AIMS | African Institute for Mathematical Sciences CAMEROON

Biomarkers and Mortality in Ebola Virus Disease Patients

Mavis Amoa-Dadzeasah (mavis.amoa@aims-cameroon.org)
African Institute for Mathematical Sciences (AIMS)
Cameroon

Supervised by: Dr. Christiana Kartsonaki, Laura Merson, Mr. Trokon Yeabah
University of Oxford, United Kingdom

29 May 2022

Submitted in Partial Fulfilment of a Structured Master's Degree at AIMS-Cameroon

Abstract

Mortality from Ebola virus disease has been a bane to the African population since the onset of the disease in 1976. Factors culminating in the death of positively tested patients have dominated studies around the disease. Research has identified clinical features such as diarrhoea, vomiting, fever, myalgia, conjuctivitis, among other symptoms to be predictive of death. However, fewer studies have been conducted which investigate blood kinetics, biomarkers, and other blood components like viremia of diseased patients. Therefore, the aim of this work was to investigate the association of laboratory biomarkers such as anion gap, chloride, hemoglobin, hematocrit, sodium, and potassium with mortality of diseased patients. Additionally, we wanted to examine the added effect of co-infection with malaria and high or low cycle threshold (C_t) values, to death from Ebola. Polymerase chain reaction (PCR) positively tested patients from Guinea, Sierra Leone, and Liberia were included in the analysis. Results from the survival analysis showed that age; low sodium, carbon dioxide, calcium ionized; high urea nitrogen; and, $C_t \geq 27$ jointly with malaria co-infection were predictive of death in patients. However, sex and independent C_t values were not predictive of death.

Keywords: Ebola virus, mortality, survival, laboratory biomarkers, co-infection, cycle threshold.

Declaration

I, the undersigned, hereby declare that the work contained in this essay is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Mavis Amoa-Dadzeasah, 29 May 2022.

Contents

1. Introduction

This chapter gives a brief opening and background to the topic of interest. Here, we discuss the background of the study, then state and expound on the problem this research addresses. Following this, we proceed to enumerate the objectives of this research before giving some justification on the relevance of this work.

1.1 Background of Study

Ebola Virus Disease (EVD), popularly known as Ebola, is a viral hemorrhagic fever that destroys the vascular system of its host organism, thus preventing the body from regulating itself. This viral disease was named after a river in a nearby village where the virus was first discovered, in the Democratic Republic of Congo. Among the many symptoms of EVD is hemorrhage (i.e., bleeding), which occurs in different forms including petechiae (red clustered rashes on the skin); ecchymosis (bleeding from underneath the skin); bleeding from phlebotomy sites; and, frank hemorrhages. Other symptoms of the disease are restlessness, muscle pain, joint stiffness, nausea, diarrhea, fever, headache, etc. These signs and symptoms increase as the disease progresses, into very fatal complications. The virus responsible for this disease is thought to be naturally inhabited in fruit bats, and some primates such as monkeys and chimpanzees. EVD is a zoonotic disease, implying that humans are able to contract it and transmit it to fellow humans. The known modes of transmission are mainly through body contact with the blood of an infected species or corpse; contact with bodily fluids such as saliva, sweat, urine, and semen; sex; and objects with infected liquid samples. Symptoms manifest about eight to ten days after contact with the virus [20]. As stated earlier, symptoms become complicated as the disease progresses, and patients may die between six and sixteen days of complications including multiorgan failure and septic shock. The World Health Organization (WHO) reported an average case fatality rate of 50% in February 2021 and added that these rates have varied from 25% to 90% in the past outbreaks. Less than a month after this publication, Malvy D. et al reported a case fatality of 70%, which indicates an increase in mortality rates [20]. This means that, among the population of infected persons, 70% are likely to die.

An Ebola outbreak in West Africa was declared a public health emergency of international concern by the WHO in 2014 when it rapidly spread through the populations in Guinea, Liberia and Sierra Leone. The outbreak ended in 2016 following 28,616 cases and 11,310 deaths [4]. However, Ebola remains a challenge due to the subsequent high number of cases recorded in different countries. There have been no significant changes in the case fatality rates in the past four decades according to [25]. The number of cases, and subsequently, the case fatality rate being high (above 50%) could be attributed to other factors aside the virus. In July 2020, Kamorudeen et al. [14] conducted research that identified some key health determinants associated with the high mortality rate of Ebola, and found social and economic factors such as poor sanitation, intercurrent diseases, [1] individual health status, health literacy malnutrition, poverty, and other education-related factors. They acknowledged that intercurrent diseases and individual health status are influential factors in high mortality rates. Additionally, sequelae have been reported after infection with the Zaire type of the Ebola virus. Among other sequelae, [2] the identified ones according to Choi et al. are headache, urinary frequency, fatigue, memory loss, photophobia, hyperlacrimation, and conjunctivitis [8]. There is evidence that research is actively ongoing in this area to find health-related factors and repercussions associated with the disease.

[1] An intercurrent disease is a disease that arises in the course of another disease.

[2] A sequela is a condition that arises as a result of a pre-existing illness.

Measures put in place to control the disease have helped, but many people still die and more epidemics are likely to occur in the previously affected areas as well as other newly identified areas. Additionally, the overall control of EVD requires a balance between the biological sciences including epidemiology, health systems, socio-anthropological and political science in order to enable early detection and quick responses to reduce the problem of mortality already discussed [25].

1.2 Problem Statement

The ramifications and mortality rates associated with EVD are egregious. This is being controlled by scientists in producing vaccines to mitigate the lethal effects of the virus, however, case fatality still remains high. The deadliest of the viruses that cause EVD in humans, called the Zaire ebolavirus (EBOV) is responsible for most of the outbreaks, including the two largest outbreaks recorded in history. The vaccine (Ervebo) licensed by the United States Food and Drug Administration against EBOV, was proven to be effective [28], however, it did not address the problem of increased mortality rates. This is because, the duration of protection against EBOV with Ervebo could not be determined, which implies that persons vaccinated could be infected at a later time and still die from the disease [5]. On the other hand, early supportive care including intravenous fluids, electrolyte supplementation, and nutritional support was found to significantly reduce mortality rates to about 40% in 2020 [8].

Strategies being carried out to curtail the spread and severity of this virus seem operative, even though the rates are not reducing significantly. Hence it is recommended that measures and directions given by public health agencies should be strictly adhered to. This is because, in the absence of these regulations, there will be a tremendous spread of the virus which will cause a ripple effect on economies through increased death rates. Early research works have shown that health and economic performance have a strong positive correlation. That is to say, better health will likely improve economic performance and vice versa. Thus, lower mortality rates are a powerful engine for sustained growth because reducing mortality promotes per capita GDP growth while the contrary weakens the economy [23].

Propositions and solutions to this menace remain a bother to many researchers all over. However, little attention has been given to other underlying factors that exacerbate the disease including biomarkers and comorbidities present during hospitalization. A biomarker is a natural characteristic of the blood, through which a physiological process or a disease can be identified. Rojek et al. admitted that the degree to which findings from the 2014 – 2016 epidemic can be relevant to future outbreaks is unknown, however comparing clinical findings from different outbreaks would be beneficial, although would require high-quality data [24]. Principally, clinical findings consider multiple determinants of death in patients, including biomarkers. In their study of prognostic factors for ebola patients' survival, Crowe et. al (2017) studied the behaviour of cycle threshold (C_t) values, and recommended further studies that combine C_t values with biomarkers towards providing more effective treatments [9]. It is against this background that we conduct this research to use the appropriate statistical and survival models to primarily assess the association of laboratory biomarkers such as anion gap, calcium ionized, carbon dioxide, hemoglobin, glucose, and creatinine with mortality in Ebola virus disease patients.

1.3 Motivation and Research Objectives

This research seeks to:

1. Assess the association of some laboratory biomarkers with the risk of death.
2. Determine the survival probabilities associated with positively tested malaria patients in Ebola treatment centres.
3. Investigate the risk associated with cycle threshold values combined with malaria co-infection, and mortality.

Several attempts to reduce the mortality rates around the disease have proven helpful, yet there is still more to be done. It is needful for research of this kind to be carried out to investigate underlying factors that are associated with risk of death. This is useful to effectively manage EVD. Therefore, this research is intended to inform understanding of how laboratory biomarkers and malaria status impact mortality in Ebola patients. With the results of this research, public health agencies, hospitals and patients will be educated on the risks associated with Ebola in order to take the necessary precautions to properly manage the disease during patient hospitalization.

2. Literature Review

In this chapter, we shall review literature on the topic under study. This review includes concepts, theories found in literature, articles, and empirical studies documented in journals, books, and on the internet. We shall consider the similarity in research questions, methodology, as well as different approaches to answering the same research questions.

2.1 Empirical Studies

In this section, we shall explore literature on the prognostic factors of the death of Ebola virus patients, the statistical tools and methods employed to determine these factors, and the conclusions made on these indicators. Following this, we shall select the method that best suits the data obtained and expound on this method in the next chapter.

Qureshi et al. (2015) in their study of high survival rates and associated factors among Ebola patients hospitalized at Donka National Hospital, in Guinea, used Kaplan-Meier survival methods to examine survival in patients after the onset of symptoms within a 4-month period between 25th March and 5th August in 2014. They assessed the correlation between survival and factors such as demographics and other clinical features. With a trial size of 70, of which 44 were men, 42 of the 70 were discharged alive with a survival rate of 60%, among the hospitalized patients. Additionally, they found that the survival rates were higher for 39 patients who were below the mean age of 34, and lower for 30 patients aged 35 years or above, and the rates associated with myalgia were significantly lower among the survivors. Thus, they had substantial evidence from their cohort to conclude that survival in their patients at Donka National Hospital, Conakry, were prominently higher than in other hospitalized cohorts. They recommended further studies into improving survival among patients [22] .

In studying the clinical predictors of mortality in patients with EVD, Barry et al. (2015) conducted a single-centre cohort study and collected data from the Donka National Hospital in Guinea's capital city, Conakry. The study involved 89 patients of which 32 were women. Only patients diagnosed with Ebola and hospitalized at Donka were included in the study. By the use of Mann-Whitney U test, and a multivariate logistic regression model, they determined the factors associated with death. Results showed that 73% of the patients had both EVD and malaria, and recorded a crude mortality rate of 40%, while the general crude mortality rate was 43.8%. They also found that, the main cause of death was hypervolemic shock and multiorgan failure among the 39 patients who died, and the independent factors of death according to the multivariate analysis were hemorrhage, myalgia, and difficulty in breathing. Hence, they recommended further study into myalgia and other features relating to death [2].

Similarly, Zhang et al. (2015) investigated the prognostic factors affecting the survival of Ebola virus patients. Sixty-three (63) laboratory-confirmed patients were recruited from the Jui Government Hospital in September within a three-month period between 1st October 2014 and 18th January 2015. They first investigated the correlation between a single clinical presentation and survival and later evaluated the effect of several variables on survival. Patients with a viral load higher than 10^6 copies/ml had a shorter survival time compared to those with a lesser viral load. Chest pain, coma, and viral load greater than 10^6 copies/ml were strongly associated with low survival. Also, age, and confusion, were highly associated with survival. From their analysis, they suspect that viral load is an important factor for survival from EVD [29].

In 2018, Cherif et al. studied the prognostic and predictive factors of EVD in the elderly population from the 2014 Ebola outbreak in Guinea. Their study included 2,313 positively diagnosed Ebola patients of which 309 were 60 years and above. Of the 2,004 adults (aged < 60), 49.2% were males, and 45.8% were also males in the elderly population. Findings indicated that, the median time from the onset of symptoms to hospitalization was lower in the elderly cohort, however, their median length of hospitalization and survival times were shorter. They also found that the survival curves of the two groups were different: the adults had a higher survival probability than the aged. Similarly, case fatality rate was higher in the elderly population (80.6%). Finally, there was a significant relationship between increasing age and risk of death [OR$= 1.05$]. Nevertheless, no clinical symptom was identified to be strongly correlated with the poor survival in the aged group. However they found funeral participation to be a contributing factor to the spread of the virus in the group. Cox regression showed that age was the influential factor to EVD survival, and hence was considered as an independent predictor of death in the aged group [7].

Crowe et al. (2016) assessed the time from symptom onset to ETU admission and the quantitative real-timer reverse transcription PCR cycle threshold, C_t (< 24 demonstrating high viral load), which was a substitute for viral load of 151 patients admitted at the Bo District ETU in Sierra Leone. [1] In this study which included three cohorts, they found that the average number of days from symptom onset to admission was similar for both the survivors and non-survivors in all three groups. Hence, they concluded that time from symptom onset to admission was not associated with survival. To examine C_t as a predictor of death in non-survivors, the locally weighted scatterplot smoothing (LOESS) curve was used to plot the survival curves of patient survival by C_t in R. They found that C_t was very accurate when used as a dichotomous predictor. They also used unadjusted logistic regression analysis to analyze the association between sex, age, C_t and interval from symptom onset to admission, and survival. Results showed that C_t and age were strongly associated with survival, but both male and female patients had similar odds of survival. Further research into C_t combined with other biomarkers were recommended to provide insights into the efficacy of treatment and prognosis [9].

Prior to their study in 2018, Cherif et al., conducted research investigating predictors of mortality in 2,310 Ebola adult patients in Guinea in 2017. With a case fatality rate of 68.1%, their univariate analysis identified age, history of contact with infected persons, and seven clinical features (hiccups, diarrhea, cough, fever, vomiting, sore throat, and unexplained bleeding) to be associated with mortality outcomes. Furthermore, the multivariate analysis revealed that age was an independent factor of mortality with an odds ratio of 1.06, while no other clinical feature was significant to survival [6].

Arranz et al. (2016) analyzed survival data of 75 Ebola-suspected patients admitted to the Moyamba ETC between December 2014 and March 2015, of whom 31 tested positive for EVD including more women [2]. On admission, most patients showed symptoms of fatigue, diarrhea, muscle pain, and fever. The mortality rate reported among the EVD positive group was 58%, against 11.4% in the negative group. Finally, their study detected that symptoms evolved over the length of stay at the ETC from admission date, which was associated with a poor prognosis of patients after admission [1].

With the objective of providing more accurate prognostic tools that distinctively define and group clinical severity, Hartley et al. conducted a retroactive study in 2017, using data collected from 158 positively tested Ebola patients at the GOAL-Mathaska ETC in Port Loko, Sierra Leone within an 11-month period from December 2014 to November 2015. In all, 566 patients were included in the study, of which 158 (27.9%) were positive with EVD. Kaplan-Meier estimator was used to investigate the survival in both

[1] ETU - Ebola Treatment Unit, PCR - polymerase chain reaction
[2] ETC - Ebola Treatment Centre

groups. Of the 566, 188 had malaria (34.8%) positive results. They used a univariate logistic regression to assess the association between each variable and survival, through odds-ratio and p-values. It was found that the mortality rate was 60.8% in the 158, and was highest in those aged < 5 or > 25years. The functional association between survival and the continuous variables, including age, days admitted, and Ebola viral load, was assessed using a fractional polynomial model. They considered a $C_t < 20$ as a high viral load. The proportionality assumption was violated, hence a logistic multivariable model was used. Results from their analyses showed that high viral load, asthenia, myalgia, diarrhea, and anorexia were observed in over 50% of EVD deaths. Furthermore, among the EVD positive patients, 24% were co-infected with malaria. They had a higher viral load compared to the remaining percentage that had Ebola alone. Consequently, this 24% had a significantly higher mortality rate. At the end of their work, they suggested systematic meta-analyses and other future works to confirm and fine-tune their statistical results [13].

Lanini et al. (2015), in their study of blood kinetics of Ebola virus in survivors and nonsurvivors, examined the kinetics of the Zaire Ebola viremia in 84 patients in Sierra Leone. They quantified EBOV viremia between 2 and 13 days after onset of EVD symptoms. In order to compare the viral kinetics between survivors (45%) and nonsurvivors (55%), they used time from symptom onset and clinitical outcome as the independent variables. Generally, the viral kinetics was a quadratic function of time, although in the nonsurvivors, viremia was 0.94 log copies/ml higher than in the survivors. However, survivors attained a higher peak in viremia levels at an earlier time from symptom set, and conversely a general lower mean peak as compared to the nonsurvivors (7.46 log copies/ml vs. 8.6 log copies/ml). Even though EBOV viremia increased comparably in both groups, decay after peaking was stronger in the patients who survived. Therefore, they concluded that the blood plasma concentrations of EBOV differ in both groups at the early stages after contraction, and may be a good predictor of survival in patients. They recommended that, in order to examine causal and prognostic factors of survival in EVD patients, the early phase of the disease should be properly studied [17].

Fitzpatrick et al. (2015) investigated the association between patient characteristics including viral load with mortality in 525 clinically-reported positive Ebola patients in a Case Management Centre in Sierra Leone. They determined the relationship between age, gender, occupation, whole-blood C_t value among other variables, and mortality. They also used the Kaplan-Meier methods to plot the survival curves for the various groups, including data that was censored at 36 days. Additionally, a univariate and a consequent binary multivariate analysis was carried out on the significant variables to determine their association with death. The following results were obtained: viral load, age, duration of symptom before admission, and distance traveled to the centre were significantly associated with mortality at a probability value < 0.05. Similarly, from the multivariate analysis, viral load > 10 million copies/ml were highly correlated to death, along with age and symptom duration below 5 days. Nonetheless, in the nonsurvivors, confusion, diarrhea, and conjuctivitis were significant to death [11].

Weppelmann et al. (2016) explored the trend in morbidity, mortality and indicators of survival in EVD patients in Bong County, Liberia. 391 laboratory-confirmed patients were investigated. Cox regression and other survival analysis tools were used to determine the short-term risk of death with indicators such as demographic details, clinical features, and the treatments received. The mortality rate initially recorded as 53.5% gradually decreased as treatment was administered. Infection was higher in patients who had history of contact with an infected person, and funeral attendance. Mortality was higher among the older population and higher in males, as compared to females. The significant clinical symptoms included fever, vomiting, diarrhea, and hemorrhage. After adjusting for age, there was a 74% reduction in the risk of death in hospitalized patients, as compared to those who were not receiving medical attention. Finally, supportive care including intravenous rehydration therapy significantly improved survival in these

patients [27].

Beyan Y. Sana (2019) explored the risk factors related to the contraction of Ebola Virus disease using secondary data of 1,658 patients admitted into ETUs in Liberia. It was found that funeral attendance, exposure to body fluids, and contact history were strongly associated with the contraction of the disease. However, contact history and exposure to body fluids reflected higher probabilities of contraction. No information regarding age, gender, and geographical location was inferred from the analysis [26].

Edwards et al. (2021) performed a systematic review and meta-analysis on the effect of malaria co-infection on the pathology, immune response and survival of EVD using risk ratios, odds ratios, and hazard ratios. Gathering results from 26 peer-reviewed papers, it was found that the case fatality rates were similar between the Ebola-infected patients with malaria, and those without malaria (62.8% vs. 56.7%), and that there was no significant different in survival with a relative risk measure of 1.09. One study indicated no difference in mortality by the infection status, while another produced contrasting results. Due to the inconsistencies in the results, they concluded that the literature was inconclusive, and recommended an improvement in data analysis and diagnostic methods that would promote studies in the future [10].

Multiple studies have showed some clinical features that are strongly correlated with death, such as diarrhea, vomiting, fever, fatigue, myalgia, conjuctivitis, and unexplained bleeding. It has also been identified that age, viral load, co-infection with malaria, time from symptom onset to admission, and distance traveled to ETCs significantly predict the survival of patients. However, there have been limited studies that investigate biomarkers such as hemoglobin, creatinine, glucose, calcium ionied, sodium, potassium, chloride, hematocrit, and anion gap obtained from laboratory results and their association with survival. Similarly, studies investigating threshold cycle values as a predictor of death in malaria co-infected patients, as well as threshold cycle values combined with biomarkers are on the low, and thus highly recommended. Therefore, these components will be considered as the variables of interest in this research, among other demographic characteristics. In this study, we consider a much larger sample size than the studies referenced above, gathered from multiple centers in all the three most affected countries. To arrive at a plausible conclusion, the Kaplan-Meier method will be used to determine the survival probabilities of the variables of interest. Further, Cox regression will be used to investigate the association between these variables and survival.

3. Methodology

In this chapter, we shall discuss the methods to be implemented in the data analysis which basically include survival analysis.

3.1 Survival Analysis

Survival analysis is a branch of statistics for analyzing the expected duration of time until one or more events happen. The primary point of interest in many biomedical applications is the time taken for an event of interest to occur, referred to as "time-to-event" described by the random variable T. Events investigated in medical surveys include death from a disease, relapse, response to a therapy, or the onset of a symptom or disease. In survival analysis, we are interested in studying the population that survived or experienced the event. However, many factors could prevent the whole population from experiencing the event. This results in censored data, which will be described in detail in the next subsection (3.1.1). Censoring in survival data makes the analysis of such data different from data obtained from other biomedical or statistical settings. The objectives of survival analysis are to characterize time-to-event (or survival) distributions, compare several survival distributions among study groups (e.g. control and intervention), and to examine the relationship of survival to some covariates (predictors or prognostic factors) [3, 15, 21].

3.1.1 Censoring. Censoring occurs when information on either the start or end of an event of interest is missing. For example, if the day of onset of a disease is unknown, although the patient is recorded to have contracted the disease, this observation is censored. Similarly, if a participant leaves the study without experiencing the event, this is also censored data. We may know that an individual's lifetime exceeded 7 years because the study ended at this time, but we do not have further information on whether they experienced the event at a later time or not. This is also referred to as censoring. Censoring is an important aspect of survival analysis, for this reason, survival analysis is also referred to as analysis of censored data [21].

The most pedestrian form of censoring in medical studies is *right censoring*, which occurs when an individual leaves the study without experiencing the event, or the study ends without the individual experiencing the event. *Left censoring* also happens when we have no information on the start of an event but we know the individual had the event. For example, including patients in a study when we know they have the disease but do not know which date the disease started. Some features of these types of censoring include: *informative* censoring (when participants leave the study due to reasons related to the study); *non-informative* censoring (censoring not related to reasons from the study); *interval* censoring (when we know participants experienced the event within some time interval); and *random* censoring, which occurs when participants leave by other means aside from death, and the time of leaving is not predictable. The following scenarios are also considered as censoring: death from other causes (competing risks), and loss to follow-up of participants.

We can also consider some censoring mechanisms including *Type I* and *Type II* Censoring: *Type I* Censoring describes the situation where a study is ended, so that the remaining individuals are only known to have failed up to that point. It occurs when the censoring times are pre-known, which is a degenerate case of random censoring. An example is a study investigating age that stops follow-up once 60 years is attained. On the other hand, *Type II* Censoring is said to have happened when an observation

is continued until a predetermined number of deaths have occured. This allows for a simplified analysis since the censored limit is already pre-specified, and the number of individuals who fail, is non-random.

3.1.2 Survival Function. As stated earlier, survival data requires a special type of analysis as a result of censored data. Some statistics such as mean, standard deviation, median may not have the desired statistical properties since survival data does not follow a normal distribution. The sample mean for instance, may not be an unbiased estimator for the population mean (herein, the survival time). Therefore, it is requisite to estimate the underlying distribution of the given survival data before estimating the statistics [3]:

The time-to-event random variable, T can be described using the cumulative distribution function $F(t)$, from which the survival function can be derived:

$$F(t) = \mathbb{P}[T \leq t],\ t \geq 0 \tag{3.1.1}$$

Given that T is the time-to-event or survival time, $F(t)$ is the probability that an individual in the study will die before time t. We consider T to be a continuous random variable such that its density function $f(t)$ is defined as :

$$f(t) = \frac{dF(t)}{dt} \tag{3.1.2}$$

$$F(t) = \int_0^\infty f(u)du \tag{3.1.3}$$

The survival function $S(t)$, however, is the complement of the cumulative distribution function $F(t)$ in Equation 3.1.1. The survival function, which can also be referred to as the reliability function or the complementary cumulative distribution function is a function that gives the probability that a person, device, or any object of study will survive beyond a given specific time. Hence it is defined as:

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \mathbb{P}[T \leq t] \\ S(t) &= \mathbb{P}[T > t],\ 0 < t < \infty \end{aligned} \tag{3.1.4}$$

$S(t)$ describes the probability that an individual in the study survives from the time of origin to a future time t, or beyond. In terms of Equations 3.1.2 and 3.1.3, $f(t)$ and $S(t)$ can be expressed as:

$$f(t) = -\frac{dS(t)}{dt} \tag{3.1.5}$$

$$S(t) = \int_t^\infty f(u)du \tag{3.1.6}$$

The diagram in Figure 3.1 below describes a typical survival function.

In this survival curve for instance, it can be observed that the survival probability at time 0 is 1, $S(0) = 1$, which is generally true for survival functions, and $S(\infty) = 0$. Survival functions are non-increasing functions. We can also find that the median survival time (survival probability at 0.5) is 6 years. This means that, at this time, 50% of the population were still alive (or had not yet experienced the event). Similarly, 80% survived the event approximately 4.2 years into the study, as demonstrated by the dashed lines.

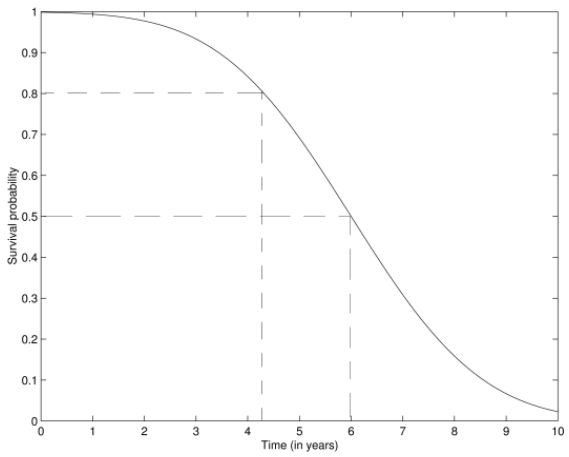

Figure 3.1: The survival function for a hypothetical population [3].

Considering two subgroups of the population of analysis, the survival distribution of the Group 2 is said to be stochastically larger than that of Group 1 if $S_2(t) \geq S_1(t),\ \forall t \geq 0$, and $S_k(t)$ is the survival function for Group *k*. Hence the time to event T_2 is also said to be stochastically larger than T_1. However, this does not mean that $T_2 \geq T_1$, but that a bigger proportion of Group 2 will survive the event at any timepoint in the study, than Group 1 [3].

3.1.3 Hazard Function. Another way of describing T is by using the hazard function. Hazard is simply the exposure or risk of having the event. This implies that the higher the hazard, the more likely it is to fail (or experience the event). Hazard rate is therefore the instantaneous rate at which the event happens, given the individual survived to time t [19]. This is also known as force of mortality, especially in actuarial or mortality studies. It is derived from the understanding that, mortality rate is the proportion of individuals among the population who die between time t and $t+1$, discretely denoted as :

$$m(t) = \mathbb{P}[t \leq T < t+1 \mid T \geq t]$$

Hazard rate $\lambda(t)$ is then the limit of the mortality rate with a rather smaller interval $(t, t+h)$, expressed as:

$$\lambda(t) \quad = \quad \lim_{h \to 0} \frac{\mathbb{P}[t \leq T < t+h \mid T \geq t]}{h} \tag{3.1.7}$$

In terms of the density and cumulative distribution functions:

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{1 - F(t)} \\ \lambda(t) &= \frac{f(t)}{S(t)} \end{aligned} \tag{3.1.8}$$

Recall Equation 3.1.5. Hence the hazard function $\lambda(t)$ becomes:

$$\begin{aligned}
\lambda(t) &= \frac{-\frac{dS(t)}{dt}}{S(t)} \\
&= -\frac{S'(t)}{S(t)} \\
\lambda(t) &= -\frac{d\log(S(t))}{dt} \qquad [3]
\end{aligned}$$

The cumulative hazard function $H(t)$ is then expressed from 0 to t as:

$$\begin{aligned}
H(t) &= \int_0^t \lambda(t)\ dt \\
&= \int_0^t -\frac{d\log(S(t))}{dt}\ dt \\
H(t) &= -\log S(t) \\
-H(t) &= \log S(t) \\
e^{-H(t)} &= e^{\log S(t)} \\
\therefore\ S(t) &= e^{-H(t)}
\end{aligned} \tag{3.1.9}$$

Thus, at $t = 0$, $\begin{cases} S(t) = 1 \\ H(t) = 0 \end{cases}$ and, at $t = \infty$, $\begin{cases} S(t) = 0 \\ H(t) = \infty \end{cases}$

Through this relation, we can compare the hazard and survival functions in a study. The higher the average hazard, the smaller the corresponding survival of the observation, and vice versa. A notable assumption considered is that both censored and uncensored populations have the same risk of failure.

3.2 Estimation of Survival Functions

Regardless of the form of survival data obtained, in order to determine survival probabilities among the sampled population, it is crucial to estimate and plot the survival function from the data. This, however, can be challenging due to censoring. As such, various approaches including parametric and non-parametic methods have been developed to effectively carry out this estimation. In the sections below (3.3 and 3.4), we discuss the various approaches to estimating survival functions of survival data.

3.3 Non-Parametric Estimations

Non-parametric methods of estimation are simple methods that do not make any distributional assumptions about the survival times. They are rather flexible methods that are used especially in medical and health applications considering multiple changes in human and animal behaviour [21]. However, they may not be able to properly describe survival in complex situations [15]. The Kaplan-Meier and Nelson-Aalen estimators shall be discussed in detail below.

3.3.1 Kaplan-Meier Estimator. Also known as the Product-limit estimator, the Kaplan-Meier (KM) method is the most widely used non-parametric approach. It determines the survival probability at any given time, t considering other small time intervals. It considers the survival times (t_i), participants who fail (d_i), number at risk of experiencing the event (n_i), and even the censored population (c_i) in $[t_i, t_{i+1})$. The product-limit estimate is mathematically expressed as:

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - q_i), \text{ where } q_i = \frac{d_i}{n_i} \tag{3.3.1}$$

where $t_i \leq t : t_1 < t_2 < ... < t_p$ and $n_i = \sum_{i=1}^{p}(c_i + d_i)$, the number at risk just before time t_i [15].

The Kaplan-Meier estimation therefore, produces a non-increasing step function. This is because it is assumed that at small time intervals, a proportion of individuals die or are lost to follow-up, whichever occurs first. See example below (Table 3.1 and Figure 3.2):

Table 3.1: Example of a hypothetical KM Table [21]

t_i	n_i	d_i	q_i	$1 - q_i$	$S_i = \Pi(1 - q_i)$
2	6	1	0.167	0.833	0.846
4	5	1	0.200	0.800	0.693
6	3	1	0.333	0.667	0.497

The corresponding KM survival curve is:

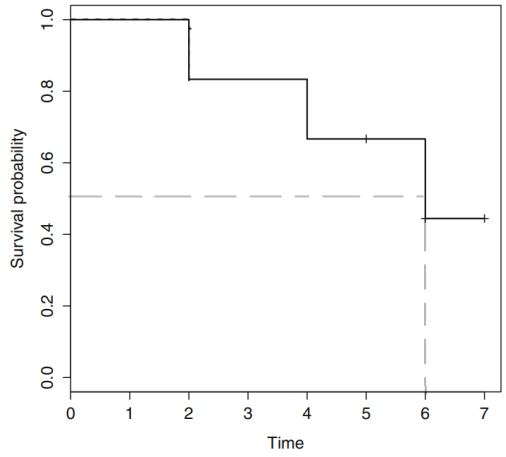

Figure 3.2: Example of a hypothetical KM Survival Curve [21]

In this example, the columns are rightly presented, and we observe that the probability of survival at $t = 6$ is the product of the probability of survival at $t = 4$ and $t = 2$, given survival at t_{i-1}. In the KM curve, the median survival time, which is the smallest time so that $S(t) \leq 0.5$ is equal to 6 time units [21]. An alternative approach to the KM method of estimation is the *actuarial life tables*, which is most useful when the data is not on an individual-level [16].

There are three fairly stringent assumptions made when using the KM estimator. The first assumption is that, the censored and uncensored population have the same chances of survival, at any time t.

Secondly, the survival probabilities are the same for all individuals regardless of the time of entry into the study. Lastly, the events happen at the times specified, and no lag is considered [12].

Standard errors associated with finding survival probabilities can be calculated in order to find the confidence intervals for these probabilities using the **Greenwood Formula** as follows:

$$\text{var}[\hat{S}(t)] \approx [\hat{S}(t)]^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{3.3.2}$$

Confidence limits estimated from this variance may however attain values above 1 or less than zero, requiring truncation at these points. The complementary log-log could be used to calculate these intervals to avoid the problem of truncation by using the following equation [21]:

$$\text{var}\Big(\log\big[-\log \hat{S}(t)\big]\Big) \approx \frac{1}{\big[\log \hat{S}(t)\big]^2} \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{3.3.3}$$

The Kaplan-Meier Estimator also allows comparison between the survival curves of two or more sub groups in an analysis. Groups of study interest could be sex (male and female), age groups, intervention and control, patients with or without a named risk factor such as diabetes or heart disease, country of origin, et cetera.

To test for statistical significance or difference between survival of the curves, several statistical tests can be used including the *Log-rank* test. In this work, we shall only focus on this method of testing.

3.3.2 Log-Rank Test. This is a non-parametric test used to investigate whether two groups or more have statistically equivalent survival curves. The log-rank test, or the *Mantel-Haenszel test* is a large-sample chi-square test that generally compares KM curves [16]. It tests the hypothesis that k survival functions: $S_0(t), ..., S_k(t)$ are not different [15].

Null hypothesis (H_0): There is no variation between KM curves : $S_0(t) = S_1(t) = ... = S_k(t)$

Alternate hypothesis (H_1): There is a real difference between KM curves : $S_0(t) \neq S_1(t) \neq ... \neq S_k(t)$

Its test statistic compares observed failures with expected number of failures over the independent classes of outcomes, mathematically expressed as:

$$\chi^2 \approx \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} \qquad OR \qquad \chi^2 \approx \frac{\sum_{j=1}^{k} (O_j - E_j)^2}{\sum_{j=1}^{k} \text{Var}(O_j - E_j)} \tag{3.3.4}$$

O_j and E_j are the observed and expected failures of group j, respectively. The sums of the observed and expected failures are computed for each event time and summed for each comparison group. The log-rank statistic χ^2, has degrees of freedom equal to $k-1$, where k is the number of comparison groups. There are numerous variations of the log rank statistic as well as tests to compare KM curves between independent groups. The modified Wilcoxon test for instance, is more sensitive to larger difference in hazards.

3.3.3 Nelson-Aalen Estimator. The estimate of the cumulative hazard function $H(t)$ can be found by using the Nelson-Aalen estimate in Equation 3.3.5 [15], also known as the empirical cumulative hazard function [18].

$$\hat{H}(t) = \sum_{i:t_i \leq t} \hat{h}_i \ , \text{ where } \hat{h}_i = \frac{d_i}{n_i} \tag{3.3.5}$$

This estimate can also be calculated from the negative logarithm of the Kaplan-Meier estimate, from the relation in Equation 3.1.9 [15].

3.4 Parametric Estimations

Another method of estimating survival functions is by using parametric approaches. These approaches make some assumptions about the distribution of the survival times prior to estimation, unlike the non-parametric methods. The patterns in survival times can be described by the use of continuous parametric survival models that assume distributions such as the *exponential*, *Weibull*, *log-logistic*, *Gompertz*, and *Makeham* distributions. In this estimation method, we assume that the hazard $\lambda(t)$ has a distinct shape that is approximated using estimated parameter(s) from the data. Parameters can be estimated by the Maximum Likelihood Method [15].

3.4.1 Exponential Distribution. With a constant hazard $\lambda(t) = \lambda$, the cumulative hazard function $H(t) = \lambda t$, and the corresponding survival function as defined by Equation 3.1.9 is:

$$\begin{aligned} S(t) &= e^{-H(t)} \\ S(t) &= e^{-\lambda t} \end{aligned} \tag{3.4.1}$$

and probability density function $f(t)$ from Equation 3.1.8 as:

$$\begin{aligned} f(t) &= \lambda(t) \cdot S(t) \\ f(t) &= \lambda e^{-\lambda t} \end{aligned} \tag{3.4.2}$$

Thus, the median survival time t_{median}, such that $(S(t) \leq 0.5)$ is:

$$\begin{aligned} 0.5 &= e^{-\lambda t} \\ t_{median} &= \frac{\log(2)}{\lambda} \end{aligned} \tag{3.4.3}$$

Although the Exponential distribution is the simplest of the parametric survival distributions, it is inefficient at describing lifetimes of humans and animals as a result of the constant hazard assumption. A relatively flexible alternative is the Weibull distribution [21].

3.4.2 Weibull Distribution. Similar to the exponential distribution, the Weibull distribution adjusts for monotonic increasing or decreasing hazard functions which has parameter α [15]. The hazard function is

$$\begin{aligned} \lambda(t) &= \alpha\lambda(\lambda t)^{\alpha-1} \\ \lambda(t) &= \alpha\lambda^{\alpha}t^{\alpha-1} \end{aligned}$$

with corresponding cumulative hazard and survival functions given by

$$\begin{aligned} H(t) &= (\lambda t)^{\alpha} \\ S(t) &= e^{-(\lambda t)^{\alpha}} \end{aligned}$$

The median survival time t_{median} is also given by

$$t_{median} = \frac{[\log(2)]^{\frac{1}{\alpha}}}{\lambda}$$

The function is monotone decreasing when $\alpha < 1$, and monotone increasing when $\alpha > 1$. When $\alpha = 1$, this is equivalent to the exponential distribution.

See illustration in Figure 3.3 below.

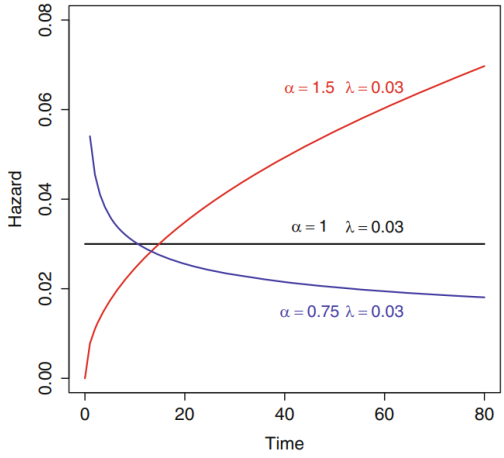

Figure 3.3: Weibull and exponential hazard functions [21]

Among many other parametric families are the *Pareto*, *log-normal*, and *log-logistic*, which adjusts for either a monotonic or unimodal hazard function [15].

Several parametric distributions have been provided through which survival functions can be estimated. However, due to vagaries in lifetime of human and animal data, it is difficult to find the appropriate parametric distribution that best fits the data. Therefore in medical applications, non-parametric or semi-parametric methods are preferred, which will be employed in the data analysis in this research.

3.5 Cox Proportional Hazards Model

As earlier stated, one of the objectives of survival analysis is to examine the relationship between covariates and survival. One of the well-known ways to perform this kind of analysis is by use of the Cox regression or the Cox proportional hazards model.

A Cox regression model has the form:

$$h(t \mid X) = h_0(t)\, e^{\beta X} \qquad [18] \tag{3.5.1}$$

where: $h(t \mid X)$ is the hazard; $h_0(t)$ represents the baseline hazard function, a function of time; X is a covariate or explanatory variable of the response variable, time-to-event, T; and β is the coefficient of the covariate, the measure of the effect size. In the case where multiple predictors are considered, X is considered as a column vector X_i, and each X has an effect size, β_i. Then, the Cox model takes the form:

$$\begin{aligned} h(t \mid X_i) &= h_0(t)\, e^{\sum_{i=1}^{p} \beta_i X_i} \qquad [16] \\ h(t \mid X_1, X_2, ..., X_p) &= h_0(t)\, e^{\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p} \end{aligned} \tag{3.5.2}$$

for p covariates. The baseline hazard is the hazard when $X = 0$ [15]. An additional covariate added to Equation 3.5.1 is called a confounder, that adjusts for the effect of other variables as in the case

on Equation 3.5.2 [16]. In this model, the hazard estimate is considered to be relative, rather than absolute. This is because, confounders are assumed to have a modification effect on the natural log of the hazard ratio. The parameters β, reflect the predicted change of the log-hazard of the event, and the estimates represent the increase in the expected log of the relative hazard for each unit increase in the predictor, holding other predictors constant. It can also be interpreted as a unit increase in X, increases the hazard by a factor of e^{β}. In the case of a dichotomous variable X, for example Sex, representing either Male or Female factored as 0 and 1 respectively, we would say the hazard of Females (1), is e^{β} times the hazard of Males (0) [15].

Cox proportional hazards model assumes a constant hazard ratio over time. This assumption is known as the *Proportional Hazards Assumption (PH Assumption)*. The hazard ratio is the ratio of the hazard function to the baseline hazard, and is represented in the model by e^{β}. Using the exponential function in hazard function makes it possible to have a positive hazard.

3.5.1 Hypothesis Testing. Generally, if a hazard ratio $e^{\beta} \approx 1$, then the predictor has no effect on survival. A hazard ratio < 1, indicates a protective covariate, meaning it is associated with improved survival, while a hazard ratio > 1 is associated with increased hazards from the predictor. Testing for statistical significance can then be built on these hypotheses:

Null hypothesis (H_0): The independent variable makes no difference in the survival of the event i.e., $e^{\beta} = 1$.

Alternate hypothesis (H_1): The explanatory variable increases or decreases the hazard of the event i.e., $e^{\beta} > 1$ or $e^{\beta} < 1$.

3.5.2 Estimation of Cox Parameters. This model has similarities with the usual linear regression and logistic regression models for modeling a continuous or binary response variable, respectively. However, Cox regression exhibits unique properties that distinguish it from other models, such as its semi-parametric nature. One of the biggest advantages of the Cox PH model is that we can estimate β_i, without making any assumptions about the shape or distribution of $h_0(t)$ although the covariates are assumed to be related to the survival times in some way. This is what makes the model semi-parametric. There are however other assumptions such as independence, changes in predictors produce proportional changes in the hazard irrespective of time, and a linear relationship between the natural logarithm of the relative hazard and the predictors.

We can estimate β_i using the partial likelihood function. The partial likelihood, also called Cox likelihood, mostly considers the order in which the events were observed, instead of the joint distribution of events. It is summarized in the formula:

$$PL(\beta) = \prod_{t_i:\text{event at time } t_i} \frac{h_0(t)e^{\beta X_{(t_i)}}}{\sum_{j \in R_{(t_i)}} h_0(t)e^{\beta X_j}}$$

where $R_{(t_i)}$ is the population at risk at time t_i. This formula can be simplified, eliminating the baseline hazards, to:

$$PL(\beta) = \prod_{t_i:\text{event at time } t_i} \frac{e^{\beta X_{(t_i)}}}{\sum_{j:t_j \geq t_i} e^{\beta X_j}}$$

This function is called partial likelihood because the likelihood formula considers probabilities only for subjects who failed. Actually, these subjects are the ones who had the event among the whole population

at risk at that time. The assumption is that only one event happens at each time [15]. The maximum likehood estimation however, maximizes this function [16].

3.5.3 Evaluating the Cox PH Assumption. Cox regression is effective when the proportional hazards assumption is fulfilled. The model assumes that the hazard ratio of any two covariates is constant over time, with a proportionality constant $\hat{\theta}$ that does not depend on time. See Equation 3.5.3

$$\begin{aligned} \frac{\hat{h}(t \mid X')}{\hat{h}(t \mid X)} &= \hat{\theta} \\ \hat{h}(t \mid X') &= \hat{\theta}\, \hat{h}(t \mid X), \end{aligned} \tag{3.5.3}$$

where X' and X are two sets of covariates. Hence, the hazard of a subject is constantly proportional to the hazard of any other subject in the cohort. Thus, the estimated hazard ratio (HR) is given by the formula:

$$\widehat{HR} = e^{\sum_{i=1}^{p} \beta_i (X_i' - X_i)},$$

where $X_i' = X_1', X_2', X_3', ..., X_p'$, and $X_i = X_1, X_2, X_3, ..., X_p$ are two sets of covariates. There are instances, however, where the proportional hazard assumption is violated by one or more variables in the Cox model. A clue to identifying a violation of the PH assumption is crossing-curves in a KM plot, although non-crossing curves do not imply that the assumption is satisfied. To efficiently assess this, a graphical, goodness-of-fit test, or a time-dependent variables' approach can be used.

The graphical approach makes use of the comparison between $\log$-$\log$ survival curves between groups of a covariate to check for parallelism. The closeness of observed versus predicted survival curves can also be used to check for this assumption graphically. Alternatively, the goodness-of-fit tests which provide large sample standard normal or chi-square test statistics, with accompanying p-values can be used to assess if the assumption is violated, by a significant p-value. An example of this is the Schoenfeld residuals. Extending the Cox model to include interaction terms of time-independent covariate(s) that depends on time is yet another way of evaluating the PH assumption [16].

Methods of resolving a violation include modelling the time-dependence and stratification [15].

Stratification implies classifying the elements of the covariate (that does the fulfill the PH assumption) into groups known as strata, where each stratum has a different baseline hazard $h_{0k}(t)$, however, assuming the β of the predictor is the same for each stratum k. This effectively drops that predictor from the model, while other predictors are included in the model.

Given a simple Cox model:

$$h(t \mid X) = h_0(t) e^{\beta_1 X_1 + \beta_2 X_2}$$

and X_1 does not satisfy PH Assumption. A "stratified" Cox model that stratifies the model by X_1 is fitted of the form:

$$h_k(t \mid X) = h_{0k}(t) e^{\beta_2 X_2},$$

where k is the categories of X_1. Assuming X_1 has $k = 2$, for the respective values of k, the stratified Cox model will be:

$$\begin{cases} h_1(t \mid X) = h_{01}(t) e^{\beta_2 X_2} & \text{for } \; k = 1 \\ h_2(t \mid X) = h_{02}(t) e^{\beta_2 X_2} & \text{for } \; k = 2 \end{cases}$$

and β_2 is the same for both models. Hence, implying that the hazard ratios $\widehat{HR}$ are the same [16]. Additionally, if there are differences between the categories of k, which is not reflected by the predictors, this stratification method can be used. Therefore, any difference is absorbed into each $h_{0k}(t)$. This stratified model however, disallows a between-strata comparison [15].

4. Data Analysis

In this chapter, we shall define the source of the data, summarize by use of flowchart, explore and describe the data, and present the results in the order in which the analysis was carried out. There will also be a brief discussion to conclude the chapter.

4.1 Data Collection and Summary

Data was collected by the Infectious Diseases Data Observatory (IDDO), an independent and multidisciplinary organization that materializes data gathered from infectious diseases to improve patient conditions worldwide. The data gathered from the 2013 – 2016 Ebola outbreak were collected from multiple Ebola Treatment Centres in the three main affected countries, Guinea, Sierra Leone, and Liberia. Some relevant criteria for inclusion in the collection were: PCR or laboratory tested positive patients; aged 18 or older; non-pregnant and non-lactating mothers; and, survivors of EVD. However, the following groups were excluded from the study: clinically recovering patients; and patients or survivors with missing data. Thus, a total of 9,472 records were captured from all three countries.

The data comprised patient demographics, disposition including final patient outcome of death or survival, laboratory results, microbiology specimens, as well as other relevant information recorded during hospitalization. The flow diagram below (Fig 4.1) gives a detailed summary of the sample extraction and criteria for the final dataset on which the analysis was performed.

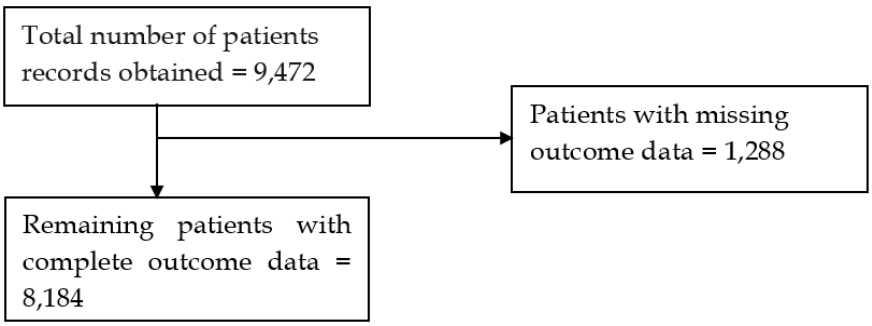

Figure 4.1: Flow diagram on sample size

Thus, the final sample size used for this analysis from the demographic data was $8,184$. All analyses carried out on this data were performed in R version 4.1.3. See below the tabulated results of the exploratory analysis carried out on the $8,184$ patients:

From Table 4.1, $6,057$ patients survived death from Ebola virus, while $2,127$ died from the disease. Of the $8,184$, there were more male than female patients $[4,437\ (54.2\%)$ vs $3,735\ (45.6\%)]$, whereas the remaining $12\ (0.1\%)$ were not identified with any of the sexes. With a mean age of 30.1, the age group $[30,65)$ were the majority age group making 44.0% of the data, and the least was $[65,100)$ which was 3.9%. $92\ (1.1\%)$ had missing ages. Additionally, 41.9% of these patients were from Sierra Leone, 30.1% from Liberia, and 27.9% from Guinea, making Sierra Leone the majority of the population. The average duration of hospitalization was approximately 6 days. The minimum length of stay was 1 day since some patients died on arrival and others within 48 hours of admission, and the maximum, 51. However, patients who died had maximum duration of hospitalisation of 32 days.

Table 4.1: Summary of Patients' Demographics

	Survived	Died	Total
	(N=6057)	(N=2127)	(N=8184)
Sex			
Male	3323 (54.9%)	1114 (52.4%)	4437 (54.2%)
Female	2731 (45.1%)	1004 (47.2%)	3735 (45.6%)
Unknown	3 (0.0%)	9 (0.4%)	12 (0.1%)
Age (years)			
Mean (SD)	29.4 (16.9)	32.1 (18.1)	30.1 (17.2)
Median [Min, Max]	28.0 [0, 102]	31.0 [0, 90.0]	29.0 [0, 102]
Age Groups			
[0,15)	1130 (18.7%)	366 (17.2%)	1496 (18.3%)
[15,30)	2106 (34.8%)	571 (26.8%)	2677 (32.7%)
[30,65)	2554 (42.2%)	1046 (49.2%)	3600 (44.0%)
[65,110)	220 (3.6%)	99 (4.7%)	319 (3.9%)
Country			
Guinea	1807 (29.8%)	478 (22.5%)	2285 (27.9%)
Liberia	1750 (28.9%)	717 (33.7%)	2467 (30.1%)
Sierra Leone	2500 (41.3%)	932 (43.8%)	3432 (41.9%)
Length of Stay (days)			
Mean (SD)	6.30 (5.79)	4.45 (3.28)	5.82 (5.32)
Median [Min, Max]	4.00 [1.00, 51.0]	4.00 [1.00, 32.0]	4.00 [1.00, 51.0]

4.2 Analysis of Patient Demographics

We examined the survival probability of all patients admitted, and then investigated their survival by sex, age, and country using Kaplan-Meier curves. Figure 4.2 below shows the overall survival of hospitalized patients. Crosses $(+)$ represent censored data.

In the data, about 74% of the population survived. This resulted in recording nil median survival times in almost all investigations. Results from the survival analysis of the demographics and outcomes of the patients are outlined in the Table 4.2.

We observed from the graphs in Figures B.3, B.1, and B.2 that, the survival curves of males and females are very close and almost indistinguishable. By age groups, patients aged 65 and above had the lowest survival, while those aged between 15 and 30 had the highest survival. Similarly, in terms of country, patients from Liberia had the lowest survival, while those from Guinea experienced a relatively higher survival.

A further analysis considered the associations between sex and age in each of the countries. All results from Kaplan-Meier, log-rank test, and the Cox regression models have been included in Table 4.2 and Table 4.3 below:

The log-rank test was run on the individual covariates namely, Sex, Country and Age groups. A p-value of 0.12 does not provide enough evidence of a difference between the survival of Male and Female

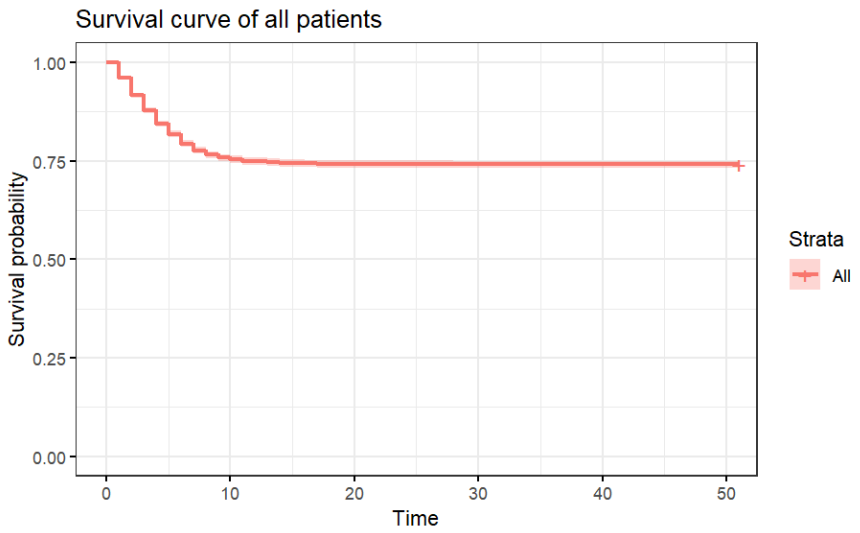

Figure 4.2: Kaplan-Meier Curve for All Patients

Table 4.2: Summary of Survival Analysis on Patient Demographics

	Univariable Analysis				
	coef	HR (95%C.I)	se(coef)	z	p
Sex					
Female	-	-	-	-	-
Male	-0.067	$0.935(0.858-1.018)$	0.044	-1.548	0.122
Country					
Guinea	-	-	-	-	-
Liberia	0.379	$1.461(1.301-1.64)$	0.059	6.398	1.6×10^{-10}
Sierra Leone	0.291	$1.338(1.198-1.494)$	0.056	5.164	2.4×10^{-7}
Age					
$0-15$	-	-	-	-	-
$15-30$	-0.140	$0.87(0.762-0.992)$	0.067	-2.080	0.038
$30-65$	0.216	$1.241(1.102-1.399)$	0.061	3.547	3.9×10^{-4}
$65-110$	0.295	$1.344(1.076-1.678)$	0.113	2.604	9×10^{-3}

patients. Whereas the opposite is true for the survival between countries and age groups. The p-values for country and age were significant, implying a difference in survival probabilities between the categories of the covariates. A Cox regression model was fitted to estimate the parameters β_i and the hazard ratio e^{β}. However, since sex and country did not satisfy the PH assumption, the model was stratified by sex and country leaving only age in the model. Refer to Table 4.3. Estimates from the model indicate that patients aged 65 and above, had 1.4 times the hazard of patients below the age of 15, while those between 15 and 30 had 1.3 times their hazard. This indicates that the older population, were at a higher risk of death than the other groups. Investigating this by country produced similar results. The aged population were still at a higher risk of death than the others with the same hazard ratio.

Table 4.3: Stratified Model Estimates of Demographics

	coef	HR (95%C.I)	se(coef)	z	p
Age $[15, 30)$	-0.12	$0.88(0.775 - 1.009)$	0.07	-1.83	0.067
Age $[30, 65)$	0.23	$1.25(1.112 - 1.412)$	0.06	3.70	2.2×10^{-4}
Age $[65, 110)$	0.34	$1.40(1.119 - 1.748)$	0.11	2.95	0.003

Likelihood ratio test $= 54.45$ on 3 df, p $= 9 \times 10^{-12}$

4.3 Laboratory Biomarkers

In addition to demographic information, 11 laboratory biomarkers recorded from blood and urine samples including Sodium, Creatinine, Glucose, Potassium, Chloride, Urea Nitrogen, Carbon Dioxide, Hematocrit, Hemoglobin, Calcium Ionized, and Anion Gap were investigated for association with risk of death. All these biomarkers were evaluated individually adjusting for age and sex. Country of origin was not included in the univariable analysis because 96% of these results were from Guinea, hence including country as a predictor would introduce bias in the model. Results from the likelihood ratio tests that were significant at the 0.05 level were included in the multivariable analysis. As such, all biomarkers were included in the multivariable analysis. Nonetheless, many factors lost significance. Therefore, the backward method of model selection was used to eliminate all insignificant variables producing a resultant model of 4 factors instead of 13. However, the resultant model was adjusted for sex and age, which by means of hazard ratios showed significance, and hence were retained in the model. The final model included Sodium, Urea Nitrogen, Carbon Dioxide, Calcium Ionized, Age, and Sex. The proportional hazards assumption was fulfilled.

Table A.1 in the Appendix details the summary of descriptive statistics for the biomarkers. Table 4.4 below shows the results obtained from the univariable analysis, while Table 4.5 shows that of the multivariable analysis. Sodium, Glucose, and Anion Gap violated the PH assumption, hence they were stratified in the individual models. This accounts for the nil values corresponding to these biomarkers in Table 4.4.

Table 4.4: Summary of Survival Analysis on Laboratory Biomarkers

	Univariable Analysis				
	coef	HR (95%C.I)	se(coef)	z	p
Sodium	-	-	-	-	-
Creatinine	0.606	$1.8(1.058 - 3.182)$	0.281	2.160	0.03
Glucose	-	-	-	-	-
Potassium	0.570	$1.77(1.046 - 2.986)$	0.268	2.128	0.03
Chloride	-0.142	$0.867(0.504 - 1.495)$	0.278	-0.511	0.609
Urea Nitrogen	1.882	$6.569(3.662 - 11.783)$	0.298	6.314	2.7×10^{-10}
Carbon Dioxide	-1.599	$0.202(0.113 - 0.362)$	0.296	-5.410	6.3×10^{-8}
Hematocrit	-0.600	$0.549(0.274 - 1.100)$	0.355	-1.691	0.091
Hemoglobin	0.476	$1.61(0.895 - 2.895)$	0.299	1.591	0.112
Calcium Ionized	-1.268	$0.281(0.154 - 0.514)$	0.308	-4.120	3.8×10^{-5}
Anion Gap	-	-	-	-	-

According to the results of the 177 patients whose biomarkers were investigated, urea nitrogen, carbon

dioxide and calcium ionized were the most significantly associated with risk of death in the univariable analysis, while glucose and anion gap were the least significant.

Table 4.5: Model estimates of Laboratory Biomarkers

	coef	HR (95%C.I)	se(coef)	z	p
Sodium $[135, 160)$	-1.19	$0.31(0.01 - 0.56)$	0.31	-3.88	1.03×10^{-4}
Urea Nitrogen $[20, 60)$	1.20	$3.33(1.54 - 7.21)$	0.39	3.05	0.002
Carbon Dioxide $[20, 1000)$	-1.00	$0.37(0.12 - 0.77)$	0.37	-2.68	0.007
Calcium Ionized $[1, 40)$	-0.95	$0.39(0.20 - 0.74)$	0.33	-2.89	0.004
Sex:Male	-0.58	$0.56(0.29 - 1.07)$	0.33	-1.75	0.080
Age $[15, 30)$	0.11	$1.11(0.42 - 2.97)$	0.50	0.22	0.829
Age $[30, 65)$	0.04	$1.04(0.40 - 2.72)$	0.49	0.08	0.935
Age $[65, 110)$	1.23	$3.44(0.78 - 15.10)$	0.76	1.63	0.103

Likelihood ratio test $= 75.97$ on 8 df, p $= 3 \times 10^{-13}$

Further analysis showed that sodium concentration of $[135, 160)$mmol/L with relative hazard ratio of 0.305 was highly significant, followed by urea nitrogen and calcium ionized. It was found that sodium concentration of 135mmol/L and above, carbon dioxide of 20mmol/L and above, and calcium ionized above 1mmol/L were protective against death. This implied that the opposing categories of these predictors were highly predictive of death. Similarly, urea nitrogen above 20mmol/L and patients aged above 65 years were at a higher risk of death with a hazard ratio of 3.33 and 3.44 respectively. The male patients however, were discovered to have had half the risk associated with death in female patients, rendering them to have a better survival than females. See Table 4.5.

4.4 Microbiology Tests' Results

This domain of tests detailed results from the Ebola, Malaria, and Typhoid tests that were conducted on the patients. From these results, we investigated the general survival of positively and negatively tested Ebola and Malaria patients on admission. We then proceeded to determine the survival probabilities associated with malaria co-infection with Ebola. Lastly, we examined cycle threshold C_t as a predictor of death in the positively tested Ebola patients as well as the malaria co-infected patients. Table A.2 and Table 4.6 contain a summary of the records and analysis, respectively. Cycle threshold, C_t, did not fulfill the PH assumption in the univariable model adjusted for sex, age, and country. Hence, it was stratified, as observed in Table 4.6.

From the survival curves that were obtained, we observed that the negatively tested malaria patients had a reduced survival as compared to the positively tested ones. Of those that were Ebola positive and had malaria results, a better survival was observed with those that tested malaria positive, than the negative patients. The negative group had the worst survival distribution. Furthermore, among Ebola positive patients, there was no significant difference between their survival with respect to cycle threshold. The survival curves overlapped. Meanwhile, malaria co-infected subjects with a $C_t < 27$ had the best survival distribution, while those with $C_t \geq 27$ had the worst survival. We recommend additional studies to confirm these results.

Results from the univariable analysis identified malaria status, age above 30, and residence in Sierra Leone as predictive of death. Patients in Sierra Leone experienced 0.4 times the hazard experienced

by patients in Guinea. This implies that patients from Guinea were correlated to a high risk of death. Microbiology results from Liberia represented 0.002% of the results, hence it was omitted from the analysis.

Table 4.6: Summary of Survival Analysis on Microbiology Results

	Univariable analysis				
	coef	HR (95%C.I)	se(coef)	z	p
Malaria					
Negative	-	-	-	-	-
Positive	-0.780	$0.462(0.316-0.676)$	0.193	-4.039	6.8×10^{-5}
Age					
$0-15$	-	-	-	-	-
$15-30$	-0.043	$0.958(0.760-1.207)$	0.118	-0.364	0.716
$30-65$	0.264	$1.302(1.053-1.610)$	0.108	2.411	0.015
$65-110$	1.096	$2.993(2.070-4.328)$	0.188	5.828	5.6×10^{-9}
Country					
Guinea	-	-	-	-	-
Sierra Leone	-1.004	$0.367(0.262-0.513)$	0.172	-5.838	5.3×10^{-9}

Although C_t violated the PH assumption in the univariable analysis, it satisfied the assumption in the multivariable model and showed significance. See Table 4.7. The multivariable analysis showed that the population within the ages of 15 and 30 experienced a reduced risk of death as compared to those above 35. Being aged over 65 was the most significant predictor of death, with a hazard ratio of 2.24. Additionally, testing positive for malaria was associated with a better survival than testing negative. A threshold cycle above 27 was associated with higher risk of death with hazard ratio 1.95. Table 4.7 summarizes the estimates of the fitted model.

Table 4.7: Model Estimates of Microbiology results

	coef	HR (95%C.I)	se(coef)	z	p
Plasmodium Positive	-0.68	$0.51(0.32-0.80)$	0.23	-2.93	0.003
C_t $[27,50)$	0.67	$1.95(1.22-3.12)$	0.24	2.81	0.005
Age $[15,30)$	-0.76	$0.47(0.29-0.76)$	0.25	-3.07	0.002
Age $[30,65)$	-0.18	$0.83(0.54-1.29)$	0.22	-0.81	0.417
Age $[65,110)$	0.81	$2.24(1.17-4.30)$	0.33	2.43	0.015

Likelihood ratio test $= 43.3$ on 5 df, p $= 3\times10^{-8}$

4.5 Discussion

The entire analysis showed that low concentration of sodium in the blood was the leading predictor of death. Specifically, sodium levels below 135mmol/L are associated with a higher risk of death in the Ebola patients. In addition to that, urea nitrogen level above 20mmol/L has a relatively higher hazard (three times the risk experienced in patients with urea nitrogen below this level). Calcium ionized values below 1mmol/L is identified to be a predictive factor of mortality in Ebola patients. Similarly,

large concentrations of carbon dioxide in the blood are protective against death, while values below 20mmol/L are associated with high risk of death. Patients with the above characteristics, in addition to being malaria negative, are associated with a higher risk of death. The hazard ratio associated with malaria positive patients who have low sodium, high urea nitrogen, and low carbon dioxide relative to malaria negative patients is 0.68. Refer to Table A.3. A patient who has malaria, and a cycle threshold value above 27, was found to experience half the risk of those without malaria. Of these patients, those aged 65 years and above were at a higher risk of death as compared to those between the ages of 15 and 30 with hazard ratios [2.24 vs. 0.47]. These results were more prominent among patients from Guinea. Overall, sex and country were not strong predictors of death. Meanwhile, age above 30 years accounted for most of the deaths resulting from the disease. Guinea was more associated with deaths from the aged population, followed by Liberia. However, we could not say this about Sierra Leone because the PH assumption for age in Sierra Leone was violated.

5. Conclusion

This project sought to perform a survival analysis on Ebola patients who were admitted to Ebola Treatment Centers during the epidemic in $2013 - 2016$. We wanted to identify the correlation between death and some variables including demographic details like sex, age, country of origin; malaria status; cycle threshold values; and, some biomarkers. The study involved $8,184$ patients of whom 54% were males and 46% were females. The dominant age group in the cohort was between the ages of 30 and 65. 41% of the data were recorded from Sierra Leone, although Liberia and Guinea were all included in the study. Generally, more than half of the sampled population survived death, while about a fifth of them died from the terrible disease [74% vs 26%]. Results from Kaplan Meier, Log-rank tests, and Cox Proportional Hazards Models indicated that biomarkers such as sodium, urea nitrogen, calcium ionized, and carbon dioxide are very significant factors that are associated with the death of Ebola virus patients. Also, malaria co-infection was not identified as a very influential predictor, since most malaria positive patients survived death than negative patients. Cycle threshold values however, were shown to predict death of malaria negative patients, although it was not independently predictive of death. At the end of the entire analysis, we found that low sodium and carbon dioxide, and high urea nitrogen concentrations in the blood were the most predictive factors of death in Guinea. Similarly, testing positive for malaria, and having cycle threshold values greater than 27 were also predictive factors of mortality in Guinea.

The main challenge encountered in the course of this work was the data cleaning process. The data was very messy due to the fact that they were collected in the heat of the epidemic. This phase, although challenging, was concurrently educative.

Within the scope of our work, we considered covariates independently in the various models. Additional terms that could be considered were interactions between the covariates. With this, we could be able to identify if the association between two or more covariates had any correlation with the risk of death. Additionally, testing influential observations or outliers could have provided more insight into the results obtained. Nevertheless, these elements were not within the scope of our work. They may however be included in subsequent studies.

5.1 Recommendations for Further Studies

Our results indicate that malaria negative patients whose C_t values were above 27 had a reduced survival as compared to the malaria positive patients. This group of patients may have exhibited or possessed other characteristics that were responsible for the high association with death, as reported. Hence, future works can consider studying the drugs and treatments administered to them, clinical features, vital signs, and other subject characteristics specific to these patients.

Lastly, investigating the risks associated with biomarkers combined with cycle threshold values in Ebola virus patients is also recommended.

Appendix A. Statistical Summaries

Table A.1: Summary of Patient Biomarkers

	Survived	Died	Total
	(N=108)	(N=69)	(N=177)
Sodium (mmol/L)			
[100,135)	24 (22.2%)	43 (62.3%)	67 (37.9%)
[135,160)	82 (75.9%)	25 (36.2%)	107 (60.5%)
Creatinine (umol/L)			
[0,500)	85 (78.7%)	45 (65.2%)	130 (73.4%)
[500,9000)	20 (18.5%)	22 (31.9%)	42 (23.7%)
Glucose (mmol/L)			
[0,5)	23 (21.3%)	31 (44.9%)	54 (30.5%)
[5,50)	82 (75.9%)	36 (52.2%)	118 (66.7%)
Potassium (mmol/L)			
[0,4)	71 (65.7%)	36 (52.2%)	107 (60.5%)
[4,10)	35 (32.4%)	29 (42.0%)	64 (36.2%)
Chloride (mmol/L)			
[0,102)	53 (49.1%)	31 (44.9%)	84 (47.5%)
[102,140)	50 (46.3%)	30 (43.5%)	80 (45.2%)
Urea Nitrogen (mmol/L)			
[0,20)	101 (93.5%)	43 (62.3%)	144 (81.4%)
[20,60)	4 (3.7%)	24 (34.8%)	28 (15.8%)
Carbon Dioxide (mmol/L)			
[0,20)	23 (21.3%)	38 (55.1%)	61 (34.5%)
[20,1000)	78 (72.2%)	22 (31.9%)	100 (56.5%)
Hematocrit (%)			
[0,35)	25 (23.1%)	22 (31.9%)	47 (26.6%)
[35,45)	54 (50.0%)	18 (26.1%)	72 (40.7%)
[45,80)	23 (21.3%)	20 (29.0%)	43 (24.3%)
Hemoglobin (g/L)			
[0,140)	58 (53.7%)	32 (46.4%)	90 (50.8%)
[140,1000)	44 (40.7%)	28 (40.6%)	72 (40.7%)
Calcium Ionized (mmol/L)			
[0,1)	29 (26.9%)	36 (52.2%)	65 (36.7%)
[1,40)	73 (67.6%)	18 (26.1%)	91 (51.4%)
Anion Gap (mmol/L)			
[0,20)	63 (58.3%)	29 (42.0%)	92 (52.0%)
[20,100)	36 (33.3%)	23 (33.3%)	59 (33.3%)

Table A.2: Summary of Microbiology Results

	Survived	Died	Total
	(N=650)	(N=198)	(N=848)
Plasmodium Test			
Negative	432 (66.5%)	158 (79.8%)	590 (69.6%)
Positive	218 (33.5%)	40 (20.2%)	258 (30.4%)
Cycle Threshold			
[0,27)	607 (93.4%)	151 (76.3%)	758 (89.4%)
[27,50)	33 (5.1%)	22 (11.1%)	55 (6.5%)
Country			
Guinea	31 (4.8%)	47 (23.7%)	78 (9.2%)
Sierra Leone	619 (95.2%)	151 (76.3%)	770 (90.8%)
Day of Test			
Mean (SD)	1.16 (0.777)	1.38 (1.21)	1.21 (0.901)
Median [Min, Max]	1.00 [1.00, 12.0]	1.00 [1.00, 17.0]	1.00 [1.00, 17.0]
Length of Stay (days)			
Mean (SD)	5.95 (4.96)	5.26 (3.69)	5.79 (4.70)
Median [Min, Max]	4.00 [1.00, 31.0]	5.00 [1.00, 32.0]	4.00 [1.00, 32.0]

Table A.3: General Model Estimates (n = 55, no. of events = 31)

	coef	HR (95%C.I)	se(coef)	z	p
Sodium [135,160)	-1.14	0.32(0.15-0.68)	0.39	-2.94	0.003
Urea Nitrogen [20,60)	1.31	3.70(1.53-8.93)	0.45	2.90	0.004
Carbon Dioxide [20,1000)	-0.88	0.42(0.19-0.90)	0.39	-2.22	0.026
Plasmodium Positive	-0.39	0.68(1.20-2.31)	0.62	-0.62	0.534

Likelihood ratio test $= 23.86$ on 4 df, p $= 9 \times 10^{-5}$

Appendix B. Kaplan-Meier Curves

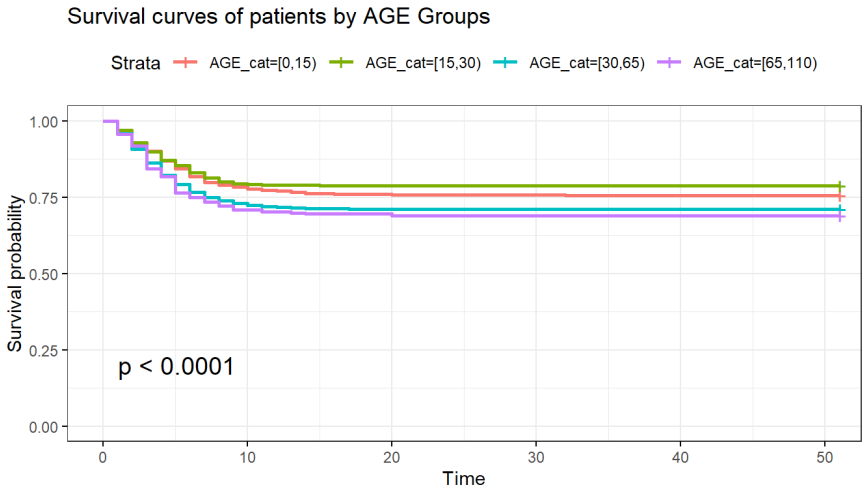

Figure B.1: Kaplan-Meier Curves of Patients by Age Groups

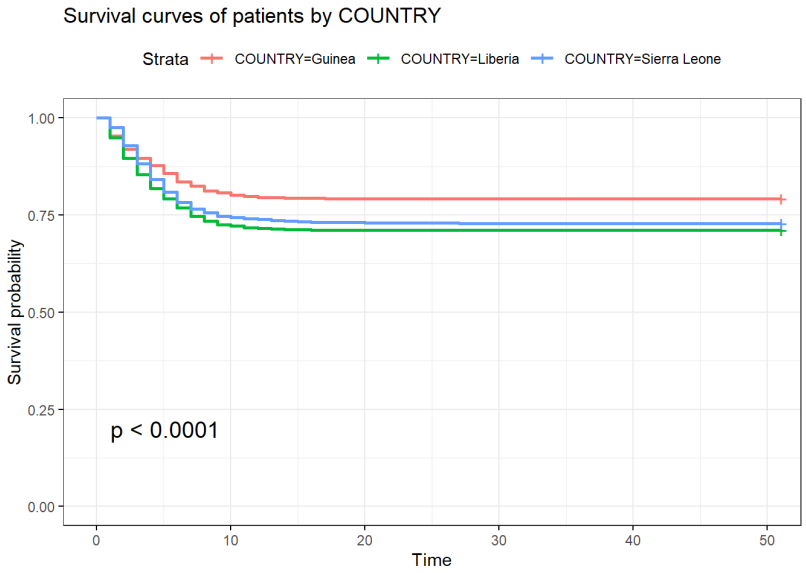

Figure B.2: Kaplan-Meier Curves of Patients by Country of Origin

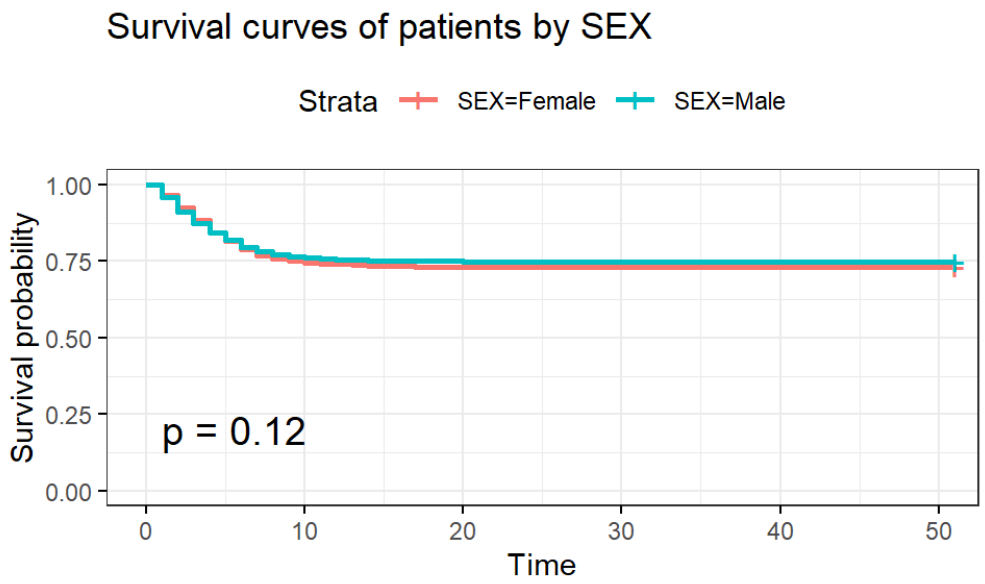

Figure B.3: Kaplan-Meier Curves of Patients by Sex

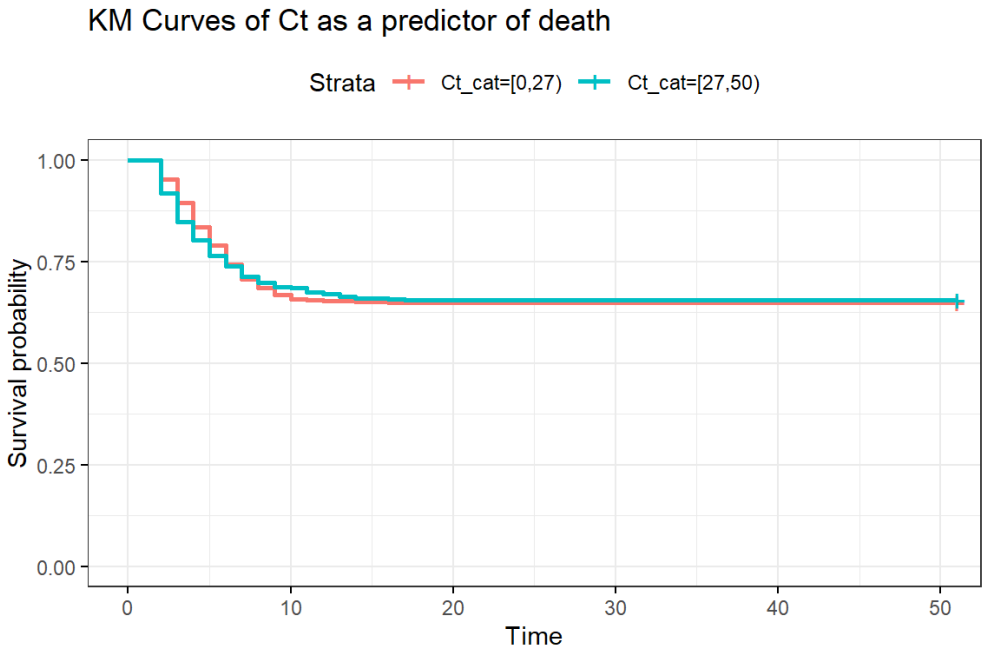

Figure B.4: Kaplan-Meier Curves of Cycle Threshold

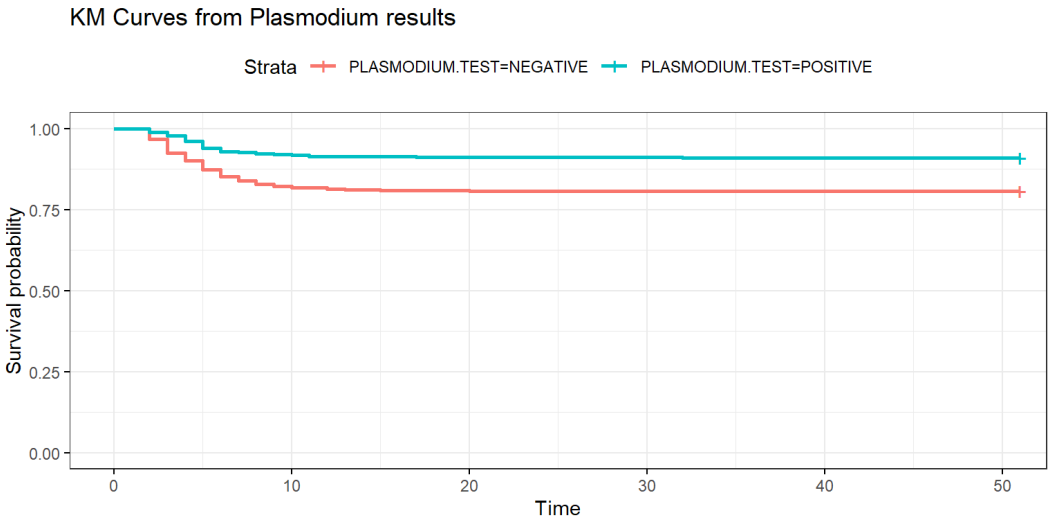

Figure B.5: Kaplan-Meier Curves of Malaria Co-infection

Acknowledgements

I avow the assistance I received in the course of completing this work. My first gratitude goes to God Almighty for the grace and sustenance during the period. The next appreciation goes to my supervisors, Dr. Christiana Kartsonaki, Laura Merson, and Mr. Trokon Yeabah, as well as my tutor James Berinyuy Njong for the immense support and guidance in making this work a success. I also appreciate my family, Dad, Mummy, Godfrey, Anthony, and Samuel, as well as my friends, for the moral and emotional support. Again, I express my profound gratitude to the authors of the various articles and books that provided relevant information, and Infectious Diseases Data Observatory for providing the data.

Without losing sight of the staff members and the entire student body at AIMS Cameroon, I sincerely thank everyone for their inputs.

References

[1] Arranz, J., Lundeby, K. M., Hassan, S., Zabala Fuentes, L. M., San José Garcés, P., Haaskjold, Y. L., Bolkan, H. A., Krogh, K. Ø., Jongopi, J., Mellesmo, S., et al. (2016). Clinical features of suspected ebola cases referred to the moyamba etc, sierra leone: challenges in the later stages of the 2014 outbreak. *BMC infectious diseases*, 16(1):1–9.

[2] Barry, M., Touré, A., Traoré, F. A., Sako, F.-B., Sylla, D., Kpamy, D. O., Bah, E. I., Bangoura, M., Poncin, M., Keita, S., et al. (2015). Clinical predictors of mortality in patients with ebola virus disease. *Clinical infectious diseases*, 60(12):1821–1824.

[3] Borovkova, S. (2002). Analysis of survival data. *Nieuw Archief voor Wiskunde*, 3:302–307.

[4] Center for Disease Control, C. D. C. and Prevention (2019). 2014 – 2016 ebola outbreak in west africa. https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html. [Online; accessed 27-05-2022].

[5] Center for Disease Control, C. D. C. and Prevention (2022). Ebola vaccine: Information about ervebo. https://www.cdc.gov/vhf/ebola/clinicians/vaccine/index.html. [Online; accessed 28-05-2022].

[6] Cherif, M., Koonrungsesomboon, N., Diallo, M., Le Gall, E., Kassé, D., Cherif, F., Koné, A., Diakité, M., Camara, F., and Magassouba, N. (2017). The predictor of mortality outcome in adult patients with ebola virus disease during the 2014–2015 outbreak in guinea. *European Journal of Clinical Microbiology & Infectious Diseases*, 36(4):689–695.

[7] Cherif, M. S., Dumre, S. P., Kassé, D., Cherif, F., Camara, F., Koné, A., Diakité, M., Cissé, M., Diallo, M. P., Le Gall, E., et al. (2018). Prognostic and predictive factors of ebola virus disease outcome in elderly people during the 2014 outbreak in guinea. *The American Journal of Tropical Medicine and Hygiene*, 98(1):198.

[8] Choi, M. J., Cossaboom, C. M., Whitesell, A. N., Dyal, J. W., Joyce, A., Morgan, R. L., Campos-Outcalt, D., Person, M., Ervin, E., Yu, Y. C., et al. (2021). Use of ebola vaccine: recommendations of the advisory committee on immunization practices, united states, 2020. *MMWR Recommendations and Reports*, 70(1):1.

[9] Crowe, S. J., Maenner, M. J., Kuah, S., Erickson, B. R., Coffee, M., Knust, B., Klena, J., Foday, J., Hertz, D., Hermans, V., et al. (2016). Prognostic indicators for ebola patient survival. *Emerging infectious diseases*, 22(2):217.

[10] Edwards, H. M., Counihan, H., Bonnington, C., Achan, J., Hamade, P., and Tibenderana, J. K. (2021). The impact of malaria coinfection on ebola virus disease outcomes: A systematic review and meta-analysis. *PloS one*, 16(5):e0251101.

[11] Fitzpatrick, G., Vogt, F., Moi Gbabai, O. B., Decroo, T., Keane, M., De Clerck, H., Grolla, A., Brechard, R., Stinson, K., and Van Herp, M. (2015). The contribution of ebola viral load at admission and other patient characteristics to mortality in a medecins sans frontieres ebola case management centre, kailahun, sierra leone, june–october 2014. *The Journal of infectious diseases*, 212(11):1752–1758.

[12] Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, 1(4):274.

[13] Hartley, M.-A., Young, A., Tran, A.-M., Okoni-Williams, H. H., Suma, M., Mancuso, B., Al-Dikhari, A., and Faouzi, M. (2017). Predicting ebola severity: a clinical prioritization score for ebola virus disease. *PLoS neglected tropical diseases*, 11(2):e0005265.

[14] Kamorudeen, R. T., Adedokun, K. A., and Olarinmoye, A. O. (2020). Ebola outbreak in west africa, 2014–2016: Epidemic timeline, differential diagnoses, determining factors, and lessons for future response. *Journal of Infection and Public Health*, 13(7):956–962.

[15] Kartsonaki, C. (2016). Survival analysis. *Diagnostic Histopathology*, 22(7):263–270.

[16] Kleinbaum, D. G. and Klein, M. (2004). *Survival analysis*. Springer.

[17] Lanini, S., Portella, G., Vairo, F., Kobinger, G. P., Pesenti, A., Langer, M., Kabia, S., Brogiato, G., Amone, J., Castilletti, C., et al. (2015). Blood kinetics of ebola virus in survivors and nonsurvivors. *The Journal of clinical investigation*, 125(12):4692–4698.

[18] Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.

[19] Lee, E. T. and Wang, J. (2003). *Statistical methods for survival data analysis*, volume 476. John Wiley & Sons.

[20] Malvy, D., McElroy, A. K., de Clerck, H., Günther, S., and van Griensven, J. (2019). Ebola virus disease. *The Lancet*, 393(10174):936–948.

[21] Moore, D. F. (2016). *Applied survival analysis using R*. Springer.

[22] Qureshi, A. I., Chughtai, M., Bah, E. I., Barry, M., Béavogui, K., Loua, T. O., and Malik, A. A. (2015). High survival rates and associated factors among ebola virus disease patients hospitalized at donka national hospital, conakry, guinea. *Journal of Vascular and Interventional Neurology*, 8(1.5):S4.

[23] Rocco, L., Fumagalli, E., Mirelman, A. J., and Suhrcke, M. (2021). Mortality, morbidity and economic growth. *Plos one*, 16(5):e0251424.

[24] Rojek, A., Horby, P., and Dunning, J. (2017). Insights from clinical research completed during the west africa ebola virus disease epidemic. *The Lancet Infectious Diseases*, 17(9):e280–e292.

[25] Rugarabamu, S., Mboera, L., Rweyemamu, M., Mwanyika, G., Lutwama, J., Paweska, J., and Misinzo, G. (2020). Forty-two years of responding to ebola virus outbreaks in sub-saharan africa: a review. *BMJ Global Health*, 5(3):e001955.

[26] Sana, B. Y. (2019). *Risk Factors Associated with the Contraction of Ebola Virus Disease in Liberia*. PhD thesis, Walden University.

[27] Weppelmann, T. A., Donewell, B., Haque, U., Hu, W., Magalhaes, R. J. S., Lubogo, M., Godbless, L., Shabani, S., Maeda, J., Temba, H., et al. (2016). Determinants of patient survival during the 2014 ebola virus disease outbreak in bong county, liberia. *Global Health Research and Policy*, 1(1):1–10.

[28] Woolsey, C. and Geisbert, T. W. (2021). Current state of ebola virus vaccines: A snapshot. *PLoS pathogens*, 17(12):e1010078.

[29] Zhang, X., Rong, Y., Sun, L., Liu, L., Su, H., Zhang, J., Teng, G., Du, N., Chen, H., Fang, Y., et al. (2015). Prognostic analysis of patients with ebola virus disease. *PLoS neglected tropical diseases*, 9(9):e0004113.