UNIVERSITY OF GHANA, LEGON

Survival Analysis Among Tuberculosis Patients: A Case Study of Adults in Kano State in Nigeria

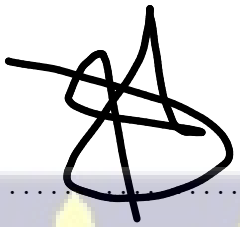
BY IBRAHIM ADAMU (10754228)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF GHANA IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF PHILOSOPHY, ACTUARIAL SCIENCE

May, 2022

Declaration

I hereby declare that this submission is my own work towards the award of the Master of Philosophy degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.



Name of Student	.....	11-05-2022 .....
Student (107514228)	Signature	Date

Certified by:		
Dr.Louis Aseidu	.....	11-05-2022 .....
Principal Supervisor	Signature	Date

Certified by:		
Dr.Samuel Iddi	Samuel Iddi	11/05/2022 .....
Co-supervisor	Signature	Date

Dedication

This research project is dedicated to Almighty Allah and his Prophet Muhammad (SAW) for showering me with knowledge, wisdom, understanding, kindness, protection and provision through the course of this study. Also, to my Mum (Hajiya Aisha Aliyu), Late Father (Alh Adamu Maikifi),Late Sister (Hauwa Adamu) and beloved Niece (Nana Aisha Hamisu) who has returned to the Almighty Allah after a brief illness.

Abstract

Tuberculosis (TB) is an infectious disease that has been considered as a significant risk factor that causes ill health. Globally, it has been found to be among the top 10 causes of death and ranks above HIV/AIDS as a single infectious agent that causes death in patient. Many researches have been documented using semi-parametric and non-parametric models to analyze survival data in Nigeria. There is dearth of studies on the use of parametric models on tuberculosis survival data. Parametric models such as Weibull, Exponential, Log-logistic, Gompertz etc have been used in various studies to analyze data and Weibull was mostly found to be suitable. The popular non-parametric and semi-parametric tests used in various studies include the K-M, Log rank and Cox Proportional hazard model. However, necessary diagnostic checks on model fitness and non-violation of assumptions were mostly ignored. This reduces the reliability of result and increase chance of estimation error. This study assessed the parametric and semi-parametric model of survival such as Cox Model, Weibull, Exponential and Gompertz Models. A retrospective cohort analysis was conducted on the tuberculosis patients receiving treatment under the Tuberculosis & Leprosy Control Program in Kano, Nigeria. The risk factors for death were assessed using the Cox proportional hazard model. The risk factors for death were assessed using the Cox proportional hazard model. The parametric models were compared, and the gompertz model was found to be the best fit for the data based on its minimum AIC & log-likelihood value. Among 2,555 the TB cases, the success rate of TB treatment was 97.06% and the mortality rate was 2.94%. Multivariate analysis showed that HIV, Age & Weight were significant factors associated with mortality in TB patients during therapy. The study recommends the use of diagnostic checks such as Martingale, Deviance Residuals in model fitness. Also, comparism of parametric models is recommended in determination of best model that fits tuberculosis data of

patients.

Key words: Survival Analysis, Kaplan Meier, Cox Proportional Hazard Model, Parametric Models, Tuberculosis.

Acknowledgment

The successful completion of any research work depends on the contribution and assistance received from many people to whom I owe a duty to express my warmest gratitude and appreciation.

Firstly, I am grateful to Federal University Dutse and National Insurance Commission (NAICOM) for finding me worthy of a scholarship to undertake this programme. My sincere gratitude goes to my Former Vice Chancellor, Prof. Fatima Batul Mukhtar, Dr Fatima Abdulkarim (Former HOD, Actuarial Science Department) and Dr Aminu Nass Ma'aruf (HOD, Actuarial Science Department), Dr Sanusi Sa'ad (Former Dean, Faculty of SMS), and my colleagues at the Department of Actuarial Science for the support towards achievement of this goal.

My sincere appreciation goes to my thesis supervisors, Dr Asiedu Louis and Dr Samuel Iddi for their constructive criticisms, invaluable advice, and thorough supervision on the content of this research work. Further gratitude to Prof K. Doku-Amponsah (HOD, Department of Statistics & Actuarial Science, University of Ghana) who has been guiding, mentoring and providing assistance towards successful completion of my studies. Also, I appreciate the moral support of Madam Charlotte Chapman-Wardy and other lecturers of the Department of Actuarial Science and Statistics.

My unreserved gratitude wholeheartedly goes to my family for their guidance, advice, prayers and financial support especially my sibblings (Hamisu, Usman, Jamila, Tijjani & Jubril). Thanks to my Lovely wife (Halima) & son (Faheem) who I cherish. To my friends, Mahmud Nura Ringim, Wilfred Nettey, Samuella Adams, Anasthasia Kukah, Isaac Essel and rest of my course mates for their support during my studies. Special appreciation to Fred Mawuli Amenu.

Contents

List of Abbreviation

PHM ..... Proportional Hazard Model

K-M ..... Kaplan Meier

WHO ..... World Health Organization

TB ..... Tuberculosis

HIV ..... Human Immunodeficiency Virus

AIDS ..... Acquired Immune Deficiency Syndrome

FoMH ..... Federal Ministry of Health

DOTS ..... Directly Observed Treatment Short Course

AFT ..... Accelerated Failure Time

ART ..... Antiretroviral Therapy

CPT ..... Cotrimoxalone Preventive Therapy

AIC ..... Akaike Information Criterion

CI ..... Confidence Interval

NTLCP ..... National Tuberculosis and Leprosy Control Program

HR ..... Hazard Ratio

TB/HIV ..... Tuberculosis and Human Immunodeficiency Virus Infection

MDR ..... Multi-Drug Resistance Tuberculosis

HAART ..... Highly Active Antiretroviral Therapy

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of the Study

Tuberculosis (TB) is an infectious disease that has been considered as a significant risk factor that causes ill health. Globally, it has been found to be among the top 10 causes of death and ranks above HIV/AIDS as a single infectious agent that causes death in patient (WHO, 2017) cited in Asgedom et al. (2018). The contagious bacillus mycobacterium tuberculosis is responsible for the disease. TB can be easily contacted when a person is exposed to the bacteria expelled through coughing by a TB infected person. Based on the affected sites, TB can be broadly categorized as either pulmonary or extrapulmonary. The Tuberculosis of the lungs is called Pulmonary TB while TB that affects other parts of the body such as abdomen, bone, spinal cord etc is referred to as Extrapulmonary TB. According WHO (2019) many people face the risk of developing TB disease as studies have found that out of the global population, mycobacterium TB has infected about 25% (WHO, 2019).

Tuberculosis is still a major public health issue around the world. Despite the fact that attempts to control the pandemic have lowered mortality and incidence, there are a number of predisposing factors that should be changed in order to reduce the disease's burden such as illicit drug abuse, alcoholism, smoking etc. this can be achieved through the evaluation of illicit drug user's behavior and establishment of policies that can intervene in controlling the spread of the disease (Silva et al, 2018).

Globally, Tuberculosis (TB) remains a significant issue that has bedeviled public health system. However, the discovery of chemotherapy has been effective in controlling the worsening effect of the disease. Studies have shown that large number of deaths recorded annually is still attributable to the Tuberculosis. Early diagnosis or notification of suspected cases (especially in less developed countries) still remains a major challenge even countries with advanced medical facilities still face such challenge (Smith et.al, 2006). Despite scientific advances aimed at reducing the disease's negative effects, millions of people contract tuberculosis each year (WHO, 2018). From the total deaths in world population recorded in 2017, about 1.6 million deaths are attributed to TB infection. TB-HIV co-infected patients accounted for only 18.75% out of the 1.6 million TB related deaths while Patients with negative HIV accounted for 81.25% of the TB related deaths (WHO, 2018).

Tuberculosis affects both male and female gender and can be present in both adults and children. Most studies found the TB to be predominant among adults (male in most cases) especially those that are above 15 years of age. These adults accounted for 57% of all TB cases while women accounted for 32% of TB related deaths in 2018. Due to advancement in research, early detection and treatment of HIV/AIDS patient, deaths among patients who are con-infected with TB accounted for less than 9% of all global TB related deaths (WHO, 2019).

In 2014, a total of 1.5 million of global deaths were TB related. Most of these deaths are found to have occurred within sub-Saharan Africa and South-East Asia (WHO, 2015). According to McNerney (2012) cited in Adamu et.al (2017), infection with HIV, rising drug resistance and co-morbidities like diabetes, as well as social deprivation (such as poverty and illiteracy), further amplified by out-dated diagnostics, and treatment are significant risk factors that leads to increase in TB infection.

There are still cases of large under reporting of TB cases globally. The gap has

been estimated at 3 million confirmed cases of TB not reported. Out of this number, 80% are found within ten countries. Nigeria represents 12% of the top four countries, others are India, Indonesia and Philippines with 25%, 10% and 8% respectively. Nigeria ranks second among countries that accounted for the highest gap in reported TB with 13% below India among top four countries identified. As a result, Nigeria has Africa's highest tuberculosis burden and one of the world's largest disparities between estimated and reported cases (WHO, 2019).

The year 2020 witnessed the outbreak of the novel Coronavirus Disease (COVID-19). COVID-19 is a respiratory infection caused by the coronavirus 2 that causes severe acute respiratory syndrome (SARS-CoV-2) according to Lai et al (2020) cited in Mustapha et al (2020). This virus mostly affects the lungs and causes illness such as common cold. The disease is highly contagious and easily transmissible. An infected person can easily transmit the virus through coughing similar to tuberculosis. The WHO in first quarter of 2020 labelled it a pandemic after a surge in cases recorded across the globe. Limited access to safe, effective, high quality, and inexpensive medicines exacerbates these illness loads (WHO, 2020).

In 2020, an estimated 1.8 million individuals would die from TB disease alone with an additional 200,000 to 400,000 deaths based on a model developed by WHO. This additional deaths projected to occur in 2020 if there is a 25% decline in detection and treatment of TB infected patients over a period of three month largely due to actions/inactions of government in prioritizing the control of COVID spread (WHO, 2020). Stop TB partnership projected 1.3 million additional deaths in TB patients to occur between 2020 and 2025 due to effect of stringent lockdown measures implemented in 2020. It has been reported that there are various declines in cases notification among 14 countries rated as high burden between January and June 2020. India, Indonesia, the Philippines, and South Africa accounted for 44% of worldwide tuberculosis infections in 2020, with a more than 25% decrease in TB notification and an 80% decline in daily

notification cases (especially in periods immediately after lockdown imposition) compared to 2019 (Ravelo, 2020). During the lockdown between period of March and April, India recorded its major decline in TB notification in 2020 before eventually picking up while that of Indonesia was between March and May of the same year. There was also a decline in TB notification cases in the Philippines from January until an eventual pick up from April 2020. On the other hand, South Africa's TB notification drop was observed from March to June 2020 (Ravelo, 2020).

According to report released by Devex (2020), people's avoidance of health facilities, insufficient number of health facilities for TB cases, disruptions in the acquisition and transportation of medical supplies, movement restrictions combined with partial and full lockdowns imposed by the government, and loss of livelihood due to a decline in economic activities are all possible reasons for a decrease in TB notification cases. COVID-19 has the potential to raise the number of persons getting tuberculosis by more than one million per year between 2020 and 2025. The report further stated that "Although physical distancing policies may help to reduce TB transmission, this effect could be offset by longer durations of infectiousness, increased household exposure to TB infection, worsening treatment outcomes, and higher levels of poverty" (Ravelo, 2020).

Inadequate infrastructure and weak healthcare systems have hampered responses to African epidemics, including a lack of proper monitoring to establish the scope of the outbreak and insufficient mechanisms to prevent, diagnose, and treat infections (Mustapha et al, 2020). The various lockdown measures taken to curb the wide spread of the disease in Nigeria has affected economic activities, create fear and anxiety among people. There were several mysterious death recorded in Kano and the exact cause is yet to be established during the lockdown period. Although many attributed the cause of death to the coronavirus disease but not scientifically proven which is largely due to inability for autopsies to be conducted

on the dead. The relationship between the "mystery" disease and COVID-19 was met with a slew of ambiguity, equivocation, and denial in Kano, with state officials dismissing any link between deaths from the strange disease and COVID-19 (Nwozor et al, 2020). Other observers blamed the deaths on a shortage of medical care due to the state's shutdown of health facilities (Kanabe, 2020). In the same vein, one may see that inability to access health facilities during this period may slow down efforts in detecting TB patients and affect their ability to access drugs. There may be increased number of loss to follow up during this period compared to the previous. The coronavirus disease outbreak and measures taken to curb it may have impacted on the ability of TB patients to easily access the hospital facility for regular drugs or diagnosis to confirm TB infections.

In epidemiology, data of participants are being analyzed using survival analysis after they have been followed up to a particular time when an event of interest has occurred or otherwise. The best statistical method usually applied in this type of survival analysis is the Kaplan Meier estimate. This methods helps to analyze the data and to also compare survival among independent groups such as control and treatment groups. To compare survival probability among two or more independent groups, the log rank test statistic is used. It is used to test the null hypothesis that the survival between two or more independent groups are equal (Ilker, Sulaiman & Rukayya, 2017).

Survival times are time intervals measured from a certain starting point to the occurrence of a specific event, such as from the time of diagnosis of a disease to the time of death (Bewick, Cheek & Ball, 2004; Goel, Khanna & Kishore, 2010). During survival studies or time to event study, some event may not have occurred (e.g. death) during the follow-up time. This may be due to Loss to Follow Up or the subject has not suffered the event of interest at end of the study period. Such subjects are considered censored and this characteristics is a unique feature in survival studies.

Survival analysis can also be applied in a business setting. Risk factors that can affect the survival of a business can be assessed using the multivariate Cox Regression Model. In a study conducted to predict risk factors associated with bank failures in Nigeria, the Cox Proportional hazards was adopted. Data were obtained from financial statement of banks for a period of 9 years. The researchers identified 12 risk factors that can lead to bank failures. The study recommended that regulators should design specific policies to address these factors to avoid collapse of banks which can adversely affect the economy (Babajide, Olokoyo & Adegboye, 2013).

To compare survival among two or more independent groups, the Kaplan-Meier Survival Estimate is used. Kaplan-Meier analyses can be used in a variety of fields, including medicine, engineer, economics, physics, and demography. Kaplan-Meier's example can be seen in cancer studies, where patients are monitored for a set period of time until they either die, relapse or drop out of the research dependent on the event of interest. Patients who drop out of the study or are considered censored. The Kaplan-Meier approach, which is non-parametric, can be used to determine the percentage of patients that lived beyond a certain time period (Ilker, Sulaiman & Rukayya, 2017). Several statistical tests have been developed to compare survival among groups such as Peto & Peto (1972), Tarone & Ware (1977), Kalbfleisch & Prentice (1980), and Cox & Oakes (1984).

The proportion of subjects living within a given period of time can be best estimated using the Kaplan Meier. The effect of an intervention or medication administered on patients is usually assessed by measuring the number of survivors during the period of the intervention (Armitage, Berry and Mathews, 2002).

1.2 Problem Statement

Over the years, there has been strenuous efforts by both local and international organisations in collaboration with various government agencies to curb the menace of TB. However, the disease continues to constitute a major health challenge that has led to several deaths in the world. More efforts are required to combat this deadly virus. The need to equip our health facilities cannot be overemphasized as it can help in accurate diagnosis of presumptive cases as can be seen in the use of GenExpert machines to detect traces of TB. Also, having qualified medical practitioners with proper training that can improve clinically confirmed cases for immediate treatment (Murray, 2018).

In Nigeria, The incidence of TB increased by almost 3% in 2018 as compared to 2017 which showed a contrast with reported improvement worldwide. Consequently, fatality rates also increased to 157,000 in 2018 from reported case of 155,000 in the previous year. There is insufficient treatment of TB infected patients which stood at less than 25% of the entire confirmed cases during these periods (Adepoju, 2020).

The management of TB patients in an effective manner is vital towards improving survival among patients (especially those with HIV/AIDS co-infection). lack of proper health care to such patient can lead to increase in spread of the disease which can hinder economic growth as infected patients may not be able to be engaged in productive activities. Identification of the health risk factors can go a long way in drug administration and prioritizing the patients based on those identified to have a severe risk factor that can easily cause death or suffering to such patient if not urgently attended to. identifying major risk factors through research can help medical practitioners in curbing the menace of the virus and saving lives.

Although several studies have been conducted on TB treatment outcomes in Nigeria (Alobu et al., 2014; Adamu et al., 2017; Dauda 2010; Fatiregun et al., 2009; Ifebunandu et.al, 2012; Ige & Akindele, 2011; Ukwaja et.al, 2014; Peters et.al, 2004; Salami & Oluboyo, 2003; Michael & Bolarinwa, 2020). Most of these studies use non-parametric and semi-parametric models for the survival analysis with only few studies applying the parametric models without employing diagnosis on the model such as assumption of hazard proportionality of the cox model, linearity of continuous variable or assess goodness of fit for the model. Failure to carryout model diagnostic checks and assumptions makes the result less reliable and increases the probability of errors. This study will also compare survival of patients using the multivariate analysis methods of both parametric (such as Gompertz, Weibull, Exponential Model) and the semi-parametric (Cox) proportional hazard models. Necessary statistical tools will be employed to select the best model that fits the data of TB patients.

1.3 Objectives

The general objective of this study is to estimate the survival probability of TB patients receiving treatment in various health facilities in Kano State Metropolis, Nigeria.

Specific objectives of the study is to:

- Estimate survival times of TB patients using non-parametric survival model.
- Compare the survival functions of TB patients in Kano State
- Assess the effect of risk factors for mortality of TB Patients in Kano State using the Cox Proportional Hazard Model.
- Compare Parametric models of the Proportional hazard by their log-likelihood

and AIC values to select the best model that fits the tuberculosis data.

1.4 Significance of the Study

It will assist policy makers, health care professionals and public in creating further awareness on the risks associated with TB that can lead to premature deaths. It will assist in using the appropriate survival model to assess risk factors associated with death in patients and make necessary policy to improve survival of patients. This study can also help policy makers in closing the widening gaps in TB reported cases so as to increase activities that can curb the menace of the contagious disease.

The results of this study may provide useful information helpful to health care professionals, clinicians, policy makers, and health educators, and enlighten the public on TB health risk factors that can adversely affect survival of patients. The appropriate model to be used in analyzing tuberculosis data will be recommended from the compared parametric and non-parametric models. This will help in adopting the best model in analyzing tuberculosis data by policy makers, healthcare professionals and researchers.

1.5 Scope of the study

This is a retrospective cohort study of adult TB patients registered for treatment at various DOTS (Directly Observed Treatment-Short Course) Unit of a licensed TB treatment facility within metropolitan area of Kano State, Nigeria for the period of January 1, 2019 to December 31, 2020. Patients record are followed up to July 31, 2021.The event of interest is time to death recorded during the period of study among the infected patients as per the TB register.

1.6 Organization of Study

The study will be broken down into five sections. The context of the study, problem statements, explanation of objectives, and significance of the study including scope of the investigation are all covered in the first chapter. Chapter two contains existing and relevant literature reviews of scholarly publications on the subject matter (conceptual and empirical reviews) as well as a discussion of the underlying theory.

Chapter three discusses the researcher's method for presenting the study's findings. The research design, sampling technique, source of data, statistical techniques in data analysis and analytical tools will all be covered.

Chapter four presents the results, analysis and discussions of findings from the research.

The fifth chapter will summarize the findings, draw pertinent conclusions, and provide recommendations based on findings. At this point, the researcher will assess the data collected and, if necessary, consult supplementary literature to back up or refute the findings.

Chapter 2

Literature Review

2.1 Introduction

This chapter discusses tuberculosis infection from conceptual standpoint, and empirical review of risk factors associated with TB deaths among patients. It also provides some explanations on the survival model to be employed in the study and other test statistics relevant in analysing the data for this study. The review also summarizes the findings of previous studies on this topic.

2.2 Tuberculosis Disease (TB) Epidemiology

Tuberculosis is a fatal disease that is highly contagious. It ranks above HIV/AIDS as a single infectious virus that has led to several deaths in the world. Once a person contracts the virus, it weakens the immune system thereby causing sickness in patients. According to WHO (2017) cited in Asgedom et al. (2018), Tuberculosis is one of the diseases that contribute to global death rates. Mycobacterium tuberculosis is the bacteria that causes tuberculosis. TB is an airborne disease that can be easily contracted through close contact with an infected person especially when such person released droplets of the bacteria into the air by coughing. It can affect several part of the body such as lungs, bone, spinal cord, neck etc. Most cases of tuberculosis reported have found to be Pulmonary (TB of the lungs) with few cases of Extrpulmonary (TB of other parts of the body excluding the lungs). About 25% of the world's population stands at risk of developing tuberculosis

as studies have shown that these estimated persons are infected with Mycobacterium tuberculosis (WHO, 2019).

According to WHO (2019) in its Global Report, transmission of TB can be curbed when infected patients are diagnosed and treated early. Treatment usually requires an administration of drugs to TB patients for a period of 6 months. In treatment of the disease, the first-line antibiotics have been found to be effective with most patients being cured after treatment completion with few deaths or relapse cases being reported. Tobacco smoking, diabetes and HIV are found to be significant health risk factors that increase TB infection among patients. Other determinants of TB infection include poverty and malnutrition. TB related deaths can be reduced when measures are taken to prevent latent TB infection, improved quality of life and reduction in risk factors that can affect health.

According to the Nigerian 2019 Annual TB Report issued by the Nigerian Federal Ministry of Health (FMoH), "Nigeria has the highest burden of TB in Africa and is among the eight countries that accounted for two thirds of the Global TB burden". Nigeria is ranked 6th out of 30 countries with high TB burdens, as well as first in Africa with the greatest TB burden. In addition, the country is one of the top three global high burden countries, with high TB, TB patients co-infected with HIV and MDR-TB (Multi Drug Resistance TB) mortality rates. The report further shows that TB is a major public health challenge in Nigeria that needs to be dealt with decisively (FMoH, 2019).

The disease can be seen in both pulmonary and extrapulmonary areas of the human body. Around 80% of TB cases, which mostly affect the lungs, are located in the former, whereas 20% of cases are found in the latter. The infection of other organs via the circulation, lymphatic vessels, or proximal transmission from one organ to another is known as extrapulmonary tuberculosis. Management of

tuberculosis has usually been issued via guidelines and recommendations of the WHO which is regularly updated in line with recent findings. The guidelines clearly states the treatment regimens, approved anti tuberculosis drugs, and the dosage of anti tuberculosis drugs (WHO, 2019).

2.3 Survival Analysis and TB related deaths

Ajagbe, Kabir and O'Connor (2014) conducted a study of adult TB patients in Ireland. Survival Analysis tools were used to analyze the survival data. It was a retrospective cohort study where data of 647 confirmed TB cases were reviewed. Medical records of adult TB patients that have been bacteriologically confirmed were obtained from two teaching hospitals. Health risk factors that may likely affect survival were also obtained for the analysis. These risk factors were used as predictor variables in the Cox regression Model to determine impact of such factors on survival and in computing the hazard ratio. In the univariate model, survival among independent groups were estimated and compared using the Kaplan-Meier method. In comparing survival curves between male and female gender, the study found no significant difference. Also, the K-M curves found no significant difference among the age groups. However, survival time of men is less than that of women. This is supported by the higher hazard ratio of men as compared to women. Proper medication for identified risk factors such as Anti-diabetic medication and other immunosuppressive drugs increases survival time. Decreased survival time is found in patients who engage in alcohol consumption & tobacco smoking (Ajagbe, Kabir & O'Connor, 2014).

Adamu et al. (2017) investigated the cause of deaths among TB Patients in Nigeria. The researchers' analysed data for 5 years from records of TB Patients

receiving treatments at a large tertiary hospital. The study found that 16.6% of TB Patients died during the period after commencing TB treatment and most of the deaths occurred within the period of 0-1 month of treatment initiation. Risk factors associated with deaths in TB patients was evaluated using Cox Proportional Hazards model. It was discovered that TB/HIV co-infected patients have higher risk of death than those who are not (aHR 1.39(CI:1.04–1.85)), Patients who reside not within the city (aHR 3.18(CI:2.28–4.45)) are at higher risk compared to those who have close proximity to the health facility in the city. Patients with previous tuberculosis therapy (aHR 3.48(CI:2.54–4.77)) have lower survival than new patients. Patients on treatment based on clinically suggestive TB have increased risk of death than those who were bacteriologically confirmed (aHR 4.96(CI:2.69–9.17)), there is 1.45 times increased risk of death in patients with both pulmonary and extrapulmonary TB than those with one of the two types of TB (aHR 1.45(CI:1.03–2.02)), and patients that were on referral from other health facilities other than one being monitored by TB programme have higher risk of death than those were receiving treatment from a health facility coordinated by the TB programme (aHR 3.02(CI:2.01–4.53)). The study concluded that patients who were earlier diagnosed with TB and placed on ART have lesser risk of deaths from the disease compared to those not receiving ART, the distance from place of residence to the treatment centre hinders patients from accessing the medication. This may be due lack of funds to transport patient to such facility.

In high burden settings, TB related deaths among patients occur early during the period of therapy. Studies have shown that deaths are mostly recorded during the first 2 months of treatment and is prevalence among adult patients (<49 years). Also, poverty, HIV co-infection in TB Patients and low body mass index (BMI) have been found to be contributory factors in increased death rates among TB patients. Prolonged illness caused by having advanced diseases in patients such as HIV (< 200 CD4 cell count) in adult patients, children below the age of 5,

acute malnutrition etc. affects survival of TB patients. Most deaths during early stage of TB-HIV co-infected patients that are receiving therapy are caused by TB infection rather than HIV (Bhargava and Bhargava, 2020).

A prospective cohort research was conducted among patients from various treatment centres in Tanzania to assess risk variables linked to deaths in patients with pulmonary tuberculosis. Survival among HIV co-infected patients and non-HIV infected patients were analyzed. Patients were followed up from time of treatment therapy to the period of treatment completion in Tanzanian health facilities by Nagu et al. (2017). The study found that two third of deaths among tuberculosis patients died during the first two months of treatment (intensive phase) with 67% of total deaths among patients occurring early in the study. During the study, 1,696 patients were enrolled out of which 58 deaths (3.4%) were recorded. The mortality risk among patients co-infected with HIV who commenced antiretroviral therapy (ART) within two weeks after initiating TB treatment was less as compared to patients who are not co-infected with HIV (RR = 3.55; 95% CI: 1.44, 8.73 $p < 0.0001$). The risk was higher in patients who commenced ART within 3 months before anti-TB therapy (RR = 10; 95% CI: 3.28, 30.54; $p < 0.0001$).

A prospective cohort study conducted in Nigeria on Tuberculosis patients with HIV co-infection receiving treatment at an Infectious Diseases Centre found that efficacy rate was low as only 40% of patients within the cohort were cured of the disease after 8th month when treatment has been completed. Additionally, most of the cured patients from the disease initially had a negative sputum for AFB after the 2 months intensive phase of regimen. Most patients are found to maintain the negative sputum status after the intensive phase with only few additional patients (not more than 6%) reporting a negative sputum during the continuation phase and at completion of treatment. After a positive sputum AFB test observed for patients at the intensive phase, the chance of getting completely

cured is low for such patients at the continuation or completion stage. Hence, there is need for improved drugs administration that can be more effective in curing TB-HIV patients (Dauda, 2010).

A study found that mortality rates in TB patients that are not engaged in tobacco smoking is less as compared to TB patients who are smokers. The risk of death among TB patients is significantly higher among smokers. Hazard rates in TB patients that have history of smoking is 9 times higher than non-smokers. An important factor that contributes to reduction in mortality cases among TB patients is ability to quit smoking (Mollel & Chilongola, 2017). The risk of TB related deaths is significantly reduced when smokers quit smoking. Termination of smoking by patients under treatment therapy has significant impact in reduced mortality among TB patients. This is supported by finding in the study which shows a 65% drop in TB related deaths among patients who quit smoking (Mollell & Chilongola, 2017).

Previous history of TB (especially among men) is found to be related with increase in risk of deaths among TB Patients. There have been reports of lung damage after TB patient has completed treatment or cured. Impaired lungs that caused difficulty in breathing after completion of TB therapy has been found among men in Brazil (Menezes et al., 2007). Most patients who reported to have experienced difficulty in breathing had a previous history of TB treatment. The study conducted in Brazil found that the risk in experiencing the event of airway obstruction among men with previous TB treatment is 4.1 times greater than those who had no prior history of being treated for tuberculosis. The result of the finding was not altered after adjusting for several predictor factors associated with airway obstruction such as literacy level, smoking, exposure to smoke or dust, history of childhood respiratory disease, gender and age of subjects in the study. The study concluded that airway obstruction is prevalent in advanced age

adults who had prior history of tuberculosis infection (Menezes et al., 2007).

Several studies have shown a correlation between risk of tuberculosis infection and diabetes mellitus. The risk of tuberculosis infection increases for patients diagnosed with Diabetes Mellitus (DM). Chances of progressing from latent to active tuberculosis is higher among Diabetes Mellitus Patients (WHO, 2016). According to case-control studies, the risk of contracting tuberculosis is 2.44 to 8.33 times higher in individuals with DM than in those without (Shetty et al., 2006 & Jabbar et al, 2006).

Diabetes Mellitus Patients have a higher risk of being infected with Tuberculosis. In a comprehensive evaluation of 13 observational studies, Jeon & Murray (2008) cited in Silva et al (2018) discovered that diabetes mellitus patients have higher risk of being infected with tuberculosis or to develop tuberculosis illness with 3 times higher risk of TB infection as compared to patients who have not been diagnosed with diabetes mellitus (relative risk = 3.11; CI: 2.27-4.26). Thus, DM is an important risk factor that can cause TB infection among patients.

In a related study, individuals with a hemoglobin A1c (HbA1c) level not lower than 7% had three times the risk of TB infection as those with a HbA1c level less than 7% (HR = 3.11; CI: 1.63-5.92). The risk of developing tuberculosis is higher among patients with increased insulin intake compared to those who use less of insulin in diabetes treatment. The risk is twice as high among patients who take more dosage of insulin (>40 units) than those who use less (Dooley & Chaisson, 2009) cited in Silva et al (2018).

Shimazaki et al (2013) examined the factors relating to deaths among hospitalized Tuberculosis patients that are HIV negative in Philippines during a 3 month study conducted in 2009. 151 out of 403 patients died while receiving treatment in the hospital (37.5%) due to tuberculosis. Bacterial pneumonia, anorexia, anaemia

and older age are found to be risk factors significantly associated with deaths among admitted patients. Poor urban area in Philippines recorded high case of TB patients' mortality. Risk factors associated with deaths among hospitalized TB patients was analyzed during the study. It was found that complications of bacterial pneumonia is a significant factor associated with deaths in hospitalized patients (aOR 4.53, 95%CI: 2.65–7.72). Pneumonia which mostly affects the lungs, is the highest factor that causes deaths in patients. Anorexia (aOR 3.01, 95%CI: 1.55–5.84), anaemia of less than 10 g/dl haemoglobin level (aOR 2.35, 95%CI: 1.34– 4.13) and older age (aOR 1.85, 95%CI: 1.08– 3.17) are the next significant factors associated with deaths. Haemoptysis was also linked to a higher chance of survival (aOR 0.44, 95%CI 0.25–0.80). Due to recall bias as exact moment of onset of disease could not be confirmed, the study could not ascertain the existence of any significant association between the duration from detection of TB infection to time of hospitalization and mortality.

Lack of proper diagnosis leads to increase in poor treatment outcome among TB patients. In a study on TB patients' treatment and determinant of outcomes in Nigeria, it was discovered that patients who were not properly diagnosed of the disease have a higher risk of poor treatment outcome, and male patients have a higher risk of poor treatment outcome than their female counterpart (Fatiregun et.al, 2009).

In a similar study, Hayibor, Bandoh, Asante-Poku, and Kenu (2020) found HIV/TB co-infection, older age, previous TB treatment and category of patients to be significant in determining patients' TB treatment outcome in a study conducted in Accra. Type of TB was not significantly related to adverse treatment outcome in TB patients.

According to Subramani et al (2008) cited in Bhargava and Bhargava (2020), TB burden and related mortality is higher in both low and middle income countries.

The study found that deaths in younger age groups in these countries is higher as compared to adult patients who are much older (> 44 years). In rural South India, TB related deaths among patients in the age group of < 44 years was found to be 12 times higher than those expected in the entire population.

Asgedom et al. (2018) conducted a study in Ethiopia to assess the survival and risk factors that affects survival among tuberculosis (TB) patients. Death was the outcome of interest. Place of residence, age, gender, pre-treatment weight, type of therapy (Antiretroviral or cotrimoxazole preventive therapy), Type of TB, TB/HIV co-infection, year of anti-TB therapy and TB category at presentation were the predictor variables. Kaplan Meier, Cox Proportional were used in the univariate and multivariate analysis respectively. To compare survival among groups, Log Rank, Tarone-Ware and Generalized Wilcoxon (Breslow) Tests were used. Using Kaplan Meier, TB patients infected with HIV on either ART (antiretroviral therapy) or CPT (Cotrimoxalone Preventive Therapy) treatment and Type of TB were found to be significantly associated with survival ($p < 0.05$). Also, Results from cox proportional hazard models showed that site of TB and CPT were significant predictors of survival. Extra-pulmonary TB patients were 17 times more likely than pulmonary TB patients to have the event of interest (HR = 17.38, 95% CI; 3.88– 77.86, $p < 0.001$). Furthermore, patients with TB/HIV co-infection who were receiving CPT had an 85% lower risk of dying from TB infection than those who were not on CPT (HR = 0.15, 95% CI; 0.03–0.74, $p = 0.02$).

A study conducted on timing and causes of deaths among TB Patients in South Africa by Field et al. (2014) found that TB related deaths was higher among patients during the first month after treatment initiation. Also, Patients with HIV Positive not on ART have higher mortality risk than TB-HIV patients who have initiated treatment. Risk factors associated with deaths in TB Patients was found to include older age, previous history TB, HIV Positive, pulmonary TB,

and uncertainty in TB diagnosis (Field et al., 2014).

Age, gender, type of tuberculosis and drug susceptibility profile affects the risk of death due to Tuberculosis disease. However, age as a predictor variable of risk of death varies according to countries. In some European and Asian countries, mortality has been found to be associated with older people as people above 45 years recorded more mortality in TB Patients (Lee et al., 2017 & Shuldiner et al., 2014).

Bajehson et al. (2019) studied the factors that are related to deaths among patients with Multidrug resistant Tuberculosis in Northern Nigeria. Survival data were analyzed using the non-parametric model (Kaplan-Meier) and semi-parametric (Cox proportional hazard models) were used to assess impact of co-variates on survival. During the study, data of 147 confirmed patients with Drug resistance were analyzed and 25% of DR-TB patients died. Probability of mortality increases among TB patients with HIV co-infection. Delays by patients in receiving treatment after initial diagnosis of confirmed TB case reduces their survival probability. There was a significant negative association between survival and delay in treatment initiation of patients. The study also found that survival probability of patients is not significantly influenced by the mode of care (facility or community based).

Many studies have considered HIV as predictor variable in assessing the incident of Tuberculosis infection (Akesa et al., 2015; Ifebunadu & Ukwaja, 2012; Adamu et al., 2017). According to a study by Pathmanathan et al. (2017) the use of ART (antiretroviral therapy treatment) for HIV infected patients reduces mortality and morbidity among Tuberculosis patients. However, the use of ART does not grant immunity to patients against TB infection. ART administration issued to HIV patients serves as treatment to such patients and they can get co-infected with TB while on ART. The study found that the rate of TB infection among HIV patients on ART was low which may be related to the awareness of such risks

among patients after ART initiation. Advanced HIV, history of previous TB infection and non-treatment of suspected TB cases are found to be predictors of TB incidence among HIV Patients (Pathmanathan et al., 2017).

Lin et al. (2014) found that most TB related mortality in patients occurred within three weeks during the study and are caused by septic shocks. The study was conducted to ascertain deaths among TB patients and classify deaths based on TB and non-TB related deaths during the 5 year retrospective study. Other co-morbidities found to be predictors of TB deaths include malignancy, liver cirrhosis, renal failure, cavitary and radiographic patterns (military and pneumonic) etc. the study concluded that most deaths recorded among tuberculosis patients are non-TB related as such timely recognition of clinical manifestations that causes septic shock can help prevent TB deaths. The study concluded that deaths in TB patients are due to septic shocks rather than the TB itself (Lin et al., 2014).

Patients with AIDS-TB con-infection have higher risk of death. The risk among co-infected patients was 1.65 times the not co-infected in a study conducted in Brazil. Factors associated with increased survival was found to be Female gender, years of education (minimum of 8 years education experience) and CD4 diagnostic criteria. On the other hand, non-use of HAART (Highly Active Antiretroviral Therapy) for AIDS infected patients, no prior investigation on Hepatitis B status of patients during enrolment, age (above 60 years) and patients with more than two opportunistic infections (such as candidiasis, pneumonia etc) are associated with lower survival (Melo, Donalisio, and Cordeiro, 2017).

Aung et al. (2019) found that survival rate among TB/HIV con-infected patient's declines during a 12 year retrospective cohort study in Myanmar (82.0% at 5 years and 58.1% at 10 years) and some patients that are co-infected with HIV but are receiving the ART died within a period of ten years from when the ART treatment has commenced. About 40% of such TB/HIV patients died.

In India, Pardeshi (2009) studied the survival trends of patients on Directly Observed Treatment-Short Course (DOTS) based on treatment categories, age, and sex. The survival of patients in three categories was compared using Kaplan-Meier plots and log rank tests. In multivariate analysis, Cox Proportional Hazard was utilized to analyze the effect of risk factors on survival. The study discovered a significant difference in patients' survival curves across the three DOTS categories (i.e new smear positive PTB, sputum smear positive for TB relapsed and new smear negative PTB & less EPTB), between pulmonary and extrapulmonary TB (HR=0.61; P= 0.45) but not between male and female sex (HR= 0.87; p = 0.69) or between new and retreatment patients (HR = 0.75, p= 0.94). Patients with age above 40 years have higher risk of death (HR=7.81, p=0.012).

Ghazal et al. (2014) investigated factors associated with mortality among in-hospital Tuberculosis patients in Pakistan. It was a retrospective cohort study of 120 patients. The patients were divided into two equal groups of 60 patients each for cases (i.e those who did not survive hospitalization) and control group (i.e those who were discharged after treatment). Positively diagnosed tuberculosis patients who were hospitalized in the same month of the year were included in the study. Late presentation of disease for treatment led to increase in early death among patients, lack of compliance with the treatment therapy, extended period of illness prior to treatment initiation, and low body weight were all found to be significant risk factors for mortality in hospitalized patients at 5% level of significance. Comorbidities such as leukocytes and low serum protein were found to be significantly related to deaths among hospitalized TB patients.

Lefebvre & Falzon (2008) conducted a study to assess the risk factors associated with Tuberculosis related deaths among patients in Europe. The study employed case based data of patients from 15 European Union Countries within the period of 2002-2004. The study found that risk of deaths varies among the countries.

However, the most significant factors of death were determined to be old age and treatment resistance. In bivariate analysis, the following characteristics were found to be significantly related with death: above 19 years old, male sex, pulmonary TB, and a previous history of TB. Multi-Drug resistance TB (MDR) was connected to death, and this connection was nearly as twice as high in individuals with a past history of tuberculosis (secondary MDR) than in cases without a prior history of tuberculosis (primary MDR). The length of time between notice and death had no bearing on the outcome. Patients with tuberculosis in Europe have a two to five time higher risk of dying than those in other parts of the world. Male gender, Age, European origin, pulmonary TB, and MDR, were all found to be significantly associated with patient death in a multivariate analysis, with the risk of death being higher for secondary MDR (OR 3.6, 95% CI 3.0–4.3) than for primary MDR (OR 2.5, 95% CI 2.0–3.1). Portugal and the Czech Republic had ORs(Odds Ratio) that were much lower than the reference country (Lefebvre & Falzon, 2008).

Advanced age, male gender, TB/HIV co-infection, first sputum positivity, TB retreatment, and delayed visit were all identified to be risk factors for mortality among pulmonary tuberculosis patients in a retrospective analysis of 7,032 TB patients in China. The forward stepwise cox model based on partial maximum likelihood was used to examine the risk factors, while Kaplan-Meier was employed to predict survival probability. The study found that visit delays (greater than 14 days after initial diagnosis) is the risk factor associated with TB-related deaths (RR=1.386, 95% C.I = 1.096-1.753) and survival probability of patients with pulmonary TB is not affected by a delay in diagnosis prior to therapy. The risk of death was 1.8 times greater among patients who are male than females (RR=1.847, 95% C.I = 1.387-2.459) while age (RR= 1.059, 95% C.I = 1.051 – 1.067) is associated with increase in risk of death by 5.9% (Yi et al., 2020).

2.4 Related Literature on Parametric Models

Saroj (2019) conducted a study on data of under-five child mortality in Indian setting. The objective of the study was to identify the risk factors which affect under-five child mortality using both parametric and semi-parametric survival models. Cox proportional hazard model and parametric models were compared to find out the best parametric model for under-five child mortality. Accelerated Failure Times (AFT) Models of the Weibull, exponential, log-normal, and log-logistic were used for this study. The weibull distribution was found to be the best of the four parametric models based on the minimum value using AIC criteria.

Michael & Bolarinwa (2020) used data of Tuberculosis Patients in Nigeria to model survival using parametric models. It is a retrospective cohort study of TB patients between the periods of 2010 to 2016. The time it took to recover from tuberculosis infection was the outcome variable. In order to model survival of TB patients, the study used the AFT Models of the Weibull, Exponential and Log-logistic. To find the best model that fits the data, the AIC criteria was utilized. The impact of covariates on survival (time to recovery) of TB patients was assessed using age, gender, type of TB and patient occupation. Age, gender and occupation were identified to be key factors of recovery for TB patients. The Weibull AFT Model with the lowest AIC value was also discovered to be the best model in assessing TB patients' survival.

Parametric models were also adopted in a breast cancer studies conducted in Malaysia. The researchers used data set of breast cancer patients during the period of December 2008 and February 2017. Age and type of treatment were the covariates modelled against the time of breast cancer infection (survival time, t). The performance of each of the 3 models were compared using the log likelihood, AIC and BIC criterion. The study found that the Weibull distribution was the best fitted model with the highest log-likelihood, lowest AIC and BIC value as

compared to both Exponential and Log-logistic Models (Amra et al., 2017).

Daniel, Lasisi & Banister (2020) conducted a study on Tuberculosis patients' data from 2015 to 2017 using parametric models, generalized gamma frailty model with a mixture of gompertz distribution and cox proportional hazard to analyze the data. The aim was to determine the best model that fits the data. Data was obtained from a hospital in Bauchi State, Nigeria. To estimate survival probability of patients, several predictor variables such as pre-treatment weight, drug usage, Smoking, level of education, age, marital status and type of TB were used in the model. Smoking and administration of drug (Treatment Therapy) are significantly associated with survival while other covariates are insignificant in predicting survival of TB patients. The result of the fitted Weibull PH, Exponential PH, Gompertz PH, Cox P H, Lognormal AFT, Log-logistic AFT and generalized frailty gamma with a mixture of gompertz were compared using AIC to determine the best that fits the data. The frailty model of the gamma with a combined gompertz distribution was discovered to have fitted the model much better than the other parametric models based on its lowest AIC value as compared to others.

Survival analysis was used to assess determinants of length of stay by tourists in Turkey. Data was collected from primary sources through the administration of questionnaire to tourist during the summer vacation of 2005. The study found that 16 variables were significant in determining the causes for the length of stay. The parametric models of Weibull, Exponential and Gompertz were used in the study. The semi-parametric model of the cox proportional hazard was also used to analyze the data. Using the AIC values of the fitted parametric models, the Weibull distribution was found to be the best fit which has the minimum AIC value (Gokovali, Bahar & Kozak, 2007).

The Weibull, Log-logistic and Exponential Models was also used in a study on survival of smokers. Time to event for smokers who initial quitted smoking to

the time when they resumed the habit was studied by Elketroussi & Fan (1991). The study was for a period of 44 months and most smokers were found to have resumed smoking within 8 months after quitting. Using the maximum likelihood estimates to compare between the 3 models, the Weibull and Log-gistic were found to have performed better in fitting the model than the Exponential.

The lognormal model was compared with cox proportional hazard in modelling the survival of ovarian and breast cancer patients by Royston (2001). The result of the lognormal distribution was compared with that of the cox model. The researcher opined that using the lognormal, the median survival may improve by 25% as compared to result obtainable using the cox model on the same data. According to the author, cox model has weakness of validating a model using a new data. This is so because, the cox model uses a partial likelihood estimate which prevents it from specifying the complete probability of an event as the baseline hazard is not estimated in the model. This makes it inefficient in estimating survival as compared to parametric models. The study fitted other parametric models in to the breast cancer data such as Weibull, Gompertz, Exponential, Loglogistic and Gamma distributions. Using the AIC values, the gamma and lognormal were very close and differences between the two was found to be insignificant at 5% level of significance when compared using chisquare test statistic. This is expected as the lognormal is a special case of the gamma distribution. However, the logistic model was a better fit for the ovarian cancer data as it performs better in terms of AIC values. When the lognormal was compared to the cox model, the lognormal performed better in predicting survival.

2.5 Summary of Literature Review

This chapter reviewed empirical studies on tuberculosis, discussed its epidemiology and risk factors associated with survival among TB patients. The non-

parametric model of Kaplan Meier and Cox Proportional Hazard along with other parametric models of survival were used in several studies. The parametric models of Weibull, Exponential, Log-normal etc were usually compared using AIC, BIC or log-likelihood to determine which of the models best fits the data where the data is assumed to follow a particular distribution. Researchers have justified the use of the parametric models as superior tool to the Cox Model in survival analysis due to the semi-parametric attribute associated with the cox model. As such, only a partial likelihood is obtainable from the cox model as a non-parametric baseline hazard is always used in the Model.

For the purpose of this study, both parametric and semi-parametric models of survival will be applied as available in the literature.

Chapter 3

Methodology

3.1 Introduction

The methods and statistical programs used in the study are discussed in detail in this section. Description of data employed for the study, models to be used in analyzing the survival data, tests of goodness of fit etc will be discussed in this section.

3.2 Description of Data and Variables of the Study

Consecutive patients with positive Tuberculosis cases managed between January 2019 and December 2020 at the Directly Observed Therapy Short Course (DOTS) Unit of the hospital were enrolled for the study. The goal of this study was to determine the survival probability of TB patients who are receiving treatment at the DOTS Unit within Kano Metropolitan Area. Four Health care facilities were selected based on the recorded high number of TB in the facility during the period. The Infections Disease Hospital (IDH), Murtala Muhammed Specialist Hospital, Umma Zaria Healthcare Centre & Gwagwarwa Health Centre accounted for most TB cases in Kano from January 1, 2019 to December 31, 2020. Patients entered the cohort on the date of enrollment for TB treatment and remained in the cohort for a minimum period of six months unless they are either cured,

completed treatment, Loss to follow up, transferred out or died within this period when treatment is being administered on them. Patients were also followed up to July 31, 2021 especially for those who commenced the 6 months therapy in December, 2020.

The records of positively diagnosed TB patients who are either clinically confirmed (such as x-ray & biopsy results) or bacteriologically confirmed cases (sputum result from GenExpert Machine) were obtained from the Tuberculosis register located at the health facility.

Survival time for TB Patients constitutes the dependent variable of the study. The independent variables include pre-treatment weight of patients, Gender, HIV Status, Age and type of TB Disease (Pulmonary and Extrapulmonary) based on available patient's medical information taken during treatment.

3.3 Research Design

This is a retrospective cohort study of adult TB patients registered for treatment at various DOTS (Directly Observed Treatment-Short Course) Unit of a licensed TB treatment facility within metropolitan area of Kano State, Nigeria for the period of January 1, 2019 to December 31, 2020. Being a secondary source of data, records of patient can be objectively and accurately obtained for the period of therapy. The medical records within the TB register are easily obtained and information used retrospectively for data analysis.

3.4 Data Collection and Technique of Sampling

Secondary source of data was used to obtain clinical and demographic information of tuberculosis patients from the TB register at the DOTS unit of the facility. The Age and Sex of patients, Type of TB disease, HIV status and initial body weight were the information obtained from the register. This register also includes details of the period when patient commenced treatment, when he/she dies or loss to follow. Patient status such as treatment completion, cured, relapse and transferred out/in patient were also available in the register. A convenience sampling technique was used based on data availability and personal judgment of the researcher. Kano Metropolis, being a city has concentration of health facilities that provides treatment to patients within and outside the city. As a result of its population and urban migration, majority of TB cases reported in the state are predominant among people living within the city. Out of the available health facilities that provides TB therapy in Kano, 4 were identified to have the largest reported cases during the period under study and were selected for this study.

3.5 Data Analysis

In terms of means and proportions of the survival data, descriptive statistics will be employed to describe the data properties. Chi-square will be employed to test for a connection between the independent factors and dependent variable (death) in bivariate analysis. To assess average failure time, associated level of risk, and difference in patient average time, Kaplan-Meier estimates and log rank test would be utilized.

Log rank test is a popular test which is used for determining whether there

is no difference in survival between two or more independent groups. The log rank test, which is closely related to the chi-square test statistic, compares the observed and expected number of events at each time point during the follow-up period. The Gehan-Breslow-Wilcoxon test is another method for comparing survival functions among independent groups. However, it places higher weight on deaths that occur at early time period. The log rank test, on the other hand, gives all time points an equal weights and is the more powerful of the two tests assuming the proportionality of hazards assumption is not violated (Avijit and Gogtay, 2017). Other variants of the Generalised log rank test include the Fleming and Harrington & Peto-Peto Prentice tests for comparing survival among independent groups. For the purpose of this study, the log rank test statistic was used.

Cox and Parametric Proportional Hazard Models of multivariate analysis was used to assess the effect of predictor variables on death. Statistical significance is determined by considering nominal p-value of less than 5% (p¡ 0.05) with a 95% confidence level and less than 10% with a 90% confidence level.

3.6 Analytical Tools

Data was analyzed using both STATA and R Statistical Package software programs. Data collected from the hospital was initially entered into Microsoft Excel Spreadsheet and variables coded appropriately for reference during analysis. This Excel document is subsequently exported to both STATA and R statistical packages for further analysis.

3.7 Ethical Consideration

The Health research ethics committee of Kano State Ministry of Health approved the application to conduct the study using the TB data (NHREC Approval Number: NHREC/17/03//2018). Written inform consent of participants could not be obtained during the course of the study, the informed consent was waived by the committee. Hence, restricted records of Patient was listed in the anonymized list prior to the study.

3.8 Concept of Survival Analysis

Satagopan et al (2004) described survival of individuals at a particular time as "the conditional probability of surviving to a specific time given that the individual is at risk for the event (such as mortality) at that time" (Satagopan et al., 2004). Survival of any individual at any given time is estimated as number of individuals that have not experienced the event (e.g death) at that time divided by number of individuals that have not experienced the event at least up to that time (Satagopan et al, 2004).

The analysis of data measured from a given moment of inception until an event of interest or a predetermined endpoint is known as survival analysis (Collett, 1994). The study of time to event data is known as survival analysis (death, relapse from a treatment, cured etc). The event of interest in most clinical trials is evaluating the chance of experiencing an event by a certain time. The Kaplan-Meier approach can be used to produce a nonparametric estimate of the cumulative incidence or probability of encountering the event of interest when the data consists of patients who experience an event and censored persons (Satagopan et al., 2004). To estimate the survival probability of patients (in clinical studies), the status (dead, alive or censored) and length of stay (time to event)

are used.

Survival analysis is a set of approaches for studying data in which the outcome variable is the time until an event of interest occurs (Viv et al., 2004). In most clinical studies, the event of interest is time to death, time between response to treatment and relapse-free period (such as in cancer studies).Some individuals may not have experienced the event of interest during or after the study has ended, as such, their survival times will not be known. This is a major difficulty in survival analysis. This is known as censoring, and it can happen in a number of ways: patient has not experienced the outcome of interest, such as death or relapse, at the end of the study, or a patient is lost to follow up during the study period, or a patient has a different event that prevents further follow-up (Clark et al., 2003). Time in which the event of interest has occurred is the major concern in survival analysis (Ajagbe et al., 2014).

Survival function S(t) is defined as:

$$S(t) = P_r(T > t) = 1 - P_r(T \leq t) \tag{3.1}$$

characteristics of Survival S(t)

(a.) $S(t)$ =1, if t=0

(b.) $S(\infty) = lim_{n\to\infty} S(t) = 0$

(c.) S(t) is non-decreasing in t

In general, the survival function S(t) provides useful summary information such as the median survival time, t-year survival rate, etc.

Density Function for S(t):

a. if T is a discrete random variable,

$$f(t) = P_r(T = t) \tag{3.2}$$

b. If T is (absolutely) continuous, the density function is:

$$\begin{aligned} f(t) &= \lim_{\triangle t \to 0^+} \frac{Pr(\text{failure occuring in}[t; t + \triangle t])}{\triangle t} \\ &= \text{Rate of occurrence of failure at t} \\ &= \lim_{\triangle t \to 0^+} \frac{F(t + \triangle t) - F(t)}{\triangle t} \\ &= \frac{dF(t)}{dt} \\ &= \frac{dP(T \leq t)}{dt} \\ &= \frac{[1 - P(T > t)]}{dt} \\ &= \frac{d[1 - S(t)]}{dt} \\ &= -\frac{dS(t)}{dt} \end{aligned} \tag{3.3, 3.4}$$

3.9 Non-Parametric Survival Model: Kaplan–Meier (K-M)Survival Estimate

In Survival Analysis studies, the major difference between Kaplan-Meier (non-parametric analysis) and the usual parametric analysis (such as using Weibull, Gompertz and other probability distribution) is Censoring. To avoid bias, researchers use the K-M method which considers censored data in analysis. To account for censored cases during the study, data must be adjusted at each point where patient(s) are lost (Rich et al., 2010). When it cannot be confirmed that a participant during a study has died or the participant drops out during or at end of study, he/she is considered censored. When the participant has not en-

countered the event of interest during or at the end of the study, and more data cannot be gathered, right censoring is said to have occurred.

According to Jager et al. (2008) cited in Dudley (2016), "patients who are censored must meet the following critical assumptions: censored patients have the same chance of survival as those who continue in the study, and survival odds are the same whether participants enroll early or late in the study". Patients who have been censored are included in probability estimations of the event up until the evaluation point prior to their censoring, but they are omitted from subsequent analysis (Blagoev, Wilkerson, & Fojo, 2012).

In survival data, censorship is an inherent phenomenon (life time data). As a result, applying parametric models to survival data is a tough task. In investigations with censored data, the non-parametric estimators Kaplan- Meier or Nelson-Aalen are used (Saranya & Karthikeyan, 2015).

The success of any clinical or community based study depends on the number of participants or patients that are prevented from having an adverse event (e.g. death) or alive. Some participants may be loss to follow-up, drop out of the study etc. Hence, not all participants will remain in the study until its completion. In order to come up with unbiased analysis and valid conclusion concerning patients' probability of survival, Kaplan–Meier estimate (also called "product-limit method") serves as a simple, reliable measure (Kalra, 2017).

The Kaplan Meier method, according to Rich et al. (2010), is the most commonly utilized survival analysis method in randomized medical clinical trials. Patients are allocated to various arms at random; they do not enter the study at the same time; and they drop out or are lost to follow up from the study at different times after they begin. The outcome variable of interest may or may not arise within the study observation period (Rich et al., 2010).

Kaplan Meier estimate is a useful non parametric estimation where there is an

incomplete observation in a data used in anaylsis. Kaplan and Meier (1958) described an event subject to random sampling with incomplete observation of all members as "death". Kaplan–Meier test is nonparametric in nature typically used for estimating the survival distribution, that is, to compute the fraction of participants who survived for a certain specified period after the intervention or treatment. Even when individuals drop out or are investigated for varying durations of time, K-M allows the estimation of survival throughout time. The inclusion of censored observations makes the K-M a superior non-parametric estimate in estimating survival than other parametric measures. Generally, when there is loss of participant, the proportion of survival decreases (Kalra, 2017).

3.9.1 Assumptions for the Kaplan-Meier (K-M) method

The following assumptions on Kaplan Meier were provided by Koletsi & Pandis (2017):

(1.) The K-M method presupposes that the likelihood of censoring is unrelated to the event of interest.

(2.) Irrespective of when an individual enters into the study, they are all assumed to have same risk of experiencing the event. The survival probabilities for all study participants are the same and no circumstances are assumed to alter the baseline survival risk of participants.

(3.) Lastly, the events are assumed to have occurred within a specific period of time. Information of the exact time may not be known in some instances but the status of the participant during the last follow up time is known prior to the event.

3.9.2 The Kaplan-Meier's Product-Limit formula

The survival probability can be estimated non parametrically from observed survival periods for both censored and uncensored failure times, using the K-M (or product-limit) approach introduced by Kaplan and Meier in 1958. Saranya & Karthikeyan (2015) define the K-M formular as follows:

Let $t_1, t_2, t_3, \cdots$ denote the actual times of death of the n individuals in the cohort. Let $d_1, d_2, d_3, \cdots$ represent the number of deaths that occured at each of these times, and

Let $n_1, n_2, n_3, \cdots$be number of people who are at risk within the cohort." (Saranya & Karthikeyan, 2015)

The survivor function's Kaplan-Meier estimator at time t ,for $t_{(k)} \leq t < t_{(k+1)}$

$$S(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right)\cdots\left(1 - \frac{d_{(i-1)}}{n_{(i-1)}}\right) \tag{3.5}$$

$$= \prod_{i;y_i<t}\left(1 - \frac{d_i}{n_i}\right) \tag{3.6}$$

From equation 3.5, it can be deduced that survival at a time t (probability of surviving past time t), is a product of conditional probability for surviving past time $t_{(k)}$. To enhance the reliability and accuracy of the K-M survival estimates confidence interval are estimated as common with other models. Several school of thoughts have proposed methods of confidence interval estimation for the Kaplan Meier such as Transformational method, Peto's method etc but the commonly used method is the greenwood's formula. The purpose is to derive a 95% confidence for a given survival at a fixed time period. The greenwood's formular is used to compute the variance for constructing the confidence interval of the K-M survival estimates. The greenwood's formular for estimation of variance is given

as:

$$Var[S(t)] = S(t)^2 \left[\sum_{i;t_i<t} \frac{d_i}{n_i(n_i - d_i)} \right] \tag{3.7}$$

and the confidence interval of the Kaplan Meier is estimated as:

$$S(t) \pm Z_{\alpha/2} \sqrt{Var[S(t)]} \tag{3.8}$$

From Equation 3.8, $\sqrt{Var[S(t)]}$ is referred to as the standard error of the estimated survival at time t, while the 95% confidence interval for the $Z_{\alpha/2}$ quantile of the normal distribution is ± 1.96.

3.10 Hazard Function

Hazard indicates the rate at which an event of interest occurs. "Hazard is usually denoted by h(t) or $\lambda(t)$ and is the probability that an individual who is under observation at a time t has an event at that time. In other words, hazard represents the instantaneous event rate for an individual who has already survived to time t" (Clark et al., 2003). Hazard function is concerned with the event of interest occuring (e.g death) whereas the survival function is concerned with the event not occurring. Survival function, denoted as S(t), is mainly concerned with the probability that an individual survives to a specified future time from a given point.

Density function of the Hazard Function $\lambda(t)$

$$h(t) = \lim_{\triangle t \to 0} \frac{P(t + \triangle t > T \geq t | T \geq t)}{\triangle t} = \frac{f(t)}{S(t)} \tag{3.9}$$

In comparing the hazard between two independent groups (treatment and control), The likelihood (or hazard) of events occurring in the treatment group as a percentage of events occurring in the control group is referred to as a hazard ratio. A hazard ratio has no dimensions and merely tells you about the data's consistency and reliability (Blagoev et al., 2012).

Spruance et al (2004) cited in Dudley (2016) opined that a hazard ratio is only relevant if two basic assumptions of constant and proportionality are met i.e the difference between groups in a survival analysis are constant and the hazard ratio is proportional. A hazard ratio of more than or less than 1 indicates that one of the groups fared better than the other in terms of survival (Dudley, 2016). The cumulative distribution function F(t) and the probability distribution function f(t) can be used to define the above hazard function in equation 3.11.

Recall from eq (3.1)

$$S(t) = Pr(T > t) = 1 - Pr(T \leq t) = 1 - F(t) \tag{3.10}$$

Therefore,

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \tag{3.11}$$

$\lambda(t)$ Which is the instantaneous event rate at a given period t equals the probability of events at time t, divided by the probability of non-occurrence of event at the specified period.

It can be seen that there exists a relationship between the survival function S(t) and the hazard function h(t), this can be expressed in terms of the calculus

formula provided below:

$$h(t) = -\frac{d}{dt}\log S(t) \tag{3.12}$$

Hence, the negative derivative of the survival function when divided by the survival function, gives the hazard function by substituting equation (3.4) into equation (3.11).

3.10.1 Relationship between Cummulative Hazard and Survival Function

The hazard can be estimated from the cumulative hazard H(t) or $\Lambda(t)$. The integral of the hazard, or the area under the hazard function between times 0 and t when T is (absolutely) continuous defines the H(t) as shown below:

$$H(t) = \int_0^t \lambda(u)du \tag{3.13}$$

From equation (3.11).

$$\begin{aligned} H(t) &= \int_0^t \frac{f(u)}{S(u)}du \\ &= \int_0^t -\frac{\frac{d}{du}S(u)}{S(u)}du \\ &= -\int_0^t \frac{\frac{d}{du}S(u)}{S(u)}du \end{aligned}$$

evaluating the integral at the limit

$= -\log S(t) + \log S(0)$ $S(0) = 1$ and $\log(1) = 0$,we have the relation below

$$H(t) = -\log S(t) \tag{3.14}$$

$$S(t) = \exp\left[-H(t)\right] \tag{3.15}$$

$$S(t) = \exp\left[-\int_0^t \frac{f(u)}{S(u)} du\right] \tag{3.16}$$

From the above, it can be concluded the survival function is used in obtaining the hazard density function. This can further be used to derive cumulative hazard using Equation (3.11). On the other hand, the survival function can also be derived from the hazard. This is done by integrating the hazard function and taking its exponent to obtain the survival function in Equation (3.16).

3.11 Cox Proportional Hazard

In the study of survival data, Cox's proportional hazards regression model and the log-rank test statistic has become a standard among statistician (Armitage, 1987) cited in Kelvin (2003). The cox model can be used for both univariate and multivariate analysis of multiple predictor variables. It can also be used to compare survival among two independent groups as applicable in the log rank test statistic. Cox proportional hazard model is very important in survival analysis; the advantage of this model is that it includes the nonparametric and parametric element (Geachew and Bekele, 2016) cited in Rakesh (2019).

As cited in Bradburn et al. (2003), Cox (1979) proposed a model for estimating the hazard based on order statistics. Mathematically, the Cox model is written

as:

$$h(t) = h_0(t) \exp\left[\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right]$$
$$h(t) = \log\left[\frac{h(t)}{h_0(t)}\right] = \sum_i \beta_i x_i \tag{3.17}$$

From the above equation, the hazard function h(t) is dependent on (or determined by) a set of p predictor variables $(x_1, x_2, \cdots, x_p)$, whose impact is quantified by the size of the relevant coefficients (β_1,β_2,$\cdots$,β_p). The baseline hazard h_0 is the value of the hazard when all the x_i are equal to zero.

From equation (3.17), a unit increase in the explanatory variables (covariates) is associated with β_i increase in the log hazard rate. The log of the hazard rate in a particular group (treatment) can be compared to the control group when quantified using the regression coefficient while adjusting for the predictors included in the model.

The widespread use of the Cox Proportional Hazard model arises from long experience with it, as well as the fact that it is distribution-free, requiring no assumptions about the underlying distribution of survival times to make inferences about relative rates of the event. The assumption has been relaxed to make it more malleable and easier to utilize in real-life situations (Kevin, 2003).

3.12 The Log Rank Test

For comparing survival among independent groups, the most widely used test by researchers is the Mantel-Haenszel, often known as log-rank test, which was first developed by Mantel in 1966 and then by Cox in 1972. As a result, some researchers refer to the process as the Cox- Mantel test (Martinez, 2007) cited in Etikan et al. (2018). The test determines the difference between expected

and observed number of events in the two groups of participants (Etikan et al., 2018). This test statistic is useful where there is an assumption of a proportional hazard. In other words, the hazard ratio between independent groups should remain constant.

The null hypothesis is for a right censored data is given by:

$$H_0 : h_A(t) = h_B(t) \quad (\text{or} \quad S_A(t) = S_B(t)), \quad \text{for all t} \tag{3.18}$$

For a right censored data, log rank test statistic is derived by:

$$Z^2 = \left[\frac{\sum_{i=1}^{k} D_{(i)} - E_0[D_{(i)}]}{\sum_{i=1}^{k} var_0(D_{(i)})}\right]^2 \sim \mathcal{X}_1^2, \quad \text{where n is large} \tag{3.19}$$

Where;

$$E_0(D_i) = m_D\left(\frac{n_A}{N}\right) = n_A\left(\frac{m_D}{N}\right)$$

and ,

$$Var_0(D_i) = \frac{n_A n_B n_A m_D m_{\bar{D}}}{N^2(N-1)}$$

Where:

m_D=Total number of failures.

n_A=Number of individuals in the risk set at i from group A.

n_B=Number of individuals in the risk set at i from group B.

N=Number of individuals in the risk set.

$m_{\bar{D}} = N - m_D$=Total Number of survivors.

D_i=failures at i from group A.

The Log Rank Tests approximately follows a Chi-square with a 1 degree of freedom.

3.13 Model Diagnosis

Residuals have been used as a tool for diagnosing the fitness of model. Weisberg (2014) developed residuals that have been applied in theory of linear regression and it has proven to be a vital tool in analyzing model's goodness of fit. It is crucial to analyze the cox regression model's suitability in characterizing the TB data utilized in this investigation, taking into account the assumptions made in the cox proportional hazard's model. Several approaches for evaluating the cox model's appropriateness are available in the literature; four of these diagnostics will be used in this study i.e. Martingale Residuals (Linearity assumption and fitness of model), Deviance Residuals (model fitness), Schoenfeld Residuals (proportionality assumption of the hazard) and Cox Snell Residuals (fitness of the cox proportional hazard model).

3.13.1 Martingale Residuals

This is used to determine whether the variables and the log-hazard have a linear relationship. Breslow and Prentice (1998) proposed the use of martingale residual plots to determine the functional form of the continuous variables (or covariates). The residuals obtained from the martingales indicates the difference between observed and the expected number of events as assumed in the Cox Model (Klein & Moeschberger,2003). It is defined as:

$$M_i = \delta - \hat{H}_0(t) \exp\left[X_i^T \bar{\beta}\right] \tag{3.20}$$

Where $\hat{H}_0(t)$ is the estimate of the baseline cumulative hazard at t_i and δ_i is the event indicator for subject i.

Since non-linearity is not usually associated with a categorical variable, the martingale residual is used to test for non-linearity in continuous variable (Age and Weight of Patients in this case). Martingales have a Mean of 0 and ranges between $-\infty$ and 1. A value close to 1 indicates early death of a patient while large negative values implies that a patient lived beyond expected time of death (lived too long). The martingale residuals in some computer application can present the plot along with a curve fitted on it. This curve is known as loess smoother which is used to show the linear assumption of the cox model has been satisfied.

3.13.2 Deviance Residuals

This is also another model to check proportionality assumption of the cox model. It is usually used to detect outliers. It is a transformation of the martingale that has been normalized. It is usually symmetrical with zero distribution and standard deviation of 1. Non-negative values implies that there is early death in patients. Patients who did not experience the event (death) or survived beyond their expected survival time are assumed to have lived for a longer period of time. However, the model poorly predicts values for extremely large or small outliers. The deviance residual is given as:

$$D_i = sign(M_i)\sqrt{[-2(M_i + \partial_i log(\partial_i - M_i))]} \tag{3.21}$$

where M_i is the Martingale Residuals and ∂_i is the hazard

3.13.3 Schoenfeld Residuals

Proportionality of hazard assumption can be assessed both graphically and by the statistical computation. The scaled schoenfeld residuals provides the graphical diagnostic of the proportional hazard assumption. The Schoenfeld residuals was invented by David Schoenfeld (1982) using expectation. By summing over all indices in the risk set at a particular time when the event of interest has occurred. It is used to confirm if each independent variable satisfies the cox proportional hazard assumption.

The schoenfeld residual is given by:

$$\mathbb{E}(X_{jm}) = \sum_{i \in R_j} (X_j)[i][j] \times P(i \text{ dies} | R_j) \tag{3.22}$$

$$P(i \text{ dies} | R_j) = \frac{\exp\left(x_i^T \beta\right)}{\sum_{k \in R_j} \exp\left(x_i^T \beta\right)} \tag{3.23}$$

X_i= mth regression variable's value ith individual,

$\mathbb{E}(X_{jm})$ = Expected value of the mth regression variable in R_j

Equation (3.22) can be used to estimate expected value for the column of each covariate in the study. From equation (3.23), we can see that Schoenfeld residuals assumes a common baseline hazard for all individuals. It is the probability of the cox hazard which also assumes similar baseline hazard for all individuals in the study. Data used to derive schoenfeld residuals are uncorrelated with each other in a large survival data set where the cox proportional assumption has not been violated (Schoenfeld, 1982). Also, Grampbsch and Therneau (1993) proved that the mean of the scaled Schoenfeld Residuals is Zero if the coefficients of the cox

regression do not vary with time.

3.13.4 Cox Snell Residuals

The residuals developed by Cox and Snell (1968) can be used to evaluate the fit of a Cox proportional hazards model. Assume you have fixed covariates that are fitted to a data using a Cox Model given by equation (3.17), if the model is properly fitted, then the integral transformation on the true death time T, results to a random variable with a uniform distribution $U = H(T_i|X_i)$ has an exponential distribution with a hazard rate of 1 and a linear cumulative hazard rate. The Nelson-Aelen cummulative hazard estimation of the deviance residuals is used to confirm this (Klein & Moeschberger, 2003). Hence, If the Cox-PH model is correctly specified then, $r_1, r_2, \cdots, r_n$ should comprise a right censored sample from an exponential distribution. The residual can be estimated as:

$$r_i = (\hat{H}_0)(T_i) \exp\left[\sum_{k=1}^{p} X_{ik}\hat{\beta}_k{}^T\right]; \mathrm{i} = 1, \cdots, n \tag{3.24}$$

Where β_k= estimated values of β_s that are close to actual β_s if model is true $(\hat{H}_0)(T_i)$= baseline hazard rate estimated by breslow, X_{ik}= kth covariate for an ith subject

In some text, the deviance residuals can be obtained from difference between the censoring variable and the martingale residuals i.e

$$r_i = \partial_i - M_i \tag{3.25}$$

Where δ_i is censoring indicator (1 if failure occurs and 0 otherwise) and M_i is the martingale residuals. The Plot of the cox snell residuals against the cumulative hazard estimated using Neslson Aelen, should be a straight line passing through

the origin with a slope of 1. If model is correct, there should be no serious deviation in the values from the straight line of 45-degree from the origin. Model is well predicted when values are on the 45-degree line. There is over and under prediction when the values are above and below the straight line respectively.

3.14 Parametric Models for Estimating Survival

When data is known to follow a certain distribution, some studies have justified the use of parametric models over non parametric or semi-parametric models in estimating survival time. David, Stanley & Sussane (2008) highlighted the benefits of using a parametric model in survival analysis as follows: "full maximum likelihood can be used to estimate the parameters, estimated coefficients or transformations can provide clinically meaningful estimates of effect, fitted values from the model can provide estimates of survival time, and residuals can be computed as differences between observed and predicted values" (David, Stanley & Sussane, 2008).

Some parametric models used in estimating survival times include Exponential Distribution, Weibull Distribution, Log-logistic, Log-normal, Gamma and Gompertz Distribution. Some are Accelerated Faliure Times Models (Log-Logistic and Log-normal) while others are both AFT and Proportional Hazard Models (Weibull and Exponential). The Gompertz is strictly a Proportional hazard Model. For the purpose of this study, the Exponential, Weibull and Gompertz Distributions will be used to model the TB data. These are common models used in modelling survival data as available in the literature. A candidate model will be adopted based on comparism of the models using Akaike Information Criterion (AIC) to select the model with the lowest value or the highest log-likelihood value.

3.14.1 Exponential Distribution

Exponential distribution is one-parameter lifetime distribution and a special case of Weibull distribution (Saroj, 2019). It has a constant hazard which is independent of time ($\lambda = \theta$). It is the simplest parametric model in modelling survival data.

If the distribution of the time-to-event data is assumed to follow an exponential distribution, this can be expressed as:

$$T \sim exp(\theta); \theta > 0 \tag{3.26}$$

The density function of the exponential distribution with the parameter $\theta > 0$ is:

$$f(t) = \theta \exp(-\theta t); for\ t > 0;\ \theta > 0 \tag{3.27}$$

The survival function is:

$$S(t) = \int_t^\infty f(u;\theta)du = \int_t^\infty \theta \exp(-\theta u)du = \exp(-\theta t),\ for\ t \geq 0 \tag{3.28}$$

The hazard function is:

$$h(t) = \frac{f(t;\theta)}{S(t;\theta)} = \theta,\ a\ constant \tag{3.29}$$

The Proportional Hazard function is:

$$h(t|x) = \exp[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p] \tag{3.30}$$

The θ is reparametarized as $\sum_i \beta_i x_i$.

The Exponential distribution is a one parameter distribution with a Constant hazard. Having a constant hazard makes it unrealistic in modelling survival data. It is difficult to maintain same probability of experiencing an event between intervals (Michael & Bolarinwa, 2020).

3.14.2 Weibull Distribution

In the field of public health, the Weibull parametric model has been found to be quite useful in modelling lifetime data. It is flexible and can be easily applied in modelling time to event data. The Weibull distribution is characterized by two parameters (β, θ) which represents the shape and scale parameters respectively. The hazard function of the weibull distribution might increase, decrease or stay constant over time, depending on the value of the shape parameter (Carroll, 2003; Lee & Go, 1997). The Weibull has both Proportional Hazards and Accelerated Failure Times Properties which "allows the simultaneous description of treatment effects both in terms of hazard ratios and also in terms of the relative increase or decrease in survival time" (Carroll, 2003). It is considered unique among other parametric distributions due to its proportional hazard and accelerated failure time property (Collette, 1994; Prentice & Kalbfleish, 1979). This improves the ability to examine proportionality of hazards and effective prediction of maturity of data over time. The estimates of the hazard ratios are similar for both the weibull and cox model. A simulation study by Carroll (2003) reveals that both hazard ratios and standard errors are similar between treatment for weibull and cox model. This is possible where the proportionality assumption of hazard holds. However, the weibull model performs better than the cox model where the assumption of hazard proportionality is violated.

Given the scale and shape parameters $\theta > 0$ and $\kappa > 0$ respectively, the density

function of the weibull distribution is defined as:

$$f(t,\theta,\kappa) = \kappa\theta t^{\kappa-1}\exp\left[(-\theta t)^{\kappa}\right],\ for\ t \geq 0 \tag{3.31}$$

It also assumed the parameterized survival function:

$$S(t) = \exp\left[(-\theta t)^{\kappa}\right] for\ t > 0 \tag{3.32}$$

The hazard function is:

$$h(t,\theta,\kappa) = \frac{f(t,\theta,\kappa)}{S(t,\theta,\kappa)} = \frac{\kappa\theta t^{\kappa-1}\exp\left[(-\theta t)^{\kappa}\right]}{\exp\left[(-\theta t)^{\kappa}\right]} for\ \theta,\kappa > 0$$

$$h(t,\theta,\kappa) = \kappa\theta(\theta t)^{\kappa-1} \tag{3.33}$$

Note that the hazard function $\lambda(t)$ of the weibull distribution reduces to the exponential distribution when the shape parameter ($\beta = 1$) is constant. Also, hazard is increasing in t if $\beta > 1$, and decreasing in t if $\beta < 1$.

The Proportional Hazard model is:

$$h(t|x) = \kappa t^{\kappa-1}\exp\left[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right] \tag{3.34}$$

θ of the hazard is reparametarized as $\sum_i \beta_i x_i$.

3.14.3 Gompertz Distribution

This is characterized by a scale parameter λ and a shape parameter γ. The density function is given as:

$$f(t) = \lambda \exp(\gamma t) \exp\{-\lambda\gamma^{-1} \exp[(\gamma t) - 1]\} \; for \; \lambda > 0, \gamma \geq 0 \tag{3.35}$$

Survival function is given as:

$$S(t) = \exp\{-\lambda\gamma^{-1} \exp[(\gamma t) - 1]\} \tag{3.36}$$

Hazard function is:

$$h(t) = \lambda \exp[\gamma t] \tag{3.37}$$

The Proportional Hazard Model is:

$$h(t|x) = \exp^{\gamma t} \exp[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p X_p] \tag{3.38}$$

λ of the hazard is reparametarized as $\sum_i \beta_i x_i$. The hazard rate increases when $\gamma > 0$ and decreases when $\gamma < 0$. Hazard remains constant at $\gamma = 0$ and becomes an exponential distribution.

The exponential, gompertz and weibull distributions have been used to compare models in several medical studies such as cancer, stroke, Tuberculosis, Child Mortality etc (Jackett & Rosenberg, 1993; Sprat et al., 1992; Michael & Bolarinwa, 2020; Suroj, 2019).

3.15 Parameter Estimation for the Proportional Hazard models

The Maximum Likelihood Estimates will be used to estimate the various parameters for both the semi-parametric and parametric models adopted for the study. However, partial likelihood estimates was used to estimate the parameter (β_i) for the Cox Proportional Hazard Model has a partial likelihood estimate. The Cox Proportional hazard's likelihood function only estimates survival probabilities for subjects who experienced the event of interest. The Parametric Model considers both censored and uncensored observations in estimate of the parameters using the full maximum likelihood function.

3.16 Model Selection Criterion

Several literature in field of statistics have presented results from data analyzed using maximum likelihood model as selection criteria among various models (Kadane & Lazar, 2004). The Akaike Information Criterion (AIC) is popular model in the field of statistics and provide a reliable means of selecting a model among list of models. AIC is a model that is being derived from maximum likelihood estimate (MLE). The likelihood model alone as a criterion for selecting a model is not advisable as studies have shown that it is known to select models that are overly parameterized (Gelfland & Dey, 1994). The AIC provides a means of measuring the goodness of fit of the estimates within a model (Akaike, 1974).

The AIC is given as:

$$AIC = 2k - 2ln(\hat{L}) \tag{3.39}$$

Where k denotes the number of estimated parameters from the model, $\hat{L}$ is the model's likelihood function. A model that best fits the data has a lower AIC value. The AIC tries to avoid the use of irrelevant variables in the data that can cause increase in the AIC value.

3.17 Summary of Statistical Modelling Technique and Analytical Tools

As mentioned earlier in this chapter, the survival analysis is a technique which measures time to event (occurrence or non-occurrence). In clinical studies, patients are followed until an event occurs such as relapse from a treatment in cancer patients, time until a person resumes smoking, time until death of a patient undergoing treatment etc. The survival probability of patients or participants within a study can be estimated from the time a patient enters the study until when he/she experienced the event of interest. Several methods such as Kaplan-Meier (1958), Cox Model (1972), Log rank etc are used. Also, the Proportional Hazard Models of the Weibull, Gompertz and Exponential Distributions will also be used in the multivariate analysis of the tuberculosis data. These statistical methods are employed to estimate survival, compare survival among independent variables and measure the effects of such variables on the survival time. Diagnostic checks will be employed in assessing the fitness of the model. The tuberculosis data of patients is initially extracted from aTB register and entered into a Microsoft Excel document for further analysis. The data is then imported into the R Software and STATA 14.0 statistical packages for analysis.

Chapter 4

Result & Analysis

4.1 Introduction

This chapter provides the analysis of tuberculosis data collected and presentation of results. Figures, tables, plots etc. were used to describe the data. It contains result obtained from various statistical models provided in chapter three of this study. The Kaplan-Meier survival model, Log rank test to compare survival among groups were both used in descriptive analysis. Multivariate analysis using Cox proportional hazard model and Parametric Models were carried out. R Software and STATA 14.0 Statistical Package were used for the analysis.

4.2 Classification of Variables

The variables of the study include some set of covariates (independent variables) and their effect on the dependent variable (death). Time until death among drug susceptible TB patients who received treatment at the various health facilities within Kano Metropolis is the event of interest in this study. Gender, type of TB, HIV status, Age and baseline weight of patients at treatment initiation were among the factors investigated in relation to mortality or censoring. The variables are being coded below for ease of analysis:

Table 4.1: Description of Variables Used in the Data

Variables	Description	Categories
Age	Age at the time of enrollment for treatment	Young (< 40 years) Old (≥ 40 years)
Initial Body Weight	Weight of Patient at start of study	Low (< 35kg) Heavy ($\geq$35kg)
Gender	Sex of Patient	Male = 0 Female = 1
Type of TB	Type of clinical or bacteriological manifestation of TB	Pulmonary = 1 Extrapulmonary = 0
HIV Status	THIV co-infection of Patient at the beginning of the study	Positive = 1 Negative = 0
Failure Time	Patient experiences the event of interest or Loss to Follow Up	Dead = 1 Censored = 0

4.3 Descriptive Analysis

The descriptive statistics for the survival data are as shown in Table 4.2 & Table 4.3. Both demographic and clinical information about patients were described in Table 4.2. A univariate analysis of the variables was carried out. It provided the total number of event of 75 which had occurred from a total of 2,555 patients enrolled in the study. The events and censored observations from the total number were further described per each category of covariate. The chi-square test (χ^2) and the corresponding P-value to test the null hypothesis of equal survival among the categorized variables were also presented. The categories were derived based on previous studies with medical justification (Field et al, 2014 & Floe et al, 2017).

In terms of gender, there were 973 female and 1,582 male adults among 2,555 patients whose data were collected during the study. There were 47 (2.97%) male

Table 4.2: Demographic and Clinical data of TB patients

Summary of Observations			Survival Status		Test Statistics	
Risk Factors	**Categories**	**Total**	**Death**	**Censored**	**p-value**	$\mathcal{X}^2$
Gender	Male	1,582 (100%)	47 (2.97%)	1,535 (97.03%)	0.83954	0.0410
	Female	973 (100%)	28 (2.88%)	945 (97.12%)		
Age	≥40 years	769 (100%)	41 (5.33%)	1,745 (94.67%)	0.0001	14.7638
	< 40 years	1,786 (100%)	34 (1.90%)	735 (98.58%)		
Type of TB	Pulmonary	2,427 (100%)	72 (2.97%)	2,355 (97.3%)	0.5986	0.2771
	Extrapulmonary	128 (100%)	3 (3.13%)	125 (96.87%)		
Weight	< 35kg	122 (100%)	9 (7.38%)	113 (92.62%)	0.8967	2.8802
	≥ 35kg	2,433 (100%)	66 (2.71%)	2,367 (97.29%)		
HIV	Positive	101 (100%)	10 (9.9%)	91 (90.1%)	0.0014	10.2137
	Negative	2,454 (100%)	65 (2.65%)	2,389 (97.35%)		

and 28 (2.88%) female among those who died during the study. Furthermore, 1,535 male (97.3%) were censored out of the total of 1,582 male patients. 945 female patients (97.12%) did not experience the event of interest and were censored accordingly. When comparing the survival among the two categories of gender using the log rank test statistic, it revealed that difference in survival of male and female patients was not statistically significant [($p = 0.8394$, $\chi^2 = 0.0410$] (Table 4.4)

From Table 4.2, patients were stratified according to two age groups. In the first category, we included 1,786 patients with age ¿ 40 years out of which 41 (5.33%) died and 1,745 (94.67%) were censored. In the second category, 769 patients were included for patients within the age of < 40 years and 34 (1.90%) had an event and 735 (98.10%) were censored. There was a significant difference between the survivals of the age groups [$p = 0.0001$, $\chi^2 = 14.7638$].

From Table 4.2, most of the patients from the study have Pulmonary TB. Out of the patients enrolled during the study period, 2,427 patients have pulmonary TB and 72 experienced the event of interest (2.97%). 128 patients had extrapulmonary TB but only 3 had an event (3.13%). 113 patients with below 35kg weight were censored (92.62%) and 2,367 (97.29%) were also censored. The difference between the two groups are not significantly different in terms of survival

[p = 0.5986, χ^2 = 0.2771].

Patients' weight were discretized into two categories. Those who weighed less than 35kg were placed in the first category, while those who weighed 35kg or more were placed in the second category. This was used in prior TB studies conducted by Santh et al (2000). There were 2,433 patients who weighed 35kg or more, and 66 (2.71%) of them had the incident (death). Only 122 patients weighed less than 35kg out of which 9 (7.38%) died. The difference in weight categories was not statistically significant [p = 0.08967), χ^2 = 2.8802] (See Table 4.2.)

Patients were also classified according to their HIV status confirmed prior to commencement of treatment. A total of 101 patients were confirmed to have tested positive for HIV while 2,454 patients are either negative or their status are not confirmed. 10 patients who are TB-HIV co-infected died (9.9%) while 65 patients with negative HIV status died (2.25%). The difference between the two groups was statistically significant in terms of survival [p = 0.0014), χ^2 = 10.2137] (Table 4.2).

Table 4.3: Summary of Survival Data

Category	Total	Mean	Minimum	Median	Maximum
No. of subjects	2,555				
No. of records	2,555	1	1	1	1
(first) entry time		0	0	0	0
(final) exit time		180.0658	1	178	313
Time at risk	4600068	180.0658	1	178	313
Failures	75	.0293542	0	0	1
Survival	2,480	.9706458			

The summary statistics of the incident and patients that were censored are presented in Table 4.3. Also, Figure 4.1 depicted the survival data using K-M Plots. The K-M Plot depicts survival probability at each time. It is a step function, the curves remains flat horizontally between event period and drops when an event of interest occurs. The estimated survival probability by the K-M curve is the same

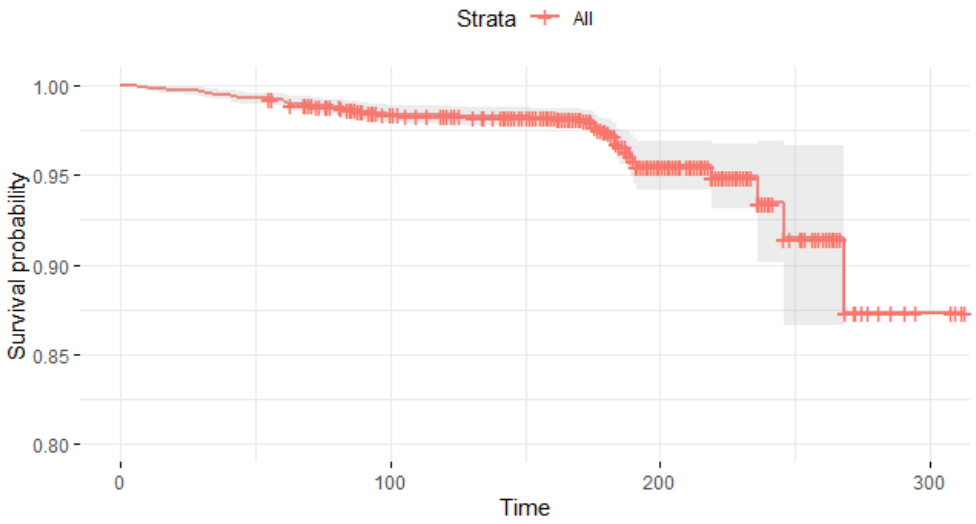


Figure 4.1: **Kaplan Meier Curve of Survival for TB Patients**

for patients who were censored between intervals from when an event occurs. The study included a total of 2,555 patients, with 75 of them experiencing the event of interest (death). 97.06% was the survival rate of the TB patients and the average time to death is 6 months (180.0658 days).

4.4 Comparing Survival of Categorical Variables using K-M Curves

By plotting the Kaplan Meier survival curves, we examined the survival function of the participants according to gender, type of disease and HIV status. Survival plots provide a pictorial view of the proportion of patients who did not experience the event declining over a given period. The shape of the distribution of the survival is displayed by the plot. It is usually displayed as a step function. The cumulative survival is plotted against the given period of event occurrence. The vertical axis represents the cumulative survival probability while the horizontal axis is the period. The K-M graph is used to display the survival and slopes downwards from left to right. It is a straight line where survival is high and

no event occurred but drops down when an event occurs to show a reduction in survival probability. When two independent groups are compared using the K-M plot, the group that has the higher survival probability has its curve is usually above that of the other group.

The KM curve shows the cumulative survival function by gender on a linear scale (Fig 4.2). From the graph displayed, we found that the survival curve of women are higher (though not significant) than their male counterpart. In comparing survival between gender category, Pardeshi (2009) & Dehghan et al (2019) also found an insignificant difference in their survival probability. In contrast, to several studies, the survival probability of women (not experiencing the event of interest) was significantly higher than that of men (Floe et al., 2017; Pathmanathan et.al, 2017 ; Adamu et al, 2017; Field et al, 2014;Yi et al, 2020). Patients with extrapulmonary tuberculosis had a greater survival rate than those with pulmonary TB, as shown in Figure 4.4. Patients weighing less than 35kg have a worse survival rate than those weighing 35kg and above as shown in Figure 4.4. The graph also reveals that those aged 40 and above had the lowest survival curves, while those aged 40 and below had the highest (see Figure 4.3).

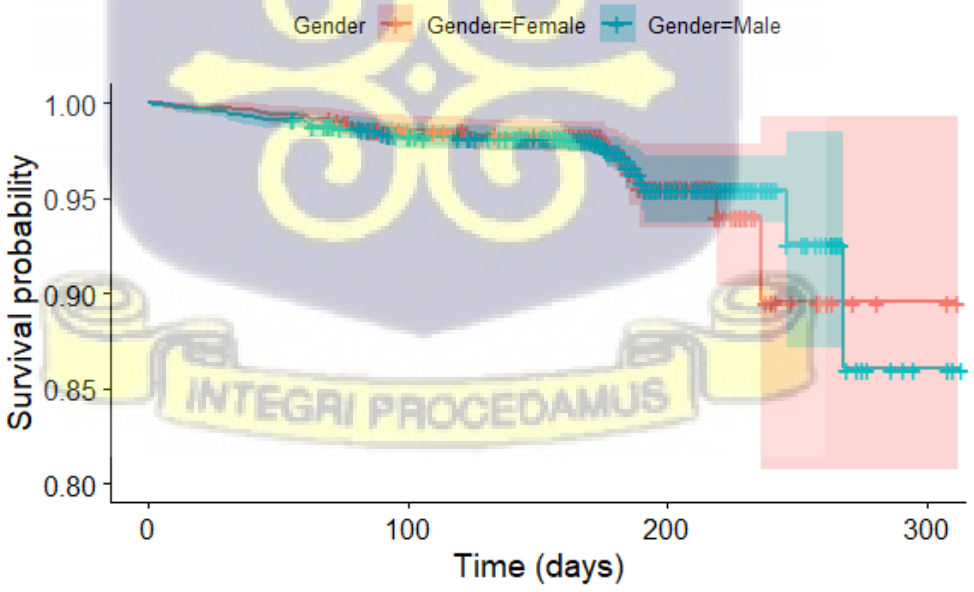


Figure 4.2: **Kaplan Meier Curve of Survival for Gender Category of Patients**

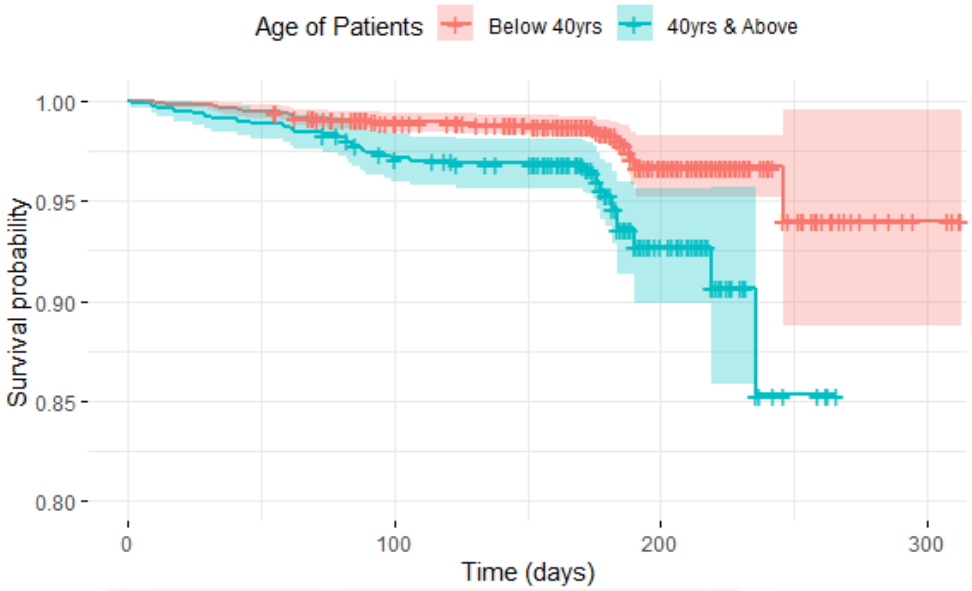


Figure 4.3: **Kaplan Meier Plot for Age Category of Patients**

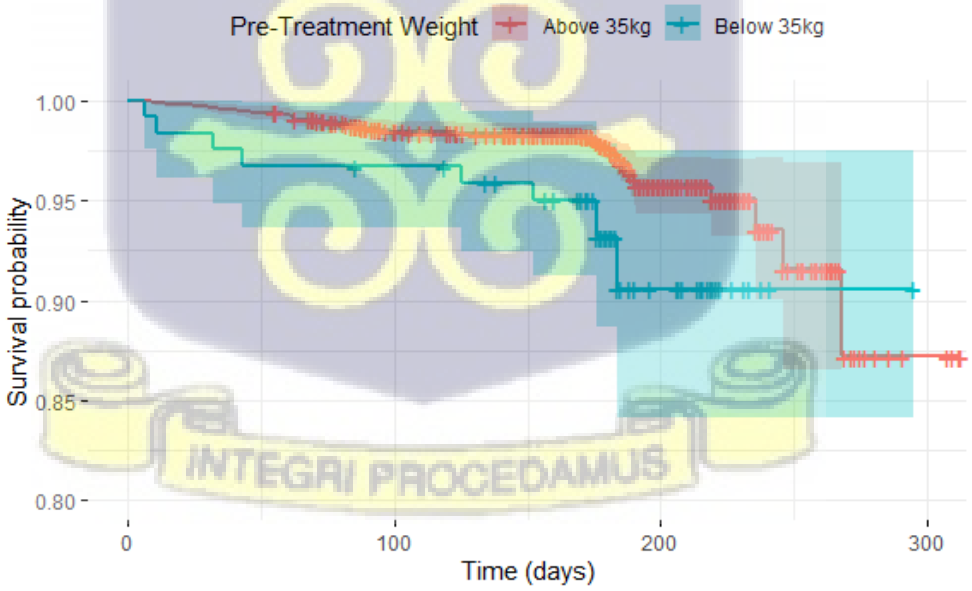


Figure 4.4: **Kaplan Meier Curve for Pre-treatment Weight of Patients**

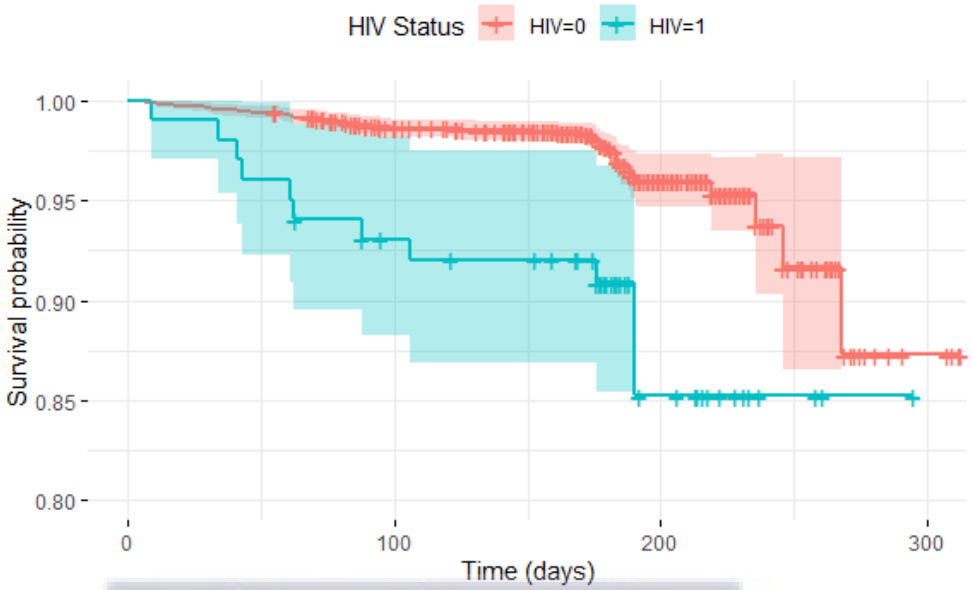


Figure 4.5: **Kaplan Meier Curve for HIV Status of Patients**

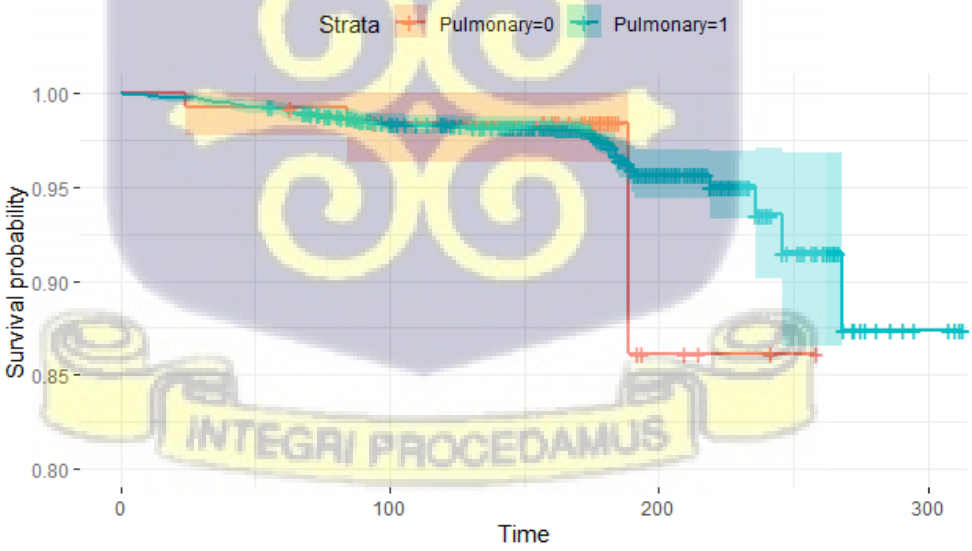


Figure 4.6: **Kaplan Meier Curve for Type of TB**

4.5 Log Rank Tests to Compare Survival among groups

The most popular test statistic in comparing survival among independent groups is the log rank test. The null hypothesis of equal survival among the groups is tested to either reject or fail to reject the result based on the p-value at a 5% level of significance. This test assigns equal weights across all periods. The other forms of the generalized log rank tests such as Gehan and Tarone-Ware Tests assigned different weights based on the period where most of the events have occurred.

These tests are used to determine whether the groups being compared have a substantial difference in survival. Regardless of the test utilized, the null hypothesis of equal survival between the groups is being tested to reject or accept at a specific point.

Table 4.4: Test statistic of the Log rank for Tuberculosis patients

Risk Factors	Categories	DF	p-value	$\mathcal{X}^2$	Decision
Gender	Male Female	1	0.9432	0.01	Fail to Reject Null
Age	40 years & above below 40 years	1	0.0000	22.60	Reject Null
Type of TB	Pulmonary Extrapulmonary	1	0.8228	0.05	Fail to Reject Null
Weight	Below 35kg 35kg & above	1	0.0054	7.74	Reject Null
HIV	Positive Negative	1	0.0001	16.40	Reject Null

Table 4.4 shows that there is statistical difference on survival among each category of Age, Weight of Patients and HIV Status as the p-values are less than 5% using any of the log rank test statistic. Hence, we reject the null hypothesis of equal survival among the category of covariates. However, there is no statistically significant difference in survival between the category of Type of TB and Patients' Gender ($p>0.05$).

4.6 Cox Proportional Hazard Model

The Cox PHM is useful in evaluating the effect of covariates (independent variables) on the survival time of patients. The hazard model aids in determining the difference in survival times of the various groups to be evaluated using the study's independent variables. The instantaneous risk of experiencing an event of interest (death) at a particular time is referred to as "hazard". It is the probability that an individual will experience an event given that he has survived up to a given time. The dependent variable in the Cox Model is the hazard. The result of the relationship between the covariates and the hazard are given in the table below:

Table 4.5: Cox Proportional Hazard Model for Univariate Analysis of Variables

Covariates	Coef	Exp(coef)	Standard Error	z	P>—z
Gender	-0.0170	0.9831	0.2388	-0.071	0.943
Age	0.027432	1.027811	0.006691	4.100	4.13×10^{-5}*
Type of TB	0.1322	1.1414	0.5901	0.224	0.823
Weight	-0.01961	0.98058	0.01216	-1.613	0.107
HIV	1.287	3.621	0.340	3.785	0.000154 *
**significant at 5% alpha level, *significant at 10% alpha level					

Gender is not statistically significant (p =0.943) and negatively associated with the event. Female Patients have less risk of experiencing the event of death. Patients aged 40 years and above were at higher risk of death [HR = 1.0278, p= 0.0000) than compared to Patients with age below 40 years. The result is statistically significant at p<0.05. Weight is statistically not significant at 5% level of significance [HR = 0.9806, P=0.107] and negatively associated with survival. HIV positive patients are 3.6 times at risk of experiencing the event than HIV negative status patients. The result is statistically significant at 5% level of significance [HR = 3.621, p = 0.0002]. Type of disease is also found not to be significantly associated with death [HR = 1.414, p = 0.823] (Table 4.5)

Table 4.6: Multivariate Cox Proportional Hazard Model

Covariates	Coef	Exp(coef)	Standard Error	Z-value	P>—z
Age	0.028357	1.028763	0.006676	4.247	2.16×10^{-5}*
Type of TB	0.228208	1.256346	0.593241	0.385	0.700475
Gender	-0.044714	0.95627	0.246667	-0.181	0.856152
Weight	-0.020616	0.979595	0.0124	-1.663	0.096391*
HIV	1.192783	3.296241	0.342645	3.481	0.000499**

****significant at 5% alpha level, *significant at 10% alpha level**

The β (Coef), $h(t)$ [exp(coef)], se(coef), z and P value is displayed for all continuous and categorical variables. Gender is not statistically significant with pvalue (0.86) and negatively associated with the event i.e death. Being female was found to be a risk factor for mortality among TB patients. Female Patients have less risk of experiencing the event of death. From hazard ratio [i.e exp(coef)], Gender (Female) is 0.95627 which can be interpreted as 1-0.95627 = 0.04373 which means the risk of Female experiencing the event is 4.4% less likely to occur compared to being a male. Patients aged 40 years and above were at higher risk of death [HR = 1.0278, p= 0.0000) than compared to Patients with age below 40 years.The immune system tends to decline in patients as they grow older and treatment may be less effective in older patients compared to the youthful age. These findings are consistent with retrospective cohort studies in India & Nigeria (Pardeshi, 2014 & Adamu et al., 2017). Patients with weight below 35kg are negatively associated with survival at 10% level of significance and are at higher risk of experiencing the event (death) as compared to patients that weighed above 35kg [HR = 0.9796, p=0.0964]. HIV positive patients are 3.3 times at risk of experiencing the event than HIV negative patients. The result is statistically significant at 5% level of significance [HR = 3.2962, p = 0.0005]. the Type of TB was found not significant in predicting survival at 5% level of significant [HR = 1.2564, p= 0.70]. However, the risk of death was higher among pulmonary TB patients by 26% than in extrapulmonary patients (Table 4.6).

Weight was almost statistically significant at 10% level (p=0.107) in the univariate analysis. Their exclusion had little impact on the results of the model as Age

Table 4.7: Multivariate Cox PHM for the significant variables

Covariates	Coef	Exp(coef)	Standard Error	z-value	P>—z
Age	0.027836	1.028227	0.006847	4.065	4.8×10^{-5}*
HIV	1.270388	3.562234	0.340099	3.735	0.000187 *

**** significant at 5% alpha level, * significant at 10% alpha level**

and HIV both show an insignificant change in their p values. Their level of significance remains the same with or without the Gender, Weight and type of TB variables. Variables used in field of medical research are important in model development even if found not to be statistically significant. For example, weight of a patient prior to treatment is a determinant factor on the dosage of drugs to be administered during tuberculosis therapy. Hence, all the variables will be maintained in the model as shown above (Table 4.7).

4.6.1 Proportionality of Hazard Assumption in the Cox Model

The assumption of constant hazard over time is a requirement in the Cox model. To test the assumption of the proportionality assumption, we used the test of Schoenfeld residual. For this assumption to hold, there should be independence between the scaled residuals and time. The statistical significance of each covariates is derived and a global test for the whole model is also computed from the schoenfeld residuals. A covariate is assumed to have satisfied the proportionality assumption if its P value is above 5% significant level.

Table 4.8: Test of Proportionality Assumption of the Cox Model

Covariates	Rho	Chisq	p-value
Age	0.0651	0.3192	0.572
Type of TB	-0.0443	0.1516	0.697
Gender	0.0468	0.1615	0.688
Weight	0.0147	0.0184	0.892
HIV	-0.1416	1.4806	0.224
Global	NA	2.2731	0.81

From table 4.8, the p values obtained from the model are not statistically significant (p>0.05), we can assume that proportionality of hazard has not been violated.

4.7 The Schoenfeld Residual Plots for the Categorical Variables

The Schoenfeld residual plots gives a graphical view of the residuals to assess the proportionality as obtained in Table 4.8.

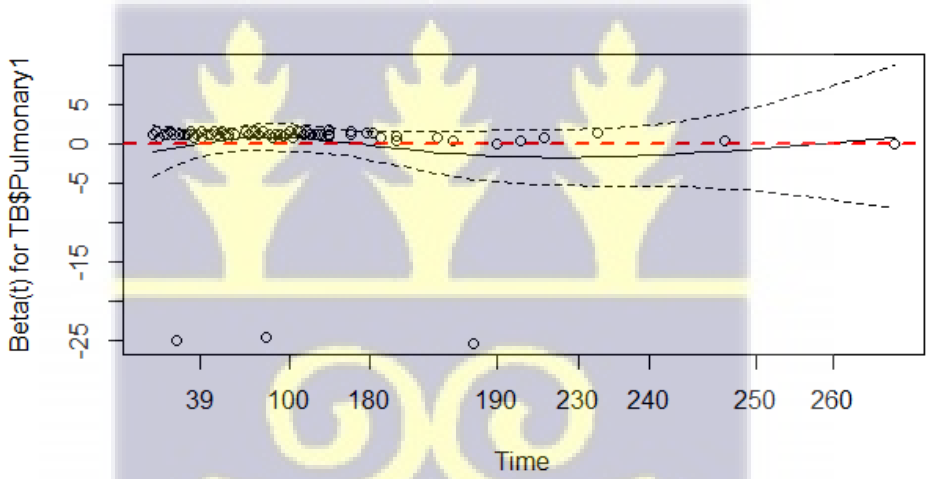


Figure 4.7: **Schoenfeld Residual Plots for Type of TB Variable**

The scaled schoenfeld residuals is plotted against time that has been transformed for each of the categorical covariate in the model fitted into the tuberculosis data. The broken lines represents a ±2 standard error bands around the fit while the thick line is a smoothing spline fitted onto the plot. The red dotted line indicates no change and falls within standard error bands almost the entire time to show that the variables are not changing over time confirming the proportionality of hazard assumption.

From the plots above in Fig 4.7, Fig 4.8 & Fig 4.9, it can be observed that the

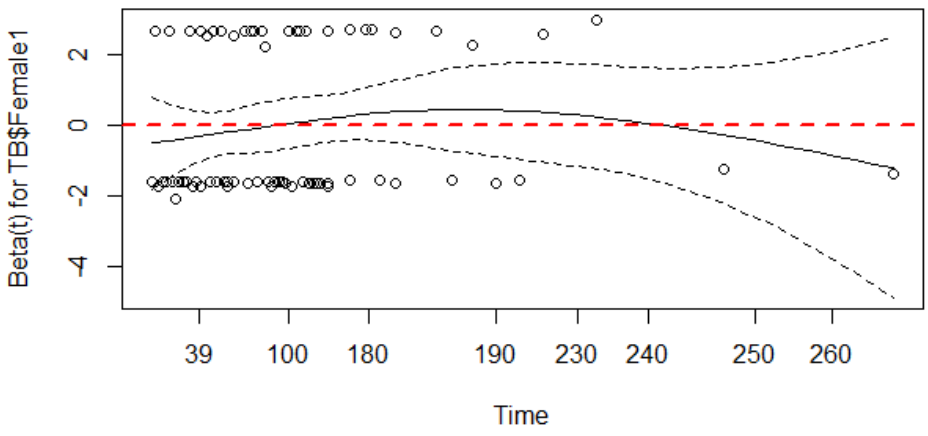


Figure 4.8: **Schoenfeld Residual Plots for Gender Variable**

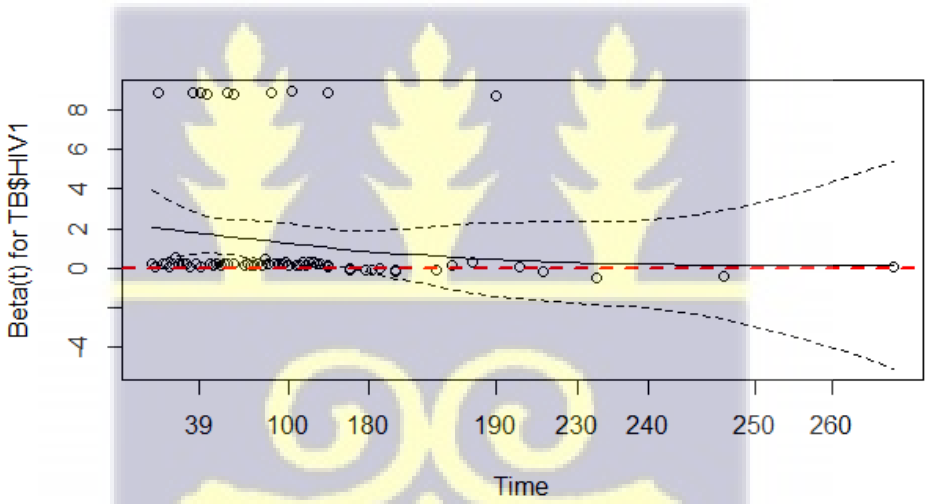


Figure 4.9: **Schoenfeld Residual Plots for HIV Variable**

assumptions of proportional hazard has not been violated by any of the categorical variables as they do not vary much with time.

4.8 Assessment of linearity for Continuous Variables in the Model

To assess linearity in the model, martingale residuals will be plotted against the covariates. Since nonlinearity is not an issue in the categorical variables, we will assess the linearity in the Weight and Age variables that are continuous. The figures 4.10 & 4.11 shows that there is evidence of linearity in the weight and Age data respectively. Martingale residuals was used to assess the form of the continuous covariates and how best they fit the model. The red line divides the data into the group of positive and negative residuals at point zero of the residuals. The blue lines are fit by the LOWESS (Locally Weighted Scatterplot Smoothing) that produces a smooth curve that fits points within the scatterplot of the martingale. The residuals above the line are positive values (very close to 1)and shows the number of individuals who died much earlier than expected. Most of the patients survived in the study and can be seen from the higher number of patients whose residuals fall below zero. Lines in both plots are horizontal and linear.

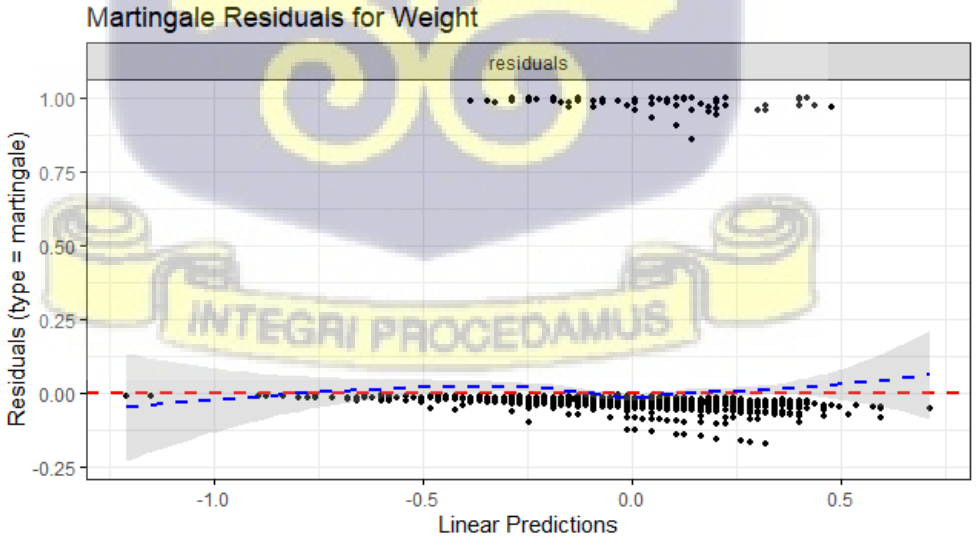


Figure 4.10: **Martingale Residuals Plots for Weight Linearity**

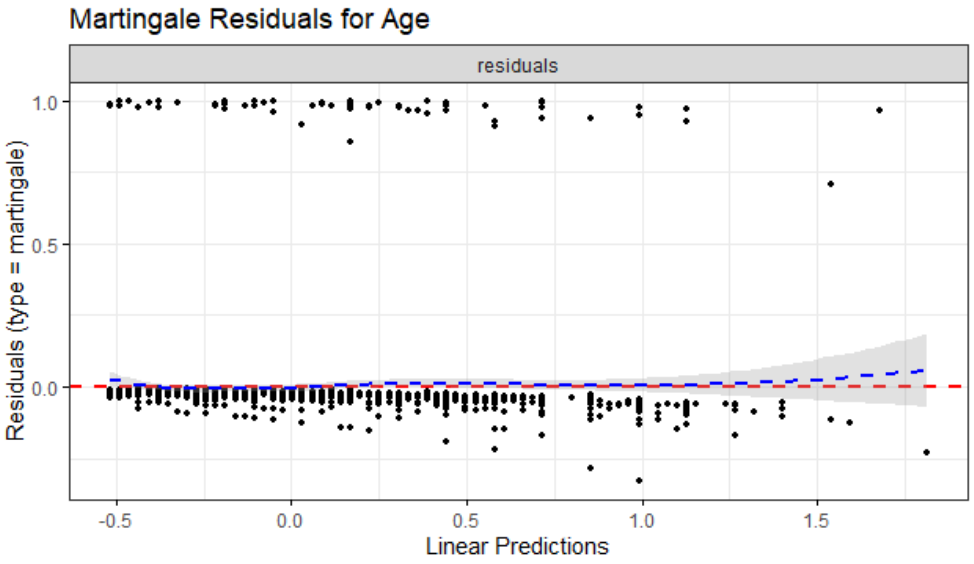


Figure 4.11: **Martingale Residuals Plots for Age Linearity**

4.9 Model Diagnostics using Residual Plots

To assess the goodness of fit for the cox model, the Martingale, Deviance and Cox Snell residuals was fitted into the data.

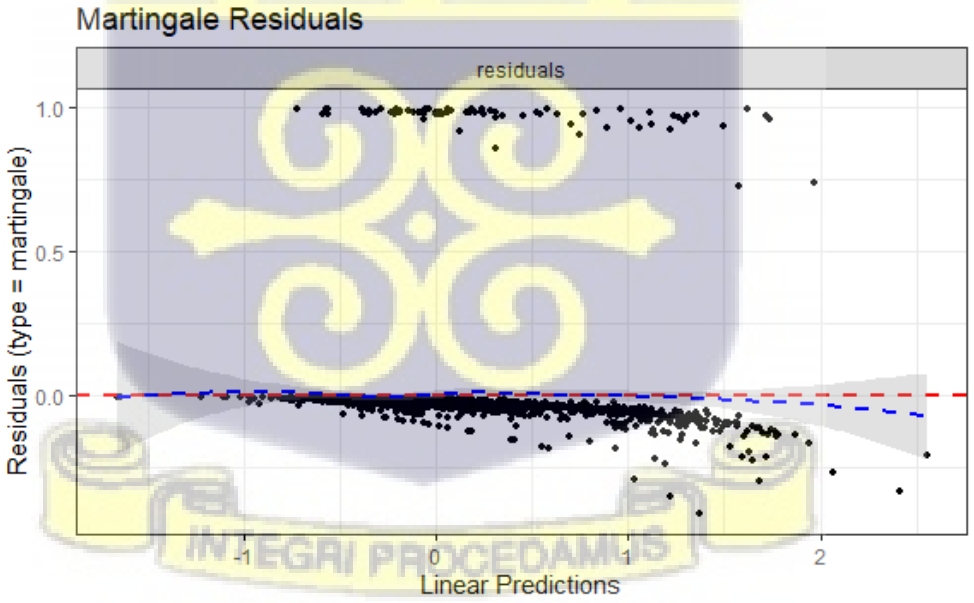


Figure 4.12: **Martingale Residuals Plots**

The martingale and deviance residual plots are used to assess the model fitness. The deviance residuals are plotted against the observations. We can see that

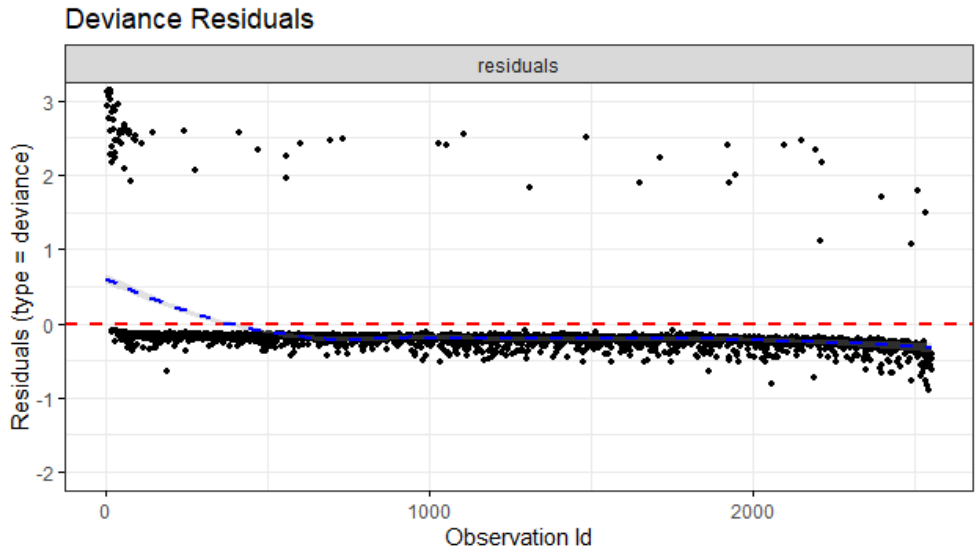


Figure 4.13: **Deviance Residuals Plot**

there are few outliers. About 6 observations were found to be outlier from the data (2547, 2539, 2061, 2538, 2493 and 2192). A further review of this individual observations shows that they are all censored observation with no event. However, these outliers represent 0.23% of the entire data and their mean survival is not too far from the average length of stay for the patients during the study. Hence, they were not excluded from the study.

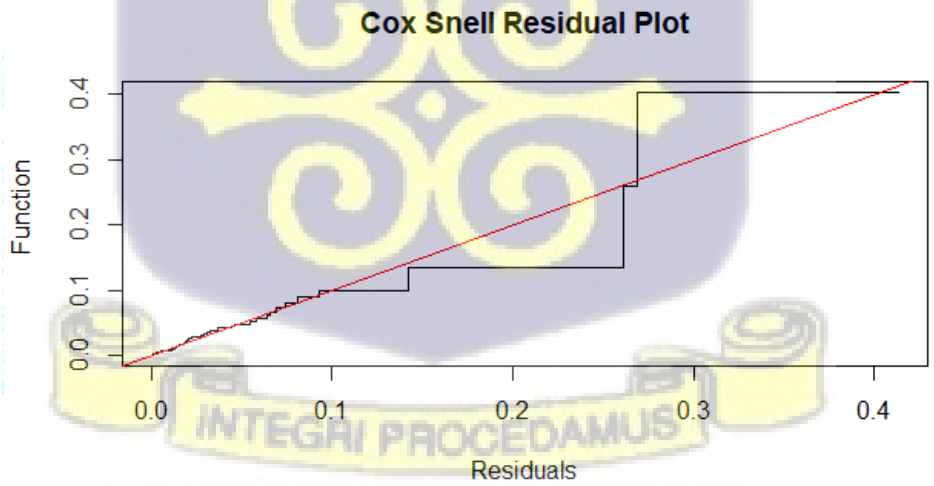


Figure 4.14: **Plot of Estimated Cumulative Hazard against Cox Snell Residuals**

From the Figure 4.14, the Cumulative Hazard function is plotted against the

residuals of the Cox Snell. A straight line is imposed on the curves with a unit slope and intercept of zero to indicate the satisfaction of the fitted survival model. The plot does not show much deviation from the 45 degree line. There is no much diversion from the line. This could suggest that the data is being fitted satisfactorily by the model.

The cox snell has basic limitation of not being able to give details on why model's assumption has been violated. It has also been criticized for its inability to assign positive or negative values that indicates the length of patient's survival as predicted by the model (Klein & Moeschberger, 2003).

4.10 Comparing the Hazard Ratios of Various Models

One of the objective of this research is to compare parametric proportional hazard models. The data will be fitted into the model and the best model that fits the data is to be selected based on the AIC and log-likelihood values obtained from each model.

Table 4.9: Cox and Parametric Models of the Proportional Hazard

Covariates	Groups	Cox	Exponential	Weibull	Gompertz
Age	Years	1.028763	1.028763	1.028797	1.028848
Type of TB	Pulmonary TB	1.256346	1.373415	1.365745	1.343054
Gender	Female	0.95627	0.9310046	0.9323225	0.9368501
Weight	Weight (kg)	0.979595	0.9752982	0.9755896	0.976474
HIV Status	HIV Positive	3.296241	3.43245	3.439982	3.430271

**** significant at 5% alpha level, * significant at 10% alpha level**

The Hazard Ratios predicted for the covariates from the various parametric models and the Cox proportional hazard model are shown in Table 4.9. The results show consistency in the hazard ratio for each of the covariates across all the models. For example, the hazard ratio for the categorical variable gender (female)

maintained a consistent result in reporting a lower risk of death compared to male gender when any of the models is being used. Likewise in Age & Weight, where the results reveal almost same values across the four proportional hazard models.

4.11 Analysis of the Models Using AIC and Log-Likelihood Values

The AIC and log-likelihood values obtained from the proportional hazard models fitted into the data in Table 4.9 will be analyzed to select the best model.

Table 4.10: Akaike Information Criterion & Log-likelihood of Parametric Models

Criteria	Exponential	Weibull	Gompertz
Log-likelihood	-386.19114	-385.13314	-382.4341
AIC	784.3823	784.2663	778.8682

To determine the best model that fits the data, the AIC and log-likelihood criterion was presented in Table 4.10. The model with the lowest AIC and highest log-likelihood is selected as the candidate model. From the table above, the values for the distributions are being extracted from the STATA 14.0 output. We can see that the Gompertz distribution is the model that best fits the data as it has the minimum Akaike's Information Criterion (AIC) and log-likelihood value when compared to the Exponentional, Weibull and Gompertz Proportional Hazard Models. This contrasts findings from previous studies that finds the Weibull as the best distribution in fitting tuberculosis data (Michael & Bolarinwa, 2020) and other similar studies such as cancer, stroke etc. (Jackett & Rosenberg, 1993; Sprat et al., 1992; Suroj, 2019). The Cox Proportional hazard being a semi-parametric model, has a partial likelihood. This explains why it can not be directly compared with other parametric models which obviously exhibit a better fit as their distributional forms are fully specified. However, the result obtained

from fitting the candidate parametric model will have a close resemblance to that of the Cox Model as shown in Table 4.8. The Log-likelihood and AIC values of the Cox Model were -538.594 and 1087.88 respectively. These values indicate that any of the parametric model will perform better in modelling the data. However, the Cox Model has relatively flexible assumptions suitable upon violation of the parametric model assumptions.

Table 4.11: Multivariate Gompertz Proportional Hazard Model

Covariates	Haz. Ratio	Std. Err.	z-value	P>z	95% Conf. Lower	Interval Upper
Female	0.9368501	0.2311745	-0.26	0.792	0.5776037	1.519533
HIV	3.430271	1.174284	3.60	0.000**	1.753624	6.709964
Age	1.028848	0.0068526	4.27	0.000**	1.015504	1.042367
Weight	0 .976474	0.0121598	-1.91	0.056*	0.9529296	1.0006
Pulmonary	1.343054	0.7954222	0.50	0.618	0.4206982	4.287623
constant	0.0000735	0 .0000704	-9.93	0.000**	0 .0000112	0.0004804
gamma	0 .0056413	0.0020562	2.74	0.006**	0 .0016112	0.0096713

**** significant at 5% alpha level, * significant at 10% alpha level**

Table 4.11 presents result of the proportional hazard using selected model of the gompertz distribution. As with the cox proportional hazard model presented earlier, the output includes the hazard ratio and the p- value at 5% significance level and confidence intervals. The coefficients of the covariates are estimated from the logarithm of the hazard function. The coefficients of the regression results quantifies the logarithm of the hazard in the reference group compared to others. It shows the size of increase or decrease in covariates that affects the logarithm of the hazard when other covariates remain constant. From the table above, the HIV Status and Age of TB patients are significant at 5% level while the Weight of Patient is significant at 10% level of significance in the model. The variables Gender and Type of TB are not significant at either 5% or 10% level of significance. This is further shown in the 95% Confidence interval which includes the null hypothesis of equal hazard i.e 1 between the gender categories [CI: 0.5776, 1.5195] and between the type of TB categories [CI: 0.4207, 4.2876].

In the case of the Weight variable, an increase in a patient's weight is related with an increase in patient's survival [HR = 0.9765, p= 0.056] with a less than 1 hazard ratio. Studies have shown that there is improvement in the health of TB patient who are on a regular diet. Malnutrition usually associated with weight loss inhibits the effectiveness of the drug administered to patients.

Chapter 5

Summary, Conclusion & Recommendation

5.1 Introduction

This chapter represents the latter part of the research. It summarizes the entire work done, concludes and make necessary recommendation. The conclusion will highlight the results obtained in the previous chapter and discuss findings with reference to what has been obtained in previous studies. Necessary recommendations will be provided to improve survival, treatment administration of TB and suggests areas that need to be further investigated in regards to this work.

5.2 Summary & Conclusion

Identification of a suitable model for the tuberculosis is an objective in this study. The data was fitted using 3 parametric models (Weibull, Exponential and Gompertz). Using the AIC and Log-likelihood criterion, the gompertz distribution was found to have performed better in modelling tuberculosis data of Kano state. furthermore, model diagnostics were carried out to ensure goodness-of-fit, non-violation of vital assumptions etc that improves the reliability of result obtained from the model. Most researches on survival analysis in Nigeria were presented without dealing with the assumptions of proportionality of the cox model used in the studies. It is important to note that non-assessment of the proportionality of the hazard model can affect the power of the log rank test as commonly used to compare survival among independent groups which may lead to an incorrect

result being reported (Abraira et al, 2013). Statistical results presented with a violated assumption can impact on decisions due to the increased probability of error (Junyong & Dong, 2019).

Of the total studied patients, 2.94% of them died over the entire follow up period. The mortality from this study was lower than what were obtained from previous studies as cited in Adamu et al. (2017) where 16.6% mortality within 5 year study periods were found in Nigeria . In African settings, mortality of 3.4%, 10% and 22% were found in Tanzania, Ethiopia & Zimbabwe respectively (Getahun et al, 2011 & Takarinda et al, 2017; Nagu et al, 2017). A total of 47.1% deaths out of 85 patients studied within two years were reported to have died according to a study in Ghana by Omari-Sasu et al. (2016). Yi et al (2020) and Pardeshi (2009) both recorded a 4.6% mortality rate among drug susceptible patients in China and India respectively. Lin et al (2014) revealed a TB related mortality rate of 2.1% among patients receiving treatment at a hospital in Taiwan. Shuldiner et al (2014) recorded 9.9% mortality in TB patients studied in Israel during 10 year period.

Estimation of survival and comparism of survival among categories of covariates is an objective of this study. To achieve this, the Kaplan-Meier curves and log-rank test were used to determine the difference between categories of covariates and their impact on survival. The study was able to estimate survival function and compare the survival curves by groups. From the Kaplan Meier curve of estimated survival probability, we noticed a higher survival among groups during the periods of the study. We also found that the mortalities are higher among men as compared to women. The study found TB to be regular among men as they account for 62% of the total population, whereas women represent 38% of the population. Men's general lifestyle may be the reason for their exposure to risk of infection such as smoking, alcohol consumption, work environment etc. In 2019, TB was reported to be among top 10 causes of death with about 10

million people being infected. Out of this infected cases, men accounted for 5.6 million, women accounted for 3.2 million while children accounted for 1.1 million. Also, TB related deaths among men were found to be higher than women (WHO, 2020).

The study also seeks to identify potential risk factors linked to survival time. We utilized the Cox proportional hazards in estimating survival in presence of covariates. It is similar to multiple regression models that it allows us to examine the difference in survival time of various groups of patients while accounting for other risk factors or covariates that can impact the survival probability of patients. The hazard is the response variable in the model. The Parametric Models of the Proportional hazards were compared with the Cox Model to determine the best model that fits the data. Using AIC and Log-likelihood value of the models, the gompertz distribution was found to be the best model in predicting hazards of the covariates for TB data.

When stratified by age, people above the age of 40 years were found to have lower survival as most mortalities were witnessed in adult > 40 years from our study. Shuldiner et al. (2014) found that death among Tuberculosis patients was prevalent in adults above 65 years in Israel. Floe et al (2017) found that risk of mortality is increased among adults within the ages of 30-39. Advanced age of patients have been identified as significant risk factors for deaths among patients in previous studies (Lee et al, 2017; Adamu et al, 2017). The result showed that Mortality was highest in participants aged 40 years or more. Age has been found to be a significant risk factor in TB related deaths and 95% of TB mortalities are found in Less and Middle Income Countries (WHO, 2019). High mortalities within these countries are found in communities that are predominantly poor people and in places where income inequality is prevalent. Many communities have little or no access to sufficient health facilities, engage in cheap labour and suffer from abject poverty. They also lack proper nutrition (Bhargava & Bhargava,

2020).

This study found that weight is a significant factor in increasing TB related mortalities and adults with less than 35kg initial weight contributed to the deaths in TB patients. The effectiveness of the anti-TB drug relies on the nutrition and quality of food intake by the patient. Body weight was significantly related to increased mortalities among TB patients. Patients with Body weight less than 35 kg body weight was found to be a major risk factor for TB mortality in South Africa & Ethiopia (Field et al., 2014 & Getahun et al., 2011).

HIV positive TB patients have increased risk of deaths than those who are not co-infected with HIV. TB/HIV co-infection increases the risk of death among TB Patients. HIV infection weakens the immune system of TB patients and slows than the recovery rate. Although the HIV therapy helps to improve recovery of patient, delay in diagnosis of HIV can impact on the TB treatment response from patients. HIV patients can easily contract TB due to the weakened immune system caused by the infection. Waitt & Squire (2011) reviewed the risk factors associated with mortality in TB patients (Pre & Post TB treatment). Their study revealed that the risk of mortality in TB patients is increased by 3-8 times when patient is co-infected with HIV (Waitt & Squire, 2011). In a related study, it was found that patients with HIV infection accounted for higher deaths as compared to those who have not been infected with HIV (Churchyard et al., 2000). This is supported by Ajagbe, Kabir and O' Connor (2014) who found HIV infection to be a relevant risk factor that affects survival of TB patients. Omari-Sasu et al.(2016) also found TB/HIV coinfection to the most significant risk factor that predicts survival of TB patients. Aung et al. (2018) & Nagu et al. (2017) both found HIV to be related to survival of TB/HIV co-infected patients. Patients who are on ART for HIV have higher survival than those who are not. This survival is found to increase in patients who have been on treatment for a period of 10 years when compared to those who have received therapy for 5 years period

(Aung et al, 2018). Michael & Bolarinwa (2020), and Adamu et a.l (2017) also found HIV to be significant risk factor in TB related deaths in Nigeria.

Type of TB (Pulmonary or Extrapulmonary) was found not to be significantly associated with deaths in TB patients. This implies that the risk of death among TB patients is similar irrespective of the type of TB. Majority of the patients in the study are pulmonary TB positive and most deaths are found among these patients. The extrapulmonary TB patients are less and recorded fewer deaths. This is consistent with previous studies by Adamu et al. (2017) where it was revealed that there was no statistically significant difference in survival in pulmonary, extrapulmonary or both in TB patient at 5% level of significance. In the same vein, Shuldiner et al. (2014) also found Type of TB not to be a significant risk factor that affects survival of TB patients in Israel. In contrast, Omari-Sasu et al.(2016) found TB type to be a significant risk factor that affects survival among drug susceptible TB patients in Ghana. Also, Michael & Bolarinwa (2020) found type of TB to be a significant risk factor that affects survival among TB patients in Nigeria.

Gender was found not to be a significant risk factor related to survival of TB patients. The survival rate among female and male gender are not significantly different. Male accounted for large number of TB patients in the study but the death rate was almost similar to that of the female patients. This is also confirmed from the log rank test which shows no significant different in survival among male and female patients. Hence, being male or female does not improve survival probability or increase hazard rate significantly. This conforms to findings by Ajagbe, Kabir and O' Connor (2014) and Pardeshi (2019) where survival between male and female patients were found not to be significantly different. Yi et al. (2020), Michael & Bolarinwa (2020), Omari-Sasu et al. (2016) found survival to be highest among female gender and it is a significant factor in predicting survival among TB patients. These findings were obtained from studies

conducted in China, Ghana and Nigeria respectively. Shuldiner et al. (2014) found male gender to be a significant risk factor for TB related deaths In Israel.

5.3 Recommendation

Diagnostic checks in assessing the proportionality hazards of the Cox assumption and model's goodness of fit is important in clinical studies. This has an impact on the decision making in the health sector to avoid administration of incorrect treatment to patients. It is therefore recommended in research related to survival analysis.

There should be increased in nutrition by TB infected patients, this will help improve the quality of treatment and improve on their overall survival. Nutrition level can be monitored by regular weight check of patients during routine adminsitration of drugs and records maintained for future reference. Although the National Tuberculosis & Leprosy Control Programme has performed better towards bringing an end to TB in Kano and Nigeria at large, there is still need to sustain the momentum and improve in treatment administration. An important aspect of TB care is assessment of patients' nutrition as recommended by WHO. This will help in moderating severe malnutrition in TB patients (WHO, 2019). Income and BMI have also found to be important risk factors that impact on mortality among TB Patients (Floe et al, 2017). BMI measurement is not a part of the routine diagnostic workup in the TB programs of Kano and should be made a part of routine evaluation in order to monitor the nutrition of patients especially during follow ups of 2nd, 4th & 6th month. Information on the height of patients need to be included. Weight of patient when available in the register, can help provide necessary advice to patient in improving nutrition. Information on Income level of patients can also help in future studies to evaluate its impact

on TB mortality and prevent mortality among low income patients.

Also, the gompertz distribution was found to be the best model in fitting the TB data after a comparism of the various parametric models (Exponential, Weibull & Gompertz) were carried out. The results of the parametric tests were compared with the widely used Cox Model after which the gompertz model was found to be the best in fitting the model. This also is in contrast with the regular weibull distribution largely applied in modelling survival of TB patients. It is also recommended that a comparism of model is made before selecting the best that suits a particular data using the AIC, log-likelihood or any relevant criterion.

HIV status was found to be significantly associated with survival of patients. Efforts should be mad to ensure that HIV diagnosis is improved to ensure early detective and treatment. HIV patients have increased risk of contracting TB. Adequate education on the danger of delayed detection of the disease and prioritizing the issuance of drugs to such individuals can improve their survival.

Older patients should be given priority in treatment of TB. the study found deaths to be associated with increased age of patients. adequate care and support during therapy can help improve survival of older patients.

5.4 Limitations

This study has a number of limitations, firstly, some patient's record were incomplete as all required information needed for the study were not available. In some cases, the treatment outcome especially after the second follow up tests were not recorded, so it becomes practically impossible to confirm their status at end of treatment period. As such, the data related to the patients with incomplete records were dropped from the study. Secondly, some vitals were either improperly recorded or were not provided for in the patient medical record. These vitals

such as height would have been used as variable in the study. Lastly, the mortality recorded may have been caused by other factors other than TB. As such, cause of death among TB patients may not have been accurately determined. Further studies on Multi-Drug Resistance TB and Length of Hospital stay of admitted patients using Accelerated Failure Time Parametric models can be conducted.

References

Aakshi Kalra (2017).The Basics of Kaplan–Meier Estimate. Journal of the Practice of Cardiovascular Sciences. Published by Wolters Kluwer – Medknow.

Abraira V, Muriel A, Emparanza J.I., Pijoan J.I, Royuela A, Plana M.N, et al (2013). Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirement proposal. Journal of Clinical Epidemiology, 2013; 66:1340-6

Adamu, A., Gadanya, M.A, Abubakar, I.S, Jibo, A.M, Bello, M.M. et al (2017). High mortality among tuberculosis patients on treatment in Nigeria: a retrospective cohort study. BMC Infectious Diseases (2017) 17:170. DOI10.1186/s12879-017-2249-4.

Adepoju Paul (2020). Nigeria's widening tuberculosis gap. https://www.thelancet.com/pdfs/jo 3099(19)30712-1.pdf

Agarwal G.G. (2012). Statistics for surgeons – understanding survival analysis. İndian Journal of Surgical Oncology. 2012;3(3):208–214.

Ajagbe OB, Kabair Z, O'Connor T (2014). Survival Analysis of Adult Tuberculosis Disease. PLoS ONE 9(11): e112838. doi:10.1371/journal.pone.0112838

Akessa, G.M, Tadesse, M Abebe G.(2015). Survival Analysis of Loss to Follow-Up Treatment among Tuberculosis Patients at Jimma University Specialized Hospital, Jimma, Southwest Ethiopia. Hindawi Publishing Corporation, International Journal of Statistical Mechanics. Volume 2015, Article ID 923025. http://dx.doi.org/10.1155/2015/923025

Alobu I, Oshi SN, Oshi DC, Ukwaja KN (2014). Risk factors of treatment default and death among tuberculosis patients in a resource-limited setting. Asian Pac J Trop Med. 2014;7(12):977–84.

Amran, S.E, Abdullah, M.A, Long, K.S., and Jamil, S.M (2017). Analysis of survival in breast cancer patients by using different parametric models. IOP Conf. Series: Journal of Physics: Conf. Series 890 (2017) 012169. doi :10.1088/1742-6596/890/1/012169.

Armiatge P, Berry G and Mathews JN. (2002). Clinical Trials: Statistical Methods in Medical Research; 4th Edition: Balckwell Science; Page 591.

Asgedom S.W, Daniel T., Yirga L. N.and Tesfay M. A (2018). Time to death and risk factors among tuberculosis patients in Northern Ethiopia. BMC Res Notes (2018) 11:696. https://doi.org/10.1186/s13104-018-3806-7

Aung Z.Z, Yu M.S, Thu N. S., New O, et al (2019). Survival rate and mortality risk factors among TB–HIV co-infected patients at an HIV-specialist hospital in Myanmar: A 12-year retrospective follow-up study. International Journal of Infectious Diseases 80 (2019) 10–15.

Avijit Hazra and Nithya Gogtay (2017). Biostatics Series Module 9: Survival Analysis. Indian Journal of Dermatology, 2017 May-Jun; 62(3): 251-257. Doi: 10.4103/ijd.IJD_201_17.

Babajide A.A., Olokoyo, F. O., and Adegboye, F. B. (2013). Predicting Bank Failure in Nigeria Using Survival Analysis Approach. International Business Information Management Conference (22nd IBIMA), Italy, Volume 22nd.

Badje A, Moh R, Gabillard D, et al.(2017). Effect of Isoniazid Preventive Therapy on risk of Death in West African, HIV-infected Adults with High CD4 cell counts: Long-term follow-up of the Temprano ANRS 12136 trial. Lancet Glob Health 2017;5:e1080–9. PMID:29025631 https://doi.org/10.1016/ S2214-109X(17)30372-

8

Bajehson M, Musa BM, Gidado M, Nsa B, Sani U, Habibu AT, et al. (2019). Determinants of mortality among patients with drug-resistant tuberculosis in northern Nigeria. PLoS ONE 14(11): e0225165. https://doi.org/10.1371/journal.pone.0225

Bhargava A. & Bhargava M.(2020). Tuberculosis deaths are predictable and preventable: Comprehensive assessment and clinical care is the key. Journal of Clinical Tuberculosis and Other Mycobacterium Disease 19 (2020) 100155.

Blagoev, K. B., Wilkerson, J., & Fojo, T. (2012). Hazard ratios in cancer clinical trials—A primer. Nature Reviews Clini¬cal Oncology, 9(3), 178–183. http://dx.doi.org/10.1038/ nrclinonc.2011.217 Christopher J.L Murray (2018). The global burden of tuberculosis: results from the Global Burden of Disease Study 2015. http://dx.doi.org/10.1016/S1473-3099(17)30703-X

Churchyard GJ, Kleinschmidt I, Corbett EL, Murray J, Smit J, De Cock KM (2000). Factors associated with an increased case-fatality rate in HIV-infected and non-infected South African gold miners with pulmonary tuberculosis. Int J Tuberc Lung Dis 2000;4(8):705–12.

Clark TG, Bradburn MJ, Love SB and Altman DG(2003). Basic concepts of survival analysis, British Journal of Cancer (2003), 89(2), 232 – 238

Collett D.(1994). Modeling survival data in medical research. London: Chapman and Hall, 1994.

Cox, D.R (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological) 34, no. 2 (1972): 187 – 220. . www.jstor.org/stable/2985181. Accessed November 20, 2020.

Dauda MM. (2010). Evaluation of the efficacy of directly observed treatment short course (DOTS) in patients with tuberculosis and HIV Co-infection in

Kano, Nigeria. Rev Infect. 2010;1:218–23.

Dehghan, M.H, Mohammad S., Shahri H., and Salari M. (2019). An Analysis of the Survival Probability of Tuberculosis Patients with Right Censored and Interval Censored Observation in Zahedan during 2014 – 2016. Zahedan J Res Med Sci. 2019 July; 21(3):e63225. doi: 10.5812/zjrms.63225.

Dooley KE, Chaisson RE. Tuberculosis and diabetes mellitus: convergence of two epidemics. Lancet Infect Dis. 2009;9(12):737- 46. https://doi.org/10.1016/S1473-3099(09)70282

Dudley, William N., Wickham, Rita, Coombs, Nicholas. (2016). An Introduction to Survival Statistics: Kaplan-Meier Analysis. Journal of Advanced Practitioner in Oncology, 7(1).

Elketroussi M. and Fan D.P (1991). Time trends of smoking cessation analyzed with six mathematical survival models. International Journal of Biomedical Comput. 27:231-44.

Etikan İ, Bukirova K, Yuval M. (2018). Choosing statistical tests for survival analysis. Biom Biostat Int J. 2018;7(5):477–481. DOI: 10.15406/bbij.2018.07.00249

Fatiregun AA, Ojo AS, Bamgboye AE (2009): Treatment outcomes among pulmonary tuberculosis patients at treatment centers in Ibadan, Nigeria. Ann Afr Med. 2009;8:100–4.

Field N., Lim M.S.C, Murray J., Robert J. D., Glynn J.R, and Sonnenberg P.(2014).Timing, rates, and causes of death in a large South African tuberculosis Programme. BMC Infectious Diseases (2014) 14:3858. DOI 10.1186/s12879-014-0679-9

Fløe A, Hilberg O, Wejse C, et al. (2017). Comorbidities, mortality and causes of death among patients with tuberculosis in Denmark 1998–2010: a nationwide,

register-based case– control study. Thorax 2017; 0:1–8. doi:10.1136/thoraxjnl-2016-209240.

Geeta Pardeshi (2009). Survival Analysis and Risk Factors For Death In Tuberculosis Patients On Directly Observed Treatment-Short Course" Indian J Med Sci, Vol. 63, No. 5, May 2009. DOI: 10.4103/0019-5359.53163

Getahun B, Ameni G, Biadgilign S, Medhin G. (2011). Mortality and associated risk factors in a cohort of tuberculosis patients treated under Dots programme in Addis Ababa, Ethiopia. BMC Infect Dis. 2011;11: 127.

Ghazal H, Ashok K, Fatima S, Shafaq I, et al (2014). Prognostic Factors in Tuberculosis Related Mortalities in Hospitalized Patients. Hindawi Publishing Corporation.Tuberculosis Research and Treatment Volume 2014, Article ID 624671, 6 pages. http://dx.doi.org/10.1155/2014/624671

Global tuberculosis report 2015 (WHO/HTM/TB/2015.22). Geneva: World Health Organization; 2015 (https://apps.who.int/iris/bitstream/handle/10665/191102/9789 pdf;jsessionid=257E179B7641F5CE7FD14BEF18488436?sequence=1, accessed 28 June 2019)

Global tuberculosis report(2019). Geneva: World Health Organization; 2019

Global Tuberculosis report (2020). World Health Organization; 2021. https: www.who.int/news-room/fact-sheets/detail/tuberculosis

Goel MK, Khanna P, Kishore J (2010). Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 1(4): 274–278.

Gokovali, U., Bahar, O. and Kozak, M, (2007). Determinants of length of stay: A practical use of survival analysis. Tourism Management 28 (2007) 736–746. www.elsevier.com/locate/tourman

Hayibor K.M, Bandoh A.D, Asante-Poku A, and Kenu E.(2020).Predictors of Adverse TB Treatment Outcome among TB/HIV Patients Compared with Non-HIV Patients in the Greater Accra Regional Hospital from 2008 to 2016. Tuberculosis Research and Treatment. Volume 2020, Article ID 1097581, 8 pages. https://doi.org/10.1155/2020/1097581

Hazra A, Gogtay N. (2017). Biostatistics series module 9: survival analysis. İndian Journal of Dermatology. 2017;62(3):251–257.

Hershkovitz I, Donoghue HD, Minnikin DE, May H, et al. (2015). Tuberculosis origin: the Neolithic scenario. Tuberculosis. 2015;95 Suppl 1:S122–6 (https://www.ncbi.nlm.nih.gov/pubmed/25726364, accessed 3 July 2019).

Ifebunandu NA, Ukwaja KN, Obi SN (2012). Treatment outcome of HIV-associated tuberculosis in a resource-poor setting. Trop Dr. 2012;42: 74 6.

Ige OM, Akindele MO (2011). Five year review of treatment outcome of directly observed therapy (DOT) for re-treatment pulmonary tuberculosis patients in UCH, Ibadan, Nigeria. Afr J Med Med Sci. 2011;40:15–21.

Jabbar A, Hussain SF, Khan AA.(2006). Clinical characteristics of pulmonary tuberculosis in adult Pakistani patients with co-existing diabetes mellitus. East Mediterr Health J. 2006;12(5):522-7.

Jeon CY and Murray MB (2008). Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. PLoS Med. 2008;5:1091–1101.

Jepchumba V., Simon K, Evans A, Lawrence M. and Hillary Kipruto (2017). Timing and Determinants of Tuberculosis Treatment Interruption in Nairobi County, Kenya. International Journal of Public Health Science (IJPHS) Vol.6, No.3, September 2017, pp. 203 212. ISSN: 2252-8806, DOI: 10.11591/ijphs.v6i3.8475

Jibrin, Y.B, Ali, A.B, Saad, S.T., and Kolo P.M (2013). Prevalence of Treatment Failure among Pulmonary Tuberculosis Patients in Federal Medical Centre, Gombe, Northeastern Nigeria. Hindawi Publishing Corporation, ISRN Infectious Diseases .Volume 2013, Article ID 461704, 4 pages. http://dx.doi.org/10.5402/2013/46

John P. Klein and Melvin L. Moeschberger (2003). Survival analysis: techniques for censored and truncated data. second edition, Statistics for biology & health. Springer-Verlag New York, Inc.

Junyong In & Dong K.L (2019). Survival Analysis: part II – Applied clinical data analysis. Korean Journal of Anesthesiology, Vol 72, No.5, October 2019

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457–481. http://dx.doi.org/10. 1080/01621459.1958.10501452

Kevin J. Carroll (2003). On the use and utility of the Weibull model in the analysis of survival data. Controlled Clinical Trials 24 (2003) 682–701. doi:10.1016/S0197-2456(03)00072-2

Kim J. (2001). Confidence intervals for the difference of median survival times using the stratified Cox proportional hazards model. Biometrical Journal 2001; 43:781–790.

Kolappan C, Subramani R, Kumaraswami V, Santha T and Narayanan PR (2008). Excess mortality and risk factors for mortality among a cohort of TB patients from rural South India. International Journal of Tuberculosis Lung Disease 2008;12 (1):81–6.

Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R (2019). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents. 2020;105924.

Lee J, Nam HW, Choi SH, et al. (2017). Comparison of early and late tuberculosis deaths in Korea. J Korean Med Sci 2017;32 (4):700–3. https://doi.org/10.3346/jkms.2017.32.4.700.

Lee, E.T & Go, O.T (1997). Survival Analysis in Public Health Research. Annual Public Health Review 1997. 18: 105 – 134. www.annualreviews.org Lefebvre and Falzon D.(2008). Risk factors for death among tuberculosis cases: analysis of European surveillance data. European Respiratory Journal 2008; 31: 1256–1260. DOI: 10.1183/09031936.00131107.

Lin, C., Kuo Y, Wang J., Hsu C & Chen J., et al (2014). Tuberculosis mortality: patient characteristics and causes. Article in BMC Infectious Diseases, January 2014. DOI: 10.1186/1471-2334-14-5 • Source: PubMed

Luka M.I, Idris S. H, Patrick N, Raymond D, et al (2014). Factors associated with interruption of treatment among Pulmonary Tuberculosis patients in Plateau State, Nigeria. Pan African Medical Journal. 2014 - ISSN 1937-8688; 17:78. doi:10.11604/pamj.2014.17.78.3464

Martinez M. (2007). Diagnostics for choosing between Log–Rank and Wilcoxon tests (Ph.D). Western Michigan University. 2007.

McNerney R, Maeurer M, Abubakar I, Marais B, et al, (2012). Tuberculosis diagnostics and biomarkers: needs, challenges, recent advances, and opportunities. J Infect Dis. 2012;205 suppl 2:S147–58.

Melo C.M, Donalisio M.R, and Cordeiro C.R (2017). Survival of patients with AIDS and co-infection with the tuberculosis bacillus in the South and Southeast regions of Brazil. Ciência & Saúde Coletiva, 22(11):3781-3792, 2017. DOI: 10.1590/1413-812320172211.26352015

Menezes AM, Hallal PC, Perez-Padilla R, Jardim JR, Mui-o A, Lopez MV, et al. (2007). Tuberculosis and airflow obstruction: evidence from the PLATINO

study in Latin America. Eur Respir J. 2007; 30 (6):1180-5. https://doi.org/10.1183/0903193

Michael, V.A & Bolarinwa, I.A (2020). Parametric Survival Modelling of Tuberculosis Data: A case Study of Federal Medical Centre, Bida, Nigeria. Modern Applied Science, Vol 14, No. 7; 2020. Canadian Center of Science & Education. Mollel E.W, Chilongola J.O (2017). Predictors for Mortality among Multidrug-Resistant Tuberculosis Patients in Tanzania. J Trop Med. 2017;2017: 9241238

Mustapha M, Surajuddeen M, Mohammed F.Z, Mustapha S., (2020). Risk Factors Associated with Mortality Among Patients with Novel Coronavirus Disease (COVID-19) in Africa. Journal of Racial and Ethnic Health Disparities. https://doi.org/10.1007/s40615-020-00888-3

Nagu, J,T., Said A, Ramadhani M, Mecky I. M, et al (2017). Tuberculosis associated mortality in a prospective cohort in Sub Saharan Africa: Association with HIV and antiretroviral therapy. International Journal of Infectious Diseases 56 (2017) 39–44.

Ifebunandu A.N., & Ukwaja N.K. (2012). Tuberculosis treatment default in a large tertiary care hospital in urban Nigeria: Prevalence, trend, timing and predictors. Journal of Infection and Public Health (2012) 5, 340—345.

Nigerian Federal Ministry of Health (2020). National Tuberculosis and Leprosy Programme: 2019 Annual TB Report. http://www.who.int/hiv/pub/guidelines/nigeria_tb.p

Omari-Sasu, A.Y., Owusu, D.A., Boanteng, M.A. & Zabogu, Z. (2016). Survival analysis of Tuberculosis patients in Upper West Region of Ghana. Int J of Stat&Apps, 6(2), 40-44. http://doi.org/10.5923/j.statistics.20160602.02.

Pathmanathan I, Dokubo EK, Shiraishi RW, Agolory SG, Auld AF, Onotu D, et al. (2017). Incidence and predictors of tuberculosis among HIV-infected adults after initiation of antiretroviral therapy in Nigeria, 2004-2012. PLoS

ONE 12(3): e0173309. doi:10.1371/journal.pone.0173309

Peters E.J, Ekott J.U, Eshiet G.A, Ayanechi C.C. (2004). Tuberculosis in Calabar : a ten-year review (1994–2003). Niger J Med. 2005;14:381–5.

Prentice RL, Kalbfleisch JD. (1979). Hazard rate models with covariates. Biometrics 1979; 35: 25–39.

Ravelo, J.L (2020). TB programs, trials pause as COVID-19 spreads. Devex News. Available at: https://www.devex.com/news/tb-programs-trials-pause-as-covid-19-spreads-97049.

Ravelo JL (2020). 1.8m people could die from tuberculosis in 2020: WHO. Devex News. Available at: https://www.devex.com/news/1-8m-people-could-die-from-tuberculosis-in-2020-who-98312.

Rich, J. T., Neely, J. G., Paniello, R. C., Voelker, C. C. J., Nussenbaum, B., & Wang, E. W. (2010). A practical guide to understanding Kaplan-Meier. Otolaryngology Head and Neck Surgery, 143(3), 331–336. http://dx.doi.org/10.1200/JCO.2008.18.8011

Salami A.K, Oluboyo P.O (2003). Management outcome of pulmonary tuberculosis: a nine year review in Ilorin. West Afr J Med. 2003; 22:114–9.

Samson Daniel, K. E. Lasisi and Jerry Banister (2020). Application of Survival Analysis of TB Patients Using Parametric Model: A Case Study of General Hospital Bayara. Asian Journal of Probability and Statistics 6(4): 54-65, 2020; Article no.AJPAS.55283 ISSN: 2582-0230 Santha T, Thomas A, Gopi PG,, Chandrasekaran V, Subramani R, et al (2005). Predictors of relapse among pulmonary tuberculosis patients treated in a DOTS programme in South India. Int J Tuberc Lung Dis, 9: 556–561

Santha T, Garg R, Frieden TR, et al. (2002). Risk factors associated with default, failure and death among tuberculosis patients treated in a Dots programme in Tiruvallur district, South India, 2000. Int J Tuberc Lung Dis. 2002;6(9):780–8.

Saranya & Karthikeyan, (2015). A Comparison study of Kaplan Meier and Nelson-Aalen Methods in Survival Analysis. International journal for research in emerging science and technology, volume-2, issue-11, nov-2015 e-issn: 2349-7610

Satagopan J.M, Ben-Porat L, Berwick M, Robson M, Kutler D, and Auerbach A.D (2004).A note on competing risks in survival data analysis. British Journal of Cancer (2004) 91, 1229 – 1235.

Schoenfeld, David (1982). Partial Residuals for The Proportional Hazards Regression Model. Biometrika, vol. 69, no. 1, 1982, pp. 239 – 241. www.jstor.org/stable/23358 Accessed 29 Nov. 2020.

Shimazaki T., Marte, S.D, Saludar N.R.D, Dimaano E.M, Salva,,E.P, et al (2013). Risk factors for death among hospitalised tuberculosis patients in poor urban areas in Manila, The Philippines. International Journal of Tuberculosis Lung Disease 17(11):1420–1426. Http://dx.doi.org/10.5588/ijtld.12.0848

Shuldiner J, Leventhal A, Chemtob D, Mor Z (2014). Mortality of tuberculosis patients during treatment in Israel, 2000-2010. Int J Tuberc Lung Dis 2014;18(7):81823. https://doi.org/10.5588/ijtld.13.0591.

Silva D.R, Muñoz-Torrico M,, Duarte R, Galvão T., et al (2018): Risk factors for tuberculosis: diabetes, smoking, alcohol use, and the use of other drugs"J Bras Pneumol. 2018;44(2):145-152,http://dx.doi.org/10.1590/S1806-37562017000000443

Stop TB Partnership (2020). The TB response is heavily impacted by the COVID-19 pandemic. Available at: http://stoptb.org/news/stories/2020/ns20_014.html.

Takarinda KC, Sandy C, Masuka N, Hazangwe P, et al. (2017). Factors associated with mortality among patients on TB treatment in the Southern Region of Zimbabwe, 2013. Tuberc Res Treat. 2017;2017 (2017):6232071.

Ukwaja KN, Oshi DC, Oshi SN, Alobu I (2014). Profile and treatment outcome of smear-positive TB patients who failed to smear convert after 2 months of treatment in Nigeria. Trans R Soc Trop Med Hyg. 2014;108:431–8.

Van Zyl Smit RN, Pai M, Yew WW, Leung CC, Zumla A, Bateman ED, et al (2010). Global lung health: the colliding epidemics of tuberculosis, tobacco smoking HIV and COPD. European Respiratory Journal 2010, 35: 27–33.

Bewick V, Cheek L and Ball J.(2004). Statistics Review 12: Survival Analysis. Critical Care 2004, 8:389-394 (DOI 10.1186/cc2955).

Waitt C.J and Squire S.B (2011). A systematic review of risk factors for death in adults during and after tuberculosis treatment. Int J Tubercul Lung Dis 2011;15 (7):871–85.

World Health Organization (2018). Latent TB infection: updated and consolidated guidelines for programmatic management. Geneva, Switzerland: World Health Organization; 2018. https://www.who.int/tb/ publications/2018/latent-tuberculosis-infection/en/

World Health Organization (2016). World Health Organization, Geneva. Tuberculosis and diabetes. Available from: http://www.who.int/tb/publications/diabetes_tb.pdf

Yi Xie, Jing Han, Weili Yu, Junping Wu, Xue Li, and Huaiyong Chen (2020).Survival Analysis of Risk Factors for Mortality in a Cohort of Patients with Tuberculosis. Canadian Respiratory Journal Volume 2020, Article ID 1654653,

9 pages. https://doi.org/10.1155/2020/1654653

Appendix A

Table 5.1: Multivariate Weibull Proportional Hazard Model

Covariates	Haz. Ratio	Std. Err.	z-value	P>z	95% Conf. Interval	
					Lower	Upper
Age	1.028797	0.006838	4.27	0.000	1.015482	1.042287
Type of TB	1.365745	.8087623	0.53	0.599	.427866	4.359446
Gender	0.9323225	0.230167	-0.28	0.777	.5746798	1.512538
Weight	.9755896	.0121604	-1.98	0.047	.9520445	.9997171
HIV Status	3.439982	1.177433	3.61	0.000	1.758765	6.728286
constant	.0000511	.0000596	-8.47	0.000	5.18e-06	.0005028
/ln p	.1701028	.1135776	1.50	0.134	-.0525052	.3927109
p	1.185427	.1346379			.9488494	1.48099
1/p	.8435781	.0958116			.675224	1.053908

Appendix B

Table 5.2: Multivariate Exponential Proportional Hazard Model

Covariates	Haz. Ratio	Std. Err.	z-value	P>z	95% Conf. Interval	
					Lower	Upper
Age	1.028763	.0068325	4.27	0.000	1.015459	1.042242
Type of TB	1.373415	.8132939	0.54	0.592	.4302753	4.383864
Gender	.9310046	.2298684	-0.29	0.772	.5738351	1.510486
Weight	.9752982	.0121636	-2.01	0.045	.9517469	.9994322
HIV Status	3.43245	1.174831	3.60	0.000	1.754938	6.713464
constant	.0001357	.0001257	-9.61	0.000	.0000221	.0008337

Appendix C

R Studio and Stata codes

R Codes

Set working directory setwd("C:/Users/user/Desktop/Survival Thesis")

#Import Excel Data #library(readxl)

#Book1< −read_excel("C:/Users/user/Desktop/Survival Thesis/Book1.xlsx")
View(Book1)

#Load Packages
#library(splines)
#library(MASS)
#library(survival)

Install Packages # install.packages (survminer)

TB< −Book1
TB$Censoring< −as.numeric(TB$Censoring)
TB$LOS< −as.numeric(TB$LOS)

Kaplan Meier Survival Estimates for each risk factor
km_TB< − survfit(Surv(LOS, Censoring) ∼ 1, data=TB, type= "kaplan-meier")
km_TBG< − survfit(Surv(LOS, Censoring) ∼ Gender, data=TB, type= "kaplan-meier")
km_TBA< − survfit(Surv(LOS, Censoring) ∼ Age, data=TB,type= "kaplan-meier")

km_TBS< − survfit(Surv(LOS, Censoring) ∼ HIV, data=TB, type= "kaplan-meier")

km_TBP< − survfit(Surv(LOS, Censoring) ∼ Pulmonary, data=TB,type= "kaplan-meier")

km_TBW< − survfit(Surv(LOS, Censoring) ∼ Weight,data=TB,type= "kaplan-meier")

Kaplan-Meier Plots comparing two groups

ggsurvplot(km_TB, data=TB, risk.table=F, conf.int=TRUE, ggtheme=theme_minimal(), ylim=c(0.8,1))

ggsurvplot(km_TBP, data=TB, risk.table=F, conf.int=TRUE,ggtheme=theme_minimal(), ylim=c(0.8,1))

ggsurvplot(km_TBW, data=TB, risk.table=F, conf.int=TRUE, ggtheme=theme_minimal(), xlab = "Time (days)", legend.labs = c("Above 35kg","Below 35kg"), legend.title="Pre-Treatment Weight", ylim=c(0.8,1))

ggsurvplot(km_TBS, data=TB, risk.table=F, conf.int=TRUE, xlab = "Time (days)", ggtheme=theme_minimal(), legend.title = "HIV Status", ylim=c(0.8,1))

ggsurvplot(km_TBG, data=TB, risk.table=F, conf.int=TRUE,xlab = "Time (days)", ggtheme=theme_minimal(), legend.title="Gender",ylim=c(0.8,1))

ggsurvplot(km_TBAg, data=TB, risk.table=F, conf.int=TRUE, ggtheme=theme_minimal(), legend.labs = c("Below 40yrs","40yrs & Above"), xlab = "Time (days)",legend.title="Age of Patients", ylim=c(0.8,1))

Log Rank Tests

loggender< − survdiff(Surv(LOS, Censoring) ∼ Gender, data = TB, rho = 0)

logHIV< − survdiff(Surv(LOS, Censoring) ∼ HIV, data = TB)

logAge< − survdiff(Surv(LOS, Censoring) ∼ Age, data = TB)

```
logweight< − survdiff(Surv(LOS, Censoring) ∼ Weight, data = TB)
logPul< − survdiff(Surv(LOS, Censoring) ∼ Pulmonary, data = TB)


# Univariate Analysis
cox_G< −coxph(Surv(LOS, Censoring) ∼ Female, data = TB)
summary(cox_G)
cox_Age< −coxph(Surv(LOS, Censoring) ∼ Age, data = TB)
summary(cox_Age)
cox_Type< −coxph(Surv(LOS, Censoring) ∼ Pulmonary, data = TB)
summary(cox_Type)
co_HIV< −coxph(Surv(LOS, Censoring) ∼ HIV, data = TB)
summary(cox_HIV)
cox_Weight< −coxph(Surv(LOS, Censoring) ∼ Weight, data = TB)
summary(cox_Weight)


# Cox Proportional Hazards (PH) Model
cox_regTBC2 < − coxph(Surv(LOS, Censoring) ∼ Age+Pulmonary+Female+Weight+HIV,
data = TB)


# MODEL DIAGNOSIS

# Proportionality Assumption Test for Cox Model # Using Schoenfeld Residuals
Plot (For Categorical Variables)
cox_phTB< − cox.zph(cox_regTBC2)
ggcoxzph(cox_phTB)


# Schoenfeld Residual Plots for Categorical Variables
cph.selected< − coxph(Surv(LOS, Censoring) ∼ Pulmonary+Female+HIV, data
```

= TB)

sf.residual¡-cox.zph(cph.selected)

ggcoxzph(sf.residual)

Testing Influential Observations or Outliers #a) Martingale Residuals

ggcoxdiagnostics(cox_regTBC2, type = "martingale", linear.predictions = TRUE, title = "Martingale Residuals")

Weight< −coxph(Surv(LOS, Censoring) ∼ Weight, data = TB)

ggcoxdiagnostics(Weight, type = "martingale", linear.predictions = TRUE, title = "Martingale Residuals for Weight", ylim=c(-0.5,1))

Age< − coxph(Surv(LOS, Censoring) ∼ Age, data = TB)

ggcoxdiagnostics(Age, type = "martingale", linear.predictions = TRUE, title = "Martingale Residuals for Age", ylim=c(-0.5,1))

b) Deviance Residuals ggcoxdiagnostics(cox_regTBC2, type = "deviance", linear.predictions = FALSE, ggtheme = theme_bw(), title = "Deviance Residuals", ylim=c(-2,3))

Linearity Assumption

Using Martingale Residuals

plot(predict(cox_regTBC2), residuals(cox_regTBC2, type = "martingale"), xlab = "fitted values", ylab = "Martingale residuals",main = "Residual Plot", las=1)

abline(h=0)

lines(smooth.spline(predict(cox_regTBC2),residuals(cox_regTBC2, type= "martingale")), col="red")

Observations that are Outliers in Martingale residuals res.mart¡-resid(cox_regTBC2,

type = "martingale") plot(res.mart) head(sort(res.mart))

Schoenfeld Residuals for Proportionality Assumption cox.zph(cox_regTBC2)

cox.zph(cph.selected)

par(mfrow=c(3,1))

plot(cox.zph(cph.selected))

par(mfrow=c(1,1))

plot(cox.zph(cph.selected)[1])

abline(h=0, col=2, lty= 2, lwd = 2)

#Cox Snell Residuals Plot

Coxsnellres< −TB$Censoring-resid(cox_regTBC2,type="martingale")

Fit< −survfit(coxph(Surv(coxsnellres,TB$Censoring)∼ 1,method='breslow'),type='aalen')

plot(fit$time,-log(fit$surv),type='s',xlab='Residuals',ylab='Estimated Cummulative Hazard Function',main='Cox Snell Residual Plot')

abline(0,1,col='red',lty=1)

Test of Proportionality Hazard Assumption Mart1< −coxph(Surv(LOS, Censoring) ∼ Age, data = TB, method = "breslow")

Mart2< −coxph(Surv(LOS, Censoring) ∼ Weight, data = TB, method = "breslow")

resid(Mart2,type='martingale')

plot(TB$Age, resid(Mart2),xlab="Age(Years)", ylab="Martingale Residuals",main='Martinga

Plot for Age')

lines(lowess(TB$Age, resid(Mart2)),col='red')

lines(x=c(0,80),y=c(0,0), lty=3)

Analysis using Stata 14.0

```
# Description of the survival data
ḋescribe # Assign Censoring to observed failure times
ṡtset LOS, failure(Censoring==1)

# Parametric Models for Multivariate Analysis
# Exponential Distribution
ṡtreg Female HIV Age PreTreatmentWeight Pulmonary, dist(exponential)
# Weibull Dsitribution
ṡtreg Female HIV Age PreTreatmentWeight Pulmonary, dist(weibull)
# Gompertz Distribution
ṡtreg Female HIV Age PreTreatmentWeight Pulmonary, dist(gompertz)
```