The Effect of Symptoms on the Survival Time of Coronavirus Patients in the Sudanese Population

Alshaikh A. Shokeralla[1,*], Mohammedelameen E. Qurashi[2], Reem Yousif Mekki[3] and Mortada S. Ali[1]

[1]Department of Mathematics, College of Science and Arts, Al-Baha University, Al-Makhwah, KSA

[2]Department of Statistics, College of Science, Sudan University of Science & Technology, Sudan

[3]Department of Statistics College of Computer Science and Mathematics, University of Bahri, Sudan

Abstract: The COVID-19 pandemic has rapidly spread worldwide, resulting in substantial rates of illness and death. Gaining insight into the various factors that impact the duration of survival among individuals diagnosed with COVID-19 is of utmost importance to inform clinical practices and public health strategies This study aims to evaluate the relationship between the acuteness of symptoms and the survival time of coronavirus patients in Sudan. The Kaplan-Meier curves and log-rank test were used to determine the symptom pattern. The results of COVID-19 and Cox regression were utilized to determine the most critical symptoms affecting coronavirus patients. The log-rank test revealed that there are differences in the pattern of age and symptoms among coronavirus patients. Cox regression revealed that symptoms affect on the survival time of coronavirus patients. The Cox proportional Hazard Model shows that the hazard of age at any time increases by 116.5%, diarrhea increases by 9%, headache increases by 62.0%, fatigability increases by 13.3%, and other symptoms increase by 47.3%. This study differs from prior studies in several ways. No current study in Sudan has used survival analysis to discover the most relevant symptoms affecting survival time.

Keywords: Survival Analysis, Kaplan –Meier, log-rank, COVID-19, symptoms.

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19), is a disease caused by a new (or emerging) type of corona virus that was first discovered when the disease outbreak occurred in December 2019 in Wuhan, the capital of China's Hubei province. Corona viruses are a large family of viruses that can cause illnesses ranging from mild illnesses, such as the common cold, to more severe diseases, such as severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). The virus can be transmitted from person to person, through small spray drops that are scattered from the nose or mouth when coughing or sneezing. The symptoms of the Cofide-19 range from simple to severe, and may appear within two days to 14 days after exposure to the virus. These symptoms may include fever, cough, shortness of breath, chills, headache, sore throat, loss of taste or smell [1].

The researchers endeavored to study the symptoms of COVID-19 patients, and several of them conducted survival analyses. Because the disease began to spread in China, Digestive Symptoms in COVID-19 Patients with Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes [2]. Another study on the symptoms of COVID-19 Time from symptom beginning to severe COVID-19 and risk variables in Southern Ethiopian patients [3]. A hospital-based study of COVID-19 patients in Ethiopia [4]. Other researchers also study symptoms and risk factors for long COVID-19 in non-hospitalized adults [5]. Clinical characteristics and risk factors for mortality in COVID-19 inpatients in Birjand, Iran: a single-center retrospective study [6]. The study aims to assess the time to death and the factors affecting death among patients who were transferred to the Alia Hospital Corona Emergency Center in Sudan.

2. METHODOLOGY

Survival analysis is concerned with studying the time before the occurrence of a particular event. The dependent variable represents the time between the beginning of the event and its end. The end of the event at the time of the occurrence of the event under study may be death or birth, or the failure to determine the final state of the singular (failure time). The survival analysis has two characteristics, the first of which is that the survival time is calculated after determining the case under study, it was a specific disease condition, and one of the statistical methods that are concerned with the study of multi-response phenomena is multiple regression. Because of the special nature of survival data, as it contains incomplete data with many types, it is inappropriate to use multiple regression, the second characteristic is that most of the traditional methods, such as regression models, analysis of variance, and (t) test, become inappropriate if the data are positively skewed, meaning that the majority of the data are concentrated on the right side of the distribution. These two characteristics force us not to use multiple

*Address correspondence to this author at the Department of Mathematics, College of Science and Arts, Al-Baha University, Al-Makhwah, KSA;
E-mail: sshokeralla@gmail.com

regression models and other statistical models, which eventually led to the extraction of the Cox regression model by the scientist, to better understand the concept of survival analysis, we must identify some concepts of survival analysis, including the basic survival functions.

2.1. Basic Survival Functions

Assuming that (t) represents a positive continuous random variable (retention time), which has a probability density function $(f(t)$ and a cumulative function $(F(t))$. To clarify the concept of the basic survival function, the following concepts must be understood.

2.1.1. Survivor Function [7, 8]

The survivor function is formulated as the probability of surviving beyond time (or at least until time t). It can be found through:

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^\infty f(t)dt \quad t \geq 0 \qquad (1)$$

$$f(t) = \frac{\partial F(t)}{\partial t} = \frac{\partial}{\partial t}\big(1 - S(t)\big) = -\bar{S}(t) = \frac{-\partial S(t)}{\partial t} \, (2)$$

$$f(t) = \frac{\partial S(t)}{\partial t} = -f(t)(3)$$

Where $S(t)$ represented survival function

$$S(t) = \begin{cases} 1 & t = 0 \\ 0 & t = \infty \end{cases}$$

2.1.2. Hazard Function [8, 9]

Also know the hazardrateor the current instantaneous rate of occurrence of the event, referred to as$h(t)$. It can be found through the following:

$$h(t) = \lim_{\nabla t \to 0} \frac{P(t < T < t + \Delta t / T > t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{-\bar{S}(t)}{S(t)} \, (4)$$

The hazard function also defines the instantaneous failure rate or the force of mortality, and conditional failure rate, and the hazard function gives the probability of hazard failure for each unit during the period, which has an important role in analyzing survival data.

2.1.3. Cumulative Hazard Function [8, 10]

It is the sum of the risks that occur until the arrival of time (t), and it can be found through:

$$h(t) = \int_0^t f(x)\,dx = \int_0^t \frac{f(x)}{S(x)} d(x) = \int_0^t \frac{1}{S(x)} \left[\frac{\partial S(x)}{\partial x}\right] dx$$
$$= -\ln(S(t)) \quad (5)$$

$$S(t) = \exp(-h(t))\,(6)$$

$$F(t) = h(t).\exp(-h(t))\,(7)$$

2.1.4. The Expectation of Life [8, 11]

Assuming that μ represents the mean or expected value of t, where

$$\mu = \int_0^\infty t.f(t)dt \, (8)$$

$$\mu = E(t) = \int_0^\infty S(t)dt \quad S(0) = 1, S(\infty) = 1 \qquad (9)$$

Where $S(t)$represents the probability that the item is alive at time t, and $E(t)$ represents the life expectancy of the item.

2.2. The Censoring

The data of survival time often contains items that did not know the time of the event, and we cannot know its final state during the period. are some items are recorded from the occurrence of the event, while others are not able to record the time of the event. There are two types of censoring [8, 9]:

2.2.1. Right-Censored

In some cases of survival studies, the specified study period ends, but some items have exceeded the duration of the study, in other words, the survival time of the item with the duration of the study. The most important reasons for the occurrence of right-censored:

1. The researcher's decision to end the study time before the event occurred.
2. The event did not occur for some items.
3. The researcher's inability to reach the final state of the item for any reason.

2.2.2. Left-Censored

Some items entered the study, but the time of occurrence of the event was not known to the researcher, for example, the time of death for people with a specific disease, but the time of infection with this disease is not known specifically.

3. COX REGRESSION MODEL

The scientist David Cox 1972 formulated the Cox model, which is one of the relative risks models. Survival analysis models are based on their basic concept of the importance of the time before the occurrence under study as a basic factor. In other words, the main feature of this method is to study the relationship between the times before the occurrence of the event. Understudy with one or more variables as independent, and the nature of the data is quantitative, descriptive, or mixed, one of the most prominent models used in this type of study is the Cox regression model. Survival analysis can be defined as a statistical method for analyzing data if the dependent variable is the time before the occurrence of the event under study, and the time may be in the form of hours, days, weeks, months, years from the beginning of the study until the occurrence of the event that is the case of death or the beginning of the disease or others [8, 9].

3.1. Properties Cox Regression Model

1. Semi-parametric model. Does not need to choose some particular probability model to represent the number of survival times.
2. The possibility of dealing with the two types of separate and continuous measurements of event times.
3. It is concerned with the effect of variables on the risk rate.
4. Items in which the values of the independent variables are equal have the same risk function.

3.1.1. Mathematical Formula of Cox Regression Model

The scientist Cox proposed the original model of Cox regression by assuming that (t) is a random variable, then the mathematical model formula is:

$$h(t/X_i) = h_0 exp \sum_{i=1}^{n} \beta_i X_i (10)$$

$h(t/X_i)$: The risk function for the occurrence of the event is conditioned by time t for the terms that represent the vector of the explanatory variables X_i .

h_0: Baseline hazard function which depends on the time $(t)$when the values of the explanatory variables (X_i).

β_i: Parameters of the model.

X_i: Explanatory variable.

Here, we would like to indicate that$\exp \sum_{i=1}^{n} \beta_i X_i$, the amount of relative hazard that does not depend on time (t), meaning that the change in the explanatory variables (X_i) from increase or decrease does not depend on a specific point in time. Also, the ratio between the hazard rates is constant and does not depend on time, and it is as follows:

$$\frac{h(t/X_1)}{h(t/X_2)} = \frac{h_0 \exp(\beta_1 X_1)}{h_0 \exp(\beta_2 X_2)} = \exp \beta (X_1 - X_2) \ (11)$$

This and the use of the logarithmic transformation can contribute to solving the difficulties facing the interpretation of the Cox regression model, by taking the logarithm of the risk function in the Cox regression model to become:

$$\log h(t) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n (12)$$

The hazard function becomes as in equation (12) a linear equation of the basic variables, knowing that the use of the logarithmic transformation works to compress the anomalous values of the data, which helps to solve many problems, including the presence of extreme observations, heterogeneity of variance and skew distribution, and the main function is to allow the application of regression Linear when analyzing data with a binary relationship to the dependent variables, which means that the change in the independent variables is offset by a change in the logarithm of the risk function, not the dependent variable itself [8,11].

3.2. Method Estimating the Cox Regression Model

In 1975 the scientist (Cox) proposed the Partial Likelihood method to estimate the parameters of the Cox regression model, which is still used for this purpose until now. This method highlights the probability of predicting or predicting the values of the dependent variable through independent variables in the absence of knowledge of the nature of the risk function [8]. The basic$(h(t))$and the partial function takes the following form:

$$L_P(\beta) = \prod_{i=1}^{m} \frac{X_i \beta}{\sum_{j=R(t_i)} e^{X_j \beta}} \qquad (13)$$

Where (m)represents the number of failures and (X_i) represents the variable values for the item whose survival time (ti) and the risk group$(R(t_i))$,that is, all the items at risk of failure, which are$t_i = \{j = t_i \geq t_i\}$. Taking the natural logarithm of equation (13) becomes:

$$log\ L_P(\beta) = \sum_{i=1}^{m} \left[X_i \beta - log \sum_{j=R(t_i)} e^{X_j \beta} \right] (14)$$

After deriving equation (14) and equating it to zero, we get the model parameters.

4. THE KAPLAN-MEIER ESTIMATE OF THE SURVIVAL FUNCTION

The Kaplan-Meier estimate is the result of a sequence of probabilistic estimates. The Kaplan-Meier method is the best approach for estimating survival data. To obtain the Kaplan-Meier estimate, a sequence of time intervals is constructed in the same way that a life-table estimate is constructed, and each interval includes one death time, however, more than one individual may die at any given death time. The Kaplan-Meier estimate is the limiting value of the life-table estimate, and when the number of intervals approaches infinity and their width approaches zero, it is known as the product-limit estimate of the survival function [8]. To evaluate whether two survival curves are statistically distinct, utilize the Log-Rank Test.

The Kaplan-Meier estimate of the survival function is given by:

$$\hat{S}(t) = \prod_{j=1}^{k}\left(\frac{n_j - d_j}{n_j}\right) \qquad (15)$$

5. RESULTS AND DISCUSSION

5.1. Frequency Distribution and Log-Rank Test for Symptoms

The frequency distribution and log-rank test for symptoms are shown in Table **1**. Chi-square test results show that some variables are statistically significant while others are not at the 5% level.

Table **1** reveals that the percentage of males who experience Coronavirus symptoms is higher than that of females, with 91.2% versus 8.8%. The majority of Coronavirus patients (64.7%) are above the age of 51 years, with the most evident symptoms being headache (93.1%), diarrhea (95.1%), anosmia (92.2%), and myalgia (88.2%). The percentage of patients with Coronavirus symptoms is higher than the percentage of patients without symptoms. The log-rank test revealed statistically significant differences at the 5% level between age groups and symptom survival times. (Fever, Cough, shortness of breath, Confusion, and Myalgia).

Figure **1** shows the KM curves toward the survival and hazard experience from time to death it represented; the survival plot starts with a strong fall and thereafter declines. This may suggest that there were more deaths when the symptoms first appeared.

Figure **2** shows a disparity between Kaplan-Meier curves and survival functions for symptomatic patients (fever, cough, dyspnea, and anosmia), with the curve estimated for those without symptoms being greater. As a result, persons who do not have symptoms should

Table 1: Frequency Distribution and Log-Rank Test for Symptoms

Variable	Categories	Frequency N= 325	Percentage%	Log-rank test	
				Chi-square test	P-value
Gender	Male	93	91.20%	3.43	0.064
	Female	9	8.80%		
Age	Lowest throw 30	9	8.80%	14.9	0.001
	31 throw 50	27	26.50%		
	51 throw highest	66	64.70%		
Fever	Yes	82	58.80%	26.29	0
	No	22	41.20%		
Cough	Yes	81	18.80%	28.5	0
	No	21	13.80%		
Shortness of breath	Yes	62	60.78%	4.3	0.038
	No	40	39.22 %		
Headache	Yes	95	93.10%	0.02	0.892
	No	7	6.90%		
Diarrhea	Yes	97	95.10%	0.25	0.616
	No	5	4.90%		
Fatigability	Yes	70	68.60%	1.13	0.289
	No	32	31.40%		
Confusion	Yes	80	78.40%	14.87	0
	No	22	21.60%		
Anosmia	Yes	94	92.20%	8.17	0.004
	No	8	7.80%		
Myalgia	Yes	90	88.20%	4.35	0.037
	No	12	11.80%		
Other	Yes	82	50.00%	0.11	0.75
	No	20	12.20%		

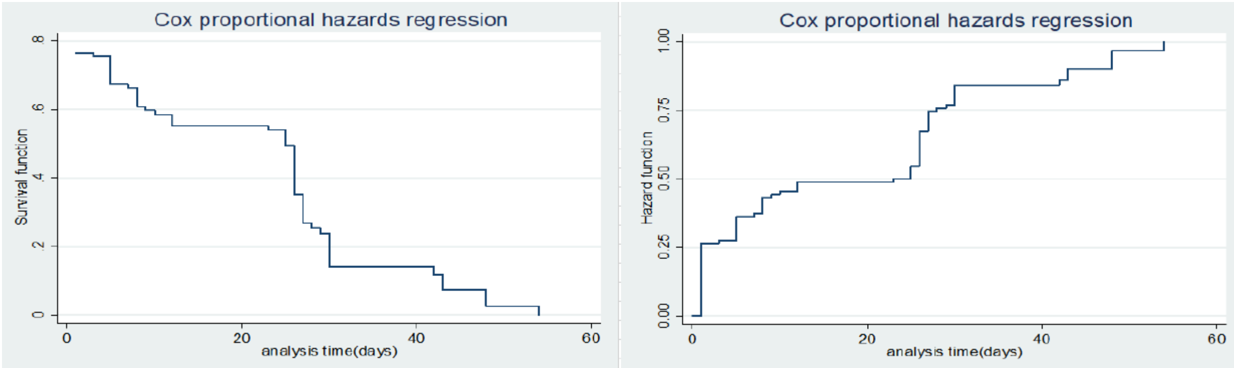

Figure 1: The KM curves survival and hazard.

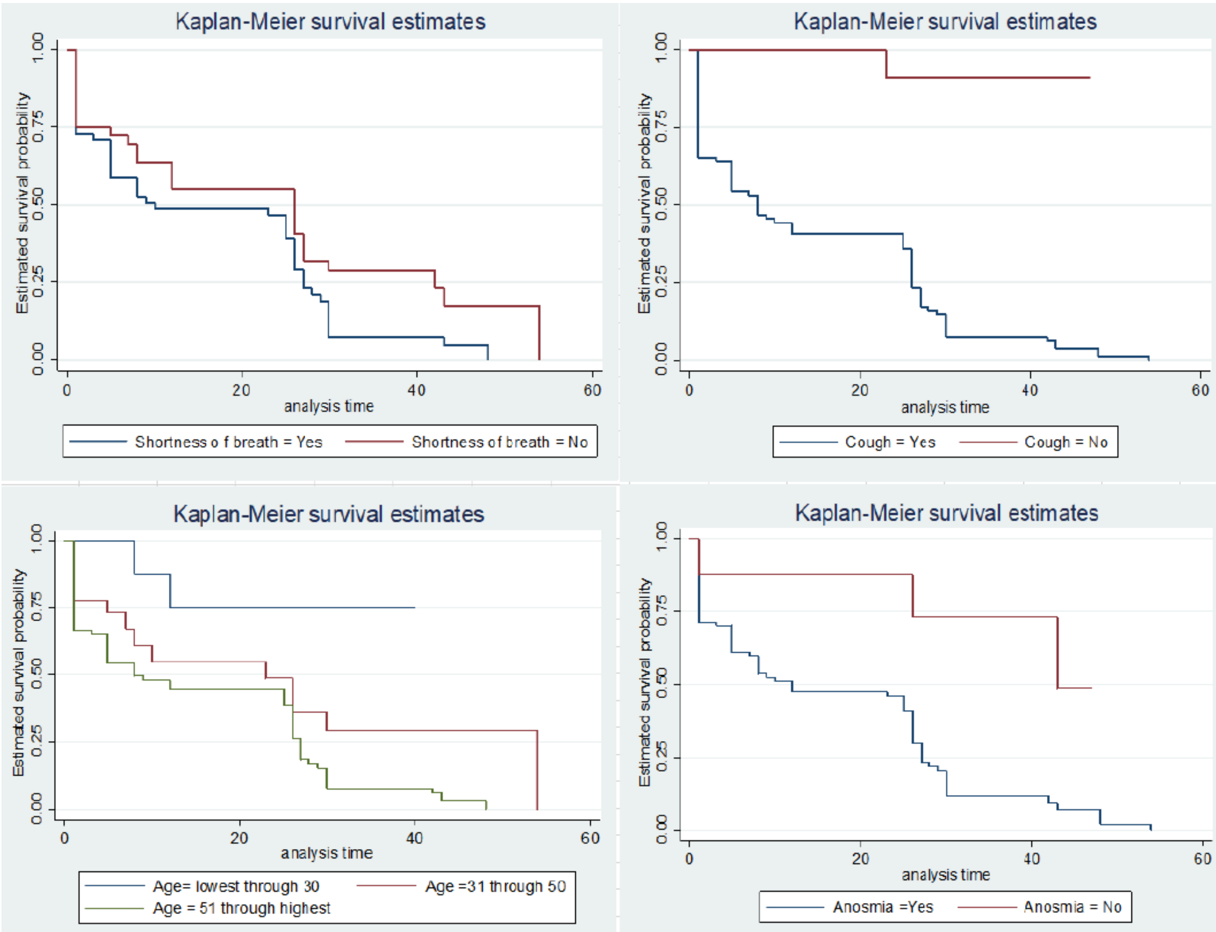

Figure 2: Kaplan-Meier curves and survival functions.

live longer. Also, the Kaplan-Meier survival curves varied per age group.

Figure **3** shows that there is a difference between the Kaplan-Meier curves of the survival functions for symptomatic patients (fever, confusion, and myalgia), and the curve estimated for those without symptoms is greater than the curve predicted in the previous figure. As a consequence, it is expected that patients who display these symptoms have a shorter survival rate.

5.2. Results of Cox Regression

Table **2** Based on multivariate analysis, it was determined that age was a risk factor (Coef=0.751, p-value=0.005 <0.05). And evaluated symptoms including fever (Coef=-2.57, p-value=0.000), cough (Coef=-3.928, p-value=0.006), Shortness of breath (Coef=-0.573,p-value=0.045), confusion Coef=--0.915, p-value=0.043), myalgia (Coef=-0.99, p-value=0.034), anosmia (Coef=-1.417, p-value=0.032 <0.0) we resignificant results from the Wald test (P-value <0.05) compared to other variables that were insignificant. It will be entered into the multivariate model.

We see from the results of Table **2** that the hazard ratio (HR) variables include age (HR=2.165), which means the hazard of age at any time in the study increases by 116.5%. Diarrhea (HR=1.090), which

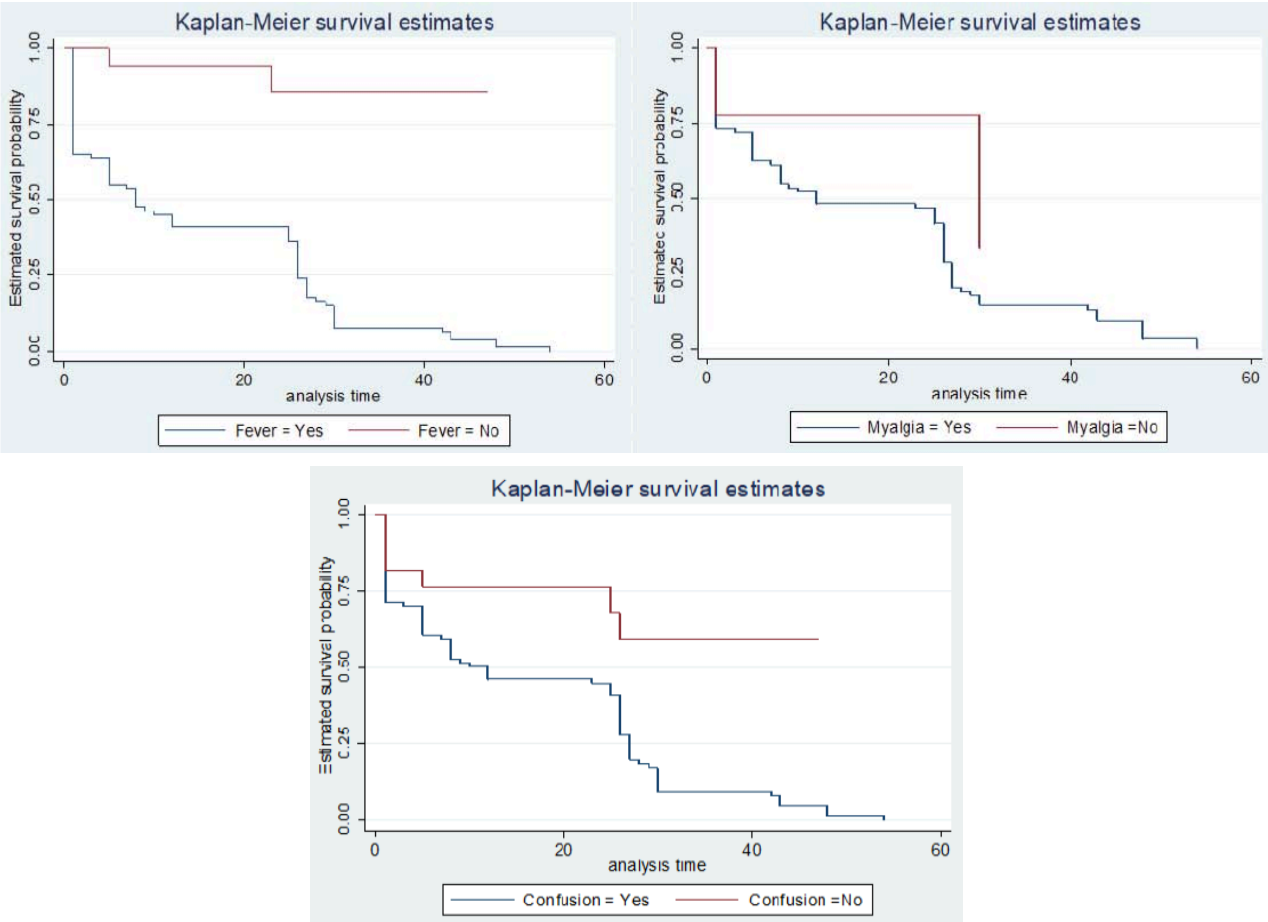

Figure 3: Survival curves of functions (fever, confusion, and myalgia).

Table 2: Estimation Results of the Cox Regression Model

Variable	Coefficient	Std. Err.	Z	P-value	[95% Conf. Interval]	
Age	0.751	0.269	2.790	0.005	0.224	1.278
Gender	-0.237	0.474	-0.500	0.617	-1.165	0.692
Fever	-2.57	0.717	-3.58	0.000	-3.975	-1.164
Cough	-3.928	1.422	-2.76	0.006	-6.714	-1.141
Shortness of breath	-0.573	0.289	-1.980	0.047	-1.139	-0.007
Diarrhea	0.086	0.498	0.170	0.863	-0.891	1.063
Headache	0.483	0.524	0.920	0.356	-0.543	1.509
Fatigability	0.125	0.295	0.420	0.672	-0.453	0.702
Confusion	-0.915	0.452	-2.030	0.043	-1.800	-0.030
Myalgia	-0.99	0.466	-2.12	0.034	-1.904	-0.076
Anosmia	-1.417	0.623	-2.270	0.023	-2.638	-0.196
Other	0.387	0.308	1.260	0.208	-0.216	0.990

means the hazard of Diarrhea at any time in the study increased by 9%. Headache (HR=1.621), which means the hazard of diabetes mellitus at any time in the study increased by 62.0%. Fatigability (HR=1.133), that means the hazard of diabetes mellitus at any time in the study increased by 13.3.8%. Other (HR=1.473), that means the hazard of diabetes mellitus at any time in the study increased by 47.3%.

6. CONCLUSIONS

The present study emphasizes the significant influence of the severity of symptoms on the duration of survival among individuals diagnosed with COVID-19. Individuals who exhibit extreme symptoms, namely [specific severe symptom], are at a substantially elevated chance of experiencing unfavorable results.

Table 3: Estimation Results of the Cox Proportional Hazard Model

Variable	Hazard ratio	Std. Err.	Z	P-value	[95% Conf. Interval]	
Age	2.165	0.581	2.880	0.004	1.279	3.665
Gender	0.789	0.374	-0.500	0.617	0.312	1.997
Fever	0.077	0.055	-3.580	0.000	0.019	0.312
Cough	0.020	0.028	-2.760	0.006	0.001	0.320
Shortness of breath	0.564	0.163	-1.980	0.047	0.320	0.993
Diarrhea	1.090	0.543	0.170	0.863	0.410	2.894
Headache	1.621	0.848	0.92	0.356	0.581	4.522
Fatigability	1.133	0.334	0.420	0.672	0.636	2.018
Confusion	0.400	0.181	-2.030	0.043	0.165	0.970
Myalgia	0.371	0.173	-2.120	0.034	0.149	0.926
Anosmia	0.243	0.151	-2.270	0.023	0.072	0.822
Other	1.473	0.453	1.260	0.208	0.806	2.691

The aforementioned results highlight the significance of timely identification, proactive treatment, and specific interventions for individuals at high risk who exhibit severe symptoms of COVID-19.

Men were more likely to develop symptoms, while adults 51 and older made up the bulk of the age groups impacted by symptoms. The study also found that the most prevalent symptoms among COVID-19 patients in Sudan include headache, diarrhea, sleeplessness, and muscle cramps. In addition to symptoms such as fever, cough, breathlessness, confusion, myalgia, and anosmia, the Cox model revealed a substantial effect on survival time by age as a risk factor in the Sudanese population afflicted with Crohn's disease. The hazard ratio model also revealed that the most severe symptoms that lead to death are Diarrhea, headache, and Fatigability, in addition to old age, which has the greatest risk rate of mortality. The study recommends that people in the community who exhibit the symptoms described in the study should seek medical assistance and diagnosis immediately, which minimizes the chance of mortality. The study also recommended that those over the age of 51, particularly men, exercise extra caution because they are more exposed to corona infection.

CREDITAUTHORSHIP CONTRIBUTION STATEMENT

Alshaikh A.Shokeralla: Conceptualization, Formal analysis, Methodology, Software, Writing - review & editing.

Mohammed Elameen E. Qurashi: Conceptualization, Writing - original draft, Software, Visualization.: Writing - original draft, Formal analysis.

Reem Y. Mekki: Writing - original draft, Visualization, Writing - review & editing.

Mortada S. Ali: Formal analysis, Visualization, Writing - review & editing.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no competing interests.

ACKNOWLEDGEMENTS

The authors are grateful to Professor Dr. EL-Nauman Said Al-Baha University for valuable comments.

CONTRIBUTION

The authors have contributed equally to this work.

REFERENCES

[1] Centers for Disease Control and Prevention. Symptoms of coronavirus 2020.

[2] Han C, *et al*. Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes. The American Journal of Gastroenterology 2020; 1-8. https://doi.org/10.14309/ajg.0000000000000664

[3] Sisay G, Mantefardo B, Beyene A.Time from symptom onset to severe COVID-19 and risk factors among patients in Southern Ethiopia: a survival analysis. Journal of International Medical Research 2022; 50(8): 1-12. https://doi.org/10.1177/03000605221119366

[4] Kasol DAW, *et al*. Survival analysis of COVID-19 patients in Ethiopia: A hospital-based study. PLOS One 2022; 17(5): e0268280. https://doi.org/10.1371/journal.pone.0268280

[5] Subramanian A, *et al*. Symptoms and risk factors for long COVID in non-hospitalized adults. Nature Medicine 2002; (28): 1706-1717. https://doi.org/10.1038/s41591-022-01909-w

[6] Ghodsiyeh A, Freshteh O. Clinical characteristics and risk factors for mortality in COVID-19 inpatients in Birjand, Iran: a single-center retrospective study. Eur J Med Res 2021; 26(1): 79. https://doi.org/10.1186/s40001-021-00553-3

[7] Lawless JF. Statistical Models and Methods for Lifetime Data, Second Edition. Second Edition ed. New York: John Wiley & Sons 2011.

[8] Collett, David. Modelling Survival Data in Medical Research. Fourth Edition. London: Chapman & Hall 2023. https://doi.org/10.1201/9781003282525

[9] Kleinbaum David G, Mitchel Klein. Survival Analysis A Self-Learning Tex. Third Edition. New York: Springer Science 2012. https://doi.org/10.1007/978-1-4419-6646-9

[10] Lee ET, Wang JW. Statistical Methods for Survival Data Analysis. Fourth Edition ed. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey 2013.

[11] Kalbfleisch JL, Prentice RL. The Statistical Analysis Of Failure Time Data,. Second Edition ed. Canada.: by John Wiley & Sons, Inc., Hoboken, New Jersey 2002. https://doi.org/10.1002/9781118032985

Received on 29-10-2023 Accepted on 25-11-2023 Published on 19-12-2023

https://doi.org/10.6000/1929-6029.2023.12.29