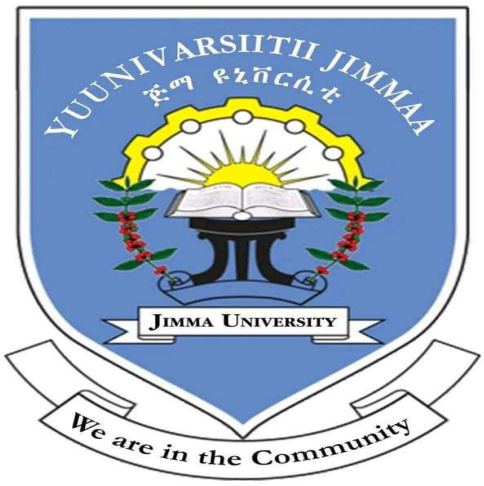

College of Natural Science

Department of Statistics

Time-to-death analysis of COVID-19 pandemic patients: A case study of Jimma Zone, Southwest Ethiopia.

By:

Getahun Taye.

A Thesis Submitted to the Department of Statistics in the College of Natural Sciences at Jimma University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biostatistics.

December,2022.

Jimma, Ethiopia.

Time-to-death analysis of COVID-19 pandemic patients: A case study of Jimma Zone, Southwest Ethiopia.

By:
Getahu Tye

Advisor: Sisay Wondaya(PhD)

Co-advisor: Gurmessa Nugussu(MSc.)

STATEMENT OF THE AUTHOR

I hereby certify that the thesis is my original work and that all sources of information utilized in the thesis have been properly cited in this document. No other institution has accepted this thesis for the purpose of awarding a diploma, certificate, or academic degree. As a partial fulfillment of the criteria for the Master of Science in Biostatistics, I have submitted this thesis to Jimma University. The thesis can be placed in the university library so that patrons can use it as a resource. This thesis may be briefly quoted without seeking special permission, so long as the source is correctly acknowledged. When, in her or his opinion, the planned use of the information is for academic interest, the director of the department of statistics may approve requirements for extensive quotes for the replication of the thesis in whole or in part. But in every other circumstance, the author's consent is required.

Getahun Taye ———————— ————————

Name Signature Date

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

MSc THESIS APPROVAL SHEET

This is to certify that the thesis titled " **Time-to-Death Analysis of COVID-19 Pandemic Patients: A Case Study of Jimma Zone, Southwest Ethiopia.**" has been submitted in partial fulfillment of the requirement for the degree of Master of Science in Biostatistics to the college of natural science at Jimma University, and is a record of original research carried out by Getahun Taye, under our supervisions, and no part of the thesis has been submitted for another degree or diploma. The assistance and the help received during the course of this investigation have been duly acknowledged. Therefore, we recommend that it be accepted as fulfilling the thesis requirement.

Sisay Wondaya (PhD) — Signature — 02/01/2023 — Date
Name of Major Advisor

Gurmessa Nugussu (M.Sc) — Signature — 02/01/2023 — Date
Name of Co-advisor Advisor

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics

Dr. Shibru Temesgen — Signature — 29/12/2022 — Date
Name of the External Examiner

Geremew Muleta (M.Sc) — Signature — 13-01-2023 — Date
Name of the Internal Examiner

Kibrealem Sisay (MSc) — Signature — 13/01/2023 — Date
Name of the Chairperson

Acknowledgment

Above all, I would like to thank the almighty God for the gift of health, wisdom, and strength throughout any steps for the achievement of my thesis and more of my entire life

I want to express a sincere acknowledgment to my advisors Dr. Sisay Wondaya(PhD.) and Mr. Gurmesa Nugussu(MSc.) for their constructive comments and guidance throughout writing this thesis that contributed to the success.

I take this opportunity to sincerely express my gratitude to my beloved family who is sources of pride and encouragement throughout my life. I am thankful to my mother, Bekelech G/Mariyam and my father, Taye Enkusilasie, for their encouragement and prayer for the success of this work.

Finally, I would like to acknowledge Jimma University and all my instructors and staffs of Statistics, Department at Jimma University for their cooperation from starting title selection up to the completion of this thesis and my special appreciation and acknowledgment goes to Bonga University for offering me the opportunity and financial support during my study at Jimma University.

Abstract

Introduction: *COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus-2. It was first reported in December 2019 in Wuhan, China, and has since spread to pandemic proportions. Since then, the virus has rapidly spread to the world and has caused over 637.351 million confirmed cases, more than 6,604 million deaths, and more than 616.952 million recoveries worldwide as of November 05, 2022. The accelerated failure time model which is an alternative to the proportional hazard model when the proportional hazard assumptions doesn't hold was used to analyze time of event, death from COVID-19 pandemic.*

Objective*: This study aimed to analyze the time-to-death of COVID-19 pandemic patients in Jimma Zone, southwest Ethiopia.*

Methodology: *A retrospective cohort study was conducted on 809 COVID-19 patients who admitted to Jimma university medical center and Shenen gibe generalized hospital from May 16, 2020 to March 9, 2022 in Jimma Zone, southwest Ethiopia. Kaplan-Meier plots and Log-Rank test were used to compare the survival experience of different categories and semi-parametric survival model and acceleration failure time models were employed to identify survival time of the patients. The performances of acceleration failure time models were compared using Akakie Information Criteria.*

Results: *From 809 patients, 135(16.7%) died in the follow-up period. Log-logistic acceleration failure time model is better fit the data than other models. The result of this model shows that the survival time of COVID-19 patients significantly affected by age, comorbidity, status at admission, HIV/AIDS, symptom at admission, intranasal oxygen use and diabetes.*

Conclusions: *The AFT model is a more valuable and realistic alternative to the Cox PH model in situations where PH assumption cannot hold and therefore should be considered as an alternative to the Cox PH for analyzing the time to death of COVID-19 patients. Older age, comorbidity, moderate or severe status at admission, HIV/AIDS, being asymptomatic at admission, intranasal oxygen use, and diabetes are factors that accelerate time to death in COVID-19 patients.*

Keywords: Loglogistic, Accelerated failure time, COVID-19 Pandemic.

Abbreviation and Acronym

AFT:	Accelerated-Failure-Time
AIC:	Akakie Informaion Criteria
COVID-19:	Corona Virus Disease-19
Cox PH:	Cox Proportional Hazard
EPHI:	Etiopian Public Health Institute
HR:	Hazard Ratio
ICU:	Intensive Care Unit
IQR:	Inter Quartile Range
JUMC-MTC:	Jimma University Medical Center-Michu Trearmet Center
KM:	Kaplan Meir
MLE:	Maximum Likilihood Estimate
PL:	Partial Likilihood
WHO:	World Health Organization
SARS:	Severe Acute Respiratory Syndrome
SARS-CoV-2:	Severe Acute Respiratory Syndrome Coronavirus-2
SHGGH-OTC:	Shenenn Gibe Generalized Hospital-Oromia Treatment Center

Contents

List of Figures

List of Tables

CHAPTER ONE

1 Introduction

1.1 Background of the Study

Coronavirus disease-19 (COVID-19) is an infectious illness caused by the SARS-CoV-2, which was initially diagnosed in December 2019 in Wuhan, China, and has subsequently grown to pandemic proportions (Rojas & Rodríguez, 2020). Most patients infected with the virus will have mild to severe respiratory sickness and will recover without needing any specific therapy. Some, though, will get very ill and require medical treatment. The virus has swiftly spread over the world as of November 05, 2022, resulting in more than 637.351 million confirmed cases, more than 6.604 million fatalities, and more than 616.952 million recoveries (Covid, 19).

Almost all African countries have been affected by the pandemic, since the first confirmed case were reported in Egypt on 14th February 2020 (Gilbert et al., 2020). On November 05, 2022, more than 12.679 million confirmed COVID-19 cases with 257,934 deaths, and more than 12.006 million recoveries have been reported from Africa (Covid, 19). Ethiopia reported its first confirmed case of COVID-19 on 13th March 2020 (Kebede & Ababor, 2020). As of November 05, 2022, the total number of confirmed COVID-19 cases in Ethiopia has reached 494,024 with 7,572 deaths and 472,117 recoveries (Covid, 19). Within less than three months after the first case of COVID-19, the virus has quickly spread to all parts of the country. By the first week of June 2020, all regions reported COVID-19 cases, with Addis Ababa and Oromia constituting about 75% and 6% of the cases, respectively (Zikargae, 2020).

The proportional hazards model (Cox) as a semi-parametric approach Cox (1972) and the accelerated failure time model or linear model representation in log time as a parametric model are the two main regression models used for survival data. The baseline hazard function's type or form is not assumed by the model, which instead assumes

that the underlying hazard rate is a function of the independent variables. Due to this, Cox's model is known as a semi-parametric model for the hazard function (D. Kleinbaum & Klein, 2005; Klein & Moeschberger, 1997). The baseline hazard is maintained in this model as an arbitrary, unspecified, and non negative function of time. Because of its simplicity and lack of assumptions about the survival distribution, it is the most widely used and well-known model among researchers in the medical sciences (Therneau & Grambsch, 2000).

A different approach for analyzing survival data is the Accelerated Failure Time (AFT) model. Numerous common parametric models, including the Weibull, exponential, log-normal, and log logistic models, are accelerated failure time models (D. Kleinbaum & Klein, 2005). Parametric methods Andersen et al. (1993) do have a number of advantages, even if the Cox regression model is the most often used methodology in survival analysis.

To develop effective and efficient preventive and treatment measures, it is crucial to have a depth sense of the epidemiological and clinical developments of COVID-19. In environments with few health care resources, this kind of proof was very important. When the epidemic first started in Ethiopia, for instance, every patient who tested positive for COVID-19 was isolated and monitored until they recovered. However, as the number of cases grew, the admission and release standards were altered to accommodate individuals who needed the most service (Ababa, 2003). Sadly, there aren't many studies on COVID-19's epidemiological and clinical development that examine survival rates, patient traits, and risk factors for critical illness and mortality in low-income settings (Kaso & Agero, 2022). The overall goal of this study is to identify factors that substantially impact the survival time of COVID-19 patients and to establish the AFT model if the assumption of the PH model fails in the analysis of time-to-death of COVID-19 patient data.

1.2 Statement of the Problem

The threat facing the entire world is the COVID-19 pandemic, which started in Wuhan City (Rojas & Rodríguez, 2020). It has spread rapidly worldwide, causing major public health concerns and economic crises having a massive impact on populations, economies, and thereby placing an extra burden on health systems around the planet (Villela, 2020; Iboi et al., 2020). Despite unrelenting global efforts to contain the spread of COVID-19, the pandemic caused unprecedented crises to the world (Roychowdhury, 2020).

Short of the most lethal weapon for prevention and treatment, every country has resorted to a trial-and-error approach to keep the balance between safeguarding the health of its citizens and saving the economy (Moti & Ter Goon, 2020). This virus has caused unprecedented morbidity and deaths mainly among older age people with underlying health conditions (Alqahtani & Oyelade, 2020). Using the right model will undoubtedly assist to uncover more accurate and reliable prognostic factors, which will help to have a more successful treatment protocol. There are two main methods for the regression analysis of censored data, such as the accelerated failure time and the proportional hazard models (Cox, 1972). According to a review of the literature, non-parametric techniques like Kaplan-Meier and Cox Proportional Hazards Model have frequently been used to compare and pinpoint the variables that affect patients with the COVID-19 pandemic's survival time. The latter technique is used when the effect of the covariate on the hazard ratio is desired.

In practice, the proportional hazards model has virtually always been employed. This is most likely caused by the fact that enables parameter estimation and inference without requiring a survival time distribution. However, the basic and most important assumption underlying this model is the assumption of proportional hazard rates, which may not be held in some situations. Where the PH assumption is not met, it is improper to use the usual standard Cox PH model as it may entail serious bias and loss of power when estimating or making inferences about the effect of predictors of mortality

(Moran & Bersten, 2008).

Recent years have seen a significant increase in interest in AFT models as parametric models, not only because they do not require the PH assumption but also because they can be easily tested and parameterized using common methods like Maximum Likelihood (Altman et al., 1995). The rare instances of AFT models being employed may be found in studies of the period to menarche and kidney transplantation (Fagbamigbe et al., 2018). The predictors of death of COVID-19 patients worldwide have not been identified using this method. In this study, AFT models were used to analyze and find factors that have a statistically significant influence on the survival of patients with COVID-19 since the assumption of PH fails. Thus, we are attempted to offer empirical support for or, in the instance of Jimma Zone, a response to the following study questions:

- What are the determinant factors to time-to-death of COVID-19 patients in the study area?
- Which model best fit the time to death analysis of COVID-19 patients data?

1.3 Objectives of the study

1.3.1 General objective

The general objective of this study is to analyze time-to-death of COVID-19 patients and identify the main factors that affect the survival time of COVID-19 patients in Jmma zone, southwest Ethiopia.

1.3.2 Specific objectives

- To estimate the survival time and compare the survival of COVID-19 patients among groups.
- To estimate the effects of the covariate on acceleration or deceleration of the survival time of COVID-19 patients.

1.4 Significance of the study

The findings or results obtained from this research are useful in many ways;

- Governmental organizations could take intervention measures and set appropriate plans to reduce mortality by giving priority in significant predictors of mortality among COVID-19 pandemic patients.
- It helps in making a decision as to which model to apply under specified conditions defined by predictor variables.
- It would have added literature on determinants of time-to-death from COVID-19 pandemic.

1.5 Limitation of the study

- The study was conducted based on secondary data, so it may contain incomplete information.
- In this study, all deaths were assumed to be caused by COVID-19.
- The study is based on baseline values of the variables of interest.

CHAPTER TWO

2 Literature Review

2.1 Coronavirus disease-19 (COVID-19)

COVID-19 is a significant global health emergency that has infected and killed millions of people worldwide (Habenom et al., 2022). The World Health Organization declared the disease a pandemic on March 11, 2020 (Ghebreyesus, n.d.). This pandemic is widespread and crosses international borders, affecting many people (Porta, 2014). It has been reported that the COVID-19 incident in Ethiopia happened in Addis Ababa, the capital city (Kassaw, 2020). Quarantine centers and international destinations are said to be the fastest-spreading places for diseases in Ethiopia. In addition, the partial closure in Addis Ababa worsened the situation.

2.2 Literature in relation to explanatory variables of the study

The most common pathological symptom observed among COVID-19 infected individual is that the virus damages the alveolar, which leads to a respiratory failure and as normal like flu, fever, cold, cough and shortness of breath, along with them the other severe symptoms observed are sputum production, haemoptysis, lymphophenia and pneumonia, in some cases increasing dyspnea and hypoxemia in the upper lobe of the lung were also observed (Rothan & Byrareddy, 2020). The silent feature of COVID-19 is its associated symptoms that they will appear during incubation period of 2–14 days (Rothan & Byrareddy, 2020). COVID-19 can infect individuals of all ages and genders and can spread easily from one person to another, but the likelihood of getting infected is higher among older population, on various medical conditions, such as, diabetes, cardiovascular diseases, hypertension, cancer and chronic respiratory diseases (Wu & McGoogan, 2020; Wang & Zhang, 2020). Severe illness due to the COVID-19 leads to death (mortality rate of 3% approximately) (Wang & Zhang, 2020).

Among COVID-19 cases reported in the WHO African area between March 21 and

October 31, 2020, the study done in Africa reveals the risk variables linked to death. Researchers determined the median time to death for the major risk variables and the average hazard ratios of death using weighted Cox regression. The study comprised 46,870 cases that were verified and reported by eight regional Member States. Using 803 deaths and 3 959 874 person-days of pertinent observation, an incidence of 20.06 per 100,000 people was calculated overall. Male sex (aHR 1.54 (95% CI 1.31-1.81)), older age (aHR 1.08 (95% CI 1.07-1.08)), residents of urban areas (aHR 1.42 (95% CI 1.22-1.65)), and those with one or more comorbidities (aHR 36.37 (95% CI 20.26-65.27) had substantially shorter time to death with $P < 0.001$. Although possibly understated, the COVID-19 mortality in the African area is comparable to that in other locations (Impouma et al., 2022).

The study conducted by Kundu et al. (2021) presents a survival analysis to determine the variation in survivorship of COVID-19 patients in India by age group and sex at various levels, including the national, state, and district levels. A total of 26,815 patients were included in the sample. The log rank test ($P < 0.001$) and Wilcoxon test ($P < 0.001$) were added to the Kaplan-Meier survival function to compare the survival functions and revealed a drop in the likelihood of survival for COVID-19 patients throughout the course of the 5-month research period. All of the survival estimates show that the age groups showed significant variation, and that the chance of dying from COVID-19 rose as age. The Cox proportional hazard model confirmed that male COVID-19 patients had a 1.14 times greater chance of passing away than female patients (Hazard ratio 1.14; SE 0.11; 95 percent confidence interval (0.93-1.38); While Eastern, North Eastern, and Southern India displayed somewhat improved outcomes in terms of survival, Western and Central India showed declining survival rates within the defined time period.

The COVID-19 survival analysis in the Mexican population was the subject of the other study, which was carried out by Salinas-Escudero & Carrillo-Vega (2020). The register of 16,752 confirmed COVID-19 cases with a mean age of 46.55 $\pm$ years, 58.02 percent males (n = 9719), and 9.37 percent deaths (n = 1569) was included in the

analysis. The risk of dying from COVID-19 was independently increased by male sex, advanced age, chronic renal disease, pneumonia, hospitalization, intensive care unit admission, incubation, and medical attention from public health services ($p < 0.001$).

Using Kaplan-Meier and Cox regression analysis, Kaso & Agero (2022) conducted a study on the survival analysis of COVID-19 patients in Ethiopia. A total of 422 COVID-19 patients who received treatment were examined; of them, 11.14 percent (or 6.35 cases per 1000 person-days) deceased. The majority (87.2%) of fatalities (with a median time-to-death of nine; IQR: 8-12) days) happened within the first 14 days of admission. Patients with HIV/AIDS (aHR = 3.66, 95%CI [1.20, 11.10]), age between 31 and 45 years (aHR = 2.55; 95%CI [1.03, 6.34]), older than 46 years (aHR = 2.59 95%CI [1.27, 5.30]), chronic obstructive pulmonary disease (aHR = 4.60, 95%CI [2.37, 8.91]), Chronic kidney disease (aHR = 5.58, 95%CI [1.70, 18.37]), admission to the Intensive care unit(aHR = 7.44, 95%CI [1.82, 30.42]), and being on intranasal oxygen care (aHR = 6.27, 95% CI [2.75, 4.30]), were independent risk factors increasing risk of death from COVID-19 disease than their counterparts.

A retrospective cohort study conducted on incidence and predictors of mortality among patients admitted with COVID-19 at Wollega University Referral Hospital (WURH), in western Ethiopia, A total of 318 patients were included in the final analysis, with a mean age of 44 (SD ± 16.7) years and a 67.9% male gender distribution. At the time of admission, more than half of patients (55,7%) were comorbidity-free. 259 (81.45%) of the patients recovered from COVID-19, and 267 (84%) were censored at the conclusion of the follow-up. The mortality incidence rate was 14.1 per 1000 (95% CI: 10.7-18.5) observational person days. Three factors were independent predictors of death among COVID-19 patients: age 59 years (HR: 5.76, 95% CI: 2.58, 12.84), low oxygen saturation (HR: 2.34, 95% CI: 2.34, 4.17), and delayed presentation (HR: 5.60, 95% CI: 2.97, 10.56) (Tolossa et al., 2022).

2.3 Cox PH versus AFT models

The effects of variables on a possibly censored response variable may be naturally expressed using the accelerated failure time model (Zeng & Lin, 2007). The semi-parametric estimators that are now available are statistically ineffective and difficult to compute. In their study, they suggested a roughly non-parametric maximum likelihood approach for the accelerated failure time model with potential time-dependent variables. By maximizing a kernel-smoothed profile likelihood function, the regression parameters were computed. Using traditional gradient-based search techniques, maximization was accomplished. The obtained estimators were asymptotically normal and consistent. The semi-parametric efficiency constraint was reached by the limiting covariance matrix, allowing for reliable estimation. Additionally, they offer a reliable estimate for the error distribution. Numerous simulations showed that the new estimators outperformed the previous ones in terms of efficiency, while the asymptotic approximations were correct in real-world settings.

The proportional hazard model and its extension were extensively employed in Ponnuraja & Venkatesan (2010) study to evaluate an intervention's impact in the presence of con-founders. They noticed that the assumptions might not hold in scenarios when the intervention accelerates poverty, hence the AFT model is also suitable in these circumstances. The goal of their research was to create a model that produces physiologically reasonable and understandable estimates of the impact of key factors on survival time. It was discovered that the AFT model provided more accurate predictions than the Cox PH model.

The proportional supposition is verified in their investigation and shown to be accurate; however, the model diagnostic for the parametric situation has not yet been established. Based on AIC, parametric and semi-parametric models were compared. The study demonstrates that there cannot be a single model in univariate analysis that is significantly superior to others. The lognormal regression was one of the parametric models that the data most strongly supported, and it can produce findings that are

more accurate than those produced by the Cox PH model.

According to the studies above, the time to death of a COVID-19-infected patient after registering for hospitalization is a function of baseline variables like age, gender, comorbidity, status at admission, symptom at admission, and so on. The purpose of this study is to analyze time-to-death of COVID-19 pandemic patients using accelerated failure time models and to investigate determinant factors for survival time of COVID-19 patients in Jimma Zone, southwest Ethiopia.

CHAPTER THREE

3 Data and Methodology

3.1 Study area

Jimma Zone is one of Oromia's regional zones, located southwest of Ethiopia. It has the most comprehensive and beautiful topography, surrounded by green areas. It has 21 districts with two referral COVID-19 care centers. The study area was around Jimma Town, which is the capital and administrative center of the zone and is located at a distance of 350 km from the capital of Ethiopia, Addis Ababa. The study area is between 1689 and 3018 meters above sea level and receives an average rainfall of between 1200 and 2400 mm per year.

3.2 Study design

This study used a retrospective cohort study design to analyze survival of COVID-19 patients that was recorded in the two treatment centers.

3.3 Data source

Data is collected from the patient's follow-up at Jimma University Medical Center-Michu Treatment Center (JUMC-MTC) and Shenen Gibe Generalized Hospital-Oromia Treatment Center (SHGGH-OTC). These two treatment centers are among public hospitals in Ethiopia, and they belong to the Jimma administrative region. Currently, both hospitals deliver COVID-19 care centers.

3.4 Inclusion and Exclusion Criteria

All COVID-19 patients registered for follow-up in SHGH-OTC and JUMC-MTC in Jimma Zone from May 16, 2020, to March 9, 2022, were included in this study, otherwise excluded. The COVID-19 data sets used in this thesis were extracted from patient

cards and pertain to patients admitted between May 16, 2020, and March 9, 2022. This study included 809 patients from 21 woredas in and around the Jimma zone.

3.5 Study Variables

3.5.1 Dependent Variable

The outcome of interest for this study is time to death since a patient hospitalized for COVID-19. The status of respondents were assigned to "1" corresponding to when the subject had developed an event (death) and "0" if they were cured, clinical improved, and discharged with consent or transferred to out of the study area. This dependent variable (in days) is measured as the length of time from treatment start date until the date of death or censor.

3.5.2 Independent Variable

The explanatory variables included in the study are factors that are assumed to affect the survival time of patients. The socio demographic factors, health-related factors, comorbidity conditions, clinical manifestation, and treatment-related factors are considered. The description of explanatory variables are given in table 3.1.

3.6 Ethical issues

The ethical clearance and permission was obtained from Research Ethical Review Board of college of Natural Sciences, Jimma, before starting data collection. An ethical clearance letter from Jimma University would be given to Manager of the Jimma Emergency Operation Center (JEOC) and Shenen Gibe General Hospital (SHGGH) and a permission letter was obtained. The researcher collected: (1) Patient history from record (hard/electronic sources) only by a trained health professional assigned by the concerned institute. (2) All data collected were treated with maximum confidentiality; the identity of the respondents/patients was never be expose to anyone at any time by any means. (3) The information (data) was never be used for any other purpose than for the scientific goal and was never be transferred to any third party with identity of the respondents/patents.

Table 3.1: Description and codes of the explanatory variables

Variables	Description	Code
Sex	Sex of patients	0=Female, 1=Male
Age	Age of patients	0=< 18,1=18-45, 2= > 45
Residence	Residence of Patients	0=Rural,1= Urban
Comorbidity	Comorbidity of patients	0= No, 1= Yes
Status at admission	Patients status at admission	0=Asymptomatic, 1= Mild, 2= Moderate, 3= Sever
HIV/AIDS	Patients having HIV/AIDS	0= No, 1= Yes
Symptom at admission	Patients having Symptom	0= Yes, 1= No
Hypertension	Patients having Hypertension	0= No, 1= Yes
Oxygen use	Patients used intranasal oxygen	0= No, 1= Yes
Diabetes	patients having diabetes	0= No, 1= Yes

3.7 Methods of Data Analysis

3.7.1 Survival Data Analysis

Basic survival analysis: In follow-up studies the exact survival time is only known for those study participants or units who show the event of interest during the follow-up period. For the others, what one can say is that they did not experience the event of interest during the follow-up period. These study participants or units are called censored observations. Individuals can be right censored, left censored or interval censored.

Censoring: Subjects are censored to the right if it is known that the relevant event occurred some time after the follow-up time that was recorded, and to the left if it is known that the relevant event occurred some time before the follow-up time that was recorded. When the actual time at which the event happened is unknown but an interval bounded by this time is known, this is known as interval censoring. It is typical to overlook this type of censorship when the period is relatively brief and choose one

end point consistently. Clinical trials and follow-up studies, including those on AIDS and cancer, commonly produce interval-censored survival data.

3.7.2 Descriptive Statistics

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. In summarizing survival data, the two common functions applied are the survivor function and the hazard function (Lemeshow et al., 2011).

3.6.2.1 Survival Function: The survivor function is the probability that the survival time of a randomly selected subject is greater than or equal to some specified time. Thus, it gives the probability of an individual surviving beyond a specified time. The distribution of survival time is characterized by survivor-ship, probability density, and hazard function. Let T be a random variable associated with the survival times, t be the specified value of the random variable T, and $F(t)$ be the underlying probability density function of the survival time T. The survivor function, $S(t)$, is

$$S(t) = pr(T > t) = 1 - F(t), t \geq 0.$$

Where $F(t) = pr(T < t) = \int_0^t f(u)du, t \geq 0$ is the $c.d.f.$ of T and
The probability density function

$$F(t) = \frac{d}{dt}F(t) = -\frac{d}{dt}S(t).$$

3.6.2.2 Hazard Function: The hazard function is a measure of the risk of the event happening at any point in time. It is the instantaneous probability of having an event at time t (per unit time) given that one has survived (i.e. not had an event) up to time t (D. G. Kleinbaum et al., 2012). It is denoted by $h(t)$ and defined as

$$h(t) = \lim_{\Delta_t \to 0} \frac{P(t \leq T < t + \Delta t / T > t)}{\Delta t} = \frac{f(t)}{S(t} = -\frac{d}{dt} \ln S(t).$$

The cumulative hazard function is given by

$$H(t) = \int_0^t h(u)du = -\ln S(t),$$

which implies that,

$$S(t) = \exp[H(t)] = \exp(-\int_0^t h(u)du).$$

3.8 Non-Parametric Estimation of Survival Function

Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e, if the estimated survival functions for two groups are approximately parallel (do not cross).

3.8.1 The Kaplan-Meier Estimator

The Kaplan-Meier Estimator is a non-parametric estimator of the survival function, which is not based on the actual observed event and censoring times, but rather on the order in which events occur. This principle of non-parametric estimation of the survival function is to assign probability to and only to event failure times. The log-rank test is utilized to test whether observed differences in survival experience between the groups are significant or not.

The Kaplan-Meier survival curve used to compare the survival of COVID-19 patients under different categories of categorical covariates. In general, patients belongs to the categories whose survival curve lays below the survival curve of the other category has a better survival time. In Kaplan Meier product limit method, survival probabilities can be obtained as:

$$S = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right), k \leq n, t_j \leq t < t_{j+1}.$$

Where; d_j = the number of failure in t_j, n_j is the number of incident cases at risk in t_j, k is the number of sequential observations, n is the total number of incident cases.

3.9 Comparison of Survival Function

The Kaplan-Meier plots are used to see whether there is a difference in survival time or not between groups of covariates under investigation. But, the KM plot cannot be

used to decide whether the survival time of patients living with COVID-19 in each covariate is significantly different or not. Instead, we use log-rank test.

3.9.1 Log-rank Test

The non-parametric log rank test was created by Mantel & Haenszel (1959) to compare two or more independent survival curves. Since it is a non-parametric test, there is no need to make any assumptions about the distributional form of the data. The strongest indication of a larger cure proportion in one group compared to the other is made by this test. Let $t_1 < t_2 < ... < t_k$ be the ordered recovery times across two groups. Suppose that d_j failures occur at t_j and that r_j subjects are at risk just prior to t_j, $j = 1, 2, 3, ..., k$. Let d_{ij} and r_{ij} be the corresponding numbers in group $i, i = (1, 2)$. The log-rank test compares the observed number of recovery with the expected number of recovers for group i . Consider the null hypothesis: $S_{1t} = S_{2t}$ i.e. there is no difference between survival curves in two groups. Given r_j and d_j the random variable d_{1j} has the hyper geometric distribution.

$$\frac{\begin{pmatrix} d_j \\ d_{1j} \end{pmatrix} \begin{pmatrix} r_j - d_j \\ r_{1j} - d_{1j} \end{pmatrix}}{\begin{pmatrix} r_j \\ r_{1j} \end{pmatrix}}.$$

Under the null hypothesis, the probability of recovery at t_j does not depend on the group,i.e., the probability of recovery at t_j is $\frac{d_j}{r_j}$.

$$\chi^2_{\text{logrank}} = \frac{\left[\sum_j^k (d_{1j} - r_{1j} \times d_j/r_j)\right]^2}{\sum_j^k \frac{r_{2j} r_{1j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}},$$

this statistic approximate χ^2 distribution with 1 df.

3.10 Survival Models

3.10.1 Cox Proportional Hazard Model

The Cox PH model is a semi parametric regression model which can be used to measure the effects of covariates on the survival time. This model is represented by the relation-

ship of the hazard function, the baseline hazard function, and one or more covariates in the form

$$\lambda(t) = \lambda_0(t) \exp(\boldsymbol{\beta' X}) \tag{1}$$

where t is survival time, $\lambda(t)$ is the hazard function, $\lambda_0(t)$ is the baseline hazard function which is left unspecified, $\boldsymbol{\beta}$ is a column vector of the regression coefficients, and $\boldsymbol{X}$ is a column vector of the covariates. Cox proportional hazard model used for the analysis of time to death of COVID-19 data in the presence of covariates or prognostic factors. The corresponding survival functions are related as follows

$$S(t, X) = [S_0(t)]^{\exp(\sum_{i=1}^{p} \beta_i X_i)} \tag{2}$$

where $S_0(t) = \exp(\lambda_0(t))$ is a baseline survival function and β_i coefficient of covariates respectively. This model, also known as the Cox regression model, makes no assumptions about the form of $\lambda_0(t)$ (non-parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model. The Cox regression model assumes that there is the proportionality of the hazard rate between any two individuals in the population. The hazard ratio is defined as the ratio of the hazard functions for two subjects with different values of covariates X_1 and X_2. The formula of the hazard ratio is given by

$$\Lambda(t) = \frac{\lambda_0(t) \exp(\beta_2 X_2)}{\lambda_0(t) \exp(\beta_1 X_1)} = \frac{\exp(\beta_2 X_2)}{\exp(\beta_1 X_1)} = \exp(\beta'(X_2 - X_1)) \tag{3}$$

It can be seen that the hazard ratio $\Lambda(t) = \exp(\beta'(x_2 - x_1))$ is independent of time. In other words, the hazard ratio for any two individuals is constant over time. This property is also known as the PH assumption. Assumptions of the Cox proportional hazards model are;

- The ratio of the hazard function for two individuals with different sets of covariates does not depend on time.
- Time is measured on a continuous scale and censoring occurs randomly.

Interpreting outputs from the Cox model involves examining the coefficients for each explanatory variable. Negative regression coefficient for an explanatory variable indicates that the hazard is lower and thus the prognosis worse. Conversely, positive a

regression coefficient implies a better prognosis for patients with higher values of that variable when time to event is recovery and conversely for death (Walters, 2012).

3.10.2 Partial likelihood estimate for Cox proportional hazards model

Fitting the Cox proportional hazards model, we wish to estimate $\lambda_0(t)$ and β. One approach is to attempt to maximize the likelihood function for the observed data simultaneously with respect to $\lambda_0(t)$ and β. A more popular approach is proposed by Cox (1975) in which a partial likelihood function that does not depend on $\lambda_0(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters $\lambda_0(t)$ in the Cox PH model). In this section, we will construct the partial likelihood function based on the proportional hazards model.

Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals. Let the ordered death time of r individuals be $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ and let $R(t_{(j)})$ be the risk set just before $t_{(j)}$ and r_j for its size. So that $R(t_{(j)})$ is the group of individuals who are alive and uncensored at a time just prior to $t_{(j)}$. The conditional probability that the i^{th} individual dies at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ dies at $t_{(j)}$ is

$$
\begin{aligned}
&P(\text{ individual } i \text{ dies at } t_{(j)} / \text{ one death from the risk set } R(t_{(j)}) \text{ at } t_{(j)}) \\
&= \frac{P(\text{ individual } i \text{ dies at } t_{(j)}}{P \text{ one individual at } t_{(j)}} \\
&= \frac{P(\text{ individual } i \text{ dies at } t_{(j)}}{\sum_{k \in R(t_{(j)})} P(\text{ individual } k \text{ dies at } t_{(j)})} \\
&= \frac{P(\text{ individual } i \text{ dies at } t_{(j)}, t_{(j)} + \Delta_t)/\Delta_t}{\sum_{k \in R(t_{(j)})} P(\text{ individual } k \text{ dies at } t_{(j)}, t_{(j)} + \Delta_t)/\Delta_t} \\
&= \frac{\lim_{\Delta_t \downarrow 0} P(\text{ individual } i \text{ dies at } t_{(j)}, t_{(j)} + \Delta_t)/\Delta_t}{\lim_{\Delta_t \downarrow 0} \sum_{k \in R(t_{(j)})} P(\text{ individual } k \text{ dies at } t_{(j)}, t_{(j)} + \Delta_t)/\Delta_t} \\
&= \frac{\lambda_i(t_{(j)})}{\sum_{k \in R(t_{(j)} \lambda_k(t_{(j)})}} \\
&= \frac{\lambda_0(t_{(j)}) \exp(\beta' X_i(t_{(j)}))}{\sum_{k \in R(t_{(j)}} \lambda_0(t_{(j)}) \exp(\beta' X_k(t_{(j)}))}
\end{aligned}
$$

$$= \frac{\exp(\beta' X_i(t_{(j)}))}{\sum_{k \in R(t_{(j)}} \exp(\beta' X_k(t_{(j)}))}$$

Then the partial likelihood function for the Cox PH model is given by

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(\beta' X_i(t_{(j)}))}{\sum_{k \in R(t_{(j)}} \exp(\beta' X_k(t_{(j)}))}, \tag{4}$$

in which $X_i(t_{(j)})$ is the vector of covariate values for individual i who dies at $t_{(j)}$. The general method of partial likelihood was discussed by Cox (1975).

Note that this likelihood function is only for the uncensored individuals. Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals and δ_i be the event indicator, which is zero if the i^{th} survival time is censored, and unity otherwise. The likelihood function in equation (4) can be expressed by

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta' X_i(t_{(i)}))}{\sum_{k \in R(t_{(i)}} \exp(\beta' X_k(t_{(i)}))} \right)^{\delta_i}, \tag{5}$$

where $R(t_i)$ is the risk set at time t_i.

The partial likelihood is valid when there are no ties in the data set. That means there is no two subjects who have the same event time.

3.11 Proportional hazard assumption checking

The main assumption of the Cox proportional hazards model is proportional hazards. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e., the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportionality.

3.11.1 Graphical method

We can obtain Cox PH survival function by the relationship between hazard function and survival function

$$S(t, x) = [S_0(t)]^{\exp\left(\sum_{i=1}^{p} \beta_i x_i\right)},$$

Where $x = (x_1, x_2, \ldots, x_p)$ is the value of the vector of explanatory variables for a particular individual. When taking the logarithm twice, we can easily get

$$\log[-\log S(t, x_1)] - \log[-\log S(t, x_2)] = \sum_{i=1}^{p} \beta_i (x_{1i} - x_{2i}).$$

This does not depend on t. This relationship is very helpful to help us identify situations where we may have proportional hazards. By plotting estimated log (-log (survival)) versus survival time for two groups we would see parallel curves if the hazards are proportional. This method does not work well for continuous predictors or categorical predictors that have many levels because the graph becomes "cluttered". Furthermore, the curves are sparse when there are few time points and it may be difficult to tell how close to parallel is close enough.

However, looking at the K-M curves and log (-log (survival)) is not enough to ascertain of proportionality since they are univariable analysis and do not show whether hazards will still be proportional when a model includes many other predictors. But they support our argument for proportionality. We will show some other statistical methods for checking the proportionality.

3.11.2 Adding time-dependent covariates in the Cox model

We create time-dependent covariates by creating interactions of the predictors and a function of survival time and including them in the model. For example, if the predictor of interest is x_j , then we create a time-dependent covariate $x_j(t), x_j(t) = x_j \times g(t)$ where $g(t)$ is a function of time, e.g., t, $\log t$ or Heaviside function of t. The model assessing PH assumption for x_j adjusted for other covariates is

$$\lambda(t, x(t)) = \lambda_0(t) \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p + \delta x_j \times g(t)\right).$$

Where $x(t) = (x_1, x_2, \ldots, x_p, x_j(t))'$ is the value of the vector of explanatory variables for a particular individual. The null hypothesis to check proportionality is that $\delta = 0$. The test statistic can be carried out using either a Wald test or a likelihood ratio test. In the Wald test, the test statistic is

$$W = \left(\frac{\hat{\delta}}{se(\hat{\delta})^2}\right).$$

The likelihood ratio test calculates the likelihood under null hypothesis, L_0 and the likelihood under the alternative hypothesis L_a. The LR statistic is then $LR = -2ln(L_0/L_a) = -2(L_a - L_0)$ where L_0, L_a are log likelihood under two hypotheses respectively. Both statistics have a chi-square distribution with one degree of freedom under the null hypothesis. If the time-dependent covariate is significant i.e., the null hypothesis is rejected, and then the predictor is not proportional. In the same way, we also assessed the PH assumption for several predictors simultaneously.

3.11.3 Tests based on the Schoenfeld residuals

The other statistical test of the proportional hazards assumption is based on the Schoenfeld residual (Schoenfeld, 1982). The Schoenfeld residuals are defined for each subject who is observed to fail. If the PH assumption holds for a particular covariate then the Schoenfeld residual for that covariate will not be related to survival time. So this test is accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranking of individual survival times. The null hypothesis is that the correlation between the Schoenfeld residuals and the ranked survival time is zero. Rejection of null hypothesis concludes that PH assumption is violated.

3.12 Cox proportional hazards model diagnostics

After a model has been fitted, the adequacy of the fitted model needs to be assessed. The model checking procedures below are based on residuals. In linear regression methods, residuals are defined as the difference between the observed and predicted values of the dependent variable. However, when censored observations are present and partial likelihood function is used in the Cox PH model, the usual concept of residual is not applicable. A number of residuals have been proposed for use in connection with the Cox PH model. For this study, three major residuals in the Cox model were used: the CoxSnell residual, the deviance residual, and the Schoenfeld residual. Then we will talk about influence assessment.

3.12.1 Cox-Snell residuals and deviance residuals

The Cox-Snell residual is given by Cox and Snell (Klein & Moeschberger, 1997). The Cox-Snell residual for the i^{th} with observed survival time t_i is defined as

$$rc_i = exp\left[\beta \hat{X}_i\right] \hat{H}_0(t_i) = \hat{H}_i(t_i) = -\log\left(\hat{S}_i(t_)\right).$$

Where $\hat{H}_0(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i , which was derived by Kalbfleisch & Prentice (1973). This residual is motivated by the following result.

Let T has continuous survival distribution $S(t)$ with the cumulative hazard $H(t) = -\log(S(t))$. Let $Z = H(t)$ be the transformation of T based on the cumulative function. Then the survival function for Y is

$$S_Z(z) = P(Z > z) = P(H(t) > z)$$

$$= P\left(T > H_i^{-1}(z) = S_T\left(H_i^{-1}(z)\right)\right)$$

$$= exp\left(-H_T\left(H_T^{-1}(z) = \exp(-z)\right)\right).$$

Thus, regardless of the distribution of T, the new variable $z = H(t)$ has an exponential distribution with unit mean. If the model is be well fitted, the value $\hat{S}_i(t_i)$ would have similar properties to those $S_i(t_i)$. So $rc_i = -\log \hat{S}_i(t_i)$ will have a unit exponential distribution with $f_r(r) = \exp(-\gamma)$.

Let $S_R(r)$ denote the survival function of Cox-Snell residual rc_i. Then

$$S_R(r = \int_r^{\infty} f_R(x)d(x) = exp(-r),$$

and

$$H_R(r) = -\log(S_R(r)) = -\log(\exp) = r.$$

Therefore, we use plot of $H(rc_i)$ versus rc_i to check the fit of the model. This gives a straight line with unit slope and zero intercept if the fitted model is correct. Note, the Cox Snell residuals will not be symmetrically distributed about zero and cannot be negative.

The deviance residual Subbaraman et al. (2007) is defined by

$$r_{Di} = sign(r_{mi}) \left[-2\{r_{mi} + \delta_i \log(\delta_i - r_{mi})\} \right]^{1/2}.$$

Where the function $sign(.)$ is the sign function, which takes the value, 1 if r_{mi} is positive and -1 if r_{mi} negative; $r_{mi} = \delta_i - r_{ci}$ the martingale residuals for the i^{th} individual and $\delta_i = 1$ for uncensored observation and $\delta_i = 0$ for censored observation.

The martingale residuals take values between negative infinity and unity. They have a skewed distribution with mean zero. The deviance residuals are a normalized transform of the martingale residuals (Sayehmiri et al., 2008). They also have a mean of zero but are approximately symmetrically distributed about zero when the fitted model is appropriate. Deviance residual can also be used like residuals from linear regression. The plot of the deviance residuals against the covariates can be obtained. Any unusual patterns may suggest features of the data that have not been adequately fitted for the model. Very large or very small values suggest that the observation may be an outlier in need of special attention.

In a fitted Cox-PH model, the hazard of death for the i^{th} individual at any time depends on the value of $\exp(\boldsymbol{\beta' x})$ which is called the risk score. A plot of the deviance residuals versus the risks core is a helpful diagnostic to assess a given individual on the model. Potential outliers will have deviance residuals whose absolute values are very large. This plot will give the information about the characteristic of observations that are not well fitted by the model.

3.12.2 Schoenfeld residuals

All the above three residuals are residuals for each individual. We described covariate wise residuals by (Schoenfeld, 1982). The Schoenfeld residuals were originally called partial residuals because the Schoenfeld residuals for i^{th} individual on the j_{th} explanatory variable x_{ij} is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j. From equation 4, this

logarithm of the partial likelihood function is given by

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^{p} \delta_i \{x_{ij} - a_{ij}\},$$

Where x_{ij} is the value of the J^{th} explanatory variable $j = 1, 2, \ldots, p$ for the i^{th} individual and

$$a_{ij} = \frac{\sum_{l \in R(t_i)} x_{jl} \exp(\beta' x_i)}{\sum_{l \in R(t_i)} \exp(\beta' x_i)}.$$

The Schoenfeld residual for i^{th} individual on $x_j$is given $rp_{ij} = \delta_i \{x_{ij} - a_{ij}\}$. The schoenfeld residuals sum to zero.

3.12.3 Diagnostics for influential observations

Observations that have an undue effect on model-based inference are said to be influential. In the assessment of model adequacy, it is important to determine whether there are any influential observations. The most direct measure of influence is $\hat{\beta}_j - \hat{\beta_{j(i)}}$. Where $\hat{\beta}_j$ is the j^{th} parameter, $j = 1, 2, \ldots, p$ in a fitted Cox PH model and $\hat{\beta_{j(i)}}$ is obtained by fitting the model after omitting observation i. In this way, we have to fit the $n + 1$ Cox models, one with the complete data and n with each observation eliminated. This procedure involves significant amount of computation if the sample size is large. We would like to use an alternative approximate value that does not involve an iterative refitting of the model. To check the influence of observations on a parameter estimate, Cain & Lange (1984) showed that an approximation to $\hat{\beta}_j - \hat{\beta_{j(i)}}$ is the j^{th} component of the vector $rSiV\left(\hat{\beta}\right)$. Where rSi is the $p \times 1$ vector of score residuals for the i^{th} observation (Collett, 2015).

Which are modifications of Schoenfeld residuals and are defined for all the observations, and $V\left(\hat{\beta}\right)$ is the variance-covariance matrix of the vector of parameter estimates in the fitted Cox PH model. The j^{th} element of this vector is called delta-beta statistic for the th j^{th} explanatory variable, *i.e.*, $\Delta_i \hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta_{j(i)}}$, which tells us how much each coefficient will change by removal of a single observation. Therefore, we were checked whether there are influential observations for any particular explanatory variable. On the other hand, the statistic, $LD_i = 2l_p(\beta) - 2l_p(\beta_{(i)})$, which is called the likelihood

displacement statistic, can be used as a measure of how the maximized partial log Likelihood changes if the i^{th} observation was deleted from the data set. Observations that influence a particular parameter estimate have a large absolute value of DFBETA than other observations in the data set. Observations that do influence the overall fit of the model are those which have large values of likelihood displacement statistics than the other observations in the data set (Collett, 2015).

3.13 Accelerated Failure Time (AFT) Model

Although parametric PH models are very applicable to analyze survival data, there are relatively few probability distribution for the survival time that can be used with these models Jiezhi (2009). In these situations, the accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data when the proportional hazard assumptions doesn't hold. The key differences between the Cox-PH model and AFT models are the baseline hazard function and ways of estimating coefficients (D. Kleinbaum & Klein, 2005). The AFT is obtained by regressing the logarithm of the survival time over the covariates and the effect of the explanatory variables on the survival time is directly measured. Some of the standard parametric AFT models are exponential, Weibull, log-normal and log-logistic (Dätwyler & Stucki, 2011).

The AFT model describes the relationship between survival probabilities and a set of covariates. For a group with covariates $(X_{1i}, X_{2i}, ..., X_{pi})$, the AFT model is written mathematically as

$$S(t|x) = S_0(t|\phi(x)) \tag{6}$$

Where $S_0(t)$ is the baseline survival function and ϕ is an acceleration factor (time ratio) i.e. a ratio of survival times corresponding to any fixed value of $S(t)$. The acceleration factor is given according to the formula

$$\phi(x) = \exp(\beta_1 X_{1i} + \beta_{2i} X_2 + ... + \beta_p X_{pi}) \tag{7}$$

According to the relationship of survival function and hazard function, the hazard function for an individual with covariate $X_{1i}, X_{2i}, ..., X_{pi}$ is given by

$$\lambda(t|x) = \left[\frac{1}{\phi(x)}\right] \lambda_0 \left[t|\phi(x)\right] \tag{8}$$

Under an accelerated failure time model, the covariate effects are assumed to be constant and multiplicative on the time scale, that is, the covariate impacts on survival by a constant factor (acceleration factor). The corresponding log-linear form of the AFT model with respect to time is given by

$$\log T_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \sigma\epsilon_i \tag{9}$$

Where μ is intercept, σ is scale parameter and is a random variable, ϵ_i assumed to have a particular distribution. For each distribution of ϵ_i, there is a corresponding distribution for T . The members of the AFT model class include the exponential AFT model, Weibull AFT model, loglogistic AFT model, and log-normal AFT model. The AFT models are named for the distribution of T rather than the distribution of ϵ_i or $\log T$. The survival function of T_i can be expressed by the survival function of ϵ_i :

$$S_i(t) = P(T_i \geq t) = P(\log T_i \geq \log t)$$

$$= P(\mu + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \sigma\epsilon_i \geq \log t)$$

$$= P\left(\epsilon_i \geq \frac{\log t - \mu - \boldsymbol{\beta' x}}{\sigma}\right)$$

$$= S_i(t) = S_{\epsilon_i}\left(\frac{\log t - \mu - \boldsymbol{\beta' x}}{\sigma}\right) \tag{10}$$

The effect size for the AFT model is the time ratio. The time ratio comparing two levels of covariate x_i ($x_i = 1$ vs $x_i = 0$) ; after controlling all the other covariates is $\exp(\beta_i)$, which is interpreted as the estimated ratio of the expected survival times for two groups. A time ratio above 1 for the covariate implies that this covariate prolongs the time to event, while a time ratio below 1 indicates that an earlier event is more likely. Therefore, the AFT models can be interpreted in terms of the speed of progression of a disease. The effect of the covariates in an accelerated failure time model is to change the scale, and not the location of a baseline distribution of survival times.

Table 3.2: Commonly used distributions and parameters in AFT models

Distribution	$f(t)$	S(t)	$\lambda(t)$
Exponential	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ
Weibull	$\lambda\rho t^{-1}e^{-\rho t}$	$e^{-\rho t}$	$\lambda\rho t^{-1}$
Log-logistic	$\frac{\lambda\rho t^{-1}}{[1+\lambda\rho t^{\rho}]^2}$	$\frac{1}{1+\lambda\rho t^{\rho}}$	$\frac{\lambda\rho t^{-1}}{1+\lambda\rho t^{\rho}}$
Log-normal	$\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{[\log t-\mu]^2}{2\sigma^2}$	1- $\Phi[\frac{\log t-\mu}{\sigma}]$	$\frac{\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{[\log t-\mu]^2}{2\sigma^2}}{1-\Phi[\frac{\log t-\mu}{\sigma}]}$

Where

- λ and ρ denotes scale parameter and shape parameter, respectively, for Exponential, Weibull and Log-logistic distribution.
- σ and μ denote scale parameter and shape parameter, respectively, for Log-normal distribution.
- $\Phi(.)$ denotes the standard normal distribution function.

3.13.1 Methods of Parameter Estimation

The parameters of semi-parametric Cox PH model is estimated by using partial likelihood estimation method. The partial likelihood estimation is a technique used to make an inference about the regression parameters, β, in the presence of nuisance parameters $\lambda(t|x)$ Cox (1972). Whereas, the fully likelihood estimation method is used to estimate the regression parameters and baseline hazard functions in AFT models (Collett, 2015).

3.13.2 Weibull and Exponential AFT Models

Suppose the survival time T has W($\gamma;\delta$) distribution with scale parameter and shape parameter, under AFT model, the hazard function for the ith individual is

$$\lambda_i(t) = \frac{1}{[\phi_i(x)]\,\gamma}\delta\gamma(t)^{\gamma-1} \tag{11}$$

Where $\phi_i = \exp(\beta_1 X_1+\beta_2 X_2+...+\beta_p X_{pi})$ for individual i with p explanatory variables, so the survival time is given as the Weibull distribution has the AFT property. If T_i has

a Weibull distribution, then ϵ_i has an extreme value distribution (Gumbel distribution). The survival function of Gumbel distribution is given as

$$S_{\epsilon_i}(\epsilon) = \exp(-\exp(\epsilon)) \tag{12}$$

The AFT representation of the survival function of the Weibull model is given by

$$S_i(t) = \exp\left[-\exp\left(\frac{-\mu - \beta_1 X_{1i} - \beta_2 X_{2i} - ... - \beta_p X_{pi}}{\sigma}\right) t^{\frac{1}{\sigma}}\right] \tag{13}$$

The AFT representation of hazard function of the Weibull model is given by

$$\lambda_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp\left(\frac{-\mu - \beta_1 X_{1i} - \beta_2 X_{2i} - ... - \beta_p X_{pi}}{\sigma}\right) \tag{14}$$

The median survival time is $t(50) = \exp\left[\sigma \log(\log 2) + \mu + \beta' x_i\right]$
The exponential distribution can be derived from Weibull distribution, that is by taking $\sigma = 1$ or $\beta = 1$, so that the equation (16) become
$S_i(t) = \exp\left[-\exp\left(-\mu - \beta_1 X_{1i} - \beta_2 X_{2i} - ... - \beta_p X_{pi}\right) t\right]$, where $S_i(t)$ are the survival functions for exponential AFT model.

3.13.3 The Log-logistic AFT model

One limitation of the Weibull hazard function is that it is a monotonic function of time. However, the hazard function can change direction in some situations. We will describe the log-logistic model in this section. The log-logistic survival and hazard function are given by

$$S(t) = \frac{1}{1 + e^{\theta} t^k} \qquad\qquad h(t) = \frac{e^{\theta} k t^{k-1}}{1 + e^{\theta} t^k} \tag{15}$$

Where θ and k are unknown parameters and $k > \theta$. When $k \leq 1$, the hazard rate decreases monotonically and when $k > 1$, it increases from zero to a maximum and then decreases to zero.

Suppose that the survival times have a log-logistic distribution with parameter and k, under the AFT model, the hazard function for the i^{th} individual is

$$\lambda(t) = \frac{e^{\theta - k \log \phi_i} k t^{k-1}}{1 + e^{\theta - k \log \phi_i} t^k} \tag{16}$$

Where, $\phi_i = \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$ for individual i^{th} with p explanatory variables. Therefore, the survival time for the i^{th} individual has a log-logistic distribution with parameter $\theta - k \log \eta$ and k, log-logistic distribution has AFT property.

If the baseline survival function is $S_0(t) = \{1 + e^{\theta} t^k\}^{-1}$, where θ and k are unknown parameters then, the base line odds of surviving beyond time t are given by

$$\frac{S_0(t)}{1 - S_0(t)} = e^{-\theta} t^{-k}$$

. The survival time for the i^{th} individual also has a log-logistic distribution, which is

$$S_i(t) = \frac{1}{1 + e^{\theta - k \log \phi_i} t^k} \tag{17}$$

The AFT representation of survival function of the log-logistic model is given by

$$S_i(t) = \left[1 + t^{1/\sigma} \exp\left(\frac{-\mu - \beta_1 X_{1i} - \beta_2 X_{2i} - ... - \beta_p X_{pi}}{\sigma}\right)\right]^{-1} \tag{18}$$

Comparing the formula (17) and (18), we can easily find $\theta = \frac{-\mu}{\sigma}, k = \sigma^{-1}$.
According to the relationship of survival and hazard function, the hazard function for the i^{th} individual is given by

$$\lambda_i(t) = \frac{1}{\sigma t}\left[1 + t^{-1/\sigma} \exp\left(\frac{\mu + \beta_1 X_{1i} + \beta_2 X_{2i} - ... + \beta_p X_{pi}}{\sigma}\right)\right]^{-1} \tag{19}$$

The median survival time is $t_i(50) = \exp(\mu + \beta' x_i)$.

3.13.4 The Log-normal AFT model

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by

$$S_0(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right) \qquad\qquad h_0(t) = \frac{\Phi\left(\frac{\log t}{\sigma}\right)}{1 - \Phi\left[\frac{\log t - \mu}{\sigma}\right] \sigma t} \tag{20}$$

Where μ is intercept, σ is scale parameter and is a random variable; $\Phi(x)$ is the cumulative density function of the standard normal distribution. The survival function for the i^{th} individual is

$$S_i(t) = S_0(t|\phi_i) = 1 - \Phi\left(\frac{\log t - \boldsymbol{\beta' X_i}}{\sigma}\right) \tag{21}$$

Where $\phi_i = \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$. Therefore the log survival time for the i^{th} individual has normal $(\mu + \beta' X_i, \delta)$. The log-normal distribution has the AFT property.In a two-group study, we can easily get

$$\Phi^{-1}(1 - S(t)) = \frac{1}{\sigma}(\log t - \beta X_i - \mu),$$

Where X_i is the value of a categorical variable, which takes the value one in one group and zero in the other group. This implies that a plot of $\Phi^{-1}[1 - S(t)]$ versus $\log t$ will be linear if the log-normal distribution is appropriate.

3.14 Model building

The methods of selecting a subset of covariates in Cox-PH and AFT models are essentially similar to those used in any other regression models. Hos-mer and Lemeshow recommended the following steps in selecting the variables by Lemeshow et al. (2011):

- The first step is to fit model that contain each of the variables one at a time.
- We begin by fitting a multivariable model containing all variables significant in the univariable analysis at the 10 percent level.
- Use backward selection to eliminate non-significant variables and examine the effect of remaining variables.
- Starting with step (3) model, consider each of the non-significant variables from step (2) using forward selection and do the analysis.
- Fit the final model by omitting variables that are non-significant and adding variables that are significant.

3.15 Checking the Adequacy of AFT models

The graphical methods can be used to check if a parametric distribution fits the observed data. Specifically, if the survival time follows an exponential distribution, a plot of $\log[-\log S(t)]$ versus $\log t$ should yield a straight line with slope of 1. If the plots are parallel but not straight, then PH assumption holds but not the Weibull. If the lines for

two groups are straight but not parallel, the Weibull assumption is supported but the PH assumption is violated. The log-logistic assumption can be graphically evaluated by plotting $\log[\frac{(1-S(t))}{S(t)}]$ vs $\log t$. If the distribution of survival functions is log-logistic, then the resulting plot should be a straight line. For the log-normal distribution, a plot of $\Phi^{-1}(1 - S(t))$ versus $\log t$ should be linear. All these plots are based on the assumption that the sample is drawn from a homogeneous population, implying that no covariates are taken into account. So this graphical method is not very reliable in practice. There are other methods to check the fitness of the model.

3.15.1 Quantile-Quantile plot

An initial method for assessing the potential for an AFT model is to produce a quantile-quantile plot. For any value of p in the interval $(0; 100)$, the P^{th} percentile is

$$t(p) = S^{-1}\left(\tfrac{100-p}{100}.\right)$$

Let $t_0(p)$ and $t_i(p)$ be the P^{th} percentiles estimated from the survival functions of the two groups of survival data. The percentiles for the two groups may be expressed as

$$t_0(p) = S_0^{-1}\left(\tfrac{100-p}{100}\right); t_1(p) = S_1^{-1}\left(\tfrac{100-p}{100}\right).$$

Where $S_0(t)$ and $S_1(t)$ are the survival functions for the two groups. So we can get $S_1[t_1(p)] = S_0[t_0(p)]$. Under the AFT model , $S_1(t) = S_0(t * exp(-\beta' x))$, and so $S_1[t_1(p)] = S_0[t_1(p) * \exp(-\beta' x)]$. Therefore we get $t_0(p) = t_1(p) * \exp(-\beta' x)$.
The percentiles of the survival distributions for two groups can be estimated by the K-M estimates of the respective survival functions. A plot of percentiles of the K-M estimated survival function from one group against another should give an approximate straight line through the origin if the accelerated failure time model is appropriate. The slope of this line will be an estimate of the acceleration factor $\exp(-\beta' x)$.

3.15.2 Statistical criteria

We can use statistical tests or statistical criteria to compare all these AFT models. Nested models can be compared using the likelihood ratio test. The exponential model,

the Weibull model and lognormal model are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) were used instead, which is defined as

$$AIC = -2l + 2(k + c),$$

Where l the log-likelihood, k is is the number of covariates in the model and c is the number of model-specific ancillary parameters. The addition of $2(k+c)$ can be thought of as a penalty if nonproductive parameters are added to the model. Lower values of the AIC suggest a better model. But there is a difficulty in using the AIC in that there are no formal statistical tests to compare different AIC values.

3.15.3 Residual Plots

Residual plots can be used to check the goodness of fit of the model. Procedures based on residuals in the AFT model are particularly relevant with the Cox PH model. One of the most useful plots is based on comparing the distribution of the Cox-Snell residuals with the unit exponential distribution. The Cox-Snell residual for the i^{th} individual with observed time t_i is defined as

$$\hat{S}(t) = S_{\epsilon_i}\left(\frac{\log t - \hat{\mu} - \beta' x_i}{\hat{\sigma}}\right),$$

Where $\hat{\mu}, \beta'$ and $\hat{\sigma}$ are the maximum likelihood estimator of μ, β and σ respectively. $S_{\epsilon_i}(\epsilon)$ is the survival function of ϵ_i in the AFT model, and $\frac{\log t - \hat{\mu} - \beta' x_i}{\hat{\sigma}} = r_{\epsilon_i}$ is referred to as standardized residual.

The Cox-Snell residual can be applied to any parametric model. The corresponding form of residual based particular AFT model can be obtained. For example, under the Weibull AFT model, since $S_{\epsilon_i(\epsilon)} = \exp(-e^{\epsilon})$, the Cox-Snell residual is then

$$r_{\epsilon_i} = -\log\left\{\hat{S}(t_i)\right\} = -\log S_{\epsilon_i}(r_{s_i}) = \exp(r_{S_i}).$$

Under the log-logistic AFT model, since $S_{\epsilon_i(\epsilon)} = (1 + e^{\epsilon})^{-1}$, the Cox-Snell residual is then

$$r_{\epsilon_i} = \log\left[1 + \exp(r_{s_i})\right].$$

If the fitted model is appropriate, the plot of $\log(-\log S(r_{\epsilon_i}))$ versus r_{ϵ_i} is a straight line with unit slope through the origin. These residuals lead to the deviance residuals for the particular AFT model. A plot of deviance residuals against the survival time or explanatory variables is used to check whether there are particular times, or particular values of explanatory variables, for which the model is not a good fit.

CHAPTER FOUR

4 Result and Discussion

4.1 Results

Descriptive statistics of socioeconomic, demographic, and biological characteristics on survival of patients with the COVID-19 pandemic are shown in Tables 4.1 and 4.2. According to table 4.2, a total of 809 patients with COVID-19 participated in this study. Of the 809 patients, 135 (16.7%) died while 674 (83.3%) were censored during the follow-up period.

Out of all patients about 326 (40.3%) are females and 483 (59.7%) are males. Of these, 43 (5.3%) female patients and 92 (11.4%) male patients died, leaving 283 (35%) female and 391 (48.3%) male patients censored. In the table, 192 (23.7%) patients were under 18 years, 376 (46.5%) patients and 241 (29.8%) patients were between 18 to 45 and above 45 years, respectively. Of the patients, 484 (59.8%) were from urban areas and 325 (40.2%) were from rural areas from this residence, 58(7.2) and 77(9.5) died of covid-19 respectively.

Table 4.2 also displays the situation of patients at admission; out of 212 (26.2%) asymptomatic patients, 1 (0.1%) died, 103 (12.7%) mild patients, 4 (0.5%) died, 155 (19.2%) moderate patients, 24 (3.3%) died, and 339(41.9%) severe patients 106 (13.1%) death cases are recorded. Regarding comorbidity, about 375(46.4%) of patients had no comorbidity whereas 434(53.6) are patients with comorbidity. Among the patients, 484 (59.8%) patients had no symptoms when admitted to the hospital, while 325 (40.2%) did. Of a total of 809 individuals, 193 (23.9%) had HIV/AIDS while 586 (76.1%) did not. Of them, 105 (13.0%) and 30 (3.7%) passed away with covid-19 respectively.

In addition, 86(10.6%) of the total deaths were observed in COVID-19 patients who received intranasal oxygen use. Of those COVID-19 patients, 451 (55.8%) were hyper-

tensive, with a higher mortality rate of 96 (11.9%) compared to the non-hypertensive group. Finally, 451 (55.8%) were diabetic patients with a higher mortality rate of 79 (9.8%) compared to non-diabetic patients.

Table 4.1, suggests that patients were followed up for a minimum of one day and a maximum of forty days. During the study period, the overall mean and median predicted survival times for patients were 11 and 10 days, respectively, with a standard deviation of 6.43 days and inter-quartile range [7,14] days. The median follow-up time of the death was 6.0 days (Inter quartile range: 3-10 days) and censored patients was 10 days (Inter quartile range: 7-15 days).

Table 4.1: Summary statistics for time

Variable	Status	Mean	S.deviation	Median	Q_1	Q_3	Min	Max
Time	Death	7	5	6	3	10	1	27
	Censored	12	6	10	7	15	1	40
	Over all	11	6.43	10	7	14	1	40

Source: Jimma Zone COVID-19 Treatment Centers, southwest Ethiopia from May 16, 2020 through March 9, 2022

Table 4.2: Descriptive summaries of patient's characteristics diagnosed for COVID-19.

		Patients Status		
Covariates	Categories	Censored n(%)	Death n(%)	Total n(%)
Age	< 18	190(23.5)	2(0.2)	192(23.7)
	18-45	321(39.7)	55(6.8)	376(46.5)
	> 45	163(20.1)	78(9.6)	241(29.8)
Sex	Female	283(35.0)	43(5.3)	326(40.3)
	Male	391(48.3)	92(11.4)	483(59.7)
Residence	Rural	267(33.0)	58(7.2)	325(40.2)
	Urban	407(50.3)	77(9.5)	484(59.8)
Status at admission	Asymptomatic	211(26.1)	1(0.1)	212(26.2)
	Mild	99(12.2)	4(0.5)	103(12.7)
	Moderate	131(16.2)	24(3.0)	155(19.2)
	Sever	233(28.8)	106(13.1)	339(41.9)
Comorbidity	No	350(43.3)	25(3.1)	375(46.4)
	Yes	324(40.0)	110(13.6)	434(53.6)
Symptom at admission	Yes	260(32.1)	65(8.0)	325(40.2)
	No	414(51.2)	70(8.6)	484(59.8)
HIV/ AIDS	No	586(72.4)	30(3.7)	616(76.1)
	Yes	88(10.9)	105(13.0)	193(23.9)
Oxygen use	No	403(49.8)	49(6.1)	452(55.9)
	Yes	271(33.5)	86(10.6)	357(44.1)
Hypertension	No	319(39.4)	39(4.8)	358(44.2)
	Yes	355(43.9)	96(11.9)	451(55.8)
Diabetics	No	499(61.7)	56(6.9)	555(68.6)
	Yes	175(21.6)	79(9.8)	254(31.4)
Total		674(83.3%)	135(16.7%)	809(100%)

Source: Jimma Zone COVID-19 Treatment Centers, from May 16, 2020 through March 9, 2022

4.1.1 The Kaplan-Meier Estimator

As shown in Figure 4.1, as the survival time of the patient increases, the survival probability decreases.

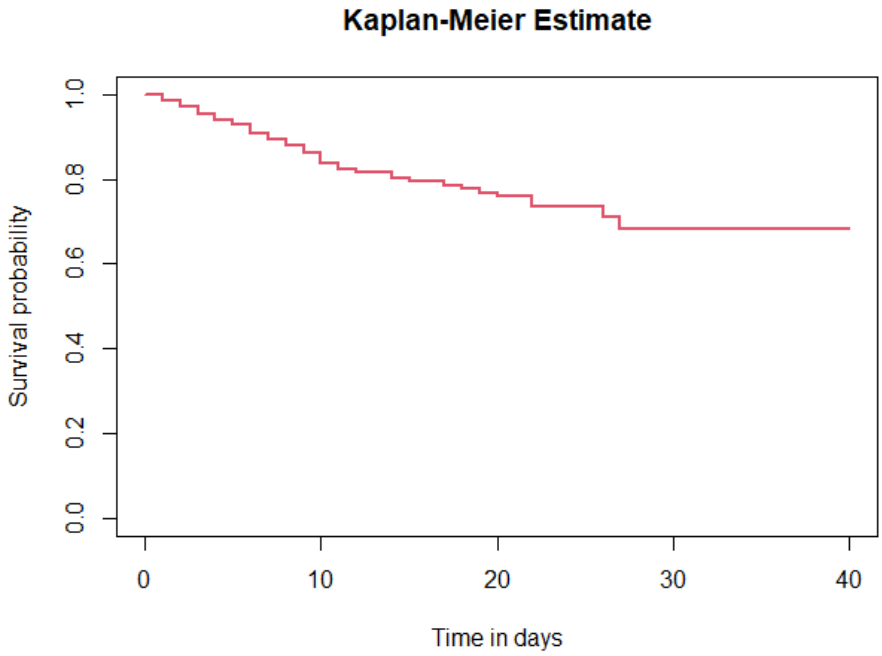

Figure 4.1: Estimated survival function of COVID-19 patients

4.1.2 Comparison of Survival function of COVID-19 patients

The survivor-ship estimate curve for age, status at admission, comorbidity, symptom at admission, HIV/AIDS, oxygen care, hypertension, and diabetes are displayed (see appendix A). It shows that there was a difference in survival times among the levels of covariates. For comparing the survival experiences between groups, the log-rank test was applied to all categorical variables. From table 4.3 the log-rank results suggest that comparison of survival function for each independent categorical variables. The table shows there is a significant difference in survival functions between the categories of age, sex, comorbidity, status at admission, HIV/AIDS, symptom at admission, intranasal oxygen use, hypertension and diabetes at 5% level of significance while residence does not have a different survival experience. The K-M curves also show the same result as the log-rank test (see appendix A).

Table 4.3: Results of the log-rank test for each categorical variables of COVID-19 patients

Covariates	Chi-square	df	P-value
Age	76.1	2	<0.001
Sex	5.7	1	0.02
Residence	0	1	0.9
Comorbidity	49.1	1	<0.001
Status at admission	94.9	3	<0.001
HIV/AIDS	24.2	1	<0.001
Symptom at admission	9.5	1	0.002
Intranasal Oxygen use	26.2	1	<0.001
Hypertension	13.4	1	<0.001
Diabetics	61.8	1	<0.001

4.2 Standard Cox PH model

This study used uni-variable analysis to check all the risk factors before proceeding to more complicated models. And also it used a uni-variable Cox proportional hazards regression for every potential risk factor. The Wald test is considered in each univariable Cox PH model. Variables are identified as significant using a 25% significance level in the univariable model. We then fit the full multi-variable Cox PH model including all the potential risk factors. Consequently, in the univariable Cox proportional hazard models the model with a single covariate, Age, sex, comorbidity, status at admission, HIV/AIDS, symptom at admission, hypertension, intranasal oxygen use and diabetes show a statistically significant association with the survival time. But the predictor variable place of residence is not statistically significant, suggesting that these variable is not associated with the survival time revealing that these variable is not included in the multivariable model. Therefore, this study considered the model that includes all the significant predictors.

Table 4.4: Uni-variable and multi-variable Cox PH model

Covariates	Univariable Analysis				Multivariable Analysis			
	β	HR	P-value	95%CI[HR]	β	HR	P-value	95%CI[HR]
Age 18-45	2.61	13.64	0.0003	[3.33,55.95]	1.66	5.26	0.028	[1.19,23.12]
>45	3.54	34.44	0.000	[8.46,140.16]	2.42	11.30	0.001	[2.59,49.25]
Sex Male	0.44	1.55	0.02	[1.02,2.23]	0.30	1.36	0.102	[0.94,1.96]
Comorbidity Yes	1.43	4.17	0.000	[2.7,6.4]	1.48	4.38	0.0007	[1.86,10.32]
St.at.admn. Mild	2.16	8.74	0.0525	[0.97,78.2]	1.65	4.68	0.145	[0.56,48.51]
Moderate	3.62	37.5	0.0004	[5.1,277.2]	2.79	16.32	0.007	[2.12,125.60]
Sever	4.31	744.4	0.0002	[10.4,533.1]	3.38	29.55	0.001	[3.86,225.70]
HIV/AIDS Yes	0.97	2.66	0.000	[1.8,3.4]	1.16	3.19	0.006	[1.41,12.62]
Hypertension Yes	0.68	1.98	0.0003	[1.36,2.88]	0.44	1.56	0.026	[1.05,2.31]
Symptom at admission No	0.53	1.70	0.002	[1.12,2.89]	0.82	2.27	0.0004	[1.28,4.63]
Oxygen use Yes	0.89	2.43	0.000	[1.71,3.45]	0.67	1.95	0.0050	[1.02,2.85]
Diabetics Yes	1.29	3.63	0.000	[2.57,5.12]	0.52	1.68	0.017	[1.10,2.58]
LRT								205.8
AIC								1502.507

AIC: Akaike Information Criterion; LRT: Likelihood Ratio Test; β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI[HR]: 95% confidence interval for HR

From table 4.4, a predictor variable sex is significant in the Uni-variable model at 5% level of significance but not significant in the multivariable model. Then by removing the variable sex from table 4.4, we get the model containing age, comorbidity, status at admission, HIV/AIDS, symptom at admission, oxygen use, hypertension, and diabetes variables that are significant.

Table 4.5: Uni-variable and multi-variable Cox PH model

Covariates	Uni-variable Analysis				Multi-variable Analysis			
	β	HR	P-value	95%CI[HR]	β	HR	P-value	95%CI[HR]
Age 18-45	2.61	13.64	0.0003	[3.33,55.95]	1.66	5.26	0.028	[1.20,23.12]
>45	3.54	34.44	0.000	[8.46,140.16]	2.46	11.67	0.001	[2.68,50.73]
Comorbidity Yes	1.43	4.17	0.000	[2.7,6.4]	1.49	4.43	0.0005	[1.90,10.34]
St.at.admn. Mild	2.16	8.74	0.0525	[0.97,78.2]	1.56	4.75	0.170	[0.51,43.92]
Moderate	3.62	37.5	0.0004	[5.1,277.2]	2.79	16.29	0.007	[2.12,125.34]
Sever	4.31	744.4	0.0002	[10.4,533.1]	3.36	28.84	0.001	[3.77,220.27]
HIV/AIDS Yes	0.97	2.66	0.000	[1.8,3.4]	-1.17	0.31	0.005	[0.14,0.70]
Hypertension Yes	0.68	1.98	0.0003	[1.36,2.88]	0.42	1.52	0.035	[1.03,2.26]
Symptom at admission No	0.53	1.70	0.002	[1.12,2.89]	0.83	2.27	0.0003	[1.28,4.63]
Oxygen use Yes	0.89	2.43	0.000	[1.71,3.45]	0.68	1.97	0.0045	[1.04,2.87]
Diabetics Yes	1.29	3.63	0.000	[2.57,5.12]	0.52	1.68	0.017	[1.10,2.57]
LRT								203.0
AIC								1503.261

AIC: Akaike Information Criterion; LRT: Likelihood Ratio Test; β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI[HR]: 95% confidence interval for HR

After developing a multi-variable model of the major impacts of the variables, we assessed the model's appropriateness by examining its goodness of fit and PH assumption. The proportional hazards assumption (PH) of the variables and their interaction with the log of survival time and Schoenfeld residuals were statistically tested, and the PH assumption was checked using a graphical technique. The PH assumption for all of the categorical variables was tested using the log (-log (survival)) against the survival time plot. The graphs in Appendix B for each of the categorical variables show lines

that appear to be parallel, suggesting that the proportional-hazards assumption among covariates such as comorbidity, status at admission, HIV/ADS, and diabetes has not been violated but that it has appeared to be violated for age, symptom at admission, hypertension, and intranasal oxygen use (see appendix B). The proportional hazards assumption (PH) for intranasal oxygen use is fails, according to a test of the variables and their interaction with the log of time in table 4.6 (p-values for intranasal oxygen use*log(time) are less than 0.05, which is 0.046).

The PH assumption in Table 4.6 is further examined using the Schoenfeld residuals. The ranked survival time and Schoenfeld residual for this covariate are correlated, and the p-value tests whether this association is zero. The p-values for status at admission and intranasal oxygen usage are less than 0.05, whereas all other variables are more than 0.05, indicating that the PH assumption is violated for status at admission and intranasal oxygen use but is valid for all other covariates. The proportional hazard assumption was not found to be generally acceptable by the global test (p-value = 0.0043). Additionally, this shows that the intranasal oxygen consumption and covariate status at admission both contradict the PH assumption at the level of 0.05.

A plot of the Cox-Snell residuals against the cumulative hazard of Cox-Snell residuals is presented in Figure 4.2. There is some evidence of a systematic deviation from the straight line, which gives us some concern about the adequacy of the fitted model. The plot of deviance residual against the risk score shows that the deviance residuals seem not to be symmetrically distributed about zero. There are very high or very low deviance residuals which suggest that these observations may be outliers (Figure 4.3). Therefore, we have some concern about the adequacy of the fitted Cox PH model.

We also use delta-beta statistic to measure the influential observations on the model as a whole. It shows that the coefficients do not change too much when the observations corresponding to the largest delta-beta statistics are removed. Therefore, we do not remove them from the dataset and conclude that there are no influential observations (Figure 4.2). Lastly, we can say that applying Cox proprtional hazards for COVID-19

Table 4.6: Statistical test for proportional hazards assumption (PH) of the covariates and their interaction with log of time (time to death) and Schoenfeld residual.

Covariates	Covariates interaction with log(time)				Schoenfeld Residual
	β	HR	P-value	95%CI[HR]	P-value
Age:log(time)					
18-45	-26.29	3.815e-12	0.061	(6.2e-15, 2.3e-09)	0.348
>45	-25.81	6.149e-12	0.069	(1.0e-14, 3.7e-09)	0.300
Comorbidity:log(time)	0.83	2.30	0.213	(1.34,3.95)	0.944
Status at admission:log(time)					
Mild	-0.28	0.75	0.632	(0.24,2.38)	0.035
Moderate	-0.78	2.18	0.100	(0.86,5.51)	0.007
sever	1.08	2.95	0.202	(1.17,7.42)	0.002
HIV/AIDS:log(time)	-0.71	0.49	0.080	0.29,0.83	0.963
Hypertension:log(time)	0.26	1.30	0.401	(1.10,1.66)	0.813
Symptom at admission:log(time)	-0.28	0.76	0.054	(0.57,1.00)	0.805
Oxygen use:log(time)	-0.30	0.74	0.046	(0.55,0.99)	0.025
Diabetics:log(time)	0.33	1.39	0.104	(1.07,1.82)	0.454
Global test					0.0043

β: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%CI[HR]: 95% confidence interval for HR

data is not suggested because the basic assumptions of Cox prortional hazards are violated. The residuals also support the violation of assumptions. So we don't need to talk about the hazards ratio and the relation of covariates with time to death of COVID-19 patients. It is better to apply AFT models.

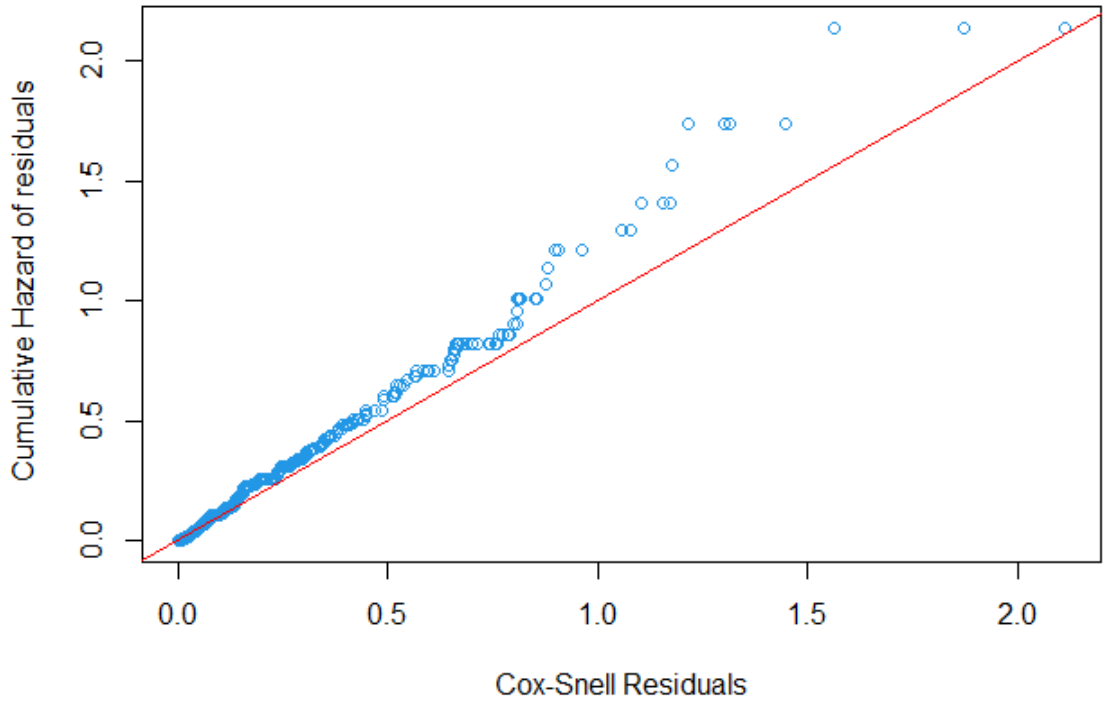

Figure 4.2: Cumulative hazard plot of the Cox-Snell residual for Cox PH model

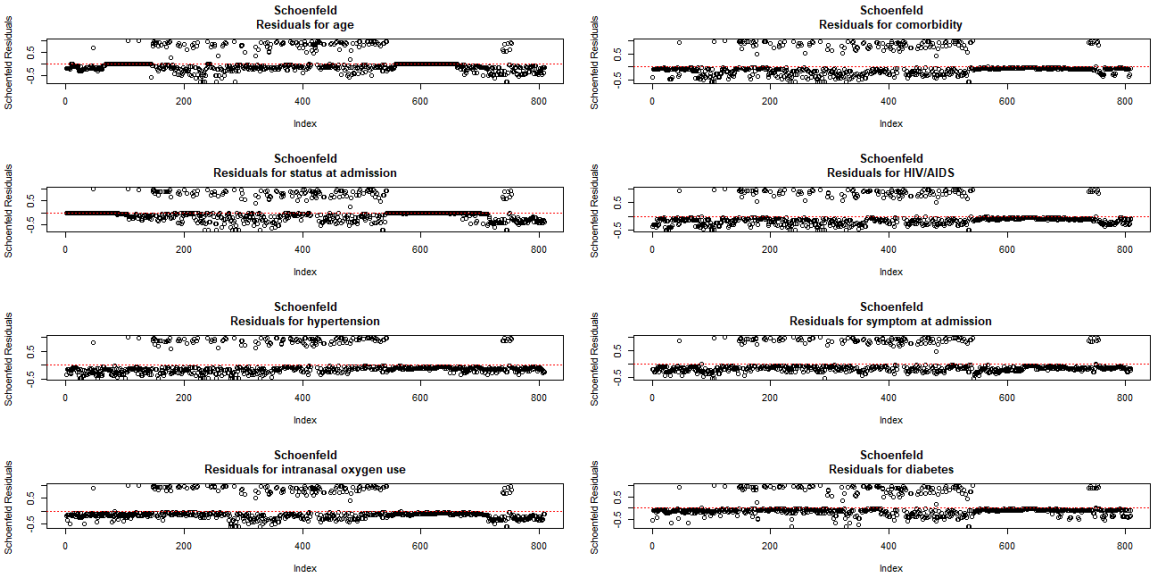

Figure 4.3: Index plots of dfbeta for the multivariate Cox regression model

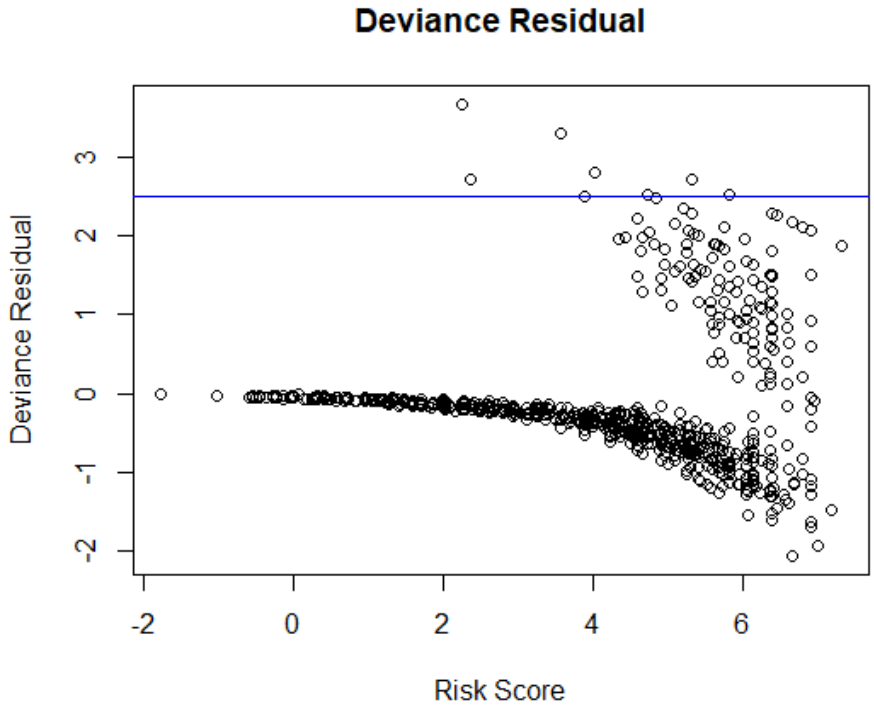

Figure 4.4: Deviance residuals plotted against the risk score for Cox PH model

4.3 Accelerated Failure Time Models

This study used uni-variable analysis before preceding the multi-variable analysis. The uni-variable analyses were fitted for each covariate with a p-value less than 0.25 by using different AFT models such as Exponential, Weibull, Log-logistic, and Log-normal distributions. As shown in table 4.7, the univariable AFT model shows that all covariates except residence are found to be significant with survival times of COVID-19 patients at 25% level. The multi-variable analysis of AFT models was done by using all significant covariates in uni-variable analysis at a 25% level. In this study the backward elimination method is used to select the final significant covariates. The covariates such as age, sex, comorbidity, status at admission, HIV/AIDS, hypertension, symptom at admission, intranasal oxygen use, and diabetes were significant in uni variable analysis of all AFT models at 25% level. The model comparison was done using those significant covariates for each AFT models.

Table 4.7: Uni-variable AFT models

		Exponential	**Weibull**	**Log-normal**	**Log-logistic**
Covariates	Categories	$\hat{\beta}[P\text{-value}]$	$\hat{\beta}[P\text{-value}]$	$\hat{\beta}[P\text{-value}]$	$\hat{\beta}[P\text{-value}]$
Age	<18				
	18-45	-2.58[0.0003]	-2.31[0.0006]	-2.15[0.0004]	-1.62[0.0005]
	>45	-3.51[0.000]	-3.16[0.0000]	-3.02[0.0000]	-2.54[0.0000]
Residence	Rural				
	Urban	0.008[0.96]	-0.003[0.98]	-0.001[0.994]	-0.003[0.98]
Sex	Female				
	Male	-0.45[0.015]	-0.41[0.014]	-0.41[0.017]	-0.41[0.021]
Comorbidity	No				
	Yes	-1.416[0.0000]	-1.28[0.0000]	-1.27[0.0000]	-1.24[0.0000]
Status at admission	Asymptomatic				
	Mild	-2.18[0.05]	-1.94[0.053]	-1,75[0.0499]	-144[0.019]
	Moderate	-3.63[0.0004]	-3.23[0.0005]	-2.95[0.0004]	-2.46[0.0001]
	Sever	-4.32[0.0001]	-0.12[0.0000]	-3.60[0.0000]	-3.15[0.0000]
HIV/AIDS	No				
	Yes	-0.95[0.0000]	-0.86[0.0000]	-0.88[0.0007]	-0.90[0.0000]
Hypertension	No				
	Yes	-0.64[0.0007]	-0.58[0.0012]	-0.61[0.0007]	-0.59[0.0011]
Symptom at admission	Yes				
	No	-0.56[0.001]	-0.51[0.001]	-0.52[0.0013]	-0.51[0.0031]
Oxygen use	No				
	Yes	-0.91[0.0000]	-0.82[0.0000]	-0.81[0.0000]	-0.78[0.0000]
Diabetics	No				
	Yes	-1.31[0.0000]	-1.16[0.0000]	-1.16[0.0000]	-1.16[0.0000]

Source: Jimma Zone COVID-19 Treatment Centers, southwest Ethiopia from May 16, 2020 through March 9, 2022

4.3.1 Model Selection

The value of AIC for all AFT models are displayed in table 4.8. A model with smaller value of AIC can be considered as a better model compared to other models under consideration because it is valid for comparing models that are not nested. The AIC value for log-logistic AFT model is the smallest compared to other AFT models. This indicates that the log-logistic AFT model better fits the COVID-19 patients data.

Table 4.8: Comparisons of AFT models using AIC

Distribution	**AIC**
Exponential	1220.387
Weibull	1218.267
Log-logistic	**1215.298**
Log-normal	1216.916

4.3.2 Log-logistic Accelerated Failure Time Model

Table 4.9 displays the estimated regression coefficient values for the log-logistic AFT model using multivariable analysis and backward elimination. Age, comorbidity, status at admission, HIV/AIDS, symptom at admission, intranasal oxygen usage, and diabetes all had a significant impact on the survival times of COVID-19 patients.

As shown in table 4.9, the estimated acceleration factor for patients whose ages 18 up to 45 and above 45 compared with age under 18 is estimated to be 0.28 with 95% CI [0.09, 0.86] and 0.14 with 95% CI [0.05, 0.43] respectively. By holding all other model parameters constant, this suggests that the expected survival time for COVID-19 patients falls by 72% and 86% for age 18-45 and above 45 respectively as compared to age less than 18. With a 95% confidence interval of [0.114, 0.4255], the estimated acceleration factor for patients with comorbidity is 0.29. Accordingly, individuals with comorbidities had a 71% shorter predicted survival time than those without comorbidities by holding all other model parameters constant. Patients who experienced moderate or severe status at the time of their admission were estimated to have an

acceleration factor of 0.12 with 95% CI [0.2664, 0.9062] and 0.06 with 95% CI [0.01, 0.29] respectively, compared to asymptomatic status at admission. Consequently, as compared to patients who were asymptomatic status at admission and all model parameters are held equal, the anticipated survival time of COVID-19 patients decreases by 88% and 94% for moderate and severe status, respectively.

The estimated acceleration factor for patients who had HIV/AIDS is 0.37, with a 95% CI of [0.261, 1.76]. This indicates that, while other variables in the model remain constant, the estimated survival time of COVID-19-infected individuals falls by 63% when compared to those without HIV/AIDS. For patients with no symptoms at admission, the estimated acceleration factor was 0.47, with a 95% confidence interval [0.13, 1.15]. Accordingly, when all model parameters were held constant, patients who were without symptom at entry had a 53% shorter survival time than those who were with symptom.

With a 95% confidence range of [0.15, 1.03], the estimated acceleration factor for patients who took intranasal oxygen is 0.49. When comparing patients who utilized intranasal oxygen to those who did not, the expected survival time of COVID-19-infected patients was reduced by 51% while other model parameters remained the same. The estimated acceleration factor for diabetic patients is 0.62, with a 95 percent confidence range of [0.43, 0.90]. This suggests that, when all model parameters are maintained constant, the estimated survival time of COVID-19-infected people with diabetes is reduced by 38% compared to those without diabetes.

We now derive model-based predictions. Using equation (18 and 19), the fitted Loglogistic survival function and hazard function for the i^{th} individual is

$$\hat{S}_i(t) = \{1 + t^{1/\hat{\sigma}} \exp(\hat{\psi}_i)\}^{-1} = \{1 + t^{1/0.714} \exp(\hat{\psi}_i)\}^{-1}, \text{ and}$$

$$\hat{\lambda}_i(t) = \frac{1}{\hat{\sigma}t}\left\{1 + t^{\frac{-1}{\hat{\sigma}}} \exp(-\hat{\psi}_i)\right\}^{-1} = \frac{1}{0.714t}\left\{1 + t^{\frac{-1}{0.714}} \exp(-\hat{\psi}_i)\right\}^{-1}.$$

Where $\hat{\psi}_i = \frac{-\mu - \beta\hat{X}_i}{\hat{\sigma}}$

$$= \frac{1}{0.714}\{-7.214 + 1.26\text{Age(18-45)} + 1.95\text{Age(above 45)} + 0.25SexMale + 1.22ComorbdtyYes$$
$$+ 1.27Mild + 2.15Moderate + 2.73sever - 0.99HIVyes + 0.32HyprtensionYes$$
$$+ 0.75SymptomNo + 0.70oxgenuseYes + 0.48DiabetesYes\}$$

Table 4.9: The fitted Multivariable Loglogistic AFT model

Covariate	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	$\hat{\phi}$	p-value	95% CI[$\hat{\phi}$]
	<18					
Age	18-45	-1.2599	0.5661	0.28	0.02605	[0.09,0.86]
	>45	-1.9475	0.5698	0.14	0.00063	[0.05,0.43]
Sex	Female					
	Male	-0.2474	0.1645	0.78	0.13262	[0.56,1.08]
	No					
Comorbidity	Yes	-1.2219	0.3998	0.29	0.00224	[0.13,0.64]
Status	Asymptomatic					
at	Mild	-1.2736	0.8320	0.28	0.12583	[0.05,1.43]
admission	Moderate	-2.1462	0.7684	0.12	0.00522	[0.02,0.53]
	Sever	-2.7323	0.7725	0.06	0.00040	[0.01,0.29]
	No					
HIV/AIDS	Yes	-0.9923	0.3868	0.37	0.01032	[0.26,0.76]
	No					
Hypertension	Yes	-0.3247	0.1782	0.72	0.06839	[0.51,1.02]
Symptom at	Yes					
admission	No	-0.7549	0.2008	0.47	0.00017	[0.13,0.75]
Oxygen	No					
use	Yes	-0.7007	0.2090	0.49	0.00080	[0.15,0.83]
	No					
Diabetics	Yes	-0.4764	0.1916	0.62	0.01289	[0.43,0.90]
Intercept		7.2137	0.9303	1357	0.0000	[219.29,8409.11]
		Scale(σ) =	0.714			

$\hat{\beta}$: coefficient estimate; $\hat{\phi}$: indicates Acceleration factor (time ratio); 95%CI [$\hat{\phi}$]: 95% confidence interval for acceleration factor; SE: standard error.

4.4 Adequacy of Accelerated Failure Time

4.4.1 Quantile-Quantile plot

The Q-Q plot of the fitted model with adequacy fit for accelerated factor for the failure time has fitted linear or not. By plotting different prognostic covariates with q-q plots has checked the adequacy for fit failure-time of COVID-19 data set using log-logistic AFT model. The adequacy of the failure time model is well-fitted within the significant prognostic covariate groups shown in the figure 4.5.

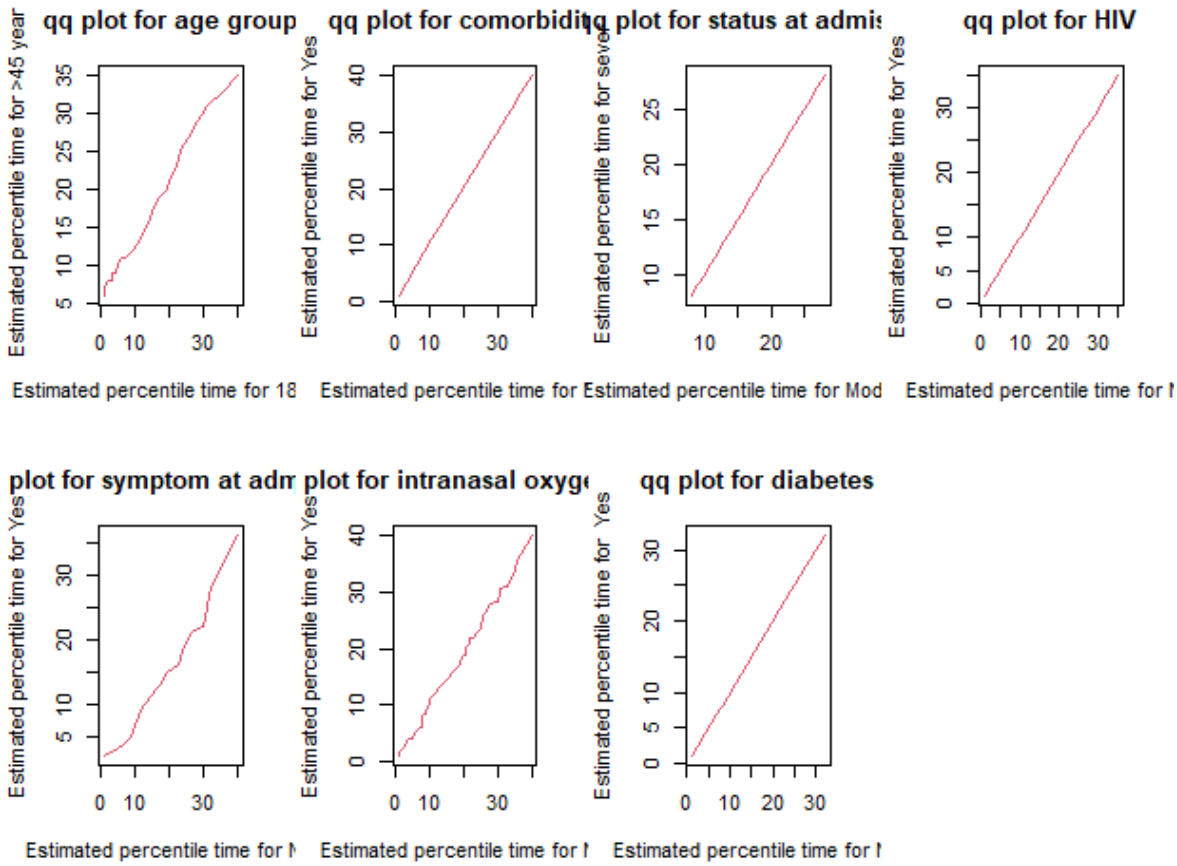

Figure 4.5: Quantile-Quantile plot of fitted log-logistic AFT model

4.4.2 The Cox Snell Residual Plots

The log-logistic AFT model appears to be an appropriate AFT model according to AIC compared with other AFT models in multi-variable analysis, although it is only slightly

better than exponential and Weibull model. However the Log-normal AFT model is best in the uni-variable analysis but has poor fit according to AIC in multi-variable analysis. Furthermore, we check the goodness of fit of the model using residual plots. Cumulative hazard plot of the Cox-Snell residuals in AFT model is presented in figure 4.6. The plotted points lie on a line that has a unit slope and zero intercept for log-logistic model. So there is no reason to doubt the suitability of this fitted log-logistic model. At last,we conclude that the Log-logistic model is the best fitting the AFT model based on AIC criteria and residuals plot containing the statistically significant covariates age, comorbidity, status at admission, HIV/AIDS, symptom at admission, intranasal oxygen use and diabetes.

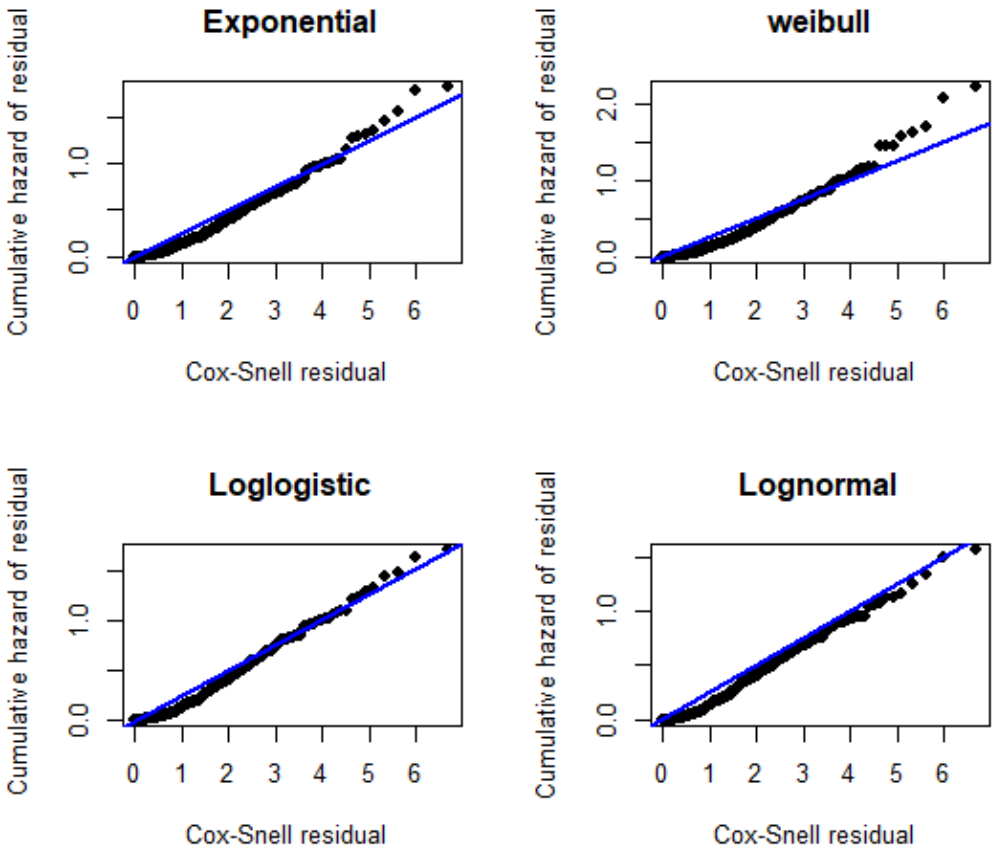

Figure 4.6: *Cox-Snell residual plots of staging subgroup for Exponential model; Weibull model; Weibull model ,Log-logistic model and Log-normal comparing these graphs, the straight line in the Log-logistic plot appears to provide the best fit to the COVID-19 patients data.*

4.5 Discussion of the result

The main goal of this study is to investigate the time-to-death of COVID-19 patients at Jimma Zone, southwest Ethiopia. The COX-PH model was first used to analyze the data. However, because the proportionality assumption of the Cox-PH model was violated, baseline distributions including exponential, Weibull, log-logistic, and log-normal were considered using AFT models. To evaluate several AFT models, AIC has been used, and the log-logistic AFT model was found to fit the time-to-death analysis of COVID-19 patients better than the others. Multivariable analysis showed that age, comorbidity, status at admission, HIV/AIDS, symptoms at admission, intranasal oxygen use, and diabetes were significantly associated with time-to-death in COVID-19 patients.

As a result of this study, it was discovered that the age of the patients is strongly correlated with the predictor variable for time till death. Accordingly, this study found that patients over the age of 45 had a nearly double-increased chance of dying compared to those under the age of 18, which was corroborated by a number of other investigations (Noor et al., 2020; Palaiodimos & Kokkinidis, 2020). Additionally in line with previous findings that reported old-aged peoples had an increased risk of death due to COVID-19 disease (Zhou et al., 2020). This may be due to the deterioration of physiologic functioning, pre-existing age-related immunosuppression, and increasing risk of comorbidities with advancing age, which further complicates the prognosis of COVID-19 patients' treatments (Kang & Jung, 2020).

According to this study patients who had comorbidity also have a greater risk of death than their counterparts, which is supported by studies Ayana et al. (2021); Bello-Chavolla & Bahena-López (2020). Because COVID-19 patients with these comorbidities are more prone to acquire serious health issues that result in immunosuppression and poor treatment results, a cytokine storm and the creation of an immediate hyper-inflammatory response might make these fatalities worse (Bhaskar et al., 2020). This study also found that status at admission is significantly associated with time

to death of patients from COVID-19 pandemic. Thus, moderate and sever status at admission had decreased the survival rate of COVID-19 infected patients as compared to asymptomatic state at admission. This finding is in agreement with the study of factors associated with death outcome in patients with severe coronavirus disease-19 (Pan et al., 2020). Because patients with moderate and severe status had more prominent laboratory abnormalities than those with asymptomatic status (Guan et al., 2020).

Similarly, the time to death of HIV-infected COVID-19 patients was significantly lower than that of HIV-negative individuals in this analysis. The same finding was revealed in a study conducted among patients admitted to Wuhan pulmonary hospital in China (Du et al., 2020). This might be explained by the adverse effects of comorbidities on autoimmune response and metabolic stress that characterizes systemic diseases and decreases the ability to respond against pathogenic agents (Zhou et al., 2020).

According to this study, patients with no symptoms have a greater risk of dying than those with symptoms. A Korean study that found that initially asymptomatic patients are the best predictors of patient death lends credence to this finding (Park et al., 2021). This may be because people with symptoms are more likely to go to a medical facility sooner than those without. Additionally, patients with early symptoms may be more likely to require ICU care Park et al. (2021) than patients without early symptoms (Dessie et al., 2022). Early detection and treatment of the disease have reduced the death rate of COVID-19, helping patients avoid complications. It may also be related to how these asymptomatic individuals cope in terms of immunity.

This study also revealed that the risk of death from COVID-19 infections was greater in people who received intranasal oxygen therapy than in people who did not. This result is in line with the study conducted in Ethiopia on the survival analysis of patients with COVID-19, which found that patients admitted to the ICU and receiving intranasal oxygen supplementation had a higher probability of dying from COVID-19 infections than non-ICU patients did (Kaso & Agero, 2022). Additionally, this result is supported by the earlier study Hu et al. (2020) that found patients getting these services had a

higher likelihood of poor treatment results also lends weight to this conclusion. The observed mortality may be attributable to patients who were given intranasal oxygen therapy having an unregulated autoimmune response and being at a high degree of disease severity.

Finally, those who had diabetes were more likely to pass away than their non-diabetic counterparts. This finding is consistent with a study from China that looked at the clinical trajectory and mortality risk factors of adult inpatients with COVID-19 and discovered that those with diabetes had a greater chance of dying from COVID-19 infections than people without diabetes (Zhou et al., 2020). This could be due to the fact that the activity of cytokines depending on type I helper T cells is disrupted by glycosylation of cytokines because innate immunity is reduced in diabetes individuals as a result of high blood glucose levels (Rashedi & Poor, 2020). One limitation of the present study is its high percentage of right-censored observations. The right censored observations in this study is about 83.3 percent. To reach an appropriate fit for parametric models, it is better not to have right-censored observations more than 40 to 50 percent (Nardi & Schemper, 2003).

CHAPTER FIVE

5 Conclusion and Recommendations

5.1 Conclusions

Various parametric acceleration failure time models were used for this thesis to analyze time to death of COVID-19 patients. From that log-logistic acceleration failure time model was best fitted model for COVID-19 data set. Based on this result the variables age, comorbidity, admission status, HIV/AIDS, symptoms at admission, intranasal oxygen use, and diabetes were found to be associated with time-to-death in COVID-19 patients. From this finding elders, individuals with comorbidity, patients with moderate or severe status at admission, those who were asymptomatic at admission, patients who have HIV/AIDS, individuals who received intranasal oxygen therapy and diabetes were more likely to die from COVID-19 disease than their counterparts.

5.2 Recommendations

Based on the findings of the study, the recommendation is as follows

- To increase the survival time of COVID- 19 patients, it is important to closely monitor aged patients, patients with comorbidity, moderate and severe patients at admission, patients with HIV/AIDS, asymptomatic patients, and patients with diabetes.
- As a result, when the Cox model's assumption of proportionality is not satisfied, researchers studying COVID-19 patients should consider the AFT model as an alternative.
- Further studies should be conducted to identify other factors that are not identified in this study.

References

Ababa, A. (2003). Federal democratic republic of ethiopia ministry of health. *Ethiopia: Postnatal Care*.

Alqahtani, J. S., & Oyelade. (2020). Prevalence, severity and mortality associated with copd and smoking in patients with covid-19: a rapid systematic review and meta-analysis. *PloS one*, *15*(5), e0233147.

Altman, D., De Stavola, B., Love, S., & Stepniewska, K. (1995). Review of survival analyses published in cancer journals. *British journal of cancer*, *72*(2), 511–518.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). Multivariate time scales. In *Statistical models based on counting processes* (pp. 675–708). Springer.

Ayana, G. M., Merga, B. T., Birhanu, A., Alemu, A., Negash, B., & Dessie, Y. (2021). Predictors of mortality among hospitalized covid-19 patients at a tertiary care hospital in ethiopia. *Infection and Drug Resistance*, *14*, 5363.

Bello-Chavolla, O. Y., & Bahena-López. (2020). Predicting mortality due to sars-cov-2: a mechanistic score relating obesity and diabetes to covid-19 outcomes in mexico. *The Journal of Clinical Endocrinology & Metabolism*, *105*(8), 2752–2761.

Bhaskar, S., Sinha, A., & Banach. (2020). Cytokine storm in covid-19—immunopathological mechanisms, clinical considerations, and therapeutic approaches: the reprogram consortium position paper. *Frontiers in immunology*, *11*, 1648.

Cain, K. C., & Lange, N. T. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, 493–499.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Covid, W. (19). *Coronavirus pandemic*.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Dätwyler, C., & Stucki, T. (2011). *Parametric survival models.*

Dessie, A. M., Animut, Z., Genet, A., & Yenew, C. (2022). Incidence of death and its predictors of covid-19 in amhara region, ethiopia: A retrospective follow up study. *Infection and Drug Resistance*, *15*, 4907–4913.

Du, R.-H., Liang, et al. (2020). Predictors of mortality for patients with covid-19 pneumonia caused by sars-cov-2: a prospective cohort study. *European Respiratory Journal*, *55*(5).

Fagbamigbe, A., Obiyan, M., & Fawole, O. (2018). Parametric survival analysis of menarche onset timing among nigerian girls. *Heliyon*, *4*(12), e01105.

Ghebreyesus, T. (n.d.). *World health organization. who director-general's opening remarks at the media briefing on covid-19-25 may 2020.*

Gilbert, M., Pullano, et al. (2020). Preparedness and vulnerability of african countries against importations of covid-19: a modelling study. *The Lancet*, *395*(10227), 871–877.

Guan, W.-j., Ni, et al. (2020). Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, *382*(18), 1708–1720.

Habenom, H., Aychluh, M., Suthar, D., Al-Mdallal, Q., & Purohit, S. (2022). Modeling and analysis on the transmission of covid-19 pandemic in ethiopia. *Alexandria Engineering Journal*, *61*(7), 5323–5342.

Hu, L., Chen, S., Fu, et al. (2020). Risk factors associated with clinical outcomes in 323 coronavirus disease 2019 (covid-19) hospitalized patients in wuhan, china. *Clinical infectious diseases*, *71*(16), 2089–2098.

Iboi, E. A., Ngonghala, C. N., & Gumel, A. B. (2020). Will an imperfect vaccine curtail the covid-19 pandemic in the us? *Infectious Disease Modelling*, *5*, 510–524.

Impouma, B., Carr, et al. (2022). Time to death and risk factors associated with mortality among covid-19 cases in countries within the who african region in the early stages of the covid-19 pandemic. *Epidemiology & Infection*, *150*.

Jiezhi, Q. (2009). Comparison of proportional hazards and accelerated failure time models. *Master's Thesis, University of Saskatchewan, Saskatoon, SK, Canada*, 79.

Kalbfleisch, J. D., & Prentice, R. L. (1973). Marginal likelihoods based on cox's regression and life model. *Biometrika*, *60*(2), 267–278.

Kang, S.-J., & Jung, S. I. (2020). Age-related morbidity and mortality among patients with covid-19. *Infection & chemotherapy*, *52*(2), 154.

Kaso, A. W., & Agero. (2022). Survival analysis of covid-19 patients in ethiopia: A hospital-based study. *Plos one*, *17*(5), e0268280.

Kassaw, C. (2020). The magnitude of psychological problem and associated factor in response to covid-19 pandemic among communities living in addis ababa, ethiopia, march 2020: a cross-sectional study design. *Psychology Research and Behavior Management*, *13*, 631.

Kebede, Z., & Ababor. (2020). Rapid evidence synthesis on covid-19 pandemic to inform the ethiopian ministry of health. *Ethiopian Journal of public health and nutrition*, *4*(1).

Klein, J. P., & Moeschberger, M. L. (1997). Survival analysis: Techniques for censored and truncated data springer. *New York*.

Kleinbaum, D., & Klein, M. (2005). *Survival analysis: A self-learning text . new york: Singer science+ business media.* Inc.

Kleinbaum, D. G., Klein, M., et al. (2012). *Survival analysis: a self-learning text* (Vol. 3). Springer.

Kundu, S., Chauhan, K., Mandal, D., et al. (2021). Survival analysis of patients with covid-19 in india by demographic factors: Quantitative study. *JMIR Formative Research*, *5*(5), e23251.

Lemeshow, S., May, S., & Hosmer Jr, D. W. (2011). *Applied survival analysis: regression modeling of time-to-event data*. John Wiley & Sons.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, *22*(4), 719–748.

Moran, J. L., & Bersten. (2008). Modelling survival in acute severe illness: Cox versus accelerated failure time models. *Journal of Evaluation in Clinical Practice*, *14*(1), 83–93.

Moti, U. G., & Ter Goon, D. (2020). Novel coronavirus disease: A delicate balancing act between health and the economy. *Pakistan journal of medical sciences*, *36*(COVID19-S4), S134.

Nardi, A., & Schemper, M. (2003). Comparing cox and parametric models in clinical studies. *Statistics in medicine*, *22*(23), 3597–3610.

Noor, F. M., Islam, M., et al. (2020). Prevalence and associated risk factors of mortality among covid-19 patients: a meta-analysis. *Journal of community health*, *45*(6), 1270–1282.

Palaiodimos, L., & Kokkinidis. (2020). Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with covid-19 in the bronx, new york. *Metabolism*, *108*, 154262.

Pan, F., Yang, L., Li, Y., et al. (2020). Factors associated with death outcome in patients with severe coronavirus disease-19 (covid-19): a case-control study. *International journal of medical sciences*, *17*(9), 1281.

Park, H. C., Kim, D. H., Cho, A., Kim, J., Yun, K.-s., Kim, J., & Lee, Y.-K. (2021). Clinical outcomes of initially asymptomatic patients with covid-19: a korean nationwide cohort study. *Annals of medicine*, *53*(1), 357–364.

Ponnuraja, C., & Venkatesan, P. (2010). Survival models for exploring tuberculosis clinical trial data-an empirical comparison. *Indian Journal of Science and Technology*, *3*(7), 755–758.

Porta, M. (2014). *A dictionary of epidemiology*. Oxford university press.

Rashedi, J., & Poor, M. (2020). Risk factors for covid-19. *Infez Med*, *28*(4), 469–474.

Rojas, M., & Rodríguez. (2020). Convalescent plasma in covid-19: Possible mechanisms of action. *Autoimmunity reviews*, *19*(7), 102554.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (covid-19) outbreak. *Journal of autoimmunity*, *109*, 102433.

Roychowdhury, D. (2020). 2019 novel coronavirus disease, crisis, and isolation. *Frontiers in Psychology*, *11*, 1958.

Salinas-Escudero, G., & Carrillo-Vega. (2020). A survival analysis of covid-19 in the mexican population. *BMC public health*, *20*(1), 1–8.

Sayehmiri, K., Eshraghian, et al. (2008). Prognostic factors of survival time after hematopoietic stem cell transplant in acute lymphoblastic leukemia patients: Cox proportional hazard versus accelerated failure time models. *Journal of Experimental & Clinical Cancer Research*, *27*(1), 1–9.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241.

Subbaraman, R., Chaguturu, S. K., Mayer, K. H., Flanigan, T. P., & Kumarasamy, N. (2007). Adverse effects of highly active antiretroviral therapy in developing countries. *Clinical Infectious Diseases*, *45*(8), 1093–1101.

Therneau, T. M., & Grambsch, P. M. (2000). The cox model. In *Modeling survival data: extending the cox model* (pp. 39–77). Springer.

Tolossa, T., Wakuma, B., Ayala, D., et al. (2022). Incidence and predictors of death from covid-19 among patients admitted to treatment center of wollega university

referral hospital, western ethiopia: A retrospective cohort study. *PloS one*, *17*(7), e0267827.

Villela, D. A. M. (2020). *The value of mitigating epidemic peaks of covid-19 for more effective public health responses* (Vol. 53). SciELO Brasil.

Walters, S. J. (2012). Analyzing time to event outcomes with a cox regression model. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3), 310–315.

Wang, H., & Zhang, L. (2020). Risk of covid-19 for patients with cancer. *The Lancet Oncology*, *21*(4), e181.

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *jama*, *323*(13), 1239–1242.

Zeng, D., & Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(4), 507–564.

Zhou, F., Yu, et al. (2020). Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet*, *395*(10229), 1054–1062.

Zikargae, M. H. (2020). Covid-19 in ethiopia: assessment of how the ethiopian government has executed administrative actions and managed risk communications and community engagement. *Risk management and healthcare policy*, *13*, 2803.

Appendices

A. K-M estimate of survival of patients with COVID-19

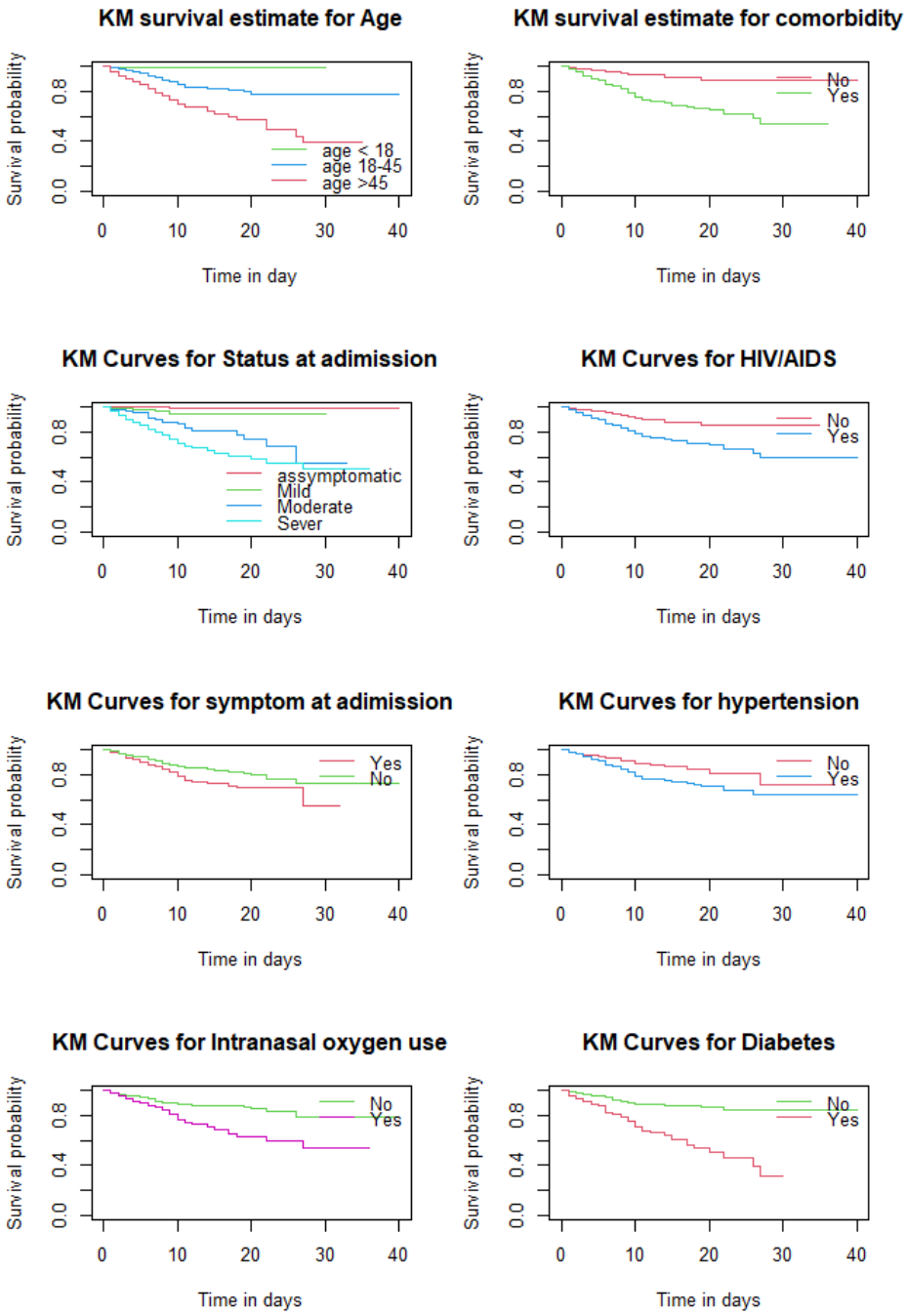

Figure 5.1: Kaplan-Meier Estimate of Survival of Patients with COVID-19

B. Log (-Log (Survival)) Versus Survival Time for Variables

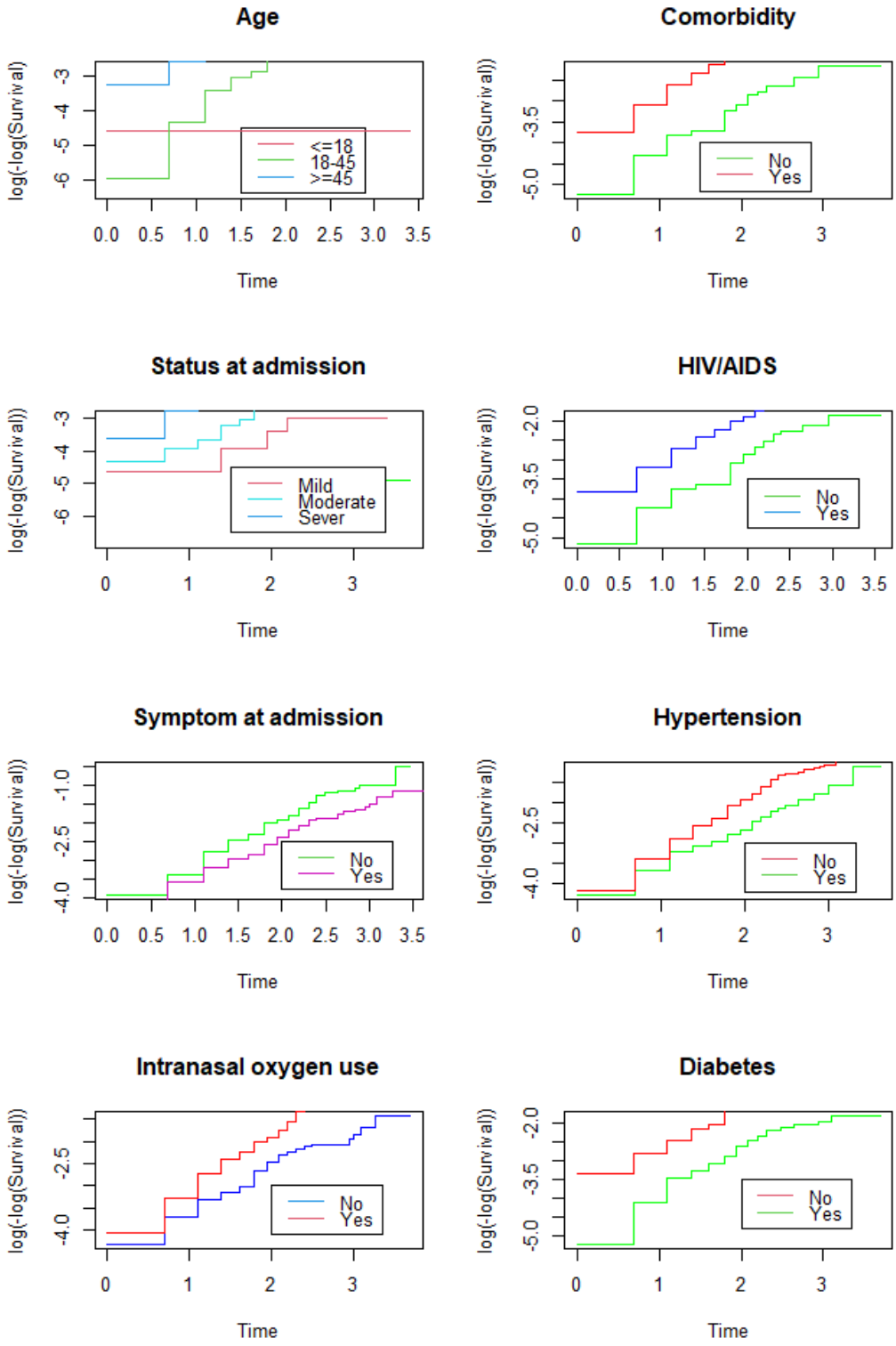

Figure 5.2: Plot of Log (-Log (Survival)) Versus Survival Time for Categorical Variables

Table 5.1: Result of Cox PH model

Covariates	Categories	β	HR	se(β)	Z	P-value	95%CI for HR
Age	<18						
	18-45	1.7057	5.5051	0.7540	2.262	0.0237	[1.25,24.13]
	>45	2.4775	11.9112	0.7493	3.306	0.0009	[2.74,51.74]
Residence	Rural						
	Urban	0.4218	1.5246	0.2126	1.984	0.0473	[1.00,2.31]
Sex	Female						
	Male	0.3192	1.3760	0.1877	1.701	0.089	[0.95,1.99]
comorbidity	No						
	Yes	1.3039	3.6837	0.4238	3.077	0.002	[1.60,8.45]
Status at admission	Asymptomatic						
	Mild	1.5431	4.6792	1.1388	1.355	0.1754	[0.50,43.60]
	Moderate	2.6428	14.0524	1.0424	2.535	0.0112	[1.82,108.40]
	Sever	3.2925	26.9094	1.0381	3.172	0.0015	[3.52,205.86]
HIV/AIDS	No						
	Yes	1.0060	2.7346	0.4103	2.452	0.0142	[1.61,4.82]
Hypertension	No						
	Yes	0.5316	1.7017	0.2041	2.605	0.0092	[1.14,2.54]
Symptom at admission	Yes						
	No	0.9683	2.6334	0.2413	4.012	0.0000	[1.24,5.61]
Oxygen use	No						
	Yes	0.6352	1.8874	0.2395	2.652	0.0080	[1.03,3.85]
Diabetics	No						
	Yes	0.6122	1.8445	0.2197	2.787	0.0053	[0.20,2.83]

Table 5.2: Result of maximum likelihood parameter estimates of the Exponential AFT model

Covariate	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	$\hat{\phi}$	p-value	95% CI[$\hat{\phi}$]
	<18					
Age	18-45	-1.623	0.756	0.20	0.03186	[0.04,0.87]
	>45	-2.361	0.751	0.09	0.00167	[0.02,0.41]
Sex	Female					
	Male	-0.305	0.187	0.74	0.10231	[0.51,1.06]
	No					
comorbidity	Yes	-1.496	0.442	0.22	0.0007	[0.09,0.53]
Status	Asymptomatic					
at	Mild	-1.683	1.135	0.18	0.13833	[0.02,1.72]
admission	Moderate	-2.791	1.041	0.06	0.00732	[0.08,0.47]
	Sever	-3.394	1.037	0.03	0.00107	[0.004,0.26]
	No					
HIV/AIDS	Yes	-1.192	0.426	0.30	0.00514	[0.04,0.76]
	No					
Hypertension	Yes	-0.431	0.200	0.65	0.0311	[0.44,0.96]
Symptom at	Yes					
admission	No	-0.825	0.232	0.44	0.00038	[0.13,0.59]
Oxygen	No					
use	Yes	-0.649	0.240	0.52	0.00682	[0.09,0.73]
	No					
Diabetics	Yes	-0.535	0.218	0.58	0.01420	[0.38,0.90]
Intercept		8.944	1.193	7665.01	0.0000	[740.22,79371.55]

$\hat{\beta}$: coefficient estimate; $\hat{\phi}$:indicates Acceleration factor; 95%CI [$\hat{\phi}$]: 95% confidence interval for acceleration factor; SE: standard error.

Table 5.3: Result of maximum likelihood parameter estimates of the Weibull AFT model

Covariate	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	$\hat{\phi}$	p-value	95% CI[$\hat{\phi}$]
	<18					
Age	18-45	-1.363	0.368	0.25	0.0327	[0.07,0.89]
	>45	-1.983	0.642	0.14	0.0020	[0.04,0.48]
Sex	Female					
	Male	-0.270	0.156	0.76	0.0838	[0.56,1.04]
	No					
comorbidity	Yes	-1.277	0.378	0.28	0.0007	[0.13,0.58]
Status	Asymptomatic					
at	Mild	-1.450	0.949	0.23	0.1265	[0.04,1.51]
admission	Moderate	-2.344	0.881	0.09	0.0078	[0.02,0.54]
	Sever	-2.851	0.885	0.06	0.0013	[0.01,0.33]
	No					
HIV/AIDS	Yes	-1.024	0.362	0.36	0.0047	[0.21,0.76]
	No					
Hypertension	Yes	-0.374	0.168	0.69	0.0257	[0.49,0.95]
Symptom at	Yes					
admission	No	-0.725	0.196	0.47	0.0002	[0.03,0.65]
Oxygen	No					
use	Yes	-0.531	0.204	0.59	0.0094	[0.15,0.69]
	No					
Diabetics	Yes	-0.452	0.184	0.64	0.0116	[0.44,0.91]
Intercept		7.915	1.064	2737.57	0.0000	[340.29,22022.89]

$\hat{\beta}$: coefficient estimate; $\hat{\phi}$:indicates Acceleration factor; 95%CI [$\hat{\phi}$]: 95% confidence interval for acceleration factor; SE: standard error.

Table 5.4: Result of maximum likelihood parameter estimates of the Log-normal AFT model

Covariate	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	$\hat{\phi}$	p-value	95% CI[$\hat{\phi}$]
	<18					
Age	18-45	-0.774	0.415	0.46	0.0622	[0.20,1.04]
	>45	-1.526	0.421	0.22	0.0003	[0.09,0.49]
Sex	Female					
	Male	-0.214	0.169	0.81	0.205	[0.58,1.12]
	No					
comorbidity	Yes	-1.302	0.432	0.27	0.0026	[0.12,0.63]
Status	Asymptomatic					
at	Mild	-1.095	0.639	0.33	0.0866	[0.09,1.17]
admission	Moderate	-1.919	0.579	0.15	0.0009	[0.04,0.46]
	Sever	-2.547	0.586	0.08	0.0013	[0.02,0.25]
	No					
HIV/AIDS	Yes	-1.078	0.422	0.34	0.0107	[0.08,0.66]
	No					
Hypertension	Yes	-0.266	0.178	0.77	0.135	[0.54,1.09]
Symptom at	Yes					
admission	No	-0.761	0.199	0.47	0.0001	[0.14,0.58]
Oxygen	No					
use	Yes	-0.741	0.210	0.48	0.0004	[0.15,0.72]
	No					
Diabetics	Yes	-0.532	0.186	0.59	0.0116	[0.41,0.84]
Intercept		6.588	0.697	726.52	0.0000	[185.16,2850.62]

$\hat{\beta}$: coefficient estimate; $\hat{\phi}$:indicates Acceleration factor; 95%CI [$\hat{\phi}$]: 95% confidence interval for acceleration factor; SE: standard error.