 (20), 10929.

[5] Kang, D., Choi, H., Kim, J. H., & Choi, J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. International journal of infectious diseases, 94, 96-102.

[6] Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53 (282), 457-481.

[7] Kaso, A. W., Agero, G., Hurissa, Z., Kaso, T., Ewune, H. A., Hareru, H. E., & Hailu, A. (2022). Survival analysis of COVID-19 patients in Ethiopia: a hospital-based study. Plos one, 17 (5), e0268280.

[8] Mbae, N. (2020). COVID-19 in Kenya. Electronic Journal of General Medicine, 17 (6).

[9] Salinas-Escudero, G., Carrillo-Vega, M. F., Granados-García, V., Martínez-Valverde, S., Toledano-Toledano, F., & Garduño-Espinosa, J. (2020). A survival analysis of COVID-19 in the Mexican population. BMC public health, 20 (1), 1-8.

[10] Sam-Agudu, N. A., Quakyi, N. K., Masekela, R., Zumla, A., & Nachega, J. B. (2022). Children and adolescents in African countries should also be vaccinated for COVID-19. BMJ global health, 7 (2), e008315.

[11] Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. Biometrika, 69 (1), 239-241.

[12] Shah, S. G. S., & Farrow, A. (2020). A commentary on World Health Organization declares global emergency: A review of the 2019 novel Coronavirus (COVID-19)? International journal of surgery (London, England), 76, 128.

[13] Sousa, G. J. B., Garces, T. S., Cestari, V. R. F., Florêncio, R. S., Moreira, T. M. M., & Pereira, M. L. D. (2020). Mortality and survival of COVID-19. Epidemiology & Infection, 148.

[14] World Health Organization. (2023, August 2). “WHO official COVID-19 information.”

[15] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. New England journal of medicine, 382 (8), 727-733.