ISSN: 2588-4115; Infection Epidemiology and Microbiology. 2024;10(2):123-139 10.52547/iem.10.2.123

Analyzing Survival Outcomes of COVID-19 Patients: A Cox Regression Approach with Schoenfeld Residual Diagnostics

ARTICLE INFO

Article Type
Original Article

Authors
Sanjith Bharatharajan Nair, *PhD*[1]
Dhananjay Yadav, *PhD*[1*]

[1] Department of Mathematical & Physical Sciences, University of Nizwa, P.O. Box 33, PC 616, Oman

*** *Correspondence***
Department of Mathematical & Physical Sciences, University of Nizwa, P.O. Box 33, PC 616, Oman.
E-mail: dhananjay@unizwa.edu.om

How to cite this article
Nair S.B., Yadav D. Analyzing Survival Outcomes of COVID-19 Patients: A Cox Regression Approach with Schoenfeld Residual Diagnostics. Infection Epidemiology and Microbiology. 2024;10(2): 123-139

Article History
Received: January 16, 2024
Accepted: April 14, 2024
Published: June 21, 2024

ABSTRACT

Background: In this study, data were collected from the Eastern Democratic Republic of Congo and analyzed by Cox regression model. In addition, hazard functions and survival outcomes in COVID-19 patients were also analyzed. .
Materials & Methods: One million simulated data on hospitalized patients' characteristics with positive SARS-CoV-2 infection were collected from the Humanitarian Data Exchange Source in the Eastern Democratic Republic of Congo from December 2020 to June 2021. Several statistical techniques were developed in this study for data analysis, including Kaplan-Meier curves, log-rank test, Schoenfeld residual diagnostics, and likelihood ratio test.
Findings: This study finding showed that there was a 4.5% increase in the expected hazard per unit year increase in age. In addition, the risk of death was higher in males than in females, and patients with no signs of anorexia, ageusia, or anosmia, no history of diabetes or tuberculosis, normal pulse rates, and no hypoxemia had a greater survival rate than those with such health conditions.
Conclusion: This study finding revealed that covariates such as age, gender, anorexia, ageusia, anosmia, diabetes, and tuberculosis were expressively connected with higher mortality rates. In addition, hypoxemia and high pulse rate were associated with higher death rates; however, anti-inflammatory and anticoagulant agents were shown to reduce mortality rates, and multivitamin or vitamin C had a substantial impact on patient survival.

Keywords: COVID-19, Mortality, Death rate, Cox regression, Kaplan-Meier curve, Log-rank test, Schoenfeld residuals, Survival function

CITATION LINKS

[1] Doocy S, et al. Clinical progression and... [2] Sepandi M, et al. Estimate of the basic reproduction... [3] Bhandari S, et al. Patient flow hynamics... [4] Liu X. Survival analysis... [5] Das D, et al. A survival... [6] Geng J, et al. Survival in pandemic... [7] Abbas J. Crisis management, global... [8] Persson I. Essays on the assumption of... [9] Kvamme H, Borgan Ø, Scheel I. Time-to-event... [10] Jullum M, Hjort NL. What price... [11] Cox DR. Partial likelihood... [12] Hu C, Lin DY. Cox regression with... [13] Anderson RN, Rosenberg HM. Age standardization... [14] Kurd DM,et al. Association of ... [15] Van Dijk PC, The analysis of... [16] Benítez-Parejo N, et al. Survival analysis and... [17] Faraggi D, Simon R. Bayesian variable selection... [18] Wang C, et al. Immediate psychological ... [19] Hoffmann M, et al. SARS-CoV-2 cell... [20] Pascarella G, et al. COVID-19... [21] Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19... [22] Zhou F, et al. Clinical course... [23] Atlam M, Torkey H, El-Fishawy N, Salem H. Coronavirus... [24] Gaskin DJ, Zare H, Delarmente BA. Geographic... [25] Ghadamgahi F,The effect of ... [26] Fung KW, et al. Effect of common... [27] Huang H, et al. Association between ... [28] Seif M, et al. Factors associated... [29] Wastnedge EA, et al. Pregnancy and... [30] Domjanović J, et al. Association of different... [31] Hastie CE, et al. Chronic pain... [32] Kojima K, et al. Increased lactate... [33] Lund LC, et al. Cox regression... [34] Bitew ZW, et al. Determinants of... [35] Schoenfeld D. Partial residuals... [36] Park S, Hendry DJ. Reassessing Schoenfeld... [37] Box-Steffensmeier JM, Jones BS. Event history... [38] Ozenne B, et al. risk Regression... [39] Grønnesby JK, Borgan Ø. A method... [40] Li H, Luan Y. Kernel Cox regression... [41] Sargent DJ. A flexible... [42] Gui J, Li H. Penalized Cox...

Introduction

The coronavirus was identified in the 1930s and initially only affected animals. Coronaviruses have undergone numerous mutational stages; they first appeared as the common cold in the 1960s and then evolved into their current form with respiratory effects. When individual talks, coughs, or sneezes, then the respiratory droplets are released, which allow the virus to transmit from one person to another. In the elderly and those with underlying medical problems, COVID-19 (coronavirus disease-2019) symptoms could be lethal and range from moderate fever, cough, and dyspnea (breathing difficulty) to severe pneumonia and acute respiratory distress syndrome. Numerous techniques have been developed to accurately forecast patient survival using symptom information and certain clinical criteria due to the increase in mortality caused by COVID-19 and the acceleration of its dissemination [1–3].

This study analyzed and modeled the survival outcomes of hospitalized COVID-19 patients in the African humanitarian environments of Juba, South Sudan, and North and South Kivu in the Eastern Democratic Republic of Congo. It should be mentioned that African regions were chosen due to the large number of infections caused by the COVID-19 pandemic, given that few published studies were available on COVID-19 data in African regions. This observational cohort study was conducted from December 2020 to June 2021 on hospitalized COVID-19 patients who received special care and treatments and were monitored daily until death or recovery [1].

In Africa, low vaccination rates, insufficient diagnostic and laboratory capacity, and restricted access to evidence-based COVID-19 treatments such ventilators, antivirals, and monoclonal antibodies all contribute to poor quality of care. Tertiary institutions are typically overcrowded and difficult to access, especially for remote communities. This study also aimed to shed light on COVID-19 impacts in resource-constrained and conflict-affected African regions.

Survival analysis has become an important tool for data analysis, especially survival data related to a specific time or from a specific time to an event. The importance of survival analysis increases in times of global outbreaks of dangerous diseases or during/ after any global pandemic, such as the COVID-19 pandemic. In addition, the analysis of failure events (deaths), as observed during this pandemic, is of great importance to communities and has social, economic, and environmental impacts [4, 5]

It is well known that any global pandemic leads to various crises, which may be economic, health, service, educational, food, or environmental and threaten the security of society in terms of sustainable energy and food. In addition, the related literature has made it clear that the threats associated with pandemics could distract society's attention from other important and unforeseen challenges, such as energy challenges, food security, and human environmental behaviors resulting from pandemics. For example, the COVID-19 crisis and the resulting closure have affected the entire world in terms of the environment, economic stagnation, low income, lack of food, etc [6, 7].

Survival analysis is often used in medical and public health investigations to determine and analyze the length of time from a defined starting point to a given event. Regression models for survival analysis with censored observations have been frequently utilized for many years. In 1972, Cox DR proposed one of the most widely used regression models, called Cox regression model (CRM) or Cox proportional hazards regression

model, which is a widely used regression model for survival data analysis [8] .

CRM simplifies survival rate analysis by defining the instantaneous mortality rate, known as the hazard function or hazard rate, which is the probability of an event (death or disease) occurring in a period of time, assuming survival or non-occurrence till that moment. A hazard ratio is the ratio of hazard rates, which corresponds to the conditions characterized by two distinct levels (groups) of a given treatment variable over a specified time period, or the ratio of hazard rates in the treatment and control groups. In this paper, it is termed as the Cox hazard ratio (CHR). It is calculated for each predictor in a CRM and used to measure the variation in the hazard rate per unit increment in the predictor variable while keeping all other variables constant [8, 9]. The proportional hazards assumption (PHA) is the major assumption in CRM, which states that CHR may remain constant across time. Prior to employing CRM, this assumption should be validated.

When using regression to analyze models for censored survival data, it is common to assess the significance of specific prognostic characteristics such as age, race, or gender in forecasting the likelihood of survival. This common issue arises in most clinical studies for cancer and AIDS research. CRMs are typically utilized to solve this issue. Based on the values of the predictors, the outcomes of a CRM could be used to forecast survival time [10, 11].

In CRM, data censoring is presumptively random or non-informative. In other words, patients who are still being followed up at time "t" (after eliminating patients who experienced an incident and were censored) would represent a random sample of the whole research population [12–14]. In fact, this crucial supposition is very difficult to be quantitatively tested. The only way to guarantee this is through active data collection (ibid).

In addition, finding the CHR and its 95% confidence interval is the major goal of CRM, and the CHR at any given time point may depend on a variety of covariates or explanatory variables, some of which may not be stated quantitatively. Other factors that may increase a person's risk rate include age, hypertension, profession, general health, existence of any other linked issues, etc. It should be mentioned that the CRM parameters are based on the estimation of covariates separately from the HR parameters at time t in addition to accounting for all response variables [10, 15, 16].

Faraggi and Simon (1998) [17] examined CRMs utilizing right-censored survival data. Most survival studies run into a more general issue, which is often resolved by a CRM utilizing Cox partial likelihood. When utilizing stochastic models to analyze data on survival, fertility, or any other population feature, CRM is utilized as an essential tool, which is based on the partial likelihood technique [11].

During the primary phases of the COVID-19 outbreak, Wang et al. (2020) performed a general public survey in China to recognize the degrees of psychological effect, anxiety, despair, and stress. Hoffmann et al. (2020) [19] demonstrated that SARS-CoV-2 infection was dependent on the host cell factors ACE2 and TMPRSS2 and could be prevented by a therapeutically effective protease inhibitor. Pascarella et al. (2020) [20] a novel coronavirus from the same family as SARS-CoV and Middle East respiratory syndrome coronavirus, has spread worldwide leading the World Health Organization to declare a pandemic. The disease caused by SARS-CoV-2, coronavirus disease 2019 (COVID-19conducted a systematic literature search using major online databases to offer an overview of COVID-19.

Shereen et al. (2020) [21] summarized and compared the occurrence and pathogenicity of COVID-19 infection to other human coronaviruses, like SARS-CoV and MERS-CoV. Zhou et al. (2020) investigated risk factors using logistic regression approaches. Atlam et al. (2021) [23] employed machine learning and artificial intelligence approaches for computing infection risks, performing survival analysis, and categorization. Gaskin and colleagues (2021) [24] investigated the relationships between COVID-19 and proximity to transportation using negative binomial regressions and CRMs.

Using propensity score matching, Ghadamgahi et al. (2021) [25] examined the clinical outcomes of diabetic COVID-19 patients. Fung et al. (2022) [26] performed Cox regressions on all COVID-19 patients with COVID-19-related hospitalization and all-cause mortality outcomes. Huang et al. (2022) [27] looked at the link between past strokes and the risk of severe coronavirus illness. Kurd et al. (2022) [14] looked at how statin medications affected COVID-19 individuals. Seif et al. (2022) [28] analyzed patient survival using CRM and a mixture cure model. Wastnedge et al. (2021) [29] investigated the considerable impact of COVID-19 on pregnancy-related changes in the body.

Domjanović and colleagues (2023) [30] compared the connection and differentiation of COVID-19 clinical risk scores in patients who underwent kidney transplantation. Hastie et al. (2023) [31] looked into whether self-reported chronic pain was linked to COVID-19 hospitalization or mortality. Kojima et al. (2023) [32] conducted a retrospective cohort analysis on 450 COVID-19 patients. Using simulation and empirical data analysis, Lund et al. (2023) [33] assessed bias in person-time based approaches with and without calendar time adjustment.

There are eight sections in this paper. Section 2 includes the objectives of the study. In Section 3, various materials and methods are presented, which describe the methodological steps taken to answer the research questions and accomplish the study goals. All the study results are included in Section 4, which is the Findings section. Section 5, Discussion, provides a clear and thorough explanation of the outcomes and procedures utilized to evaluate the validity and dependability of the findings. Sections 6 includes the conclusion of the study. Limitations and future research directions are included in Sections 7 and 8.

Objectives: The objectives of this paper were as follows: to provide information on the impact of COVID-19 in resource-constrained and conflict-affected regions in some African countries, to study available data and develop a CRM for modelling the covariates of the COVID-19 data, and to implement testing techniques in order to improve the reliability, enhance the interpretability, and ensure the generalizability of the findings.

Materials and Methods

A real data case-cohort study: It is well-known that this pandemic is hazardous considering that many people have lost their lives all over the world, especially in Africa; thus, it is critical to conduct research in Africa and the world. In order to study this pandemic, adequate data should be obtained and analyzed. The COVID-19 Pandemic-Humanitarian Data Exchange Source (https://data.humdata.org/) provided the data for a period from December 2020 to June 2021.

In this study, one million simulated data on hospitalized COVID-19 patients' characteristics in African regions including Juba, South Sudan, and North and South Kivu in the Eastern Democratic Republic of Congo were analyzed. In addition, in this

study, a CRM was developed to examine clinical progression and survival outcomes of COVID-19 patients in African regions. Several statistical techniques were executed in this study for data analysis, including Kaplan-Meier (K-M) curves, log-rank test (LRT), Schoenfeld residual diagnostics (SRD), and likelihood ratio test. Also, K-M survival analysis was used to analyze the significance of survival curves, which are generally utilized to display survival functions of predictor variables [34].

LRT is one of the most common approaches for comparing survival distributions between two or more groups. In this study, LRT was applied to compare survival distributions in K-M curves. It tests the null hypothesis that "there is no significant difference between the two survival curves." In addition, LRT provides strong statistical power when the difference in hazard rates between groups is significant. In addition, the likelihood ratio test is performed for CRM to determine the improvement in model fit when removing or adding new covariates.

R software packages 'survival', 'survminer', 'simsurv', and 'ggplot2' were used for data analysis in the study because they produce accurate results. The 'survival' package provides functions for fitting survival models like CRM and performing K-M, SRD, LRT, and likelihood ratio test. The 'survminer' package supplies visualization tools for survival analysis. Also, for simulating survival data, 'simsurv' package is utilized. The 'ggplot2' package is used to generate tailored plots of survival data and findings from CRMs and K-M estimators.

Some of the above-mentioned statistical methods like SRD and CRM are discussed in detail below.

SRD: It is well-known that one of the major limitations of CRM, is the violation of PHA, which leads to wrong parameter estimation. Therefore, it is important to perform a goodness-of-fit test to check for PHA violations. It is well-known that SRD is a representative goodness-of-fit test for validating PHA in CRMs. CHR is assumed to be stable across time under PHA. SRD assumes that the residuals of the covariates are independent of time. SRD is useful for testing the violation of PHA since plotting these residuals across time could disclose whether a covariate coefficient is time-dependent. Schoenfeld residuals are calculated using all explanatory variables included in the model. SRD incorporates essentially the observed minus predicted values of the covariates for every failure time [35, 36]"At failure time T_i, the Schoenfeld residual of the j^{th} covariate for the i^{th} individual is provided using the following equation.

$$E_{ij} = Y_{ij} - \overline{Y}_j(T_i) \quad (1)$$

Where Y_{ij} is the value of the j^{th} covariate for the i^{th} individual at failure time Ti, and is the weighted average of the j^{th} covariate, with weights based on the risk set size over all individuals who are still at risk at moment T_i. Plotting these residuals against time and observing if a systematic pattern emerges is the process of SRD. PHA violation could be suspected when the Schoenfeld residual plot presents a relationship with time [37]

CRM: CRM is a semi-parametric regression model, which is one of the most popular regression models for survival analysis. It is employed to relate several risk factors or exposures simultaneously to survival time. The basic assumption in CRM is PHA, which claims that CHR should remain constant for the two study groups throughout the study period. It may be noted here that the outcomes of CRM are examined.

In the survival analysis literature, predictor variables are frequently referred to as covariates. In CRM, the outcome is obtained as an estimate of the covariate hazard ratio, which is then expressed as the regression

coefficient in the model. If a covariate hazard ratio is more than one and statistically significant, it contributes to increasing the probability of the event occurring. If the hazard ratios of two groups do not remain constant, another analysis method should be used.

CRM in its most basic form is defined as follows:

(2)
$$h_j(t) = h_0(t)\exp\left(\beta' Z_j(t)\right)$$

Where $h_j(t)$ represents the hazard rate of the j^{th} individual at any time t; $h_0(t)$ is the hazard function (baseline hazard) that just considers time and ignores all other factors, i.e., maintaining $Z_{ij} = 0\ \forall\ i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n$; and Z_j' $(Z_{j1}, Z_{j2}, \ldots, Z_{jk})$ is the k-component covariate vector for the j^{th} individual.

T_i is a random variable that represents the time of death, it corresponds to each person (j=1, 2, ..., n). It is observed up to period dj for the j^{th} person (j=1, 2, ..., n) in a randomly censored sample. The CRM defined in Eq. (2) could be described as follows:

(3)
$$\log h_j(t) = \log h_0(t) - [\beta_1 Z_1(t) + \beta_2 Z_2(t) + \ldots + \beta_k Z_k(t)]$$

Based on an estimation of β_i (i= 1, 2..., k) and $h_0(t)$, the equation might be viewed as a single-equation log-linear model (t). However, Cox's technique is novel in that Cox suggested estimates of the regression parameters $\beta_1, \beta_2, \ldots, \beta_k$ that are independent of $h_0(t)$. $S(t) = \{j: T_j \geq t, d_j\}$ represents the risk set or the group of people at risk in the context of findings (the j^{th} person is put aside for observation between 0 and dj). Deaths are said to happen randomly.

The probability that individual j will also pass away, provided that there are n people in the sample, and that one of them passes away, is given by the following equation:

(4)
$$\frac{h_c(t)\exp(\beta' Z_j(t))}{\sum_{j\in S(t)} \lambda_c(t)\exp(\beta' Z_j(t))} = \frac{\exp(\beta' Z_j(t))}{\sum_{j\in S(t)} \exp(\beta' Z_j(t))} = P_L,$$

which extends the total over all the individuals in the risk set S, as specified by Cox.

(5)
$$L(\beta) = \prod_{T_j \leq c_j} \frac{\exp(\beta' Z_j(t))}{\sum_{j\in S(t)} \exp(\beta' Z_j(t))},$$

j=1, 2, ..., n (the product is extended over all $j\ \forall\ T_j \leq d_j$ and j=1, 2, ..., n).

The procedure produces estimates of $\beta_1, \beta_2, \ldots, \beta_p$ that have the asymptotic characteristics of maximum likelihood estimators, given the partial likelihood of identifying the parameters by the maximum likelihood technique.

According to the Cox proportional hazard model,

(6)
$$h(t \mid x) = h_0(t)\exp(x^T\beta)$$

where the undefined baseline hazard function $h_0(t)$ is used. More terminology is required to present the probability function of the observed data $\{(x_i, Z_i, \delta_i): i=1, 2, \ldots, n\}$ from the Cox proportional hazards model directly. The variables represent the observed failure times in the order $t_1^0 < \cdots < t_j^0$. In order for the covariates associated with N failures to be $x_{(1)}, \ldots, x_{(N)}$, (j) supplies the label for the item occurring at t_j^0. In addition, R_j stands for the risk that is set prior to time t_j^0: $R_j = \{i: Z_i \geq t_j^0\}$. The likelihood changes as follows:

(7)
$$L = \prod_{i=1}^{N} h_0(Z_{(i)})\exp(x_{(i)}^T\beta) \prod_{i=1}^{n} \exp\left\{-H_0(Z_{(i)})\exp(x_{(i)}^T\beta)\right\}$$

Where $H_0(.)$ denotes the cumulative baseline hazard function.

The equivalent penalized log-likelihood function is written as follows:

(8)

$$\prod_{i=1}^{N}\left[\log\left\{h_0\left(Z_{(i)}\right)\right\}+x_{(i)}^{T}\beta\right]-\sum_{i=1}^{n}\left\{H_0(Z_{(i)})\exp\left(x_{(i)}^{T}\beta\right)\right\}-n\sum_{j=1}^{d}p\lambda_n\left(|\beta_j|\right)$$

Baseline hazard and cumulative hazard function are specified but not parameterized; the panelized log-likelihood function in Eq. (8) has not yet been optimized. Nonparametric "least informative" modeling for $H_0(.)$, in line with Breslow's theory, has a potential jump h_j at the reported failure time t_j^0. More specifically:

(9)

$$H_0(t)=\sum_{j=1}^{N}h_jI(t_j^0\le t)$$

Then:

(10)

$$H_0(z_i)=\sum_{j=1}^{N}h_jI(i\in R_j)$$

The penalized likelihood function of Eq. (8) is altered by applying logarithm in Eq.

(11)

$$\sum_{i=1}^{N}\left\{\log\left\{h_j+x_{(j)}^{T}\beta\right\}-\sum_{i=1}^{n}\left\{\sum_{j=1}^{N}h_jI\left(i\in R_j\right)\exp\left(x_i^{T}\beta\right)\right\}-n\sum_{j=1}^{d}p\lambda_n\left(|\beta_j|\right)\right.$$

It is obtained by fixing the derivative to zero with regard to h_j and taking the derivative.

(12)

$$\hat{h}_j=\left\{\sum_{i\epsilon R_j}\exp\left(x_i^{T}\beta\right)\right\}^{-1}$$

By substituting into Eq. (10), the penalized likelihood is obtained as follows:

(13)

$$\sum_{i=1}^{N}\left[x_{(j)}^{T}\beta-\log\left\{\sum_{i\epsilon R_j}\exp\left(x_i^{T}\beta\right)\right\}\right]-n\sum_{j=1}^{d}p\lambda_n\left(|\beta_j|\right)\stackrel{def}{=}\ell_c\left(\beta\right)-n\sum_{j=1}^{d}p\lambda_n\left(|\beta_j|\right)$$

When $p\lambda(.)\equiv 0$, Eq. (13) is the partial probability function proposed by Cox (1975). As a result, the penalized partial likelihood is equal to the penalized likelihood. The method of obtaining it is to boost the penalized likelihood estimate of Eq. (13) with regard to β [38–42].

Findings

In this study, the survival distributions of significant covariates were displayed using K-M curves, indicating that all the associated LRT results were statistically significant at the threshold significance level of 0.05. SRD was executed to diagnose PHA violations in CRMs for all significant covariates. In this study, the likelihood ratio test was employed to extract significant covariates and thereby generate a best-suited CRM. The goodness-of-fit of all CRM covariates and the survival comparison test results of hospitalized COVID-19 patients were significant, as shown in Table 1.

It could be mentioned that the results of goodness-of-fit and survival comparison tests of all CRM covariates were significant.

Demographic covariates: Patients were grouped into four categories based on age: below 18 years, 18 to 44 years, 45 to 64 years, and above 65 years. The group over 65 years of age was taken as the baseline category for CRM, and it was observed that the survival rate in these categories decreased from 1 to 0.1, respectively, with respect to the baseline category. Fig. 1 shows the significance of survival functions of age categories. The 95% confidence interval for the CHR of age was between 1.018 and 1.073 with a CHR value of 1.045.

The CHR value of gender was 2.397, indicating that the risk of death was higher in males than in females, and the 95% confidence interval for CHR ranged from 1.13 to 6.194. The significance of gender on mortality is shown in Fig. 2**.** The estimated CRM coefficient was 0.874.

SRD was performed for demographic covariates, and the results were not significant, it was observed that the residuals did not show any relationship with time. In addition, PHA was not violated for demographic covariates. The results are displayed in Fig. 3.

Observed sign covariates: The CHR for patients with signs of anorexia (sign of

Table 1) Goodness-of-fit and survival comparison tests of all CRM covariates in hospitalized COVID-19 patients

	Covariate	Likelihood Ratio Test Statistic	*P*-Value	LRT Statistic	*P*-Value
Demographic information	Age	17.09	< .01**	14.87	< .01**
	Gender				
Observed signs	Sign of anorexia	12.78	< .01**	18.1	< .01**
	Sign of ageusia or anosmia				
Medical history	Medical history of diabetes	13.08	< .01**	18.44	< .01**
	Medical history of tuberculosis				
Clinical evaluations	Admission hypoxemia	36.74	< .01**	50.77	< .01**
	Admission high pulse rate				
Treatments provided	Noninvasive ventilation treatment	29.49	< .01**	28.01	< .01**
Medications given	Anticoagulant agents	11.8	< .01**	15.24	< .01**
	Anti-inflammatory agents				
Supplements provided	Multivitamin or vitamin C	5.21	< .05*	5.92	< .05**

(** and * indicate that the approximated regression coefficient is significant at the significance level of 1% and 5%, respectively).

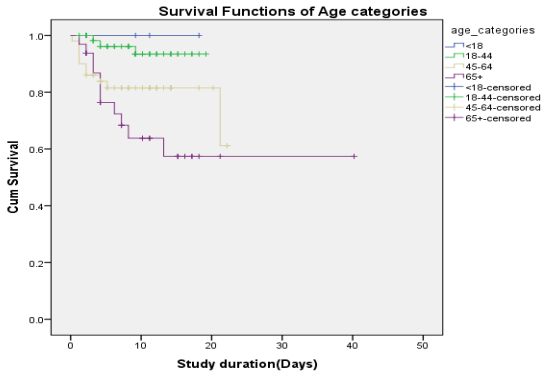

Figure 1) Survival functions of age categories in COVID-19 patients

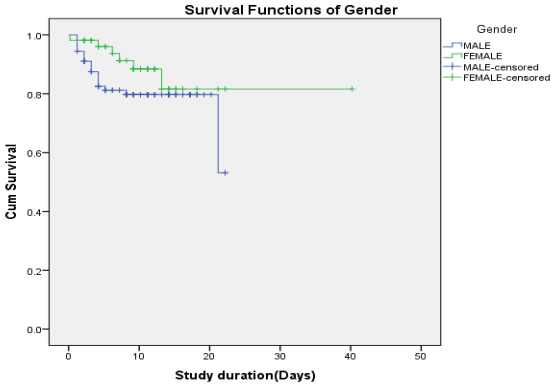

Figure 2) Survival functions of gender categories

reduced appetite) and ageusia or anosmia (loss of sense of taste or smell) was found to be 0.067 (CHR CI: 0.019 to 0.240) and 0.282 (CHR CI: 0.108 to 0.734), respectively. Figures 4 and 5 show the survival functions relevant to the signs of anorexia and ageusia or anosmia, respectively. The estimated CRM coefficients for anorexia and ageusia or anosmia were -2.697 and -1.267, respectively. SRD was executed for observed sign covariates, and the results supported the validity of the PHA. The Shoenfled residual plot is shown in Fig. 6.

Medical history covariates: The CHR values for patients with a history of diabetes and tuberculosis were 0.225 (CHR CI: 0.067 to 0.756) and 0.108 (CHR CI: 0.013 to 0.913), respectively (Table 2). The importance of history of diabetes in survival is displayed in Fig. 7. The estimated CRM coefficients for history of diabetes and history of tuberculosis were -1.494 and -2.225, respectively.

The SRD results for medical history are presented in Fig. 8. The results supported no

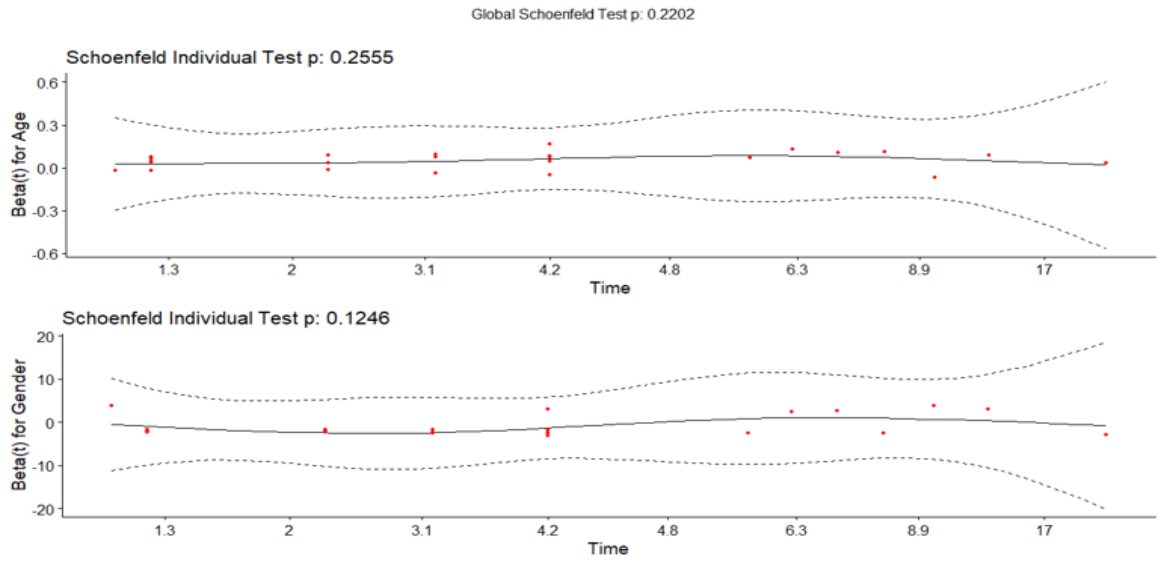

Figure 3) SRD plots for demographic covariates

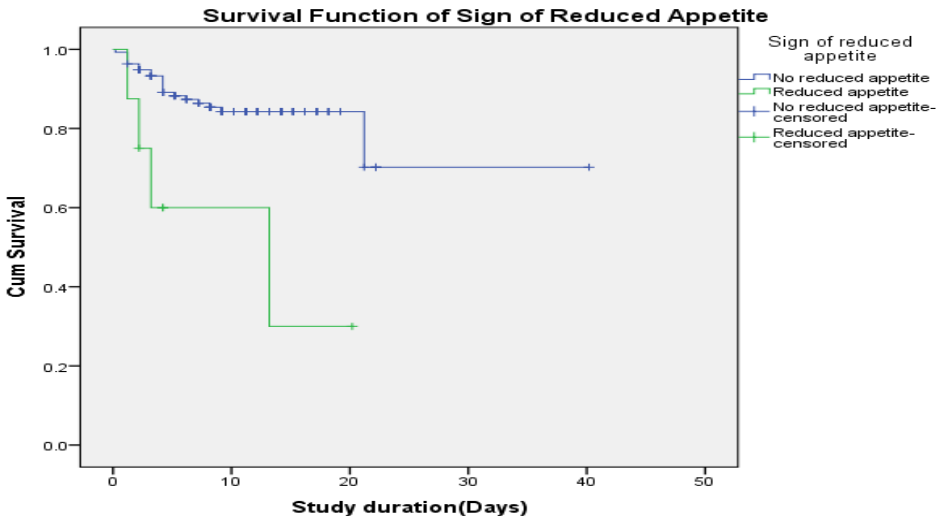

Figure 4) Survival functions of sign of anorexia

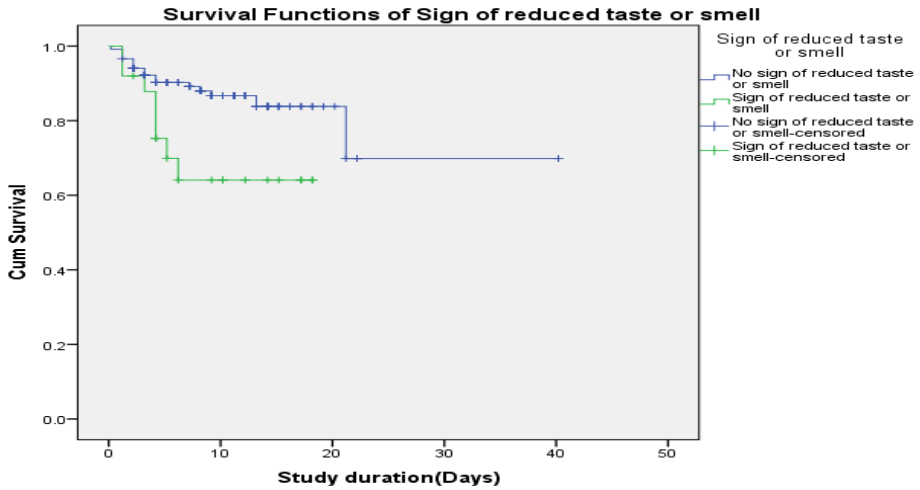

Figure 5) Survival functions of sign of ageusia or anosmia

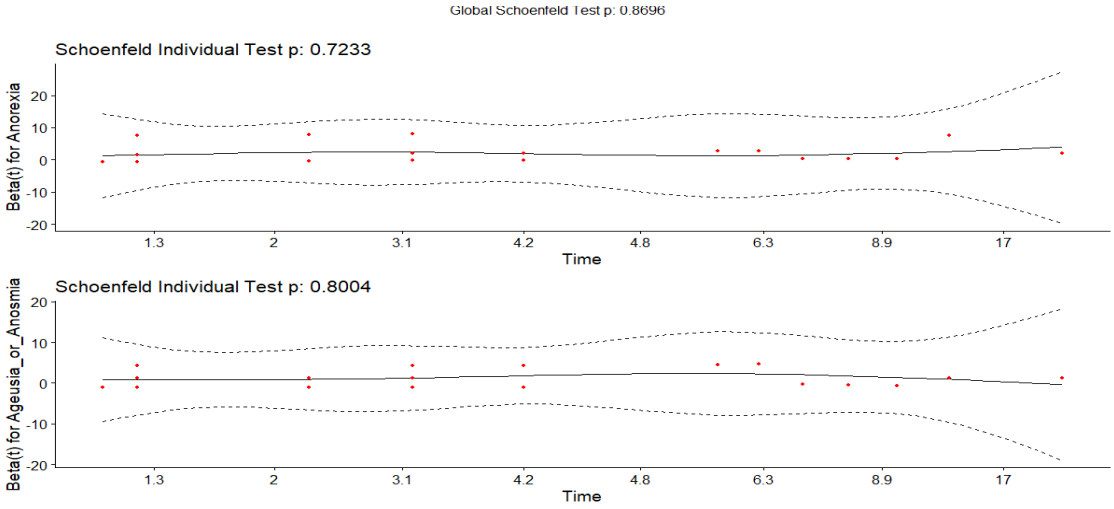

Figure 6) SRD plots for observed sign covariates

Table 2) CRM outcomes of demographic, observed sign, and medical history covariates in COVID-19 patients

	Covariate	CRM Coefficient and S. E		*P*-Value	CHR	95% CI for CHR	
		$\hat{\beta}_i$	S,E ($\hat{\beta}$i)			Lower	Upper
Demographic information	Age	0.044	0.014	.001**	1.045	1.018	1.073
	Gender	0.874	0.484	.042*	2.397	1.13	6.194
Observed signs	Sign of anorexia	-2.697	0.648	< .001**	0.067	0.019	0.240
	Sign of dyspnea	-2.410	0.754	< .001**	0.090	0.021	0.393
	Sign of ageusia or anosmia	-1.267	0.488	.011*	0.282	0.108	0.734
Medical history	Medical history of diabetes	-1.494	0.619	.016*	0.225	0.067	0.756
	Medical history of tuberculosis	-2.225	1.089	.041*	0.108	0.013	0.913

(** and * indicate that the approximated regression coefficient is significant at the significance level of 1% and 5%, respectively).

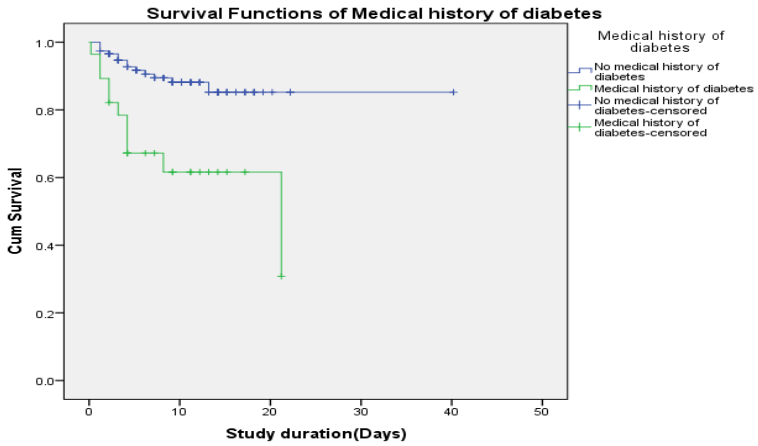

Figure 7) Survival functions of medical history of diabetes

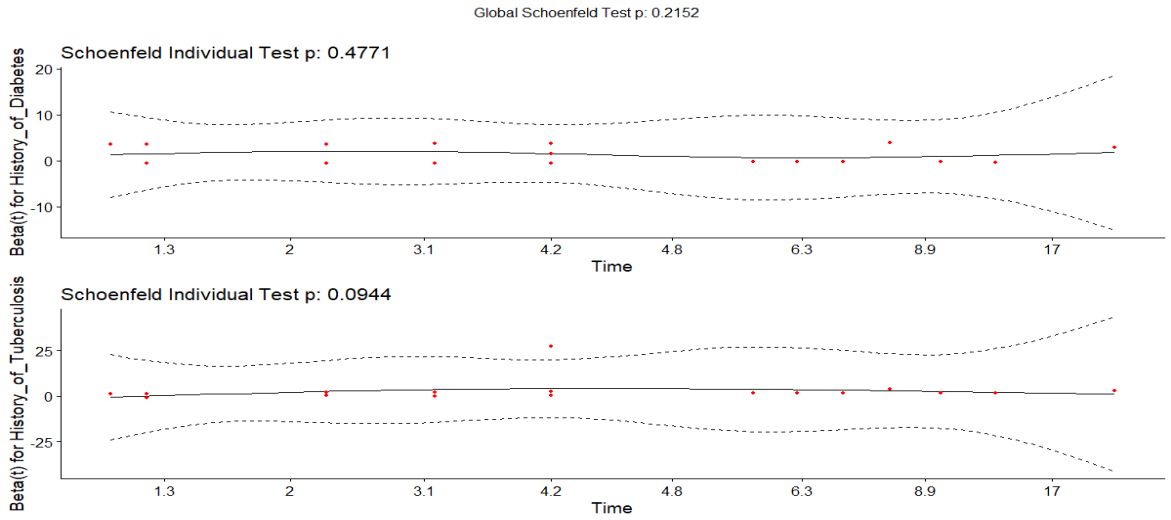

Figure 8) SRD plots for medical history covariates

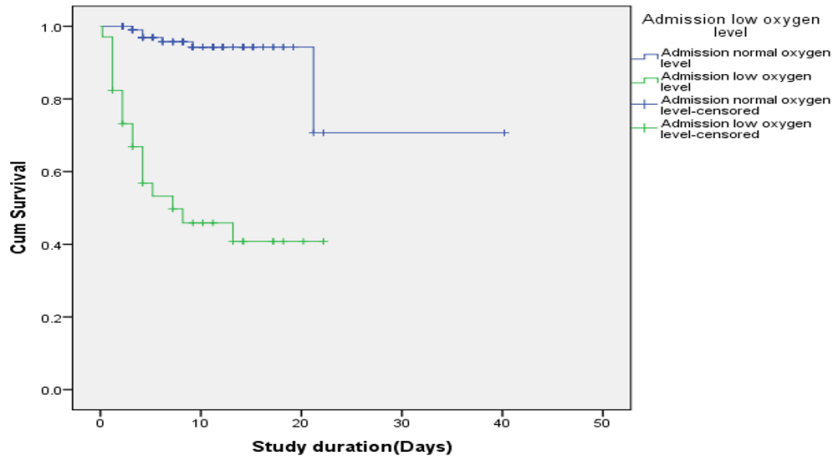

Figure 9) Survival functions of hypoxemia

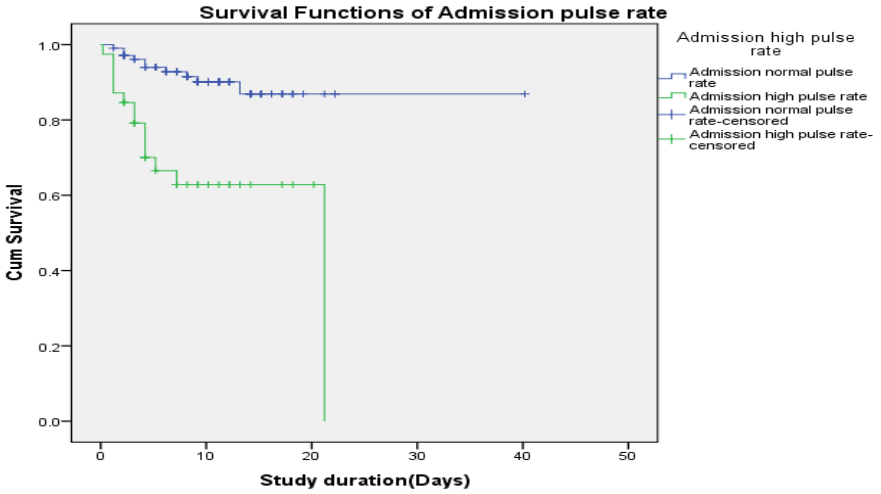

Figure 10) Survival functions of high pulse rate

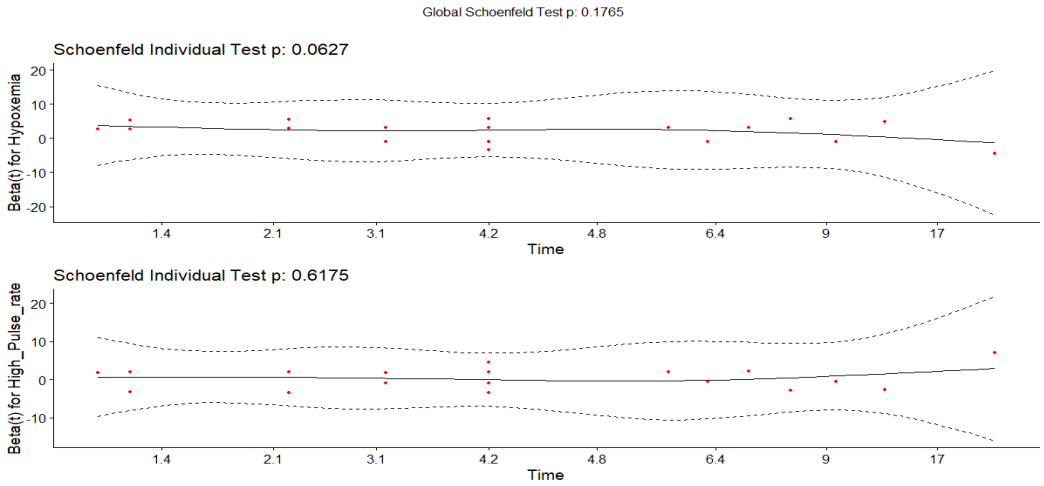

Figure11) SRD plots for clinical evaluation covariates

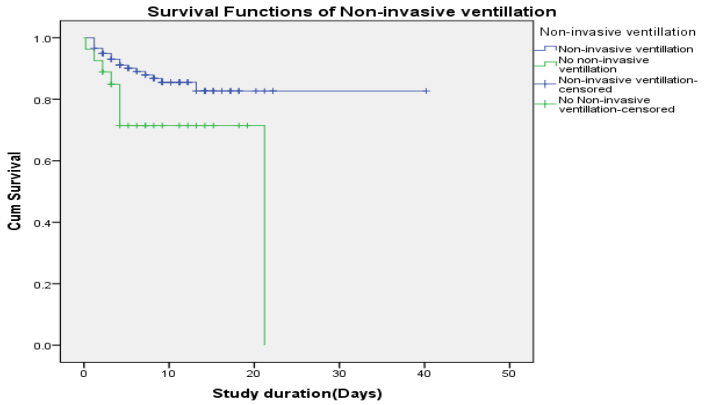

Figure 12) Survival functions of noninvasive ventilation treatment

Table 3) CRM outcomes of covariates associated with clinical evaluations, treatments provided, and medications administered in hospitalized COVID-19 patients

	Covariate	CRM Coefficient and S. E		*P*-Value	CHR	95% CI for CHR	
		$\hat{\beta}_i$	S, E ($\hat{\beta}_i$)			Lower	Upper
Clinical evaluations	Admission hypoxemia	-0.050	0.014	< .001**	0.951	0.926	0.977
	Admission high pulse rate	0.105	0.056	.046*	1.111	1.032	1.239
Treatments provided	Noninvasive ventilation treatment	-2.838	0.740	< .001**	0.059	0.014	0.250
Medications given	Anticoagulant agents	-1.254	0.495	.011*	0.285	0.108	0.753
	Anti-inflammatory agents	-1.679	0.786	.033*	0.187	0.040	0.870
Supplements provided	Multivitamin or Vitamin C	2.223	0.740	.003**	9.239	2.167	39.39

(** and * indicate that approximated regression coefficient is significant at the significance level of 1% and 5%, respectively).

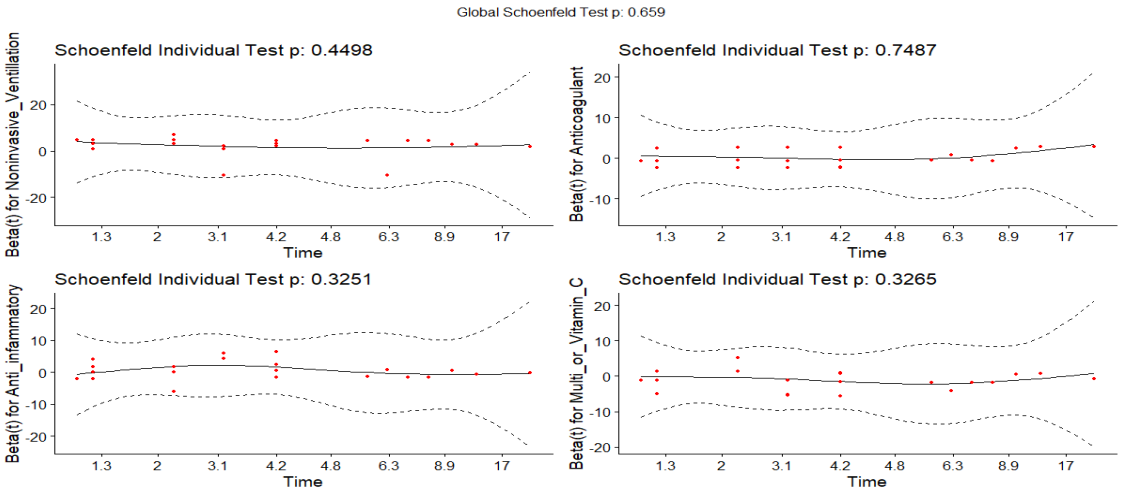

Figure 13) SRD plots for covariates including treatments provided, medications given, and supplements provided

PHA violations in CRM.

Clinical evaluation covariates: It should be noted that patients with a high pulse rate compared to those with a normal pulse rate had a higher chance of death with a CHR value of 1.111 (CHR CI: 1.032 to 1.239). Also, it was observed that the risk of death was higher in patients with hypoxemia compared to those without hypoxemia, with a CHR value of 0.951 (CHR 95% CI: 0.926 to 0.977. However, body mass index (BMI) showed no significant effect on survival rate. The survival functions of hypoxemia and high pulse rate indices are presented in Figures 9 and 10**,** respectively.

No violation in PHA was confirmed using SRD for clinical evaluations, and the results are presented in Fig. 11**.**

Covariates associated with treatments provided/ medications given/ supplements provided: The CHR value for patients with noninvasive ventilation therapy was 0.059 (CHR CI: 0.014 to 0.250). The importance of noninvasive ventilation treatment in the survival of COVID-19 patients is shown in Fig. 12. The CHR values for patients receiving anticoagulant (CHR CI: 0.108 to 0.713) and anti-inflammatory (CHR CI: 0.040 to 0.870) agents were 0.285 and 0.187, respectively. The CHR value obtained for multivitamin or vitamin C was equal to 9.239 (CHR CI: 2.167 to 39.39), indicating that patients treated with multivitamin or vitamin C had a lower risk of mortality (Table 3).

PHA violations were examined for covariates including treatments provided, medications given, and supplements provided using SRD plots. The plots are shown in Fig. 13, which clearly indicate that there is no breach for PHA in CRM.

Discussion

K-M curves graphically depict the significant survival functions of important explanatory variables as estimated from time-to-event data. LRT was used in this study to compare survival distributions in K-M curves.

SRD was used as a goodness-of-fit test to monitor for any PHA violations in CRM. SRD plots were generated for all covariates to validate the assumption. It should be noted that SRD plots for all covariates strongly supported non-violated PHA. The significance of including and excluding covariates in CRM was evaluated using the likelihood ratio test to improve CRM. Clinical assessments at admission time, patients' temperature, respiratory rate, blood pressure, pulse rate, oxygen level, and total blood count were collected in the study in

order to determine the significance of the hazard.

For each of the covariates, K-M curves graphically presented the significant survival functions of the significant groups as estimated from time-to-event data. It was observed that with each one year increase in age, the chance of death increased by 4.5%. A positive CRM coefficient was estimated for gender, which pointed out that males had an increased log hazard compared to females. The following signs were observed in patients at the time of hospitalization: anorexia, angina, shivers, cough, diarrhea, fatigue, cephalalgia, arthralgia, nausea, rhinorrhea, dyspnea, pharyngitis, ageusia or anosmia, and wheezing. The results showed that anorexia, dyspnea, and ageusia or anosmia were significantly associated with mortality.

Some of the patients' medical histories were evaluated to check their association with mortality, including asthma, cardiac disease, diabetes, HIV positivity, chronic obstructive pulmonary disease, and tuberculosis. It was found that the history of diabetes and tuberculosis had a significant impact on the death rate. The results of the following clinical evaluations were collected: temperature, respiratory rate, blood pressure, pulse rate, oxygen level, and complete blood count. Several statistical techniques were developed to analyze these results. In addition, the analysis results showed that only hypoxemia and high pulse rates had a considerable impact on death rates.

It should be noted that body mass index (BMI) had no substantial impact on survival rates. The following treatments were provided to hospitalize COVID-19 patients: oral rehydration solutions, noninvasive ventilation, and intravenous therapy. Noninvasive ventilation therapy was found to have a significant effect on survival. Anticoagulant, anti-inflammatory, antimalarial, and antipyretic agents, steroids, vasopressors, and other medications were given to patients. Treatment with anticoagulant and anti-inflammatory agents was found to have a significant impact on reducing death rates; in addition, supplements like multivitamin, vitamin C, vitamin A, and zinc each alone were given to patients during the study period. Multivitamin or vitamin C supplements were found to have a substantial impact on patient survival.

Conclusion

By developing several statistical techniques, the data were analyzed. It was shown that the chance of survival of hospitalized COVID-19 patients depended on such characteristics as demographics, observed signs, and medical histories. Many variables were analyzed in this study, and some of them were found to be significant. In addition, the analysis results showed that the covariates such as age, gender, anorexia, ageusia or anosmia, dyspnea, history of diabetes, history of tuberculosis, hypoxemia, high pulse rate, noninvasive ventilation, anticoagulant and anti-inflammatory agents, and multivitamin/vitamin C were significantly associated with survival rates. Based on the analysis done in this study, the following conclusions were obtained:

- Demographic covariates like age and gender are significantly associated with higher mortality rates.
- Signs of anorexia, ageusia, or anosmia are significantly associated with higher mortality rates.
- Medical history of diabetes and tuberculosis is significantly associated with increased mortality rate.
- Among clinical evaluations, hypoxemia and high pulse rates significantly affect the death rate.
- Noninvasive ventilation treatment is

significantly influential in reducing death rates.
- Treatment with anti-inflammatory and anticoagulant agents is significantly effective in lowering death rates.
- Supplements provided, including multivitamin or vitamin C, have a significant effect on patients' survival.

Limitations of CRM: It is well-documented that if the deviation from the PHA and the frequency of the event under study are high, then the estimated parameters are highly influenced and therefore could not be interpreted. In addition, the Cox regression technique may create a false model that does not include only time-independent predictive factors when PHA is violated. In this study, due to deleting few random variables, the analysis done did not suffer from the above problems.

Future research direction: In the future, other models like generalized multiplicative models, frailty and linear transformation models, generalized log-logistic proportional hazard models, generalized CRMs, and generalized proportional hazard models based on modified partial likelihood will be proposed and studied.

Acknowledgments

The authors are grateful to the administration of Nizwa University for continuous support and encouragement to complete this research work.

Ethical permissions: None declared by authors.

Authors' contributions: SBN collected the data, designed and analyzed the study, and drafted and reviewed thoroughly the manuscript. DY analyzed the study, read, edited, drafted, and revised the manuscript.

Conflicts of interests: The authors state that they have no conflict of interest.

Fundings: None declared by authors.

Consent to participate: None declared.

References

1. Doocy S, Bollemeijer I, Leidman E, Sebushishe A, Mbong EN, Page K. Clinical progression and outcomes of patients hospitalized with COVID-19 in humanitarian settings: A prospective cohort study in South Sudan and Eastern Democratic Republic of Congo. PLOS Glob Public Health. 2022;2(10):e0000924.
2. Sepandi M, Alimohamadi Y, Esmaeilzadeh F. Estimate of the basic reproduction number for delta variant of SARS-CoV-2: A systematic review and meta-analysis. J Biostat Epidemiol. 2022;8(1):1-7.
3. Bhandari S, Tak A, Singhal S, Shukla J, Shaktawat AS, Gupta J, et al. Patient flow hynamics in hospital systems during times of COVID-19: Cox proportional hazard regression analysis. Front Public Health. 2020;8:585850.
4. Liu X, Zhang L, Sun L, Liu R. Survival analysis of the duration of rumors during the COVID-19 pandemic. BMC Public Health. 2024;24(1):519.
5. Das D, Saikia H, Bora D, Bhattacharjee D, Das J. A survival analysis approach for identifying the risk factors in time to recovery of COVID-19 patients using Cox proportional hazard model. Decis Anal J. 2022;5:100137.
6. Geng J, Haq S, Abbas J, Ye H, Shahbaz P, Abbas A, et al. Survival in pandemic times: Managing energy efficiency, food diversity, and sustainable practices of nutrient intake amid COVID-19 crisis. Front Environ Sci. 2022;10:1–16.
7. Abbas J. Crisis management, global healthcare challenges and opportunities: The intersection of the COVID-19 pandemic and global mental health. Res Glob. 2021;3:100037.
8. Persson I. Essays on the assumption of proportional hazards in Cox regression. Sweden: Uppsala Uviversity Library; 2002.
9. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and Cox regression. J Mach Learn Res. 2019;20(129):1–30.
10. Jullum M, Hjort NL. What price semiparametric Cox regression? Lifetime Data Anal. 2019;25(3):406–38.
11. Cox DR. Partial likelihood. Biometrika. 1975;62(2):269–76.
12. Hu C, Lin DY. Cox regression with covariate measurement error. Scand J State. 2002;29(4):637-55.
13. Anderson RN, Rosenberg HM. Age standardization of death rates: Implementation of the year 2000 standard. Natl Vital Stat Rep. 1998;47(3):1-16.
14. Kurd DM, nahal AS, Noori-zadeh A, Sheikhabbasi A, Heydari F, Pakzad I, et al. Association of statin therapy on clinical outcomes in Covid-19

patients: An updated systematic review and meta-analysis on all related evidences. J Biostat Epidemiol. 2022;8(4):407-34.
15. Van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: Basic concepts and methods of Cox regression. Kidney Int. 2008;74(6):705–9.
16. Benítez-Parejo N, del Águila MM, Pérez-Vicente S. Survival analysis and Cox regression. Allergol Immunopathol. 2011;39(6):362–73.
17. Faraggi D, Simon R. Bayesian variable selection method for censored survival data. Biometrics. 1998;54(4):1475–85.
18. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health. 2020;17(5):1729.
19. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271-80.
20. Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: A comprehensive review. J Intern Med. 2020;288(2):192-206.
21. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res. 2020;24:91-8.
22. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet. 2020;395(10229):1054–62.
23. Atlam M, Torkey H, El-Fishawy N, Salem H. Coronavirus disease 2019 (COVID-19): Survival analysis using deep learning and Cox regression model. Pattern Anal Appl. 2021;24(3):993–1005.
24. Gaskin DJ, Zare H, Delarmente BA. Geographic disparities in COVID-19 infections and deaths: The role of transportation. Transp Policy. 2021;102:35-46.
25. Ghadamgahi F, Tapak L, Bashirian S, Amiri R, Roshanaei G. The effect of underlying diabetes disease on clinical outcome and survival in patients with Covid-19: A propensity score matching study. J Diabetes Metab Disord. 2021;20(2):1675-83.
26. Fung KW, Baik SH, Baye F, Zheng Z, Huser V, McDonald CJ. Effect of common maintenance drugs on the risk and severity of COVID-19 in elderly patients. PLoS One. 2022;17(4):e0266922.
27. Huang H, Chen J, Fang S, Chen X, Pan X, Lei H, et al. Association between previous stroke and severe COVID-19: A retrospective cohort study and an overall review of meta-analysis. Front Neurol. 2022;13:922936.
28. Seif M, Sharafi M, Ghaem H, Kasraei F. Factors associated with survival of Iranian patients with COVID-19: Comparison of Cox regression and mixture cure model. Trop Dis Travel Med Vaccines. 2022;8(1):4.
29. Wastnedge EA, Reynolds RM, van Boeckel SR, Stock SJ, Denison FC, Maybin JA, et al. Pregnancy and COVID-19. Physiol Rev. 2021;101(1):303-18.
30. Domjanović J, Domjanović Škopinić T, Matetic A. Association of different risk scores and 30-day mortality in kidney transplant recipients with COVID-19. Medicina. 2023;59(4):657.
31. Hastie CE, Foster HM, Jani BD, O'Donnell CA, Ho FK, Pell JP, et al. Chronic pain and COVID-19 hospitalisation and mortality: A UK Biobank cohort study. Pain. 2023;164(1):84-90.
32. Kojima K, Yoon H, Okishio K, Tsuyuguchi K. Increased lactate dehydrogenase reflects the progression of COVID-19 pneumonia on chest computed tomography and predicts subsequent severe disease. Sci Rep. 2023;13(1):1012.
33. Lund LC, Støvring H, Pottegård A, Andersen M, Hallas J. Cox regression using a calendar time scale was unbiased in simulations of COVID-19 vaccine effectiveness & safety. J Clin Epidemiol. 2023;156:127-36.
34. Bitew ZW, Ayele EG, Worku T, Alebel A, Alemu A, Worku F, et al. Determinants of mortality among under-five children admitted with severe acute malnutrition in Addis Ababa, Ethiopia. Nutr J. 2021;20(1):1-5.
35. Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika. 1982;69(1):239–41.
36. Park S, Hendry DJ. Reassessing Schoenfeld residual tests of proportional hazards in political science event history analyses. Am J Pol Sci. 2015;59(4):1072-87.
37. Box-Steffensmeier JM, Jones BS. Event history modeling: A guide for social scientists. Cambridge University Press; 2004.
38. Ozenne B, Sørensen AL, Scheike T, Torp-Pedersen C, Gerds TA. riskRegression: Predicting the risk of an event using Cox regression models. R J. 2017;9(2):440–60.
39. Grønnesby JK, Borgan Ø. A method for checking regression models in survival analysis based on the risk score. Lifetime Data Anal. 1996;2(4):315-28.
40. Li H, Luan Y. Kernel Cox regression models for linking gene expression profiles to censored survival data. Pac Symp Biocomput. 2003;76:65–76.
41. Sargent DJ. A flexible approach to time-varying

coefficients in the Cox regression setting. Lifetime Data Anal. 1997;3:13-25.
42. Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics. 2005;21(13):3001–8.