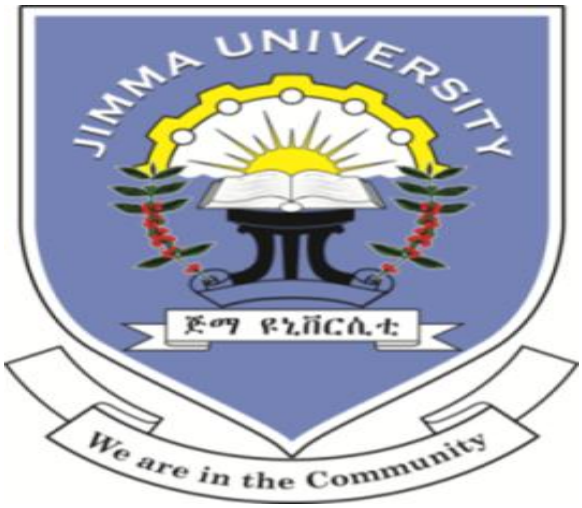


Modeling Time to Death of Patients by Multi Drug Resistant Tuberculosis at Saint Peter's Hospital, Ethiopia

By: Gosa Abebe

A Thesis Submitted to the Department of Statistics, College of Natural Science in Jimma University as a Partial Fulfillment for the Requirements of Master of Science (MSc) Degree in Biostatistics.

September 2017

Jimma, Ethiopia

Modeling Time to Death of Patients by Multi Drug Resistant Tuberculosis at Saint Peter's Hospital, Ethiopia

By:Gosa Abebe

Advisor: Zeytu Gashaw (PhD)
Co-advisor: Dechasa Bedada (MSc)

September 2017

Jimma, Ethiopia

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission if an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in her or his judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Gosa Abebe

Date: _____

Signature: _____

September, 2017

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis advisors, we here by certify that we have read the thesis prepared by Gosa Abebe under our guidance, which is entitled **"**Survival modeling time to-death of patients by multi drug resistant tuberculosis at Saint Peter's Hospital, Ethiopia**.",** in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is ready for submission to the university library.

Zeytu Gashaw (PhD)	_____	_____
Main advisor	Signature	Date
Dechasa Bedada (MSc.)	_____	_____
Co-advisor	Signature	Date

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

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

DEDICATION

I dedicate this study to the Almighty God, for his endless love, kindness, wisdom and grace that always abounds all the days of my life. .

ACKNOWLEDGMENTS

First and foremost I would like to thank the almighty God for his support and guidance in all my day activities.

Secondly I would like to express a sincere acknowledgement to my advisor Zeytu Gashaw (PhD) for his advice and guidance that contributed to the successful realization of the thesis and My deepest gratitude goes to my co-advisor Dechasa Bedada(MSc) for his valuable suggestions and comments to the successful realization of this thesis.

Thirdly I would like to thank St.Peter's Specialized TB Hospital to undertake this study with their cooperation and permission in using the data special thanks for Dr. Addisu Admasu

Fourthly I would like to thank Jimma University for providing me the financial support for this study, such advisors and instructors and for all the support.

Lastly , my deepest and warmest gratitude goes to my family that has been a source of pride and courage throughout my work, specially to my mother Abebech Daba ,my father Abebe Tulema ,my lovely sister Aberash Abebe,all my brothers(Biru,Teshome and Asnake) and all my friends who have been with me in giving constructive ideas, encouragement and motivation during this studies.

ABSTRACT

Background: World Health Organization (WHO) Annual Global TB Report, Ethiopia ranked 7th in the world for TB burden and 15th from 27 highest MDR-TB burden countries .The Ethiopian Federal Ministry of Health(FMOH) Hospital statistics data has shown that tuberculosis is the leading cause of morbidity, the third cause of hospital admission and the second cause of death in Ethiopia, after malaria.

Objective: The objective of this study was to identifying the risk factors for death of MDR-TB at Saint Peter's specialized TB Hospital.

Methods: A facility based retrospective cohort study was conducted at St.Peter's specialized TB Hospital from January 3, 2014 to October 28, 2016. 320 MDR-TB patients included in the study.Classical survival analysis like Cox PH regression model, Stratified Cox regression model and Weibull accelerated failure time model were employed to identify the risk factors MDR-TB patients. Bayesian Weibull accelerate failure time was employed to identify significant effect covariate on the survival time of MDR-TB patients.

Results: The median survival time of the patients was about 597 days with maximum and minimum survival time 969days and 1 day respectively. From Accelerated failure time result, the risk factor for the mortality of multidrug resistance tuberculosis patients were Age of patient(ETR=0.918;Cl=0.885,0.952), Body mass index(ETR=1.204;CL=1.028,1.410),smoking status(ETR=.209,CL=0.073,0.594),HIVco-infection(ETR=0.362,CL=0.137,0.953), morbidity(ETR=0.220,CL= 0.062,0.789) and TB status (ETR=7.879,CL=2.519,24.643) at 5% level of significance. From Bayesian result the risk factor for the mortality of multi drug resistance tuberculosis patient were Age of patient(μ=0.012,CRI=(0.001,0.012)),TBstatus($\mu = -0.549$,CRI=(-1.818,-0.507)),Smoking(μ=0.155, CRI =(0.074,0.570)) andComorbidity ($\mu = 0.055$, CRI = (0.011,0.323)) at 95% credible interval.

Conclusions: Mortality rate of patients was high at the earlier times of treatment. Demographic and health factors were associated with increased risk of mortality, therefore potential stakeholders like government and non-governmental organizations should pay attention to the subject.

Key word: Weibull AFT model, Bayesian survival model, MCMC, Cox Snell residual

Table of Contents

LIST OF TABLE

LIST OF FIGURE

ABBREVIATION AND ACRONYMS

AMK	Amikacin
AIC	Akaike's Information Criterion
CDC	Centers of Disease Control and Prevention
CI	Confidence Interval
DOTS	Directly Observed Treatment, Short-Course
DR-TB	Drug resistant tuberculosis
DST	Drug susceptibility test
E	Ethambutol
ETR	Event time ratio
FMOH	Ministry of Health
GHC	Global Health Committee
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
INH	Isoniazid
KAN	Kanamycin
LR	Likelihood Ratio
MDR-TB	Multidrug resistance tuberculosis
MLE	Maximum Likelihood Estimate/Estimator
RIF	Rifampicin
STM	Streptomycin
TB	Tuberculosis
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background of the problem

Multidrug-resistant tuberculosis (MDR-TB) is a form of TB caused by bacteria that do not respond to isoniazid and rifampicin, the 2 most powerful, first-line anti-TB drugs. MDR-TB is treatable and curable by using second-line drugs. However, second-line treatment options are limited and require extensive chemotherapy (up to two years of treatment) with medicines that are expensive and toxic. In some cases, more severe drug resistance can develop [1].

According WHO 2015 report worldwide Treatment outcome of MDR TB patients in 2012 cohort is 50% were treated successfully, 16% were died, 16% were lost follow up, 10% were treatment failed and 8% were had no outcome information [2]. Based on the World Health Organization (WHO) Annual Global TB Report (2009), Ethiopia ranked seventh in the world for TB burden and and 15th from 27 highest MDR-TB burden countries, with an estimated annual TB incidence (all forms) of 378 new cases per 100,000 persons and 163 new smear positive cases per 100,000 persons per year [3].

The World Health Organization (WHO) estimated 480000 new cases of MDR-TB in 2015 and an additional 100000 cases diagnosed with rifampicin-resistant TB (RR-TB). India, China and the Russian Federation accounted for almost half (45%) of the total burden [4]. Out of 580000 patients eligible for MDR-TB treatment, only 125000 (20%) were enrolled in treatment programs [4].

The study conducted in Addis Ababa from February 2010 to March 2010 GC to assess the patterns of resistance to anti-TB drugs among previously treated TB patients referred to St. Peter's TB Specialized Hospital .out 124 smear-positive pulmonary TB patients, 117 (94.4 %) were susceptible to Rifampicin, while 7 (5.7 %) were confirmed to be resistant to Rifampicin and Isoniazid. The overall prevalence of MDR-TB was 5.7 % (2.3 % among new cases and 13.9 % among previously treated cases) [5].

A cross-sectional study was conducted between March 2009 and July 2009 among smear-positive pulmonary TB patients diagnosed at the Gondar Hospital, the Gondar Health Centre, the

Metemma Hospital, the Bahir Dar Hospital and the Debre Markos Hospital, out 260 Mycobacterium tuberculosis isolates, 41 (15.8%) were resistant to at least one first-line drug, 13 (5.0%) were multidrug-resistant (MDR) and 9 (3.5%) were resistant to all first-line drugs. Any resistance to INH, RMP, SM, EMB and PZA was respectively 36 (13.8%), 15 (5.8%), 26 (10.0%), 19 (7.3%) and 12 (4.6%). Of 214 new and 46 previously treated cases, respectively 8 (3.7%) and 5 (10.9%) were MDR. All isolates were susceptible to all second-line drugs [6].
Among the 23 papers, six of them reported high prevalence of MDR-TB in the range of 3.3%-46.3%. Likewise, two studies reported XDR-TB in the range of 1% - 4.4% in Ethiopia. The most powerful predictor of the emergence of MDR-TB reported in Ethiopia is previous exposure to anti-TB drug treatment. This review indicated that MDR-TB in Ethiopia is a serious public health problem that needs to be addressed urgently [7].

The first Drug Resistance Survey, conducted between 2003 and 2006 showed that multi-drug resistance to TB drugs (MDR-TB) is present in 11.8% of previously treated cases and 1.6% of newly diagnosed TB cases, with an estimated 5,200 cases annually. The programmer's Capacity to treat MDR-TB patients is limited to two referral hospitals in Addis Ababa (St Peter & ALERT) and one in Gondar (Gondar University Hospital). At the end of February 2012, a total of 424 cases of MDR-TB patients were enrolled on treatment. The routine MDR-TB surveillance system to detect MDR-TB suspects for early diagnosis and treatment is not yet fully formed [8].

The study conducted in Oromia indicate, out 439 suspected MDR-TB cases, 265 had a confirmed M. tuberculosis infection, of whom 88 (33%) had laboratory-confirmed MDR-TB. Over two-thirds (65%) were between 18 and 39 years of age. On multivariate analysis, an occupation of farming, known TB contact history, alcohol use, HIV infection, previous known TB history, and previous TB treatment outcome were predictors of MDR-TB [9].The Ethiopian Federal Ministry of Health (FMOH) hospital statistics data has shown that tuberculosis is the leading cause of morbidity, the third cause of hospital admission (after deliveries and malaria), and the second cause of death in Ethiopia, after malaria [10]. Based on WHO report in Ethiopia, the incidence of TB of all forms and smear positives stand at 341 and 152 per 100,000 populations, respectively [11].

1.2 Statement of the problem

MDR-TB is manmade problem causes of in adequate anti-TB treatment and the emergence of extremely drug resistant TB further complicates the efforts to tackle the problem especially in Developing countries including Ethiopia. The data obtained from MDR-TB are not fully observed some data are censored. Therefore Survival analysis is one of the appropriate methods to demonstrate life time and to identify risk factors. The length of survival of MDR-TB patients depends on time from the date MDR-TB infection is confirmed until death or some observations with incomplete records(censored).Survival analysis is an appropriate analytic method for this study to assess survival/death and its risk factors. Hence, this study aims to address the following research questions:

1. Which factors significantly affect the survival/death of MDR.TB patients?
2. Which groups have better survival time among various levels of factors?
3. What is the median of survival time of the MDR-TB patient?

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study was to identify the risk factors for MDR-TB patients at Saint Peter's specialized TB Hospital from January 3, 2014 to October 28, 2016.

1.3.2 Specific Objectives

1. To identify the determinant factors that accelerates the death of MDR-TB patients.
2. To compare the survival of MDR-TB patients with respect to their categories.
3. To estimate the median of survival time of MDR-TB patients.

1.4 Significance of the study

MDR-TB is mainly occurred as results of poor treatment outcomes, poor treatment adherence, poor quality of drugs and poor infection control practices. Ethiopia is one of the high burden countries for MDR-TB. However, the extent and the magnitude of the problem is well studied but the determinant factors that accelerates the death of MDR-TB patient is not well studied . So it is important to identify risk factors for MDR-TB patient. The study would be initiate further research to control MDR-TB.

1. The results of the study might provide information to government and other concerned bodies in setting policies, strategies and further investigations for reducing death to MDR-TB patient.
2. The result may help donors and government to understand risk factors that influence the death of MDR-TB patients.
3. The study could provide base-line data for detail and further studies in the future.

1.5. Ethical Consideration

The Research Ethics Review Board of Jimma University has provided an ethical clearance for the study. The data was collected from Saint Peter's specialized tuberculosis Hospital, and to do so the department of statistics asked to write an official co-operation letter to the Hospital from where data was obtained. The study conducted without individual informed consent because it relied on retrospective data. In this research, the information obtained from patients' card kept secured.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Multidrug-resistant (MDR) TB has become a major public health problem and presents new barriers to the control of TB. It is due to human error as the result of poor supply management and quality of anti-TB drugs and inadequate or improper treatment, which is further exacerbated by human Immunodeficiency virus (HIV) [12].

Treatment of MDR-TB is more challenging. It requires use of second-line drugs that are more costly and cause more severe side-effects, and recommended regimens must be taken for up to two According to WHO 2012 report, there were an estimated 1700 and 550 MDR-TB cases among new and re-treatment pulmonary TB cases in 2011, respectively in Ethiopia [13]. When an individual who has no history of first-line TB treatment develops MDR-TB, it is termed primary. When insufficient treatment leads to selection of spontaneously resistant strains (i.e., drug resistance is acquired), the disease is termed secondary MDR-TB [14].

The emergence of MDR-TB is a threat for the populations of resource-limited countries. In Ethiopia, the low socioeconomic status of the people, high prevalence of infectious diseases and limited access to well-equipped health care facilities worsens the effect of MDR-TB. Furthermore, poor treatment outcomes, longer treatment time (about two years), higher treatment costs, and many more complications make MDR-TB a more complex disease than TB [15].
At the time of this study in Ethiopia, the LPA, or culture using Lőwenstein-Jensen media (LJ), and drug-susceptibility testing (DST) were provided only at the Ethiopian Health Nutrition and Research Institute (EHNRI) in Addis Ababa. MDR-TB occurs mostly in relation to improper treatment of drug-susceptible TB. In countries like Ethiopia MDR-TB is becoming a challenge because of poor adherence to treatment and an increase in the use of illegal and unapproved treatment regimens for MDR-TB [15].

2.2 Prevalence rate of MDR-TB

The occurrence of drug resistance has increased since the first drug treatment for TB was introduced in 1943. Mycobacterium tuberculosis is a slow- growing bacterium, resistant to most conventional antimicrobial agents partly due to its impermeable cell wall. It may persist in a

dormant or latent form, unsusceptible to agents targeting growing bacteria [16].The bacterium infects almost one-third of population globally and during the past decade there has been a resurgence of tuberculosis [17].

Drug resistant TB is confirmed through laboratory tests that show that infecting isolates of M. tuberculosis grow in vitro in the presence of one or more anti-TB drugs. The emergence of drug resistance tuberculosis, particularly MDR-TB has become a major public health problem in a number of countries and an obstacle to the global TB control efforts. Nearly half a million cases of MDR-TB emerge every year but only 3% of them get treatment globally and 100,000 die annually [18]. Multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis (TB) continue to emerge in high HIV prevalence settings, and their mortality in HIV co-infected patients remains high [19]. Each year, globally, about 440,000 MDR-TB cases are estimated to emerge, and 150,000 people with MDR-TB die [20].

Worldwide, there were 650,000 MDR-TB cases in 2010, and in 2008 World Health Organization (WHO) estimated that there were 150,000 deaths annually due to MDR-TB. Overall, the 27 high MDR-TB burden countries accounted for 85% of all MDR-TB cases. China, and India, was the top two countries accounting 50% MDR-TB cases [21]. In 2010, less than 5% of new and previously treated TB patients were tested for MDR-TB because of limited availability of the test in most developing countries [22].

According to World Health Organization 2015 report, among newly diagnosed TB cases 1.6% was found to be with MDR-TB and MDR-TB among previously treated TB cases was 11.8% [23] Another study in Ethiopia also showed that MDR-TB rate was found in the range of 3.3%-46.3% [24]. Ethiopia is 15th among the 27 MDR-TB high-burden countries, with an estimated 5,200 cases occurring each year [25].

Based on study was conducted on Rifampicin Mono-Resistance in Mycobacterium tuberculosis among patient attending at Yirgalem Hospital from August-December, 2014,Yirgalem, Ethiopia. A total of 236 participants were included under this study. Among these, males (57.6%) are slightly dominating female. Concerning to treatment history, 177 (75.0%) are new and the rest 59 (25.0%) retreated. Fifty eight (24.6%) of the total subjects were suspected for MDR tuberculosis. Twenty two (9.3%) of the subjects were smear positive. The highest positive finding of rifampicin susceptible Mycobacterium tuberculosis bacilli

observed within age group of 16-30.The overall prevalence of pulmonary tuberculosis was 16.5%. From this prevalence, 3.4% was shared by Rifampicin mono resistant Tuberculosis. Based on study finding, the overall prevalence of pulmonary tuberculosis was 16.5%. From these, 3.4% was account for Rifampicin Mono-Resistance Mycobacterium tuberculosis among study subject. Most of the affected study subjects were productive age group. Therefore, we recommend that there should be enhanced efforts in detection of MDR tuberculosis in study area to control dissemination of the disease among the community [26].

2.3 Factors associated with MDR-TB

TMDR-TB is a reflection of the mismanagement of TB cases, which includes wrong diagnosis, delay of diagnosis, wrong/interrupted treatment, misuse of TB medicines, and poor adherence to standardized treatment, unregulated supply of anti-TB drugs and utilization of TB drugs of unknown quality [27].The risk factor for MDR-TB are ;

2.3.1 Gender and MDR-TB

Female was risk factors for MDR-TB [28]. But a study finding in Nigeria shows gender was not significantly associated with MDR-TB [29]. And study findings in Thailand also shows male gender as risk factors for MDR-TB [30] and in Ethiopia (AOR =2, 95% CI [1.4-5]) showed male was a risks factor for MDR-TB [31**]**.

2.3.2 TB site and MDR-TB

According to the 2011 health and health related report of the Federal Ministry of Health (FMOH) of Ethiopia, TB is the third leading cause of death in Ethiopia. During the year 2010/11, a total of 159,017 TB cases were identified in Ethiopia. Among these 151,866 (95.5%) were new cases all forms of TB. The proportion of new smear-positive, smear-negative and extra-pulmonary TB among all new cases is 32.7%, 34.8%, and 32.5%, respectively. Re-treatment (after failure or relapse of first treatment) cases represented about 2.9% of all TB cases identified. According to the anti-TB drug resistance survey conducted in Ethiopia in 2012/13 FMOH, among 804 newly diagnosed TB cases 13 (1.6%) were found to be infected with MDR-TB. The rate of MDR-TB among specimens from 76 previously treated TB cases was 11.8%.

2.3.3 Age patient and MDR-TB

Age group at 25-44 years in Ethiopia (AOR=2.8, 95% CI [1.7–6.4]) [31] and in Bangladesh (AOR=1.72, 95% CI [1.12–2.66]) [32] and [33] was a risks factor of MDR-TB. The study included 204 patients diagnosed with MDR-TB in England, Wales and Northern Ireland. The objective of the study was to describe the clinical characteristics of patients and to examine factors associated with a successful treatment outcome; loss to follow up and death of MDR-TB patients completing treatment between 2004 and 2007.The study used logistic regression. Age, sex, TB site, ethnicity, social risk factors, comorbidities, previous diagnosis of TB, drug susceptibility test (DST) and HIV status were associated with a successful treatment outcome, mortality, loss to follow up and treatment stopped. The result of the study suggested that having any co-morbidity, particularly HIV and diabetes, were strongly associated with death of patients [34].

2.3.4 Residence of patient and MDR-TB

In Bangladesh urban residence was associated with MDR-TB occurrence [33] whereas, a study finding in southern Ethiopia shows, there was no significant association between residence and MDR-TB [35].

2.3.5 HIV co-infection and MDR-TB

HIV was a risk factor for TB/MDR-TB accordingly to, WHO report on at California; US during 2011which shows HIV contribute 4.5% MDR-TB cases [20].Study finding in southern Ethiopia shows there was no statistically significant association of HIV status with MDR-TB [35]. Whereas other study shows HIV is associated with increased risk of acquired MDR-TB (OR=1.24, 95% CI [1.04–1.43]) and (OR=2.28, 95% CI [1.52–3.04]) for primary MDR-TB [30]. And a study by Birhanu and his colleagues showed MDR-TB and HIV significant association (OR=3.7, 95%Cl [1.90, 7.22].Tuberculosis (TB) is a disease caused by Mycobacterium tuberculosis and is one of the deadliest diseases in the world. It is mostly spread from person to person through the air and usually affects the lungs, but it can also affect other parts of the body such as the brain and kidneys. About one-third of the world's population has latent TB, which means people have been infected by TB bacteria but are not (yet) ill with the disease and cannot transmit the disease. People infected with TB bacteria have a lifetime risk of falling ill with TB of 10%. However, persons with compromised immune systems, such

as people living with HIV, malnutrition or diabetes, or people who use tobacco, have a much higher risk of falling ill [36].

2.3.6 TB status and MDR-TB

Multidrug-resistant tuberculosis (MDR-TB) is a type of TB that is resistant to at least the first line anti-TB drugs, Rifampicin and Isoniazid. MDR-TB occurs either when a person is infected with a resistant strain or when improper treatment leads to drug selection of the resistant strain [33].From a nationwide survey conducted in China; the estimated MDR-TB rate was 5.7% for new cases and 25.6% for previously treated cases [22]. And a study finding in Uganda shows, MDR-TB of 1.4% from new cases and 12.1% from previously treated cases [37]. For instance, globally, more than half million new MDR-TB cases are estimated to emerge annually as a result of inadequate treatment and subsequent transmission. Although some individuals who have had previous TB treatment are infected by MDR-TB, many new cases of MDR-TB are also created each year by a combination of physician error and poor patient compliance with treatment and poor quality drugs [38].

Study conducted in republic of Georgia shows previously treated for TB were more likely to have MDR-TB than patients who were new (OR=5.27, 95% CI [3.75-7.41]). Likewise study in the Community of Madrid shows significant association with a history of previous TB treatment (OR=5.94, 95% CI [1.46-24.18]) [39]. Nationwide study in China shows previous treatment history had a more than 7 fold increased risk of MDR-TB, compared with those never previously treated [22]. MDR-TB is a reflection of the mismanagement of TB cases, which includes wrong diagnosis, delay of diagnosis, wrong/interrupted treatment, misuse of TB medicines, and poor adherence to standardized treatment, unregulated supply of anti-TB drugs and utilization of TB drugs of unknown quality [40]. For instance, globally, more than half million new MDR-TB cases are estimated to emerge annually as a result of inadequate treatment and subsequent transmission. Although some individuals who have had previous TB treatment are infected by MDR-TB, many new cases of MDR-TB are also created each year by a combination of physician error and poor patient compliance with treatment and poor quality drugs [38]. In countries where drug resistance has been identified, specific measures need to be taken within TB control programs to address the problem through appropriate management of patients and adoption of strategies to prevent the propagation and dissemination of DR-

TB [41]. Treatment of MDR-TB is more challenging. It requires use of second-line drugs that are more costly and cause more severe side-effects, and recommended regimens must be taken for up to two year [23].

2.3.7 Comorbidity and MDR-TB

Showed co-morbidities had influence on mortality among MDR-TB patients. a study of MDR-TB in Estonia, showed that co-morbidities (OR, 2.62; 95% CL, 1.00-6.87) were independent risk factors for treatment failure [42].Study in St. Peter TB specialized Hospital, Addis Ababa, Ethiopia. The study was conducted from October, 2011 up to may, 2012 among cohorts of MDR-TB patients that started treatment in February 2009. a total of 188 patients were followed for a total of 79,600 person-days. Median follow up time was 466.5 days or 1.28 years. The independent variables included were: sex, age, weight, region, HIV status, number of anti-TB drug taken, MDR category, presence of chronic disease, clinical complication, radiological findings, number of resistant drugs at initiation, therapeutic delay, smoking status and smear positivity with time of death.

Survival trend over the follow up time was studied using the Kaplan-Meier method and the covariates were fitted to Cox proportional hazard regression model. Smoking, therapeutic delay of at least one month, HIV serpositivity, and clinical complication were found to be factors significantly associated with death in the multivariate analysis. The study revealed that survival of patients under MDR-TB treatment was not associated with age, sex, baseline weight, radiological findings, previous TB treatment, number of first line resistant drugs and co-morbidity [43].

2.3.8 Smoking status and MDR-TB

Retrospective national cohort study on MDR-TB cases (n=1809) reported from 2002 to 2008 in Lithuania. Sex, age, rural/urban residence, contact with TB, smoking, alcohol use, drug abuse, homelessness, employment status, education level, HIV status, co-morbidity, TB type, smear positivity and cavitary disease were considered as predictors. The result revealed that age, rural residence, alcohol use, employment status, lower levels of education, positive or unknown HIV status, cavity disease and being smear positive at the time of MDR-TB diagnosis were associated with survival. There was no difference in survival of patients with primary MDR –TB compared with those who developed drug resistance during treatment [44]. The study conducted ALERT Hospital, Addis Ababa and Gondar University Teaching and Referral Hospital, Gondar,

Ethiopia. The descriptive analysis indicates that out of the total 342 individuals, 37(10.8%) died; 11 and 12 deaths occurred in the first and the second three months of MDR-TB treatment follow up, respectively. The median survival for MDR-TB patients was 16 months. Factors associated with increased risk of mortality were: having clinical complication (HR=4.7161; 95%Cl; 2.1861 – 10.1740), resistance to INH, RIF and at least one of other drugs (E, STM, KAN, AMK & CPM) (HR=2.9771; 95%Cl; 1.3586 −6.5238), smoking (HR=3.17; 95%; 1.32 − 7.64), weight (HR= 0.9093; 95%Cl; 0.8760 − 0.9440)and age (HR= 1.2199; 95%Cl; 1.0681 − 1.3933)[45].

2.4 MDR-TB detection status

Health education regarding spread of disease, early detection of MDR-TB by Strengthened laboratory support, effective therapy, implicating innovative control measures, and applying them specially among immigrants, would interrupt the ongoing transmission and control emerging epidemic [46]. In Eastern Europe, prisons have had to deal with substantial caseloads of MDR-TB patients. So close monitoring was mandatory for group of TB patients [36].

The MDR-TB treatment strategy in Ethiopia combines standardized and individualized treatment based on second line culture and Drug Susceptibility Testing (DST) (kanamycin and Ofloxacin) in all confirmed MDR-TB patients. Standardized regimens have given to all confirmed MDR-TB cases under daily Directly Observed Therapy (DOT). The initial phase is at least six months, and then the continuation phase is at least 12 months. In Ethiopia, the standard regimen for MDR-TB uses the combinations of: [Ethambutol- Pyrazinamide –Kanamycin (Amikacin)– Levofloxacin– Ethionamide –Cycloserine] for six months, and [Ethambutol –Pyrazinamide – Levofloxacin – Ethionamide – Cycloserine] for 12 months. The total duration may be extended by clinicians according the findings of culture conversion [47].

CHAPTER THREE

METHODOLOGY

3.1 Back ground of study area

The study was conduct at St. Peter's TB Specialized Hospital. St. Peter's TB Specialized Hospital is a governmental Hospital under Federal Democratic Republic of Ethiopia- Ministry of Health (FMOH). The Hospital provides various services especially in tuberculosis diagnosis and treatment. It serves as a referral TB Hospital in Addis Ababa, Ethiopia and has a vision to become Center of excellence for diagnosis and treatment of TB in East Africa.

3.2 Source of data and Study design

The MDR-TB data set were collected from Saint Peter's specialized tuberculosis Hospital. The dataset for this study is extracted from the card of MDR-TB patients admitted from January 3, 2014 to October 28, 2016.Total of 320 MDR-TB patients included in this thesis. A facility based retrospective study design employing medical records review of MDR-TB registration cards .Data was obtain from a cohort of MDR-TB patients enrolled in St. peter TB hospital in Addis Ababa, the capital city of Ethiopia. All patients who were diagnosis with a first MDR-TB episode and admit to one of the MDR-TB treatment centers was included in the study.

3.3 Data collection procedures

Ethical clearance and institutional permission was obtained from Saint Peter's Specialized Hospital research committee .The study incorporate secondary data. Data extracted from the medical records of patients with MDR-TB by health professionals working at the Hospital through a uniform checklist containing socio-demographic factors, clinical factors and time of the event (death/censored) occurred. The statistical software used in this research was WinBugs and R.

3.4 Inclusion criteria

All MDR-TB patients registered from January 3, 2014 to October 28, 2016 for MDR- TB treatment or those who start MDR -TB treatment at St. Peter's Hospital.

3.5 Exclusion criteria

All TB smear negative after the first line treatment.

3.6 Variable in study

3.6.1 Dependent variable

The response variable is time to event of the patients and it is defined status variable (event or Censoring variable).This is to say that the response variable is a censored survival time represented by variable time and event/death. Survival time measures the follow-up of time from a defined starting point to the occurrence of a given event. This observation time has two Components, the beginning point of the study time and the observation of time to the end. In Survival analysis, the outcome of interest (death in this study) is the duration of time until death occurs. The status variable is coded as 0 for censored and 1 for death. The dependent (outcome) variable in this study the survival time measured (in days) from the date MDR-TB treatment's start until the date of the patient's death or censor.

Starting time; the entry of the survival data was considered from the day that the patient start multi drug resistance after diagnosis.

Ending time; the ending time of this study was on October 28, 2016.

Event; event were occur when the patient died by MDR-TB between January 3, 2014 to October 28, 2016.The status variable coded 1 for event.

Censored; The censored observation in this study were the patient who died by the other case ,cured from the disease, loss follow up and shift to the other hospital The status variable coded 0 for censored .

3.6.2 Independent Variables

Several predictors were considered in this study to investigate the risk factor for the death of MDR TB patient. Some of these variables are categorical and others are continuous. A set of variables was selected for the analysis. Considering the potential importance, the following socio-demographic factor and clinical factor were considered in this study.

Independent covariate

Possible covariate	Candidate covariate
1. Sex (Male, Female)	1) Sex (Male, Female) of patient
2. Age of patient	2) Age of patient
3. Region (Address they came from)	3) Region (Address they came from)
4. Base line Weight in kilograms	4) Baseline Weight in kilograms
5. TB treatment status (Yes, No)	5) TB treatment status (Yes, No)
6. TB treatment Adherence status	6) TB site (Pulmonary and Extra pulmonary)
7. TB site (Pulmonary and Extra pulmonary)	7) Level of Education (Illiterate, Primary Secondary and Above secondary
8. Co-morbidities (No, Yes)	8) Co-morbidities (Yes, No)
9. HIV co-infection (Positive, Negative	9) HIVco-infection (Positive, Negative)
10. Clinical complication(No, Yes)	10) Smoking status (Yes, No)
11. Level of Education (Illiterate, Primary, Secondary and Above)	11) Registration group (New, Relapse, Failure treatment and Other)
12. Drug susceptibility test results	12) Alcohol use (Yes, No)
13. Therapeutic delay	
14. Smoking status (Yes, No)	
15. Marital status	
16. Registration group (New, Relapse, Failure treatment and Other)	
17. Alcohol use (Yes, No)	
18. Race/ethnicity	
19. Type of resistance	
20. Socioeconomic status	
21. Poor adherence	
22. Category of treatment	

3.7 Survival data analysis

Survival analysis is generally a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. In other words, survival analysis is an important statistical technique used to describe and model time–to–event data. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks and years. The purpose of survival analysis is to model the underlying distribution of the failure time variable and to assess the dependence of the failure time variable on covariates. Survival time then describes the time from a certain origin to the occurrence of an event. The method of survival analysis is used in other fields

1. Deaths in Biological Science: (Survival Analysis)
2. Mechanical Breakdown in Engineering: (Reliability Analysis)
3. Insurance Claim in Actuarial Science (Time to Event Analysis)
4. Events such as Divorce in Social Science: (Duration Analysis)

Several methods have been developed for the analysis of survival data. Some of these are:

1. Descriptive statistics which include life tables, survival distribution, and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from a sample.
2. Nonparametric tests are available for comparing the survival experience between two or more groups. The most common and widely used of these tests are the log-rank test, Generalized Wilcoxon test and Peto-Prentice test.
3. The multivariate Method uses Cox-proportional hazards model. It is considered as the most interesting survival modeling in the interest of examining the relationship between survival time and predictors.
4. Survival and one or more predictors. Covariates may be categorical or continuous. In addition the model has the capability of including both time-dependent and time independent variables.

3.8 Kaplan-Meier (KM) estimator

The Kaplan-Meier (KM) estimator is the standard nonparametric estimator of the survival function S(t), proposed by Kaplan and Meier (1958), is also called the Product-Limit estimator

KM estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included .It is extremely popular as requires only very week assumption and yet utilizes the information content of both fully observed and right censored data. The Kaplan-Meier estimator of the survivorship function (or survival probability) at time t, is defined as;

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} = \prod_{t_i \leq t} 1 - \frac{d_i}{n_i}$$

where $t_1, t_2, \ldots, t_n$ a set of survival time of n independent observations and $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(m)}$, $m \leq n$ be the m distinct ordered death times.

d_i is the number of individuals who failed (died) at time t_i

n_i is the number of individuals who are at risk of dying at time t_i, and The variance of the KM survival estimator which is also known as the Greenwood's formula is;

$$\text{var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum \frac{n_i}{n_i(n_i - d_i)}$$

3.9 Log rank test

A common problem in clinical studies is to compare two or more survivor functions. There are a few statistical tests for such a comparison. The log rank test is in fact a chi-squared test for a large sample. The log rank statistic compares the observed with an expected number of events. The expected number of events is calculated by the method assuming that the null hypothesis is true. The null hypothesis assumes that the compared curves are the same. The comparison is performed at every time point the observed event occurred.

H_0 : There is no significant difference between the survival curves..

$$\text{Log-Rank} = \frac{(O_i - E_i)^2}{var\,(O_i - E_i)} \sim \chi_{(1)}^2$$

Where O_i is observed and E_i is expected value in group i

3.10 Modeling Survival Data

In most medical studies which give rise to survival data, supplementary information referred to as covariates or independent variables needs to be collected on each individual, so that the relationship between survival experience of individuals and various explanatory variables have to be investigated. In order to explore the relationship between the survival experience of a patient and explanatory variables, an approach based on statistical modeling can be used. Through a modeling approach to the analysis of the survival data, we can explore how the survival experience of a group of patients depends on the values of one or more explanatory variables, whose values have been recorded for each patient at the time origin. In the analysis of survival data, interest centers on the risk or hazard of death at any time after the time origin of the study. As a consequence, the hazard function is modeled directly in survival analysis. The resulting models are somewhat different in form from linear models encountered regression analysis and in the analysis of data from designed experiments, where the dependence of the mean response. Or some function of it, on certain explanatory variables is modeled. The median survival time could then be estimated for current or future patients with particular values of these explanatory variables. The resulting estimate could be particularly useful in devising a treatment regimen, or in counseling the patient about their prognosis. A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches. One of the most popular types of regression models used in survival analysis is the proportional hazard model.

3.11 Model development

In any applied setting, performing a proportional hazard regression analysis of survival data requires a number of critical decisions. It is likely that we will have data on more covariates than we can reasonably expect to include in the model, so we must decide on a method to select a subset of the total number of covariates. When selecting a subset of the covariates, we must consider such issues as clinical importance and statistical significance.

3.12 Selection of covariates

The methods available to select a subset of covariates to include in a proportional hazards regression model are essentially the same as those used in any other regression model. There are

three methods of selection of influential covariates. These are purposeful selection, stepwise selection (forward selection and backward elimination) and best subset selection. Survival analysis using Cox regression method begins with a thorough univariate analysis of the association between survival time and all important covariates. Recommendable procedure in selecting variables is;

1. Include all variables that are significant in the univariable analysis at the 25 percent level.
2. The variables that appear to be important from step 1 are then fitted together in a multivariable model. In the presence of certain variables others may cease to be important. Consequently, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.
3. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method. This process may result in terms in the model determined at step 2 ceasing to be significant.
4. A final check is made to ensure that neither significant variable is eliminated from the model nor non-significant variable is included in the model. At this stage the interactions between Can of the main effects currently in the model can be considered for inclusion if the inclusion significantly modifies the model.

3.13 Cox Proportinal Hazard Regresssion Models

When we have several prognostic variables, we must use multivariate approaches. But we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations. We need another method to model survival data with the presence of censoring. One very popular model in survival data analysis is the Cox proportional hazards model. The Cox Proportional Hazard (PH) Model is a multiple regression method and is used to evaluate the effect of multiple covariates on the survival . The set of values of the explanatory variables in the PH model represented by vector x, so that x =(x_1,x_2,…,x_p).The Cox Proportional Hazards model is given by:

$$h(t,x) = h_o(t)exp(\hat{\beta}x) \tag{3.1}$$

Where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero;

$x = (x_1, x_2, ..., x_p)'$ is the values of the vector of explanatory variables for a particular individual, a $\beta' = (\beta_1, \beta_2, ..., \beta_p)$ is a vector of regression coefficients.

$e^{\beta' x_i}$ characterizes how the hazard function changes as a function of subject covariates is called the linear component of the model, also known as the risk score or prognostic index for the i[th] individual. The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients. The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates x and x^* is

$$\hat{HR} = \frac{h_0(t)\exp\left(\hat{\beta}' x\right)}{h_0(t)\exp\left(\hat{\beta}' x\right)} = \exp\left(\sum \beta' \left(\hat{x} - x^*\right)\right)$$

This hazard ratio is time-independent, that is why this is called the proportional hazards

3.14 Checking Assumption of Cox Proportional Hazard Model

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

3.14.1 Graphical method

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

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

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

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

This does not depend on t. By plotting estimated log (-log (survival)) versus survival time for two groups we would see parallel curves if the hazards are proportional. This method does not work well for continuous predictors or categorical predictors that have many levels because the graph becomes "cluttered". Furthermore, the curves are sparse when there are few time points and it may be difficult to tell how close to parallel is close enough. We will show some other statistical methods for checking the proportionality.

3.14.2 Adding time-dependent Covariates in the Cox model

We create time-dependent covariates by creating interactions of the predictors and a function of survival time and including them in the model. The model assessing PH assumption for x_j adjusted for other covariates is

$$h\left(t, x\left(t\right)\right) = h_0\left(t\right)\exp\left[\beta_1 x_1 + \beta_2 x_2 + ... + \beta_j x_j + ... + \beta_p x_p + \delta\, \mathrm{x}_j \times g\left(t\right)\right], \qquad (3.2)$$

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

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

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

3.14.3 Tests based on the Schoenfeld residuals

The other statistical test of the proportional hazards assumption is based on the Schoenfeld residual .The Schoenfeld residuals are defined for each subject who is observed to fail. If the PH assumption holds for a particular covariate then the Schoenfeld residual for that covariate will

not be related to survival time. Rejection of null hypothesis concludes that PH assumption is violated.

3.14.4 Cox Proportional Hazards model diagnostics

A number of residuals have been proposed for use in connection with the Cox PH model. For this study, three major residuals in the Cox model will be used: the Cox-Snell residual, the deviance residual, and the Schoenfeld residual. Then we will talk about influence assessment.

3.14.5 Cox-Snell residuals and Deviance residuals

The Cox-Snell residual is given by Cox and Snell .The Cox-Snell residual for the i^{th} individual with observed survival time t_i is defined as;

$$\gamma c_i = \exp\left[\hat{\beta} X_i\right] \quad \hat{H}_i(t_i) = \hat{H}_i(t_i) = -\log\left(\hat{S}_i(t_i)\right)$$

Where $\hat{H}_i(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i .

Thus, regardless of the distribution of T, the new variable $y = H(t)$ has an exponential distribution with unit mean. If the model will be well fitted, the value $\hat{S}_i(t_i)$ would have similar properties to those $S_i(t_i)$. So $\gamma c_i = -\log\left(\hat{S}_i(t_i)\right)$ will have a unit exponential distribution with $f_r(\gamma) = \exp(-\gamma)$. Let $S_R(\gamma)$ denote the survival function of Cox-Snell residual γc_i. Then

$$S_R(\gamma) = \int_{\gamma}^{\infty} f_R(x)dx = \int_{\gamma}^{\infty} \exp(-x)dx = \exp(-\gamma).$$

and

$$H_R(\gamma) = -\log(S_R(\gamma)) = -\log(\exp(-\gamma)) = \gamma$$

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

The deviance residual is defined by:

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

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

3.14.6 Schoenfeld residuals

All the above three residuals are residuals for each individual. We will describe covariate wise residuals.The Schoenfeld residuals were originally called partial residuals because the Schoenfeld residuals for i^{th} individual on the j^{th} explanatory variable x_{ij} is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j : logarithm of the partial likelihood function is given by

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

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

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

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

3.15 Extension of Cox Proportional Models

Cox regression model is applicable only to time- invariant predictors with time-constant effects only. We can extend a linear regression model in a variety of ways, so, too, can we extend the Cox regression model. Suppose that statistic tests or other diagnostic techniques give strong evidence of non-proportionality for one or more covariates.

Another way to consider is to use a different model. A parametric model such as an AFT model might be more appropriate for the data.

3.15.1 Stratified Cox Model

We should split the whole sample into subgroups on the basis of categorical variable (stratification variable) and re-estimate the model. Then we let the baseline hazard function differ between these subgroups. It makes sense to choose covariate if it interacts with time (i.e.

proportional hazard assumption is not satisfied for this covariate) .One method that we can use is the stratified Cox model, which stratifies on the predictors not satisfying the PH assumption. The data are stratified into subgroups and the model is applied for each stratum. The model is given by ;

$$h_g(t) = h_{o_g}(t)\exp(\hat{\beta} z_{ig}) \quad (3.3)$$

Where g represents the stratum. Note that the hazards are non-proportional because the baseline hazards may be different between strata. The coefficients β are assumed to be the same for each stratum g. The partial likelihood function is simply the product of the partial likelihoods in each stratum. A drawback of this approach is that we cannot identify the effect of this stratified predictor. This technique is most useful when the covariate with non-proportionality is categorical and not of direct interest.

3.15.2 Cox Regression Model with time-dependent variables

Until now, we have assumed that the values of all covariates did not change over the period of observation. However, the values of covariates may change over time t. Such a covariate is called a time-dependent covariate. The second method to consider is to model non-proportionality by time-dependent covariates. The violation of PH assumptions is equivalent to interactions between covariates and time. That is, the PH model assumes that the effect of each covariate k2is the same at all points in time. If the effect of a variable varies with time, the PH assumption is violated for that variable.

$$h_i\big(t, x_{in}(t)\big) = h_o(t)exp\big(\beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \beta_j x_{jn}\big) \quad (3.4)$$

Where $h_i(t)$ is the hazard function for individual *i* at time *t*.

$\beta_1 x_{i1} + \cdots + \beta_k x_{ik}$ are still time invariant co-variate

$\beta_j x_{jn}$ Time dependent co-variates.

$h_0(t)$ is the baseline hazard

3.16 Accelerated Failure Time (AFT) Models

AFT is an alternative to the PH model for the analysis of survival time data. Under AFT models, we measure the direct effect of the explanatory variables on the survival time instead of hazard, as we do in the PH model. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. Similar to the PH model, the AFT model describes the relationship between survival

probabilities and a set of covariates. Log-linear form of the AFT model with respect to time is given by:

$$\log T_i = \mu + \alpha_1 x_{1i} + \alpha_2 x_{2i} + ... + \alpha_p x_{pi} + \delta\varepsilon_i \quad (3.5)$$

Where μ the intercept, δ is scale parameter and ε_i is a random variable, assumed to have a particular distribution. For each distribution of ε_i, there is a corresponding distribution for T. The members of the AFT model class include the exponential AFT model, Weibull AFT model, log-logistic AFT model, log-normal AFT model, and gamma AFT model .The AFT models are named for the distribution of T rather than the distribution of ε_i or log T.

Distribution of ε	*Distribution of T*
Extreme value (1 parameter)	*Exponential*
Extreme value (2 parameters)	*Weibull*
Logistic	*Log-logistic*
Normal	*Log-normal*
Log-Gamma	*Gamma*

The survival function of T_i can be expressed by the survival function of ε_i

$$S_i(t) = S_{\epsilon i}\left(\frac{\log t - \mu - \alpha x}{\delta}\right)$$

The effect size for the AFT model is the time k2ratio. The time ratio comparing two levels of covariate $x_i (x_i = 1 \, v_s \; x_i = 0)$; after controlling all the other covariates is $\exp(\alpha_i)$ which is interpreted as the estimated ratio of the expected survival times for two groups.

3.16.1 Estimation of AFT model

AFT models were fitted using the maximum likelihood method. The likelihood of the n observed survival times $t_1, t_2, ..., t_n$ is given by

$$L(\alpha, \mu, \sigma) = \prod_{i=1}^{n} \{ f_i(t_i) \}^{\delta_i} \{ S_i(t_i) \}^{1-\delta_i}$$

Where $f_i(t_i)$ and $S_i(t_i)$ are the density and survival functions for the i^{th} individual at t_i and δ_i is the event indicator for the i^{th} observation.

3.16.2 Model checking for AFT models

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

3.16.3 Using Statistical Criteria

We can use statistical tests or statistical criteria to compare all these AFT models. Nested models can be compared using the likelihood ratio test. The exponential model, the Weibull model and lognormal model are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) was used instead, which is defined as;

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

Where l is specific ancillary parameters the log-likelihood, k is the number of covariates in the model and c is the number of model-specific ancillary parameters. Lower values of the AIC suggest a better model. But there is a difficulty in using the AIC in that there are no formal statistical tests to compare different AIC values. When two models have very similar AIC values, the choice of model may be hard and external model checking or previous results may be required to judge the relative plausibility of the models rather than relying on AIC values alone.

3.16.4 Using Residual Plots

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

$$\gamma_{c_i} = \hat{H}(t_i / x_i) = -\log\left[\hat{S}(t_i / x_i)\right],$$

Where t_i is the observed survival time for individual i, x_i is the vector of covariate values for individual i, and $\hat{S}(t_i)$ is the estimated survival function on the fitted model. The estimated survival function for the i^{th} individual is given by:

$$\hat{S}_i(t) = S_{\varepsilon_i}\left(\frac{\log t - \mu - \alpha\, x_i}{\hat{\sigma}}\right)$$

Where $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\sigma}$ are the maximum likelihood estimator of μ, α and σ respectively, $S_{\varepsilon_i}(\varepsilon)$ is the survival function of ε_i in the AFT model, and $\frac{\log t - \hat{\mu} - \hat{\alpha} x_i}{\sigma} = \gamma_{\varepsilon_i}$ is referred to as standardized residual. The Cox-Snell residual can be applied to any parametric model. The corresponding form of residual based particular AFT model can be obtained. For example, under the Weibull AFT model, since $S_{\varepsilon_i}(\varepsilon) = \exp\left(-e^{\varepsilon}\right)$, the Cox-Snell residual is then

$$\gamma_{\varepsilon_i} = -\log\left\{\hat{S}(t_i)\right\} = -\log S_{\varepsilon_i}\left(\gamma_{s_i}\right) = \exp\left(\gamma_{s_i}\right)$$

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

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

If the fitted model is appropriate, the plot of $\log\left(-\log S(\gamma_{c_i})\right)$ versus γ_{c_i} is a straight line with unit slope through the origin. These residuals lead to the deviance residuals for the particular AFT model. A plot of deviance residuals against the survival time or explanatory variables will

be used to check whether there are particular times, or particular values of explanatory variables, for which the model is not a good fit.

3.17 Bayesian Survival Analysis

Bayesian analysis offers a way of dealing with information conceptually different from all other statistical methods. It provides a method in which observations are used to update estimates of the unknown parameters of a statistical model. The Bayesian method is based on specifying a probability model for the observed data X, given a vector of unknown parameters θ, leading to the likelihood function L(θ/X).

3.17.1 Prior Distribution

The prior distribution $\pi(\theta)$ expresses our uncertainty about θ before seeing the data. Bayesian probability measures the degree of belief that you have in a random event. The prior distribution is a probability distribution that represents the prior information associated with the parameter of interest. It is a key aspect of a Bayesian analysis. There are two types of prior distribution, informative priors and non-informative priors.

An informative for θ prior is a prior distribution that is used when information about the parameter of interest is available before the data is collected, and this information is to be included in the analysis. Typically, informative prior distributions are created from historical studies, pure expert knowledge (experience) and a combination of both. Even if there is prior knowledge about what we are examining, in some cases we might prefer not to use this and let the data speak for themselves.

A non-informative prior distribution that is used to express complete ignorance of the value of before the data is collected. They are non-informative in the sense that no value is favored over any other and are also described as diffuse or flat at prior due to this reason and their shape. The most common non-informative prior is the uniform distribution over the range of the sample space for θ.

3.17.2 Likelihood Function

A likelihood functions is a function that gives the probability of observing of the sample data given the current parameters. Suppose we observe n independent vectors of (T_i, δ_i, x_i), where T_i is time to the event and δ_i is indicator variable telling us whether (T_i is uncensored or

censored.

$$\delta_i = \begin{cases} 0 & censored\ obsevation \\ 1 & uncensored\ observation \end{cases}$$

The likelihood function of the set of unknown parameters, θ in the presence of right censoring can be written as;

$$L(\theta) = \prod_{i=1}^{n} [f(t_i/x_i, \theta)^{I(\delta_i=0)} * S(t_i/x_i, \theta)^{I(\delta_i=1)}]$$

Log –likelihood would be as follows:

$$l(\theta) = log\{\prod_{i=1}^{n} [f(t_i/x_i, \theta)^{I(\delta_i=0)} * S(t_i/x_i, \theta)^{I(\delta_i=1)}]\}$$

$$l(\theta) = \sum_{i=1}^{n} [log(f(t_i/x_i, \theta)^{I(\delta_i=0)}) + log(S(t_i/x_i, \theta)^{I(\delta_i=1)})]$$

Where f($t_i/x_i, \theta$) and $S(t_i/x_i, \theta)$ are the density and survival distribution ,respectively
In these models, when both of β and σ^2 are unknown, no joint prior is available.

3.17.3 Posterior Distribution

The posterior distribution $\pi(\theta/X)$, expresses our uncertainty about θ after seeing the data. The posterior distribution is obtained by multiplying the prior distribution over all parameters, θ by the full likelihood functions, L (θ/X). All Bayesian inferential conclusions are based on the Posterior distribution of the model generated. The inference is performed by sampling from Posterior distribution until the convergence to the posterior distribution is achieved [48].
The major problem in the Bayesian approach is that in most cases the full form of the posterior distribution cannot be obtained in closed form, that is, the posterior density may not belong to standard distribution. Such problem cannot be solved easily. In order to solve such problems we will use MCMC simulations. Then we assume that θ is a random variable and has a prior distribution dented by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution, which is obtained by Bayes' theorem. Then posterior distribution of θ is given by;

$$\pi(\theta/X) = \frac{L(X/\theta)\pi(\theta)}{\int L(X/\theta)\pi(\theta)d\theta}$$

Combining the likelihood function with the prior distribution on (β, σ^2) and the full conditional distributions for unknown parameters, the posterior distribution can be written as:

$$\pi(\beta/\sigma^2, t, x) \propto \prod_{i=1}^{n} [f(t_i/x_i, \theta)^{I(\delta_i=0)} * S(t_i/x_i, \theta)^{I(\delta_i=1)}] * \pi(\beta/\sigma^2)$$

$$\pi(\sigma^2/\beta, t, x) \propto \prod_{i=1}^{n} [f(t_i/x_i, \theta)^{I(\delta_i=0)} * S(t_i/x_i, \theta)^{I(\delta_i=1)}] * \pi(\beta/\sigma^2) * \pi(\sigma^2)$$

The posterior distribution for the model specification above does not have closed form solution for the parameters. For these models, MCMC-Gibbs sampler is implemented using the Win BUGS package

3.17.4 Markov Chain Monte Carlo

Markov Chain: a stochastic process that generates conditional dependent samples according to some target distribution.

Monte Carlo: a numerical integration technique that finds an expectation:

MCMC is a method that generates a sequence of dependent samples from the target distribution and computes quantities by using Monte Carlo based on these samples. MCMC techniques generate a Markov chain that ultimately provides a sample from the posterior distribution and that the summary measures calculated from this chain consistently estimate the corresponding true posterior summary measures.

3.17.5 MCMC Estimation methods

The Bayesian approach applies probability theory to a model derived from substantive knowledge and deal with realistically complex situations; the approach can also be termed 'full probability modeling'. The MCMC simulation is to do the integration numerically rather than analytically by sampling from the posterior distribution of interest even when the form of that posterior has no known algebraicform [49]. This will yield all posterior summary statistics (approximately).

Gibbs Sampler

Gibbs sampler is applicable in general when the joint parameter is not known explicitly but the conditional distribution of each parameter given the other is known .let p(θ) =p($\theta_1, \theta_2, \theta_3, \dots, \theta_k$) denote the joint parameter and let p($\theta_i|\theta_{-i}$) denote the conditional density for the i^{th} component θ_i given the other k-1 components.

Gibbs sampler algorithm begin by picking the arbitrary starting point value

$\theta^0 = \{ {\theta_1}^{(0)}, {\theta_2}^{(0)}, \dots, {\theta_K}^{(0)}$. it then sample randomly from the conditional densities p($\theta_i|\theta_{-i}$) for i=1,…,k successively as follows;

I. For i=0, 1,2,…,K-1 generate each component of θ as follows
   - Draw ${\theta_1}^{(i+1)}$ $from$ $P({\theta_1}/{\theta_2}^i, {\theta_3}^i, \dots, {\theta_k}^i)$
   - Draw ${\theta_2}^{(i+1)}$ $from$ $P({\theta_2}/{\theta_1}^i, {\theta_3}^i, \dots, {\theta_k}^i)$

- Draw ${\theta_3}^{(i+1)}\ from\ \ P({\theta_3}/{\theta_1}^{i}, {\theta_2}^{i}, \dots, {\theta_k}^{i})$
- Draw ${\theta_K}^{(i+1)}\ from\ \ P({\theta_K}/{\theta_1}^{i}, {\theta_2}^{i}, {\theta_3}^{i}, \dots, {\theta_{k-1}}^{i})$

II. Repeat step 2 until convergence

III. Return $\theta^{b+1} = \left({\theta_1}^{(b+1)}, {\theta_2}^{(b+1)}, \dots, {\theta_K}^{(b+1)} \right), \theta^{(b+2)}, \dots, \theta^{(N)}$

The means of the posterior samples provide point estimates for the model parameters, while the standard deviations provide measures of precision. The 95% intervals (calculated using the 2.5th and 97.5th percentiles of the posterior samples) provide an alternative indication of the covariates effects along with estimation precision. The MC error is an estimate of the difference between the mean of the sampled values (which we are using as our estimate of the posterior mean for each parameter) and the true posterior mean. As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation.

3.15.6 Model Diagnostic

Once a model has been developed, we now would like to know how effective the model is in describing the outcome variable. This is referred to as goodness of fit. The most common ways of checking goodness of fit are: diagnosis for convergence and mixing, and posterior-predictive check. Diagnosis of the convergence is important to answer the questions of how to determine whether the sampler has reached its stationary distribution. The Markov chain must be started somewhere, and initial values are selected for the unknown parameters. In theory the choice of initial values will have no influence on the eventual samples from the Markov chain, but in practice convergence will be improved and numerical problems avoided if reasonable initial values can be chosen [49].

Convergence diagnostics are widely used to determine how many initial "burn-in" iterations should be discarded from the output of a MCMC sampler in the hope that the remaining samples are representative of the target distribution of interest. The best method is choosing the number of burn-in iterations *r* by applying convergence diagnostics to one or more pilot chains, and then basing estimation and inference on a separate long chain from which the first r iterations have been discarded. To use summary statistics of the estimated posterior distributions for inference the realized value of the parameters (the MCMC value) should converge. To check this we have to use suitable diagnosis to evaluate mixing and convergence of a sampler. From different

methods of checking convergence Gelman and Rubin (BGR) diagnostic, trace and history plots, kernel density plot and autocorrelation are among the common.

The GR diagnostic compares the within-chain and the between-chain variability, and if the ratio (converges approximately to one or if lines for each chain on the GR are nearly together, this implies that the statistics converge. A trace or time-series plot of the values in the chain or of values derived from them show that if the chain is drifting, perhaps indicating that the burn-in was not long enough, and it will illustrate the speed of mixing, which is how quickly the chain moves across the distribution. Chains that mix slowly will produce long, slow cycles, and they take longer to converge. Mixing can sometimes be improved by re-parameterizing the model.

Once we confirmed that convergence has been achieved, we will need to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference. The more samples you save, the more accurate will be your posterior estimates. One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Baseline Characteristics

Table 1: Demographic and health factors of categorical covariate by MDR-TB in Saint Peter's Specialized TB hospital from January 3, 2014 to October 28, 2016

Covariate	Categories	No of patient	No of death (%)	Noof censored (%)	Median of survival time in day
Sex	Male	155(48.4)	37(55.2)	118(46.6)	596
	Female	165(51.6)	30(44.8)	153(53.4)	598
TB status	Yes	288(90)	52(77.6)	236(93.3)	600
	No	32(10)	15(24.4)	17(6.7)	505.5
HIV status	Positive	90(28.1)	33(49.3)	57(22.5)	599
	Negative	230(71.9)	34(50.7)	197(77.5)	597
Co-morbidity	Yes	197(61.6)	56(83.6)	141(55.7)	597
	No	123(38.4)	11(16.4)	112(44.3)	599
Smoking status	Yes	42(13.1)	20(29.85)	22(13.1)	592
	No	278(86.9)	47(70.15)	231(86.9)	598
Alcohol use	Yes	38(11.9)	14(11.9)	24(9.5)	597.5
	No	282(88.1)	53(88.1)	229(90.5)	597
TB type	Pulmonary	279(87.2)	65(97)	214(84.6)	597
	Extra pulmonary	41(12.8)	2(3)	39(15.4)	605
Education level	Illiterate	49(15.3)	12(17.9)	37(14.6)	463
	Primary	128(40)	26(38.8)	102(40.3)	597
	Secondary	77(24.1)	14(20.9)	63(24.9)	601
	Above secondary	66(20.6)	15(22.4)	51(20.2)	600
Region	Addis Ababa	128(71.3)	46(68.4)	182(71.9)	599
	Oromiya	49(15.3)	13(19.4)	36(14.2)	596
	Amhara	10(3.1)	3(4.5)	7(2.8)	594.5

	SNNP	14(4.4)	2(3)	12(4.7)	601.5
	Other	19(5.9)	3(4.5)	16(6.3)	388
Registration Group	New	36(11.25)	17(25.4)	19(7.5)	505.5
	Relapse	103(32.2)	19(28.4)	84(33.2)	598
	Failure of new regime	119(37.2)	22(32.8)	97(38.3)	601
	Other	62(19.4)	9(13.4)	53(20.9)	598.5
Total		320	67(20.94)	253(79.06)	

Source: Saint Peter's Specialized TB hospital, Ethiopia; from Jan 3, 2014 to October 28, 2016

Table 2: Demographic and health factors of Continuous covariate by MDR-TB in Saint Peter's Specialized TB hospital from January 3, 2014 to October 28, 2016.

Patient status	Continuous covariates	Mean	Median	Standard deviation	Minimum	Maximum
Death	Age	40.19	38	15.3	2	75
	BMI	17.07	16.9	2.72	11.17	23.1
	Time	155.64	63	190.56	1	612
Censored	Age	30.48	29	10.73	6	65
	BMI	17.62	17.4	3.18	10.3	28.3
	Time	545.89	604	197.60	28	923
Over all	Age	32.52	30	12.45	2	75
	BMI	17.51	17.3	3.1	10.3	28.70
	Time	464.18	597	252.29	1	969

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016

The baseline characteristics were presented in the Table 1 above. In this study, a sample of 320 MDR-TB patients was considered. The medical cards of those patients were reviewed. Of these 155 (48.4 %) were males and 165(51.6%) were females and the median of survival time 596 and 598 respectively. The Region of the MDR-TB patients were, 128(71.3%) from Addis Ababa, 49(15.3%) from Oromiya, 10(3.1%) from Amhara, 14(4.4%) from SNNP and 19(5.9%) from other region .Median of survival time were 599, 596, 594.5, 601.5 and 3388 days respectively. There were 67(20.94%) known deaths; out of this death 33(49.3%) in HIV positive and

34(50.7%)) in the HIV negative group. The median survival time for HIV positive was 599 days and 597 days for HIV negative MDR-TB patients.

Similar finding conducted at St. Peter's specialized TB Hospital showed that from total study units, 87 (46.28%) of them are male and the rest 101 (53.72%) are female. Among the total subjects, there were 29 (15.43 %) known deaths, of them 30.3 % (10/33) in HIV positive and 12.3 % (19/155) in the HIV negative group occurred. The majority of MDR-TB patients were younger aged less than 35 years (81.38 %) with median age of 27 years living in Addis Ababa (75.27 %), Oromia (11.29 %), Tigray (5.38 %). All patients were tested for HIV infection as testing is imposed by national policies and 33 (17.55 %) were positive. Only 7 (3.72 %) were not treated for anti-TB drugs the rest 181 (96.28 %) were treated previously for anti-TB drugs [43]. Of the total study units, 195 (57 %) were males and the rest 147 (43 %) were females. There were 37 (10.8 %) known deaths; 12.7% (9/71) in HIV positive and 10.3 % (28/271) in the HIV negative group occurred [45].

Among those four registrations group 36(11.25%) were new, 103(32.2%) were relapse, 119(37.2%) were failure of new regime and 62(19.4%) were other registration group and the median survival time were 505.5,598,601 and 589.5days respectively. Out of the patients 197 (61.6%) were have no co-morbidity and 123(38.4%) were have comorbidity. The proportion of the patients who were died among this groups were 56(83.6%) and 11(16.4%) respectively. The median of the survival time for a patient with comorbidity and with no comorbidity were 597 and 599 day respectively. Out of the entire subject integrated in this study; 288(90%) have been previously treated for TB and 32(10%) had no history of previous TB treatment and median of survival time 600 and 505.5 respectively. Similar study conducted at Saint Peter's Specialized TB show that 7 (3.72 %) were not treated for anti-TB drugs the rest 181 (96.28 %) were treated previously for anti-TB drugs. In addition to resistance to isoniazid and rifampicin 65.96 % were resistant to other first line [43] and nationwide survey conducted in China; the estimated MDR-TB rate was 5.7% for new cases and 25.6% for previously treated cases [22].

Out of the total MDRTB patients included in this study 42(13.1%) and 278(86.9%) were smokers and nonsmokers respectively. The proportion of patients who were died among this 20(29.85%) were smoker and 47(70.15%) were nonsmokers. The median of survival time for a patient with smoking status and with no smoking status were 592 and 598 day respectively.

similar study conducted reveal that out of the total MDRTB patients included in this study10(5.32%) and 178(94.68%) were smokers and nonsmokers respectively[43].

Out of the entire subjects integrated in this study, 14(11.9%) of the patients were Alcohol users whereas 53(88.1%) were Non-Alcohol users. The death proportion was higher for those Non-Alcohol users 53(88.1%) while lower for those Alcohol user patients 14 (11.9%).The survival times were higher for those Non-Alcohol user 997.5 day while lower for those Alcohol user patients 997 day. The sample data also revealed that 279(87.2%) patients were Pulmonary TB type and 41(12.8%) were extra pulmonary TB type. Out of total death in this study 65(97%) were Pulmonary TB type patients and whereas 2(3%) extra pulmonary TB type patients.

Regarding to Educational level 49(15.3%) Illiterate, 128(40%) had primary school, 77(42.1%) had secondary school and 66(20.6%) had above secondary school. From this, the death proportion were highest for those patients who had Primary which is 26(38.8%), followed by those who had above secondary education which was 15(22.4%), while the lowest proportion of death 14(20.9%) and 12(17.9%) were patients who had secondary school and illiterate respectively. The median of survival time for patient illiterate, primary, secondary and above secondary school were 463day, 597day, 601day and 600 day respectively.

From a sample of 320 MDR-TB patients was considered in Saint Peters specialized TB Hospital form July 3,2014 to October 28,2016 253(79.06%) were censored and 67(20.94%) were died.
A total of 400 patients with MDR-TB were treated in ALERT and Gondar hospital during the study period from August, 2011 to September, 2014. However, the study included 342 MDR-TB patients for whom data for variables of interest were complete. Of these 89.2% were censored and 10.8% were died (uncensored)[45].

From the Table 2 above : The overall median estimated survival time, age and body mass index patients under the study were 596 days, 30 years and 17.3kg/m^2 respectively. The minimum follow-up time was 1day and the maximum was 969 days as well as the mean of follow up time was 464.18 days and standard deviation was 252.29 days. The mean of survival time for the status death was 155.64 days and 545.89 days for the status censored. Similar study conducted at Gonder and ALERT Hospital reveal the overall median estimated survival time and age patients under the study were 8 month , 22 years respectively. The minimum follow-up time was 1months and 42 months [45].

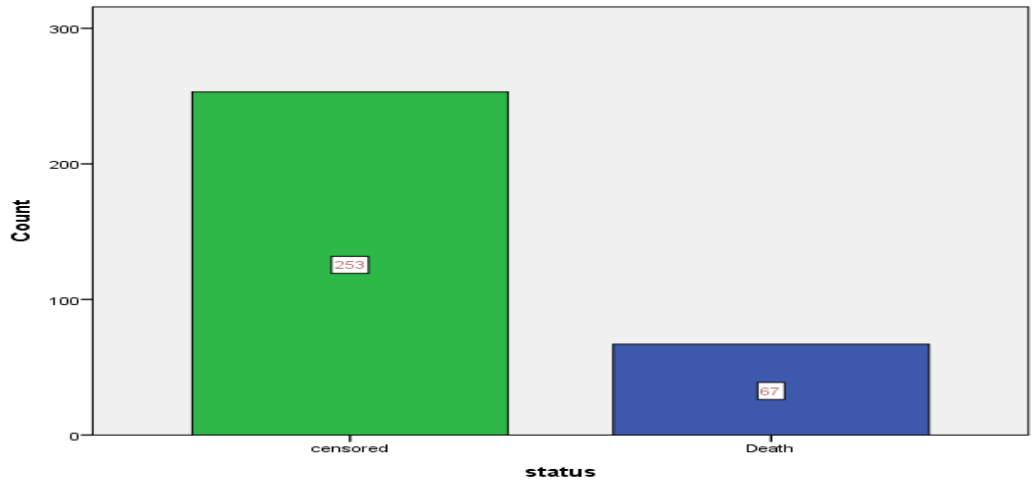


Figure 1: Bar graph for status variable (Death and Censored)

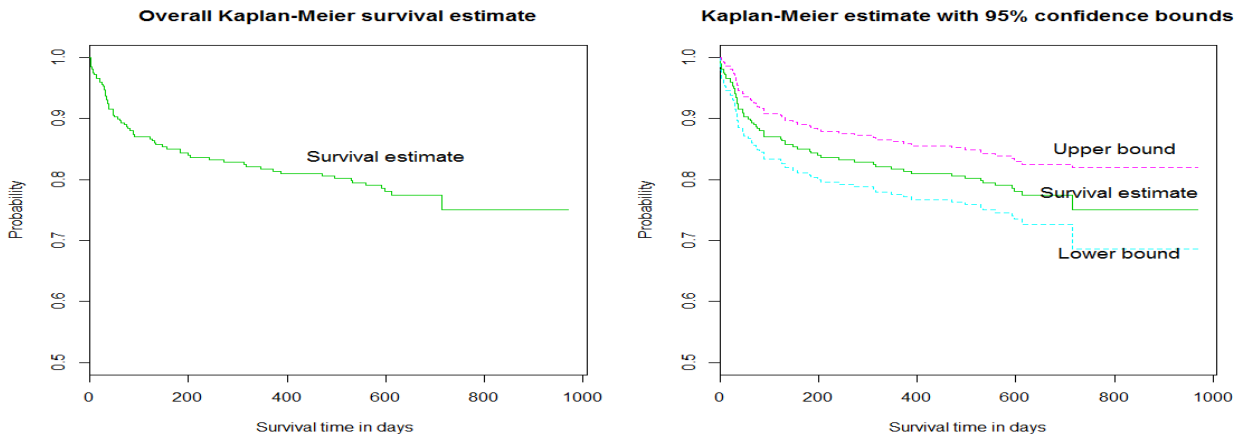


Figure 2: The plot of the overall Kaplan-Meier survival estimate and 95% confidence bound function of MDR-TB patients in St.Peter's Specialized Hospital, Ethiopia.

The patients were followed up for a median period of 597days. The minimum follow-up time was 1 day and the maximum was 969 days. From figure 2 the overall Kaplan- Meier survival estimate decline as the survival time increase. The following graph of the estimate of overall Kaplan-Meier survivor function reveals that most of the deaths occurred in the earlier day of MDR-TB treatment. The 95% confidence bounds for overall survival estimate was narrow at the beginning survival time and became wider until end of the study.

4.2 Comparison of Survival Experience

Log-rank and Peto to test the significance difference in survival experience among different categories.

Table 3: Comparison of Survival Experience of MDR-TB Patients Using Log-rank and Peto test (at St. Peter's specialized hospital January 3, 2014 to October 28, 2016)

Categorical Covariate	Degree freedom	Log-rank test		Peto test	
		Chi-square	p-value	Chi-square	p-value
Sex	1	1.4	0.245	1.3	0.245
TB type	1	5.3	0.022	5.2	0.023
TB status	1	18.4	0.000	18.9	0.000
Smoking status	1	19.2	0.000	17.6	0.000
Alcohol use	1	5	0.025	4.1	0.042
HIV status	1	16.8	0.000	16.1	0.000
Co-morbidity	1	15.5	0.000	14.8	0.000
Education level	3	1.7	0.630	1.9	0.585
Registration group	3	22.1	0.000	22.6	0.000
Region	4	2.2	0.698	2.4	0.663

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016

The Kaplan-Meier estimator survival curve used to estimate survivor function among different covariates so that one can make comparison. The survivorship pattern of one is lying above another means the group defined by the upper curve has a better survival time than the group defined by the lower curve.

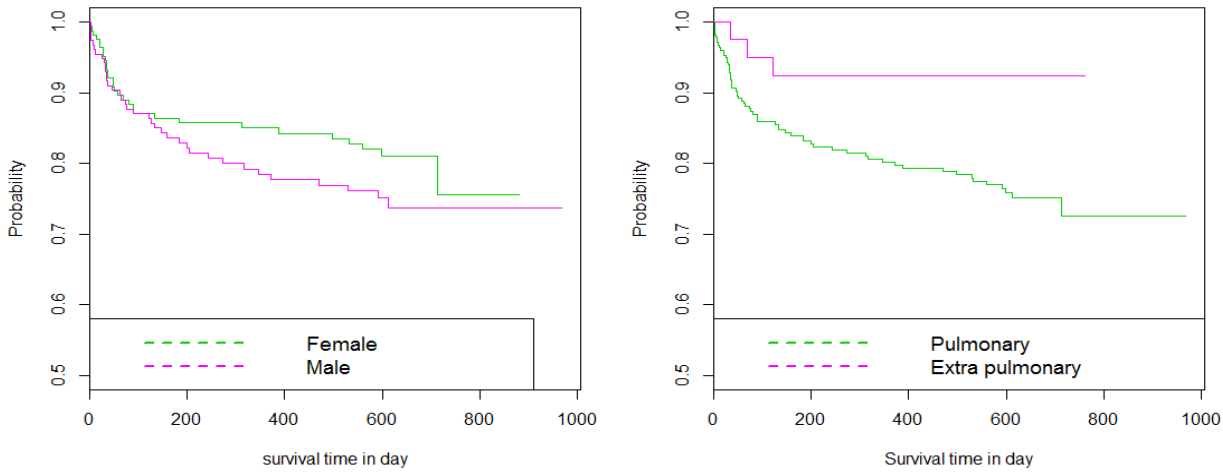


Figure 3: Kaplan-Meier survivor estimates for Sex and TB site

The Log-rank and Peto test results show that Tuberculosis type (TB site), tuberculosis status (previous history of TB), Smoking status, Alcohol use, Co-morbidity and registration group were significant difference between the survival experience at 5% level of significance, whose different levels have an impact in the survival time of MDR TB patients; whereas Sex, Education level and Region did not significant difference between the survival experience at 5% level of significance. The study conducted Gonder and ALERT Hospital reveal that, there were statistically no significant differences in survival/death experience for sex, marital status, educational level and HIV [45].

From figure 3: The survival time for Sex variable is not clear difference between the categories of male and female, it mean that at some point of time one categories better survival time than other, have the same survival time and at some point lower survival time than other categories and but clear difference between the categories for covariate TB site .Extra pulmonary TB site patient have better survival time than .pulmonary TB site.

On the (Appendix B); the Kaplan-Meier survival curve revealed that patients who are not use Alcohol and non-smoker had high survival time as compared to the other groups which did use Alcohol and Smoke cigarette patients. Similarly, had previous TB status, had no co-morbidity and negative HIV co-infection had high survival time as compared, had no previous TB status, and had comorbidity and positive HIV co-infection respectively. Again From appendix B; The

survival time for Education level and region were not clear difference between the categories, it mean that at some point of time one categories better survival time than other, have the same survival time and at some point lower survival time than other categories. General the long rank and Peto test and Kaplan-Meier survivor estimates reveal the same conclusion.

4.3 Single Covariate Cox Regression Analysis

Single covariate cox proportional hazards model analysis is an appropriate procedure that is used to screen out potentially important variables before directly included in the multivariate model. The factors which are significant at 25% level of significance in univariate cox proportional hazard analysis were included in multivariate cox proportional hazard analysis.

The relationship between each covariates and survival time of MDR-TB patients are presented in (Appendix A, table 1) from this table, survival time of the MDR-TB patients was significantly related with covariate Age, Body mass index, TB status, HIV co-infection, TB site, Smoking Status, Alcohol use, Co-morbidity and Registration group at 25% significance levels. The confidence interval of the hazard ratio for all covariates does not include 1 except covariate Body mass index and alcohol use in univariate models, at 0.25 level of significance.

However; Sex, Educational level and Region of patient were not a significant factor for the death of MDR-TB patient at 25% level of significance.

Table 4: Comparison of Univarate and Multivariable Cox Proportional Hazard Model.

Covariate	Categories	Univarite model		Multivariable model	
		(95% CI)	p-value	(95% CI)	p-value
Age		1.053[1.034,1.071]	0.000	1.044[1.025,1.063]	0.000**
BMI		0.952[0.881,1.029]	0.215	0.915[0.841,0.996]	0.040*
TB type	Pulmonary(ref)				
	Extra pulmonary	0.281[0.088,0.896]	0.032	0.455[0.140 1.478]	0.190
TB status	No(ref)				
	Yes	0.303[0.170,0.540]	0.000	0.510[0.110,2.372]	0.391
HIVco-	Negative (ref)				
infection	Positive	2.624[1.625,4.237]	0.000	1.727[1.044, 2.858]	0.024*
Smoking	No(ref)				

	Yes	3.040[1.8 ,5.132]	0.000	2.807[1.299 ,6.064]	0.009*
Alcohol	No (ref)				
	Yes	1.939[1.076,3.496]	0.028	0.878[0.311 ,1.690]	0.457
Comorbidity	No (ref)				
	Yes	3.390[1.776,6.472]	0.000	2.130[1.088 ,4.173]	0.027*
Registration	New (ref)				
group	Relapse	0.316[0.164,0.609]	0.001	0.626[0.136,2.883]	0.548
	Failure new regime	0.318[0.169,0.601]	0.000	0.767[0.169,3.474]	0.730
	Other	0.239[0.106,0.536]	0.001	0.450[0.090,2.340]	0.349

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
**P-value<.005 was statistically significant for multivariate model.*
Ref=Reference, CI=Confidence interval

To further optimize the Cox model, the variable with the highest P-value and over threshold of significance are removed from the predictive model one by one until all the rest variables were shown significant impact on the prediction of hazard rate. From Table 4 the variable Registration group was the variables with highest p-value, so it is removed first. The result is shown in (Appendix A).

4.4 Multivariable Cox Proportional Hazard Regression Analysis

Multivariable Cox PH including all the potential risk factors that had a P-value of less than or equal 0.25 in single covariate Cox PH analysis. To select the best subgroup of variables in our model, the approach of stepwise was applied as seen in (Appendix A).In order to decide whether or not a variable is significant, the p-value associated with each parameter has been estimated and variables that have p-value less than 0.05 cut point or 5% significance level were considered as important variables to predict survival time of MDR-TB patients.

Table 5: Final Multivariable Cox Proportional Hazard models on time to death of MDR-TB Patient at St.Peter's specialized hospital from January 3, 2014 to October 28, 2016

Covariate	β	HR	SD	Z-value	P-value	95% CL of HR
Age	0.044	1.045	0.009	4.804	1.56e-06 ***	[1.027 , 1.064]
BMI	-0.095	0.909	0.043	-2.237	0.025318 *	[0.837 , 0.988]
HIVco-infection	0.526	1.692	0.255	2.065	0.038892 *	[1.027 , 2.786]
Smoking	0.853	2.347	0.275	3.103	0.001916 **	[1.369 , 4.022]
comorbidity	0.822	2.276	0.339	2.427	0.015230 *	[1.171 , 4.423]
TB status	-1.132	0.322	0.303	-3.736	0.000187 ***	[0.178 , 0.584]

R square= 0.201 (max possible= 0.902) Likelihood ratio test= 71.67 on 6 df, p=1.856e-13
Wald test = 76.08 on 6 df, p=2.298e-14 Score (logrank) test = 84.01 on 6 df, p=5.551e-16

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
**P-value<.005 was statistically significant. Hazard ratio =HR,*
CI=Confidence interval, β=Parameter of estimate, SD =Standard deviation

The last step in model development strategy is consideration of interaction terms that may be useful in the improvement of the model. The researcher do not have any prior knowledge of specific interactions that we must include so we will consider all the possible bivariate interactions to see if the interaction effects can increase or decrease the survival time of MDR-TB patients but there no interaction effect for this datasets. As a result, six covariates were significant at 5% level of significance .Hence, the final multivariate model which includes the six covariates namely: Age, Body mass index, Smoking status, HIV co-infection, Co-morbidity and TB status are the risk factor for the death of MDR-TB patient or these variables significantly affects the survival of MDR-TB patients.

4.5 Model Checking

A final fitted model to be assessed after a model has been constructed. It is desirable to determine whether a fitted Cox PH regression model adequately describes the data set or not. This includes a test for violation of the assumption of proportional hazards, check influence observation and measuring the overall goodness of fit of the model.

4.5.1 Test of the assumption of proportional hazards

Table 6: Test of assumption of constant proportional hazard by using scale Schoenfeld residual.

Covariates	*Rho*	*Chi-square*	*p-value*
Age	0.06096	0.29835	0.5849
Bmi	-0.05151	0.14547	0.7029
HIV	0.04643	0.15551	0.6933
Smoking	0.24426	4.09879	0.0429*
Comorbidity	0.12007	1.03257	0.3096
Status	0.00623	0.00272	0.9584
GLOBAL	NA	6.57925	0.3615

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
rho =the estimated correlation between scale Schoenfeld residual and time

From the table 6: The variable smoking status was, hence p-value less than 5% this departure from proportional hazards. This occurs when regression coefficients are dependent on time that is when time interact with smoking status .Therefore hazard ratio for smoking status variable is not constant over a time. However the variables Age ,Body mass index , HIV co infection, Comorbidity and TB status were satisfy the assumption of constant hazard ratio over time (p >0.05) or there is no enough evidence to reject the null hypothesis that the covariates Age ,Body mass index , HIV co infection, Comorbidity and TB status satisfy the assumption of proportional hazard. Similar study conducted at Gonder and ALERT Hospital showed that smoking was not satisfy the assumption of constant proportional hazard[45].

On (Appendix C); shows the plots of scaled Schoenfeld residuals vs. the covariates. From figure the line on each plot is a smoothing spline; the broken lines give a point-wise 95-percent confidence envelope around this fit. The tendency for the effect of smoking status rise with time is clear in these plots and the tendency for the effect of Body mass index, HIV co infection, and Comorbidity and TB status constant over time. There is no evidence of a departure from the proportional hazards assumption for the covariates that are included in the model except for the covariate smoking status. The global is non-significant at 5% level of significance it means that the proportional hazard assumption is satisfied.

4.5.2 Assessment of linearity of covariates in the model

It is necessary to check whether the correct functional form of a continuous covariate holds in the model proposed to describe the data. A number of techniques are available, which are designed to determine whether the data support the hypothesis that the effect of the covariate is linear in the log hazard. In this study a graphical technique of the plots of the martingale residuals is use.

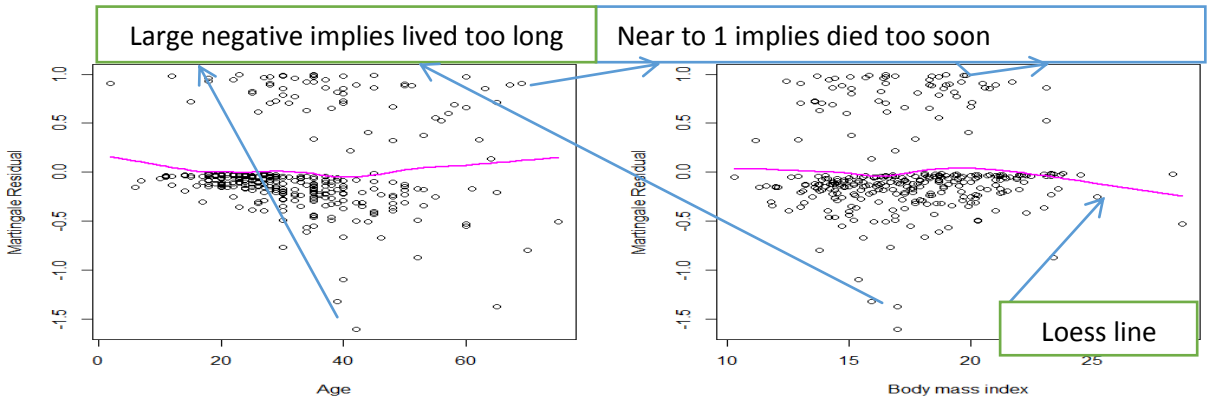


Figure 4: Plots of Martingale residuals against Age and body mass index

If martingale near to one the patients died too soon (earliest death) and large negative implies lived too long (longest survival time) .The figure 4 reveal that MDR-TB patients died too soon and lived too long. The loess line not wiggles (up and down in short movement) around zero, such a pattern was indicate non-proportional hazards (non-PH). Therefore the plots of martingale residual not confirm that Age and Body mass index of a patient have an approximate linear relationship with the survival time. Again the appendix C reveal that the covariate smoking status was not met the assumption of the constants PH and there is also an outlier form the figure. General cox proportional hazard is not good fit for this dataset, therefore other model need to accommodate the assumption of constant hazard proportion like Stratified Cox Proportional model and AFT models.

4.6 Stratified Cox PH regression

Cox regression model is applicable only to time- invariant predictors with time-constant effects only. We should split the whole sample into subgroups on the basis of categorical variable (stratification variable) and re-estimate the model [50].

Table 7: Estimate of the parameter for Stratified Cox PH regression

Covariate	β	HR	SD	Z-value	P-value	95% CL of HR
Age	0.0456	1.047	0.009	4.819	0.000	1.027 , 1.066
BMI	-0.095	0.909	0.042	-2.258	0.024	0.837 , 0.988
HIV	0.477	1.611	0.258	2.850	0.034	1.972 , 2.669
Comorbidity	0.851	2.342	0.340	2.501	0.012	1.202 , 4.562
TB status	-1.155	0.315	0.306	-3.772	0.000	0.173 , 0.574

Source: Saint Peter's Specialized hospital, Ethiopia; from January 3, 2014 to October 28, 2016

From table 6: The variable smoking status violate the assumption of proportional hazard, therefore the stratified Cox PH model is applied for this data set. Figure 4 shows that clear differences between the non-smoker and smoker survival curve. All covariate satisfies the assumption of constant proportional hazard; this supports the Cox PH model. The result of stratified Cox PH models to detect a non-proportional hazards trend for the smoking status variables.

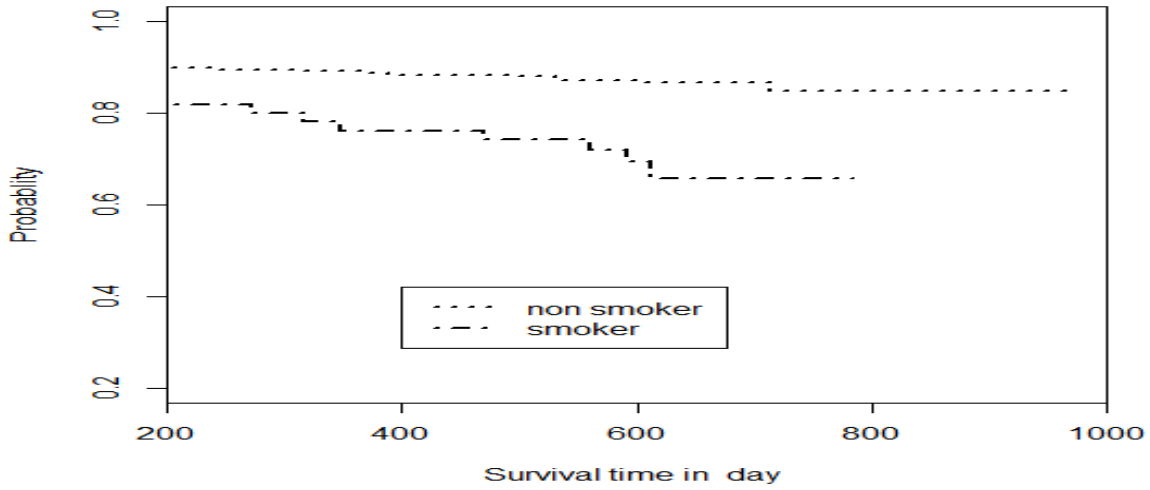


Figure 5: Stratified survivor plots to check for PH assumption

4.7 Accelerated Failure Time Models

When PH assumptions were not satisfied, the parametric AFT model should be used instead of the Cox model. The assumption of cox proportional hazard model was violated for covariate

smoking status in this case AFT model is appropriate. AFT models follow a known distribution; Exponential, Weibull, Lognormal, Log-logistics and Gamma.

4.8 Comparison of Accelerated Failure Time Models

The common applicable criterion to select the model is the Akaikie information criterion (AIC) Table 8, revealed that the Weibull regression model was the small AIC value, this indicate that the Weibull AFT model is a better fit the data of MDR-TB patient.

Table 8: Comparison of AFT model based on AIC

Distribution	*AIC*	*-2Loglik*
Exponential	*1094.472*	*1080.472*
Weibull	*1055.969*	*1039.969*
Lognormal	*1062.494*	*1046.494*
Log-Logistic	*1058.382*	*1042.382*
Gamma	*1212.687*	*1196.687*

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
AIC= Akaikie information criterion, Small AIC relatively to the other is best

4.9 Multivariable Weibull AFT Regression model

The Weibull model is a slight modification of the exponential model. We retain the assumption that ε has a standard extreme-value distribution, but we relax the assumption that $\sigma = 1$. When $\sigma > 1$, the hazard decreases with time. When $0.5 < \sigma < 1$, the hazard is increasing at a decreasing rate. When $0 < \sigma < 0.5$, the hazard is increasing at an increasing rate. And when $\sigma = 0.5$, the hazard function is an increasing straight line with an origin at 0.

For MDR-TB patient data set the value of $\sigma = 1.87$ which is greater than one therefore hazard decrease with time.

In AFT models, the sign of the coefficient indicates how a covariate affects the log survival times. Thus, a positive coefficient increases the log survival time and, hence, the expected duration. A negative coefficient decreases the logged survival time and, hence, the expected duration. Hence Age of patient, HIV co-infection, Smoking status and Comorbidity have negative coefficient, therefore logged survival time decreased , but Body mass index and TB status have positive coefficient ,it indicate the logged survival time is increased.

Table 9: Comparison of HR and ETR for Weibull PH and AFT model.

	PH model		AFT model	
	HR	95% Cl of HR	ETR	95% Cl of ETR
Age	1.047	[1.028 ,1.066]	0.918	[0.885,0.952]
BMI	0.905	[0.832,0.984]	1.204	[1.028,1.410]
HIVco-infection	1.724	[1.048 ,2.836]	0.362	[0.137,0.953]
Smoking status	2.314	[1.351,3.961]	0.209	[0.073,0.594]
Comorbidity	2.246	[1.156,4.366]	0.220	[0.062,0.789]
TB status	0.331	[0.183,0.599]	7.879	[2.519,24.643]

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
HR=Hazard ratio , ETR = Event time ratio ,PH= Proportional hazard

From table 9:Show that PH models, exponentiate the coefficients to obtain hazard ratios, these hazard ratios use to calculate the factor change or percentage change in the baseline hazard associated with a one unit increase in a covariate. All 95% CI for Weibull PH mode does not include one; it shows that all covariate were statistically significant effect for the death of MDR-TB patients. In AFT models exponentiate the coefficients to obtain time ratios. These time ratios use to calculate the factor change or percentage change in the expected survival time associated with a one unit increase in a covariate. The 95% confidence interval of ETR of all covariate does not include one .For AFT model TB status covariate have wide confidence interval relative to the other interval at 5% level of significance. General both PH model and AFT model reveal that the same conclusion.

Table 10:Variance –Covariance for the parameter of significant effect covariate.

	Intercept	Age	Bmi	HIV	Smok	com	TBs	Log(scale)
Intercept	2.7128							
Age	-0.0119	0.0003						
Bmi	-0.0918	-0.0003	0.0065					
HIV	-0.0246	0.0004	-0.0069	0.2444				
Smoking	-0.0041	-0.0005	-0.0056	0.0362	0.2849			
Comorbidity	-0.4208	0.0017	0.0010	-0.0421	-0.0064	.4235		
TB status	-0.2213	0.0006	0.0002	0.0014	-0.0148	-0.0360	0.4385	
Log(scale)	0.0399	-0.0008	0.0017	-0.0157	-0.0168	-0.0168	0.0176	.0125

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
Variance =diagonal of the matrix , covariance = Off diagonal of the matrix

Table 10 above: Reveal that the variance increase diagonal expect intercept and log(scale).The intercept of AFT model have large variance relatively to the other estimate of parameters and the parameter estimate of age patient have low variance relative to the other parameters. The parameter intercept have a negative correlation with all parameters expect log (scale).

4.11 Final interpretation of weibull AFT model

Table 11: Final parameter estimate of Weibull AFT model.

Covariate	β	Std	Z-value	P-value	ERT	95% Cl of ERT
(Intercept)	8.8174	1.6471	5.35	8.63e-08		
Age	-0.0859	0.0185	-4.65	3.35e-06	0.918	[0.885,0.952]
BMI	0.1856	0.0808	2.30	2.15e-02	1.204	[1.028,1.410]
HIVco-infection	-1.0172	0.4944	-2.06	3.96e-02	0.362	[0.137,0.953]
Smoking status	-1.5671	0.5338	-2.94	3.33e-03	0.209	[0.073,0.594]
Comorbidity	-1.5120	0.6508	-2.32	2.02e-02	0.220	[0.062,0.789]
TB status	2.0642	0.5818	3.55	3.88e-04	7.879	[2.519,24.643]
Log(scale)	0.6250	0.1117	5.60	2.17e-08		

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
Scale= 1.87

From the Table 11: The variables Age, Body mass index, HIV co-infection, Smoking status and Comorbidity were statistically significant at 5% level of significance on the survival time of MDR-TB patients. The 95% confidence interval of event time for all covariate does not include one this implies that all covariate were significant at 5% level of significance.

From Weibull AFT regression model, when the effect of all other factor kept constant, one unit(one year) increase Age of patient decrease log of survival time by 0.0859 time and The estimated acceleration factor for Age patient is estimated to be 0.918 with 95% CI: [0.885, 0.952].The confidence interval for the acceleration factor did not include one and P-value is smaller than 0.05, This implies as Age of patient increase the survival time become decreasing or older patient have less survival time than younger patient for MDR-TB patients. The study conducted in Bangladesh show that age was a risk factor of MDR-TB (AOR=1.72, 95% CI [1.12–2.66]) [33].

When the effect of other factor kept constant one unit increase Body Mass Index of patient increase log of survival time by 0.1856 times and One unit increase Body Mass Index of patient is increased survival time by a factor of 1.204.The acceleration factors for patients who smoke cigarette were estimated to be 0.209 with 95% CI (0.073, 0.594). This implied that non-smoker had longer survival time than smoker. Non-smoker MDR-TB patients survived 79.01% longer than smoker MDR-TB patients. The study conducted ALERT Hospital, Addis Ababa and Gondar University Teaching and Referral Hospital, Gondar, Ethiopia indicated that , smoking status (HR=3.17; 95%; 1.32 − 7.64) and weight (HR= 0.9093; 95%Cl; 0.8760 − 0.9440) were a risk factor for the death of MDT-TB patients .

Having HIV positive accelerates the time to event by a factor of exp (-1.0172) = 0.362 (0.362 times shorter survival time compared to the HIV negative).This indicate that HIV positive patient were significant effect for the death of MDR-TB patients, negative HIV co-infection MDR-TB patients survived 63.8% longer than positive HIVco-infection MDR-TB patients. The study conducted in Oromiya indicated HIV infection, previous known TB history, and previous TB treatment outcome were risk factor death of MDR-TB patient [9]. Study finding in southern Ethiopia shows there was no statistically significant association of HIV status ((OR=1.24, 95% CI [1.04–1.43]) with MDR-TB [35]. Whereas other study shows HIV (OR=2.28, 95% CI) [1.52–3.04]) is associated with increased risk of MDR-TB patients [30]..

As shown in Table 11: The estimated acceleration factor for TB status was estimated to be 7.879 with 95% CI lies between (2.519, 24.643). The confidence interval for the acceleration factor did not include one and P-value is smaller than 0.05,This indicate that patient who start MDR-TB treatment without showing TB status were significant effect for the death of MDR –TB patients or patient who were registered as new group were significant effect for the death of MDR –TB patients. The survival time for no previous TB history group is decreased by a factor of 7.879 times had pervious TB history group. Study findings in republic of Georgia shows previously treated for TB were significantly associated more likely to have MDR-TB than patients who were new (OR=5.27, 95% CI [3.75-7.41]). Likewise study in the Community of Madrid shows significant association with a history of previous TB treatment (OR=5.94, 95% CI [1.46-24.18]) [39].

Lastly the estimated acceleration factor for comorbidity status is estimated to be 0.220 with 95% CI lies between 0.062 and 0.789.The interval of event time ratio did not include one and P-value is smaller than 0.05, This indicate that a patient who did not have comorbidity status (had no chronic disease) have prolonged survival time than patient who had comorbidity status (had chronic disease).MDR-TB patients who have no chronic disease survived 78% longer than patient who had chronic disease. The conducted Estonia in showed that co-morbidities had influence on mortality among MDR-TB patients (OR, 2.62; 95% CL, 1.00-6.87) [42].

4.10 Fitted Weibull AFT model

Using the regression equation (3.5) and with the parameters estimated in final model the following regression model and distribution of the survival time of MDR-TB patk2ients derived as follow,Weibull distribution which can be expressed as;

$$T{\sim}Weibull(\lambda,\alpha)\ with\ \lambda = exp^{\left(\frac{-8.8174}{1.87}\right)} = 008958\ and\ \alpha = \frac{1}{1.87} = 0.535$$

$$Time{\sim}weibull(0.008958,0.535)$$

Final by substituting all parameter to Weibull distribution

$h_o(t) = (0.008958) * (0.535)t^{0.535-1}$ $is\ baseline\ hazard\ funtion\ with\ out\ covariate$

$h(t;X,\beta) = (0.008958) * (0.535)t^{0.535-1}\exp\left(\frac{-\beta^{t}{}_{aft}}{\sigma}X\right)$ is Hazard function with covariate

Additive AFT model

$$Log(T) = 8.8174 - 0.0859(Age) + 0.1856(Bmi) - 1.5671(smokingstatus) - 1.0172(HIV) + 2.0642(TBstatus) - 1.5120(comorbidity)$$

Multiplicative AFT model

$$T = \exp(8.8174 - 0.0859(Age) + 0.1856(Bmi) - 1.0172(HIV) - 1.5671(Smoking) + 2.0642(TBstatus) - 1.5120(comorbidity))$$

$$T = \exp(8.8174) * \exp(-0.0859Age) * \exp(0.1856Bmi) * \exp(-1.5671smokingsta) * \exp(-1.0172HIV) * \exp(2.0642TBstatus) * \exp(-1.5120comorbidity))$$

Survival function of AFT model for a given covariate

$$S(t; X, \beta) = S_0[t * exp\{(0.0859(Age) - 0.1856(Body\ mass\ index) + 1.0172(HIV) + 1.5671(Smoking\ status) + 2.0642(TBstatus) + 1.5120(comorbidity)))\}]$$

Hazard function of AFT model for a given covariate

$$h(t; X, \beta) = h_0[texp\{(0.0859(Age) - 0.1856(Bmi) + 1.0172(HIV) + 1.5671(Smoking) - 2.0642(TBstatus) + 1.5120(comorbidity)))\}] * \exp(-0.0859(Age) + 0.1856(Bmi) - 1.0172(HIV) - 1.5671(Smoking) + 2.0642(TBstatus) - 1.5120(comorbidity)))$$

4.11 Assessment of Adequacy of the Weibull AFT Model

Adequacy of a final fitted model to be assessed after a model has been constructed. It is desirable to determine whether a fitted Weibull AFT regression model adequately describes the data set or not.

4.11.1 Quantile-Quantile Plot

A quantile-quantile plot is made to check if the AFT provided an adequate fit to the data by using two different groups of population. We shall graphically check the adequacy of the model by comparing the significantly different groups of patients by, smoking, TB status, HIV co-infection and comorbidity. From the figure 6,The quantile-quantile plot approximately linear for all covariate. Therefore a Weibull accelerated failure time model is a best fit for this data. A Weibull accelerated failure time model is best to describe survival time of MDR-TB patients.

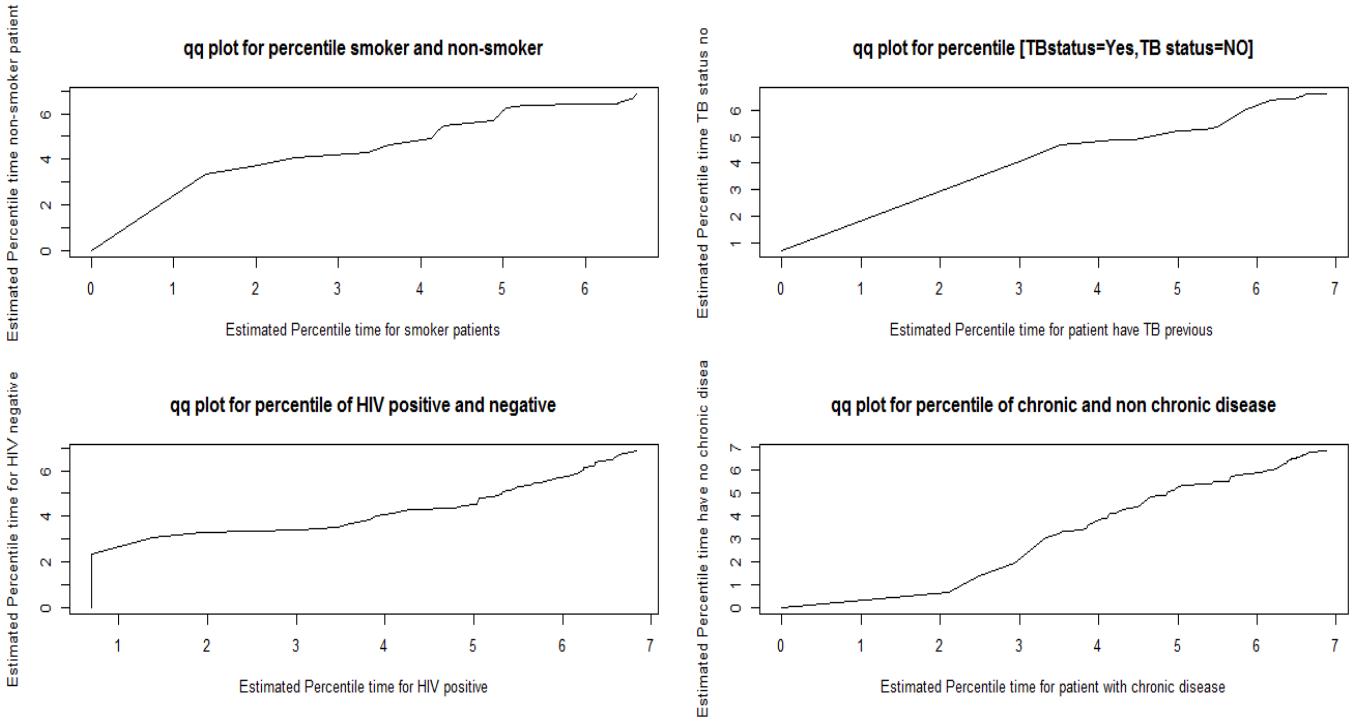


Figure 6: Quantile- Quantile plot to check the adequacy of the AFT model.

4.11.2 Log time versus the log of the estimated cumulative hazard

One of the diagnostic plots for Weibull AFT regression is plot of log time versus the log of the estimated cumulative hazard estimate. If the Weibull model has adequate fit, then the plots for each of the covariates should be roughly linear and parallel. Hence the figure 7 show that log time versus the log of the estimated cumulative hazard estimate is roughly linear and parallel for the covariate TB status, comorbidity, smoking status and HIV co-infection. Therefore a Weibull accelerated failure time model is best to describe survival time of MDR-TB patients

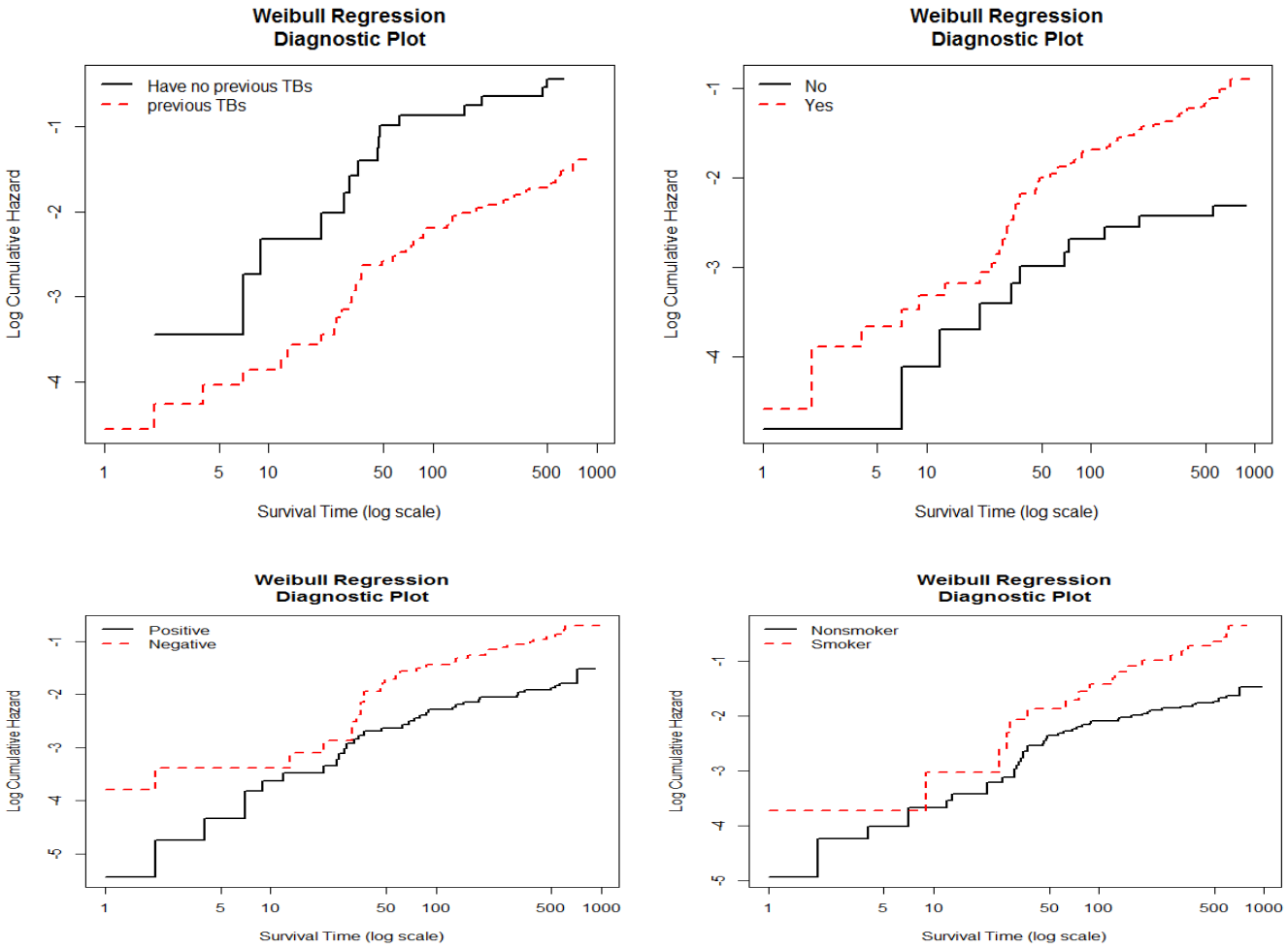


Figure 7: Plot of log Time versus the log of the estimated cumulative hazard.

4.11.3 Cox –Snell residual

The plot of the cumulative hazard function of the Cox-Snell residuals against maximum likelihood estimation with cumulative hazard functions is given in figure 8.is presented in below suggest that Weibull ATF model is an appropriate for modeling time to death of MDR-TB patients. If the model fits the data, the plot of cumulative hazard function of residuals against Cox-Snell residuals should be approximately a straight line with slope one and zero intercept.

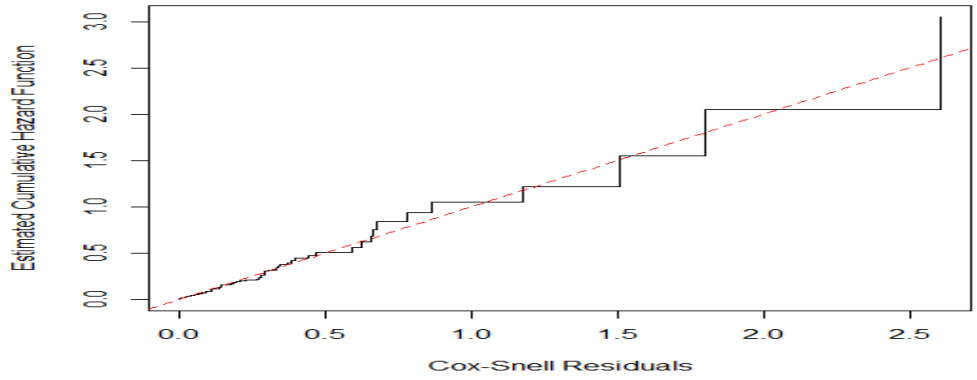


Figure 8: Cox- Snell residuals plots of Weibull baseline distribution for survival time of MDR-TB patient.

4.11.4 Over all Goodness Fit of Weibull AFT Model

The final step in the model assessment is to measure the overall goodness of fit. For this objective the study use the Cox-Snell residuals, R^2 and Likelihood Ratio, A perfectly adequate model has low R^2 due to the present of censored data. Thus, the model fitted in this study has a value of R^2statistic of 0.21, implying that 21% of full log likelihood is explained by this AFT model. Weibull AFT regression model with covariate Age of patient, Smoking status, Body mass index, comorbidity and TB status a good fit rather than null model.

$$R^2=1-\left\{\exp\frac{2}{320}[L_O - L_P]\right\}$$

L_P=log likelihood for fitted model with six covariate

L_o = log likelihood for fitted model with no covariate

$$R^2=1-\left\{\exp\frac{2}{320}[-556.7-(-520)]\right\}= 0.21$$

From the likelihood ratio test Table 12 below, revealed that Weibull AFT regression model is significant (p=8.3e-14), using the log likelihood values of the null model and the full model. The model with minimum AIC value is best, therefore weibull AFT regression with all covariate is best fit, and this implies that the model was improved after covariates were added to the model

Table 12: The likelihood ratio and AIC of the Weibull AFT regression model.

Loglik (Intercept only)	AIC (Null model)	Loglik (Full model)	AIC (Full model)	Df	Chisq	p-value
-556.7	1113.4	-520	1052	6	73.37	8.3e-14

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016
AIC=Akaike's information criteria Df=Degree freedom

4.11 Bayesian Survival Analysis

In Bayesian inference, uncertainty with respect to parameters is at any point in time quantified by probability distributions. This means that a distribution needs to be specified for all parameters in advance. These prior distributions reflect the prior expectations with respect to the parameter values. The researcher use normal distribution with mean 0 and variance 1000 (huge variance) for the parameter beta in the AFT model and gamma distribution for sigma with scale = 0.01 and shape = 0.01 parameters. The Gibbs sampler algorithm was implemented with 20100 iterations in three different chains and 60300 samples from the full posterior distribution.

Table 13: Parameter estimation for Bayesian Weibull AFT model

Covariate	Categories	Nodes	mean	Sd	MC error	2.5%	median	97.5%
Constant		beta1	-9.081	0.8743	0.03722	-10.54	-9.026	-7.715
Age		beta[2]	0.012	0.0055	9.544E-5	8.052E-4	0.0117	0.0225
Sex	Male	beta[3]	-0.047	0.1273	0.001318	-0.2967	-0.04776	0.2036
Bmi		beta[4]	-0.013	0.0216	7.147E-4	-0.05555	-0.01264	0.0282
TB site	Pulmonary	beta[5]	-0.176	0.1719	8.937E-4	-0.5235	-0.1728	0.1499
Tb status	Yes	beta[6]	-0.549	0.5872	0.02344	-1.818	-0.5176	-0.5066
Region	Oromiya	beta[7]	0.065	0.1663	0.001725	-0.2681	0.06645	0.3815
	Amhara	beta[8]	-0.081	0.3505	0.002532	-0.8093	-0.06602	0.5601
	SNNP	beta[9]	-0.250	0.2926	0.002554	-0.8563	-0.2389	0.2973
	Other	beta[10]	0.301	0.263	0.003131	-0.2332	0.3085	0.8001
HIV	Positive	beta[11]	-0.012	0.1383	0.001037	-0.287	-0.01154	0.2562
Smoking	Yes	beta[12]	0.155	0.2154	0.001541	0.0744	0.1575	0.5695
Alcohol use	Yes	beta[13]	-0.069	0.2318	0.00189	-0.5321	-0.06693	0.3772
Level of	Primary	beta[14]	-0.173	0.1776	0.002978	-0.513	-0.1754	0.1856

education	Secondary	beta[15]	-0.328	0.1938	0.002736	-0.7051	-0.3294	0.0598
	Above secondary	beta[16]	-0.219	0.2108	0.003137	-0.6286	-0.2205	0.1992
Comorbidity	Yes	beta[17]	0.0558	0.1365	0.001965	0.0106	0.05531	0.3233
Registration group	Relapse	beta[18]	0.1806	0.5665	0.02231	-0.8243	0.1408	1.431
	Failure of new regime	beta[19]	0.1268	0.5636	0.02227	-0.8684	0.08609	1.378
	Other	beta[20]	0.0431	0.569	0.02215	-0.9681	0.002915	1.294

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016

Based on the sample obtained from posterior distribution, summary statistics of all parameters for joint posterior distribution were present in Table 13 above. The variables Age, TB status Smoking status and comorbidity were statistically significant predictor for survival time of MDR-TB patients at 5% level of significance and the 95% credible interval of this covariate does not include zero. Where 2.5% is an approximation of lower endpoint of the 95% credible interval and 97.5% is approximation of lower endpoint of the 95% credible interval. The point estimate for the covariate Age was 0.012, with 95% credible interval (8.052E-4, 0.0225).

4.11.1 Graphical approaches to assess convergence

History Plots: Are commonly used to assess convergence of the parameter estimates in Bayesian analysis. The WinBUGS software gives the plot with number of iterations on the x-axis and parameter values on the y-axis for each significant parameter. If the plot looks like a horizontal band, with no long upward or downward trends, then the researcher have evidence that the chain has converged. For all simulated parameters, history plot indicates a good convergence since three independent generated chains are mix together (See Figure 9 and Appendix D Figure 5)

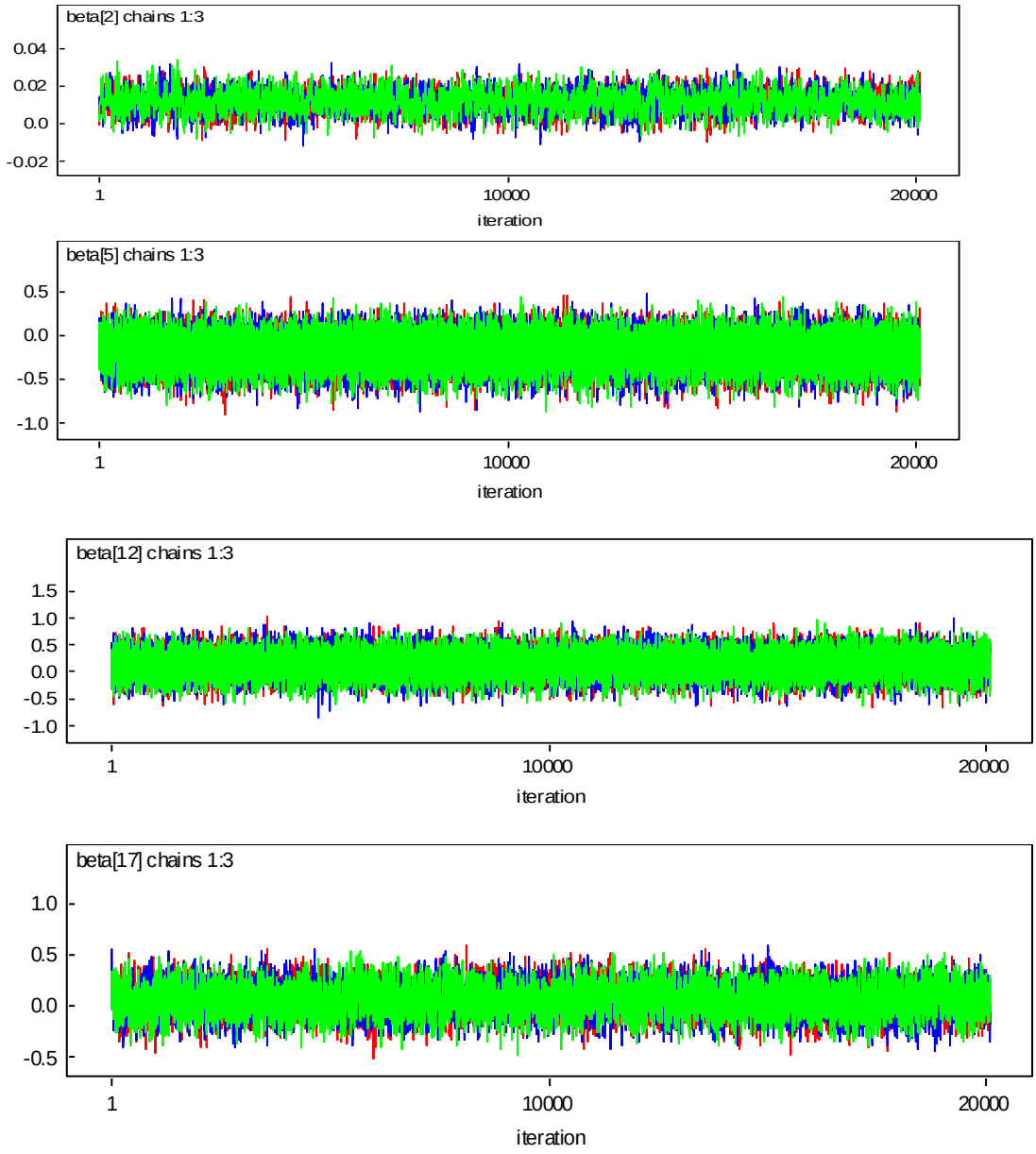


Figure 9 : History plot for variable Age,TB Status ,Smoking status and co-morbidity.

Density Plot: Use to smooth kernel density estimate for the parameter. For this study the coefficients for most of the independent variables were normally distributed. Thus, this indicates that the Markov chain has attained its posterior distribution (See Figure 10 and Appendix D Figure 6)

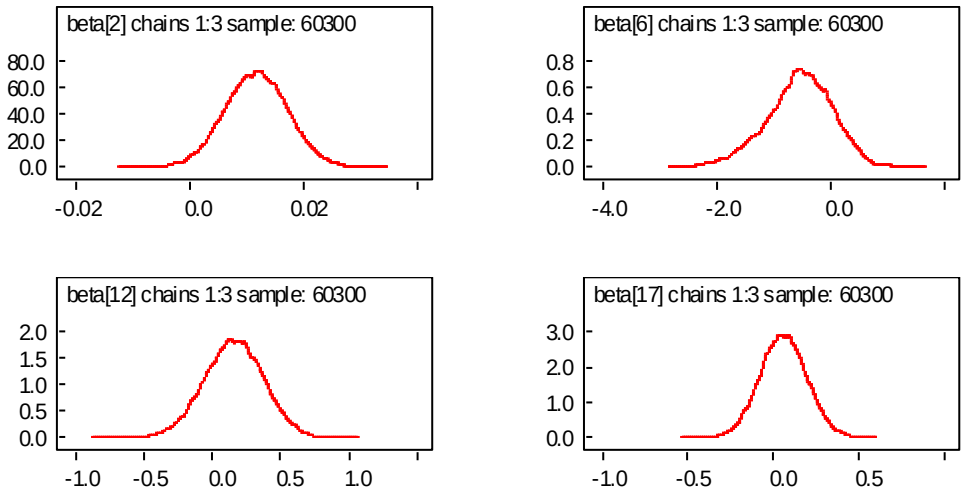


Figure 10: Density Plot for variable Age, TB Status, Smoking status and Co-morbidity.

Autocorrelation Plot: High auto correlations with chains indicate slow mixing and usually slow convergence but all significant parameter have less autocorrelation with chains this implies that high mixing and high (See Figure 11 and Appendix D Figure 8)

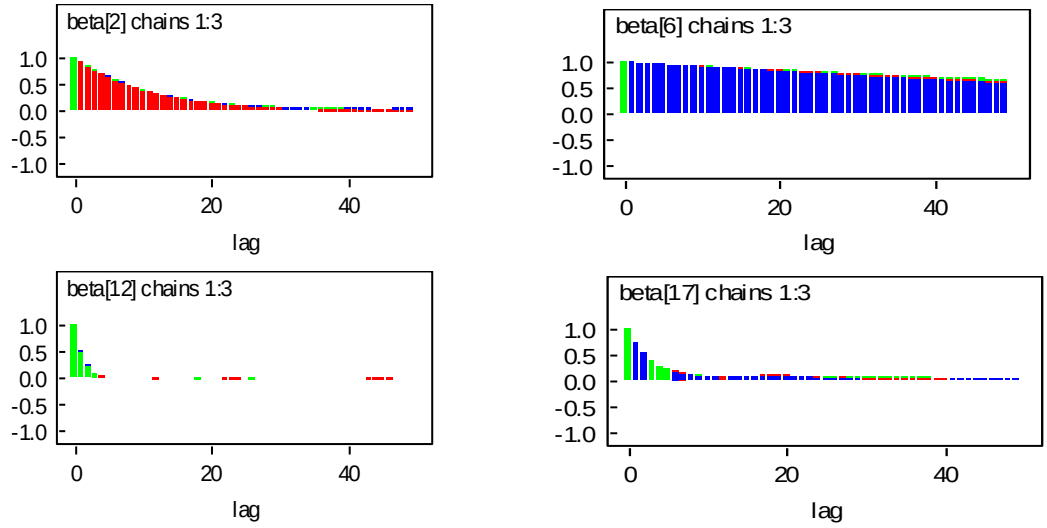


Figure 11: Convergence Analysis using Autocorrelation for covariate Age, TB Status, Smoking status and Co-morbidity

Trace plot: Running more than one chain simultaneously, the trace and history plots shows each chain in a different color. In this case, The researcher have reasonably confident that convergence has been achieved if all the chains appear to be overlapping one another (See Figure 12 and Appendix D Figure 6)

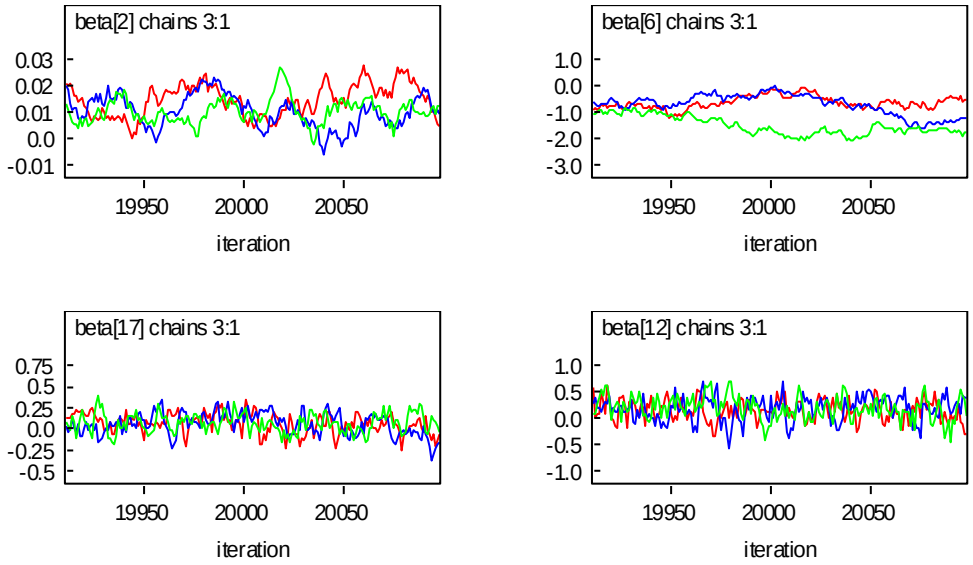


Figure 12: Convergence Analysis using Trace plot for covariate Age, TB Status, Smoking status and Co-morbidity

Gelman-Rubin Statistics: It is another way of assessing convergence for Bayesian analysis. The model is judged to have converged if the ratio of between to within variability is close to 1. The green line represents the between variability, the blue line represents the within variability, and the red line represents the ratio. Evidence for convergence comes from the red line being close to 1 on the y-axis and from the blue and green lines being stable (horizontal) across the width of the plot. Hence the Gelman-Rubin statistic of this study emphasis that one should be concerned convergence of ratio close to one (see Figure 13)

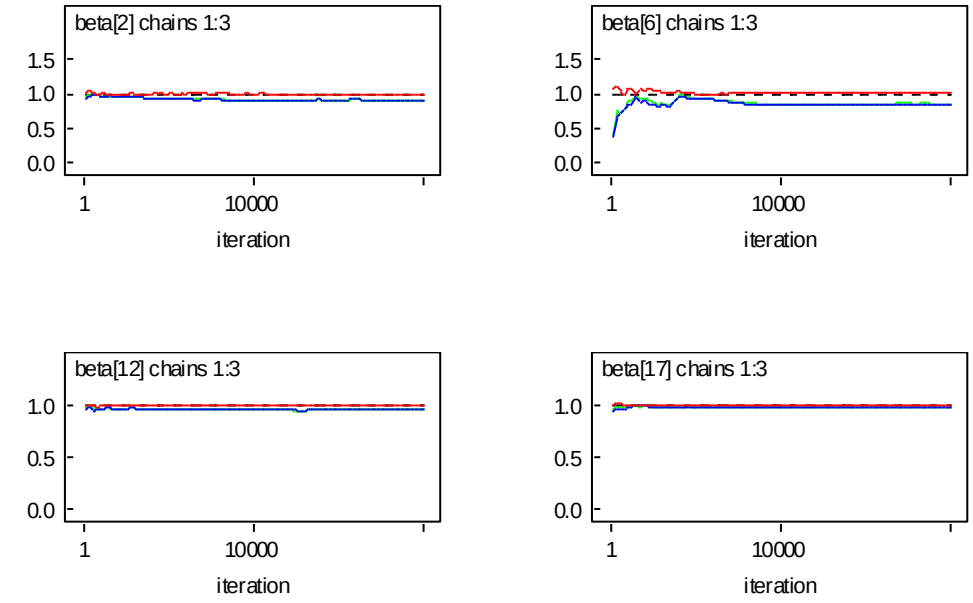


Figure 13: Convergence Analysis using Gelman-Rubin plot for covariate Age, TB Status, Smoking status and Co-morbidity

4.11.2 Assessing Accuracy of the Bayesian Survival Analysis

One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter. This is an estimate of the difference between the mean of the sampled values and the true posterior mean.As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation. This study, MC error for each significant variable is less than 5% of its standard deviation. The parametric Weibull AFT model in classical approach and Bayesian approach were fitted. Parameters in Bayesian analysis had smaller standard error than the corresponding classical AFT model but the result is not consistent. In classical (Weibull AFT) approach the variable Age, HIV co-infection ,body mass index ,TB status ,smoking and comorbidity were significant effect for the death of MDR-TB patient at 5% level of significance but in Bayesian analysis the variable HIV co- infection and Body Mass Index are not significant effect for the death of MDR-TB patients at 5% level of significance.

4.14 Discussions

This research was conducted to identify predictors that accelerate the mortality of MDR-TB patients. The analysis revealed that the proportional hazards assumption of the Cox regression model was violated. Therefore, further survival analyses like Stratification, Accelerated Failure Time and Bayesian Analysis have been incorporated to overcome the assumption of proportional hazard. Exponential, Weibull, Log-logistic and Log-normal distributions were applied on similar data. The Covariate included in the study were Age,Sex,BMI ,Smoking status, Region of patient, TB status, TB site,HIV co-infection, Alcohol use ,Level of education, Comorbidity and registration group. The outcome of the interest was survival time of MDR-TB patient measured in days.

The study was conducted from October, 2011 up to May, 2012 among cohorts of MDR-TB patients that started treatment in February 2009. A total of 188 patients were followed for a total of 79,600 person-days. Median follow up time was 466.5 days or 1.28 years and the percent [43] and The study included 342 MDR-TB patients (142 from ALERT and 200 from Gondar) who had been under treatment from August 2011 to August 2014 The median survival for MDR-TB patients was 16 months(1.33 years) and the percent of death was 10.8(10.8%) [45]. In the current study revealed that the median survival of MDR-TB patients was about 597days

(1.64years) which is greater than two studies and the percent of death for the current study was 20.78 which is greater than the study conducted in ALERT and Gonder.

Age group at 25-44 years in Ethiopia (AOR=2.8, 95% CI [1.7–6.4]) [31] and in Bangladesh (AOR=1.72, 95% CI [1.12–2.66]) [32] and [33] was a risks factor of MDR-TB and the study conduct in ALERT and Gonder show that Age was a risk factor for the survival time of MDR-TB patients[45].The current finding revealed that Age is risk factor for the survival time of multidrug resistance tuberculosis patients [ETR=0.918,Cl=0.885,0.952].

WHO report that HIVco-infection was a risk factor for survival time of MDR-TB patients [26] ,the study conducted in Ethiopia by Birhanu and His colleagues showed MDR-TB and HIV significant association (OR=3.7, 95%Cl [1.90, 7.22][18] and other study shows HIV is associated with increased risk of acquired MDR-TB (OR=1.24, 95% CI [1.04–1.43])[30]. But Study conducted in southern Ethiopia shows there was no statistically significant association of HIV status with MDR-TB [35]. The current study showed that HIV co-infection was statistically significant for survival time of MDR-TB patient [ETR=.362, Cl =.137, 0.953].

The research conducted in Nigeria shows gender was not significantly associated with MDR-TB [29]. And study findings in Thailand also shows male gender as risk factors for MDR-TB [30] and in Ethiopia (AOR =2, 95% CI [1.4-5]) showed male gender was a risks factor for MDR-TB [31**]**.but the current finding showed that gender of MDR-TB patient was not statistically significant factor for survival of MDR-TB patient. The study conduct in Nigeria and the current study support each other gender was not statistically significant for the death of MDR-TB patients but two studies conducted in Thailand and Ethiopia contradict the current studies.

The study of MDR-TB in Estonia, showed that co-morbidities (OR, 2.62; 95% CL, 1.00-6.87) were a risk factor for the death of MDR-TB patients [42] but the Study in St. Peter TB specialized hospital, Addis Ababa, Ethiopia, revealed that survival of patients under MDR-TB treatment was not associated with co-morbidity [43].The recent finding showed that variable co-morbidity was statistically significant factors for survival time of MDR-TB patients(for the death of MDR-TB Patient)[ETR=0.220,Cl=0.062,0.78].

The Study findings in republic of Georgia shows previously treated for TB were significantly associated more likely to have MDR-TB than patients who were new (OR=5.27, 95% CI [3.75-7.41]). Likewise study in the Community of Madrid shows significant association with a history of previous TB treatment (OR=5.94, 95% CI [1.46-24.18]) [39]. Nationwide study in China shows previous treatment history had a more than 7 fold increased risk of MDR-TB, compared with those never previously treated [41].The current study revealed that TB status was a risk factor for survival time of MDR-TB patient [ETR=7.879, Cl =2.519,24.643].

A nationwide survey conducted in China; the estimated MDR-TB rate was 5.7% for new cases and 25.6% for previously treated cases [22]. And a study finding in Uganda shows, MDR-TB of 1.4% from new cases and 12.1% from previously treated cases [37]. According to World Health Organization 2015 report, among newly diagnosed TB cases 1.6% was found to be with MDR-TB and MDR-TB among previously treated TB cases was 11.8% [23]. The current study show that 16.25% for previously treated cases and 4.69 for new cases (had no previously treated TB case).The study conducted in Bangladesh shows urban residence of TB patients was significantly associated with MDR-TB occurrence [33]. But this study revealed that region was not statistically associated with MDR-TB patients.

The study conducted ALERT Hospital, Addis Ababa and Gondar University Teaching and Referral Hospital, Gondar, Ethiopia revealed that Smoking status was risk factor for survival time of MDR-TB patients (HR=3.17; 95%Cl; 1.32 − 7.64) [45] and the study conducted Addis Ababa, Ethiopia from October, 2011 up to may, 2012 among cohorts of MDR-TB patients showed that Smoking status had a significant effect on survival time of MDR-TB patients [43]. This study showed that smoking status was significant effect for the death of MDR-TB patients or The results of this study suggested that smoking was significant factor for survival time of the patients. Non-smokers had longer survival time than smokers [ETR=0.209, Cl =0.073, 0.594].

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The result of Cox PH and Weibull AFT model revealed that the same predictor had significant effect on multi drug resistance tuberculosis patients but the covariate smoking status not satisfy constant proportional hazard for Cox PH model .Stratified proportional hazard model were fitted but no estimate are obtained for stratified variable , thus no test for the main effect of stratified variable ,therefore Stratified Cox PH model is not good fit for this dataset even if the model satisfy the assumption of proportional hazard model because smoking status is variable of interest. Using Weibull AFT model covariates that significantly influence the survival of MDR-TB patients are identified at 5% level of significance ,this were Age, Body mass index ,HIV co-infection ,Smoking status, Comorbidity and TB status.

Bayesian accelerated failure time model also used to analyze survival time of MDR-TB patients. The variable that significantly influence the survival of MDR-TB patients were Age, TB status Smoking status and comorbidity, at 95% credible interval. From classical and Bayesian model, Bayesian survival analysis had smaller standard error than that of the classical survival analysis but two model are not give consistent result, the variable that was significant in Weibull AFT model did not significant in Bayesian survival model. Even if Bayesian survival model is small standard error it is not best fit because the interesting variable like HIV Co-infection and Body Mass Index were not significant in Bayesian survival model. Weibull AFT is better fit for this data set.

5.2 Recommendation

On the basis of the findings of this study the following recommendations are made for ministry of health, policy makers, the community, Saint Peter's specialized TB hospital worker and researcher.

1. Health workers and data clerks, working with patients under MDR-TB, should be given special training to improve the quality of data records of patients.
2. Regular medical check-up for MDR TB is very necessary especially for aged people and HIV co-infection patients which helps to detect the problem before it become complicated
3. Potential stakeholder likes government and non- government can give attention for this subject.
4. The Weibull AFT model give better predictions to the survival probability of MDR-TB Patients. So, future studies should have to use Weibull AFT models.

REFERENCE

1) World Health Organization. "WHO treatment guidelines for drug-resistant tuberculosis 2016 update." (2016).
2) World Health Organization. (2016). *World malaria report 2015*. World Health Organization.
3) Tessema, B. (2012). *Molecular epidemiology and drug resistance of Mycobacterium tuberculosis among HIV positive and HIV negative tuberculosis patients in Amhara region, Northwest Ethiopia* (Doctoral dissertation, Institute of Clinical Immunology, Faculty of Medicine, University of Leipzig)
4) World Health Organization. (2016). Global tuberculosis report 2016.
5) Tessema, B., Beer, J., Emmrich, F., Sack, U., & Rodloff, A. C. (2012). First-and second-line anti-tuberculosis drug resistance in Northwest Ethiopia. *The International Journal of Tuberculosis and Lung Disease*, *16*(6), 805-811.
6) Fiseha, D., & Demissie, M. (2015). Assessment of Directly Observed Therapy (DOT) following tuberculosis regimen change in Addis Ababa, Ethiopia: a qualitative study. *BMC infectious diseases*, *15*(1), 40
7) Biadglegne, F., Sack, U., & Rodloff, A. C. (2014). Multidrug-resistant tuberculosis in Ethiopia: efforts to expand diagnostic services, treatment and care. *Antimicrobial resistance and infection control*, *3*(1), 31.
8) Edlin, B. R., Tokars, J. I., Grieco, M. H., Crawford, J. T., Williams, J., Sordillo, E. M., ... & Jarvis, W. R. (1992). An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. *New England Journal of Medicine*, *326*(23), 1514-152
9) Mulisa, G., Workneh, T., Hordofa, N., Suaudi, M., Abebe, G., & Jarso, G. (2015). Multidrug-resistant Mycobacterium tuberculosis and associated risk factors in Oromia Region of Ethiopia. *International Journal of Infectious Diseases*, *39*, 57-61
10) Abate, D., Taye, B., Abseno, M., & Biadgilign, S. (2012). Epidemiology of anti-tuberculosis drug resistance patterns and trends in tuberculosis referral hospital in Addis Ababa, Ethiopia. *BMC research notes*, *5*(1), 462.
11) World Health Organization, & Global Tuberculosis Programme. (2008). *Global tuberculosis control: WHO report*. Who

12) Snider DE, Jr..Raviglione M, Kochi A. Tuberculosis in Bloom DR. Editor pathogenesis protection and control. Washington DC: AM Soc microbiology 1994. Pp 2-11

13) Loddenkemper, R., Sagebiel, D., & Brendel, A. (2002). Strategies against multidrug-resistant tuberculosis. *European respiratory journal*, *20*(36 suppl), 66s-77s.

14) Hirpa, S., Medhin, G., Girma, B., Melese, M., Mekonen, A., Suarez, P., & Ameni, G. (2013). Determinants of multidrug-resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study. *BMC Public Health*, *13*(1), 782.

15) Hordofa, M. W., & Adela, T. B. (2015). Prevalence of Refampcin Mono Resistant Mycobacterium Tuberculosis among Suspected Cases Attending at Yirgalem Hospital. *Clinical Medicine Research*, *4*(3), 75-78.

16) Abate, D., Taye, B., Abseno, M., & Biadgilign, S. (2012). Epidemiology of anti-tuberculosis drug resistance patterns and trends in tuberculosis referral hospital in Addis Ababa, Ethiopia. *BMC research notes*, *5*(1), 462.

17) World Health Organization, & Global Tuberculosis Programme. (2008). *Global tuberculosis control: WHO report*. Who.

18) Heifets, L. B., & Cangelosi, G. A. (1999). Drug susceptibility testing of Mycobacterium tuberculosis: A neglected problem at the turn of the century [State of the Art]. *The international journal of tuberculosis and lung disease*, *3*(7), 564-581.

19) Agonafir, M., Lemma, E., Wolde-Meskel, D., Goshu, S., Santhanam, A., Girmachew, F., ... & Van Soolingen, D. (2010). Phenotypic and genotypic analysis of multidrug-resistant tuberculosis in Ethiopia. *The International Journal of Tuberculosis and Lung Disease*, *14*(10), 1259-1265. .

20) World Health Organization. (2011). Towards universal access to diagnosis and treatment of multidrug-resistant and extensively drug-resistant tuberculosis by 2015: WHO progress report 2011..

21) Hirpa, S., Medhin, G., Girma, B., Melese, M., Mekonen, A., Suarez, P., & Ameni, G. (2013). Determinants of multidrug-resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study. *BMC Public Health*, *13*(1), 782.

22) Chan, E. D., Laurel, V., Strand, M. J., Chan, J. F., Huynh, M. L. N., Goble, M., & Iseman, M. D. (2004). Treatment and outcome analysis of 205 patients with multidrug-resistant tuberculosis. *American journal of respiratory and critical care medicine*, *169*(10), 1103-1109

23) World Health Organization. (2011). World Health Organization global tuberculosis control report 2009. *Global tuberculosis control*

24) Multi-drug resistant tuberculosis, a critical global concern, 1-2 available at:http://www.who.int/tb/publications/MDRFactSheet2012.pdf,Accesseddate[05/07/201].

25) Ameha, K., & Gashaw, T. (2013). *Prevalence of Pulmonary Tuberculosis and Coinfection with Hiv from Patients of Selected Public and Government Hospitals of Diredawa Town* (Doctoral dissertation, Harmaya University)

26) Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis Part III: multivariate data analysis–choosing a model and assessing its adequacy and fit. *British journal of cancer*, *89*(4), 605.

27) Shorter treatment regimens for multidrug-resistant tuberculosis, treatment outcomes observed in Bangladesh for Multidrug resistant-TB cases treated with a9 month's regimen, 1-2, available at; www.who.int/tb/challenges/mdr accessed date [12/07/2014].

28) Fikadu T (2015), Risk Factors for Multi-drug Resistant Tuberculosis, Universal Journal of Public Health, vol 3(2) 1-6, http://www.hrpub.org: accessed date [21/02/2015].

29) Casal, M., Vaquero, M., Rinder, H., Tortoli, E., Grosset, J., Rüsch-Gerdes, S., ... & Jarlier, V. (2005). A case-control study for multidrug-resistant tuberculosis: risk factors in four European countries. *Microbial drug resistance*, *11*(1), 62-67.

30) Kapadia Vishakha, K., & Tripathi Sanjay, B. (2013). Analysis of 63 patients of MDR TB on DOTS plus regimen: An LG hospital, TB Unit, Ahmedabad experience. *Dr Jitendra N Patel*, *52*

31) Mesfin, Y. M., Hailemariam, D., Biadglign, S., & Kibret, K. T. (2014). Association between HIV/AIDS and multi-drug resistance tuberculosis: a systematic review and meta-analysis. *PloS one*, *9*(1), e82235

32) Belén and Teresa (2010), Resistant Mycobacterium tuberculosis Strains in theCommunity of Madrid, Arch Bronconeumol,vol 43(6), 1-10, available at, http://www.archbronconeumol.org

33) World Health Organization. (2010). Multidrug and extensively drug-resistant TB.

34) Anderson, L. F., Tamne, S., Watson, J. P., Cohen, T., Mitnick, C., Brown, T., ... & Abubakar, I. (2013). Treatment outcome of multi-drug resistant tuberculosis in the United Kingdom: retrospective-prospective cohort study from 2004 to 2007. *Euro Surveill*, *18*(40), 1028-1035

35) FMoH, E. (2008). Tuberculosis, Leprosy and TB/HIV Prevention and Control Programme. *Addis Ababa, Ethiopia*, *20*

36) Flora, M. S., Amin, M. N., Karim, M. R., Afroz, S., Islam, S., Alam, A., & Hossain, M. (2013). Risk factors of multi-drug-resistant tuberculosis in Bangladeshi population: a case control study. *Bangladesh Medical Research Council Bulletin*, *39*(1), 34-41

37) Fentahun B, Ulrich S and Arne R (2014), Multidrug-resistant tuberculosis in Ethiopia:efforts to expand diagnostic services, treatment and care, Antimicrobial Resistance and Infection Control vol 3 (31), 1-10, available at,http://www.aricjournal.com/content/3/1/31 [accessed date, 24/02/2015].

38) Seung, K.J., Mateo, D.B., Keshavjee, S., Furin, J.J., Farmer, P.E., et al. (2009) Early Outcomes of MDR-TB Treatment in a High HIV-Prevalence Setting in Southern Africa. PLoS ONE, 4, Article: e7186.

39) Anderson, L. F., Tamne, S., Watson, J. P., Cohen, T., Mitnick, C., Brown, T., ... & Abubakar, I. Tuberculosis notification rate decreases faster in residents of native origin than in residents of foreign origin in the EU/EEA, 2010 to 2015.

40) Dey, T., Brigden, G., Cox, H., Shubber, Z., Cooke, G., & Ford, N. (2012). Outcomes of clofazimine for the treatment of drug-resistant tuberculosis: a systematic review and meta-analysis. *Journal of antimicrobial chemotherapy*, *68*(2), 284-293

41) WHO (2012) Tuberculosis Global Facts. www.who.int/tb

42) Anderson, L. F., Tamne, S., Watson, J. P., Cohen, T., Mitnick, C., Brown, T., ... & Abubakar, I. (2013). Treatment outcome of multi-drug resistant tuberculosis in the United Kingdom: retrospective-prospective cohort study from 2004 to 2007. *Euro Surveill*, *18*(40), 1028-1035

43) Getachew, T., Bayray, A., & Weldearegay, B. (2013). SURVIVAL AND PREDICTORS OF MORTALITY AMONG PATIENTS UNDER MULTI-DRUG RESISTANT TUBERCULOSIS TREATMENT IN ETHIOPIA: ST. PETER'S SPECIALIZED TUBERCULOSIS HOSPITAL, ETHIOPIA. *International Journal of Pharmaceutical Sciences and Research*, *4*(2), 776

44) Balabanova, Y., Radiulyte, B., Davidaviciene, E., Hooper, R., Ignatyeva, O., Nikolayevskyy, V., & Drobniewski, F. A. (2011). Survival of drug resistant tuberculosis patients in Lithuania: retrospective national cohort study. *BMJ open*, *1*(2), e000351

45) Molalign, S., & Wencheko, E. (2016). Risk factors of mortality in patients with multi-drug resistant TB. *The Ethiopian Journal of Health Development (EJHD)*, *29*(2).

46) Liu, Q., Zhu, L., Shao, Y., Song, H., Li, G., Zhou, Y., ... & Lu, W. (2013). Rates and risk factors for drug resistance tuberculosis in Northeastern China. *BMC public health*, *13*(1), 1171.

47) Misganaw, A. S., Abebe, M. T., Lulie, A. D., & Bika, A. T. (2016). External Quality Assessment (EQA) of Randomly Blinded Rechecking Slides on TB AFB Microscopy Laboratories: A Retrospective Study, Addis Ababa, Ethiopia. *American Journal of Laboratory Medicine*, *1*(2), 9-15

48) Dezfuli, H., Kelly, D., Smith, C., Vedros, K., & Galyean, W. (2009). Bayesian inference for NASA probabilistic risk and reliability analysis.

49) Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation* (Vol. 13). John Wiley & Sons.

50) Borucka, J. (2014). Extensions of Cox model for non-proportional hazards purpose. *Ekonometria*, (3 (45)), 85-101

APPENDICES

Appendix A: Univar ate and multivariate cox proportional hazard model

Table 1: Single covariate analysis of Cox proportional hazards on time to death of MDR-TB patient at St.Peter's specialized hospital from January 3, 2014 to October 28, 2016.

Covariate	Categories	Parameter	Hazard ratio	se(coef)	p-value	95% CI of Hazard ratio
Age		0.051	1.053	0.009	0.000	[1.034,1.071]
BMI		-0.049	0.952	0.040	0.215	[0.881,1.029]
Sex	Female(ref)					
	Male	0.285	1.329	0.246	0.250	[0.821,2.152]
TB type	Pulmonary(ref)					
	Extra pulmonary	-1.2684	0.281	0.5908	0.0320	[0.088,1.896]
TB status	No(ref)					
	Yes	-1.193	0.303	0.295	0.000	[0.170,0.540]
	Addis Ababa(ref)					
	Oromiya	0.331	1.392	0.314	0.293	[0.752,2.577]
Region	Amhara	0.475	1.607	0.596	0.426	[0.500,5.172]
	SNNP	-0.360	0.698	0.722	0.619	[0.170,2.875]
	Other	-0.190	0.827	0.5962	0.750	[0.257,2.661]
HIV co-	Negative (ref)					
infection	Positive	0.965	2.624	0.245	0.000	[1.625,4.237]
Smoking	No(ref)					
Status	Yes	1.112	3.040	0.267	0.000	[1.8 ,5.132]
Alcohol	No (ref)					
use	Yes	0.6624	1.939	0.301	0.028	[.076,3.496]
	Illiterate (ref)					
Level of	Primary	-0.347	0.707	0.3493	0.321	[0.357,1.402]
education	Secondary	-0.481	0.618	0.3939	0.222	[0.286,1.337]
	Above secondary	-0.203	0.816	0.3878	0.600	[0.382,1.745]
Co-	No(ref)					

morbidity	Yes	1.221	3.390	0.3299	0.000	[1.776,6.472]
	New (ref)					
Registrati	Relapse	-1.153	0.316	0.335	0.001	[0.164,0.608]
on group	Failure of new regime	-1.081	0.340	0.318	0.001	[0.182,0.632]
	Other	-1.642	0.194	0.4503	0.000	[0.080,0.468]

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016

Table 2 Comparison of univariate and multivariate cox proportional hazard model

Covariate	Categories	Univarite model		Multivariate model	
		Hazard ratio(95% CI)	p-value	Hazard ratio(95% CI)	p-value
Age		1.053[1.034,1.071]	0.000**	1.044[1.025,1.063]	0.000**
BMI		0.952[0.881,1.029]	0.215*	0.915[0.841,0.996]	0.040*
Sex	Female(ref)				
	Male	1.329[0.821,2.152]	0.250		
TB type	Pulmonary				
	Extra pulmonary	0.281[0.088,0.896]	0.032*	0.455[0.140 1.478]	0.190
TB status	No(ref)				
	Yes	0.303[0.170,0.540]	0.000**	0.510[0.110,2.372]	0.391
	Addis Ababa(ref)				
	Oromiya	1.392[0.752,2.577]	0.293		
Region	Amara	1.607[0.500,5.172]	0.426		
	SNNP	0.698[0.170,2.875]	0.619		
	Other	0.827[0.257,2.661]	0.750		
HIVco-	Negative (ref)				
infection	Positive	2.624[1.625,4.237]	0.000**	1.727[1.044, 2.858]	0.024*
Smoking	No(ref)				
	Yes	3.040[1.8 ,5.132]	0.000**	2.807[1.299 ,6.064]	0.009**
Alcohol	No (ref)				

	Yes	1.939[1.076,3.496]	0.028*	0.878[0.311 ,1.690]	0.457
	Illiterate (ref)				
Level of	Primary	0.707[0.357,1.402]	0.321		
education	Secondary	0.618[0.286,1.337]	0.252		
	Above secondary	0.816 [0.382,1.745]	0.600		
Comorbidity	No (ref)				
	Yes	3.390[1.776,6.472]	0.000**	2.130[1.088 ,4.173]	0.027*
Registration	New (ref)				
group	Relapse	0.316[0.164,0.609]	0.001**	0.626[0.136,2.883]	0.548
	Failure new regime	0.318[0.169,0.601]	0.000**	0.767[0.169,3.474]	0.730
	Other	0.239[0.106,0.536]	0.001**	0.450[0.090,2.340]	0.349

Source: Saint Peter's Specialized TB hospital, Ethiopia; from January 3, 2014 to October 28, 2016

**P-value <0.25 was statically significant for univaraite model.*

**P-value<.005 was statistically significant for multivariate model.*

Ref=Reference, CI=Confidence interval

Table 3: Step wise variable selection procure

Covariate	*β*	*HR*	*SD*	*z-value*	*p-value*	95 % Cl of HR
Age	*0.043*	*1.044*	*0.009*	*4.539*	*5.64e-06* ***	[1.025 1.063]
BMI	*-0.089*	*0.915*	*0.043*	*-2.054*	*0.03998* *	[0.841 0.996]
as.factor(HIV)1	*0.546*	*1.727*	*0.257*	*2.127*	*0.03343* *	[1.043 2.858]
as.factor(smok)1	*1.032*	*2.807*	*0.393*	*2.626*	*0.00863* **	[1.299 6.064]
as.factor(com)1	*0.756*	*2.130*	*0.343*	*2.205*	*0.02747* *	[1.088 4.173]
as.factor(TBs)1	*-0.673*	*0.5103*	*0.784*	*-0.858*	*0.39082*	[0.110 2.372]
as.factor(alco)1	*-0.321*	*0.725*	*0.431*	*-0.744*	*0.45670*	[0.311 1.690]

as.factor(TBt)1	*-0.788*	*0.455*	*0.601*	*-1.310*	*0.19032*	[0.140 1.480]
as.factor(grou)2	*-0.468*	*0.626*	*0.779*	*-0.601*	*0.54790*	[0.136 2.883]
as.factor(grou)3	*-0.266*	*0.767*	*0.771*	*-0.344*	*0.73048*	[0.169 3.474]
as.factor(grou)4	*-0.778*	*0.460*	*0.830*	*-0.936*	*0.34907*	[0.090 2.340]

Concordance= 0.771 (se = 0.036) , Rsquare= 0.212 (max possible= 0.902) , Likelihood ratio test= 76.44 on 11 df, p=7.177e-12 , Wald test = 79.49 on 11 df, p=1.856e-12Score (logrank) test = 88.71 on 11 df, p=2.986e-14

Source: Saint Peter's Specialized TB hospital, Ethiopia; January 3, 2014 to October 28, 2016

**P-value<.005 was statistically significant. Hazard ratio =HR,*

CI=Confidence interval , β=Parameter of estimate, SD =Standard deviation

Table 4: Stepwise after registration group is removed

Covariate	*β*	*HR*	*SD*	*z-value*	*p-value*	95 % Cl of HR
Age	0.042	1.043	0.009241	4.559	5.13e-06 ***	[1.0243 1.0621]
Bmi	-0.088	0.916	0.042545	-2.059	0.039488 *	[0.8428 0.9958]
as.factor(HIV)1	0.539	1.715	0.256972	2.099	0.035843 *	[1.0363 2.8376]
as.factor(smok)1	1.088	2.968	0.377430	2.883	0.003943 **	[1.4166 6.2199]
as.factor(com)1	0.782	2.186	0.340739	2.295	0.021732 *	[1.1209 4.2624]
as.factor(TBs)1	-1.090	0.336	0.304331	-3.583	0.000339 ***	[0.1851 0.6102]
as.factor(alco)1	-0.361	0.697	0.419375	-0.861	0.389154	[0.3063 1.5854]
as.factor(TBt)1	-0.826	0.438	0.600465	-1.375	0.169086	[0.1350 1.4207]

Concordance= 0.768 (se = 0.036) Rsquare= 0.208 (max possible= 0.902)

Likelihood ratio test= 74.56 on 8 df, p=6.029e-13 Wald test = 77.84 on 8 df, p=1.327e-13

Score (logrank) test = 86.24 on 8 df, p=2.665e-15

Source: Saint Peter's Specialized TB hospital, Ethiopia; from Jan 3, 2014 to October 28, 2016

**P-value<.005 was statistically significant. Hazard ratio =HR,*

CI=Confidence interval , β=Parameter of estimate, SD =Standard deviation

Table 5: After variable alcohol use is removed

Covariate	*β*	*HR*	*SD*	*z-value*	*p-value*	95 % Cl of HR	
Age	0.042962	1.043898	0.009242	4.649	3.34e-06 ***	[1.0252	1.0630]
Bmi	-0.090560	0.913419	0.042487	-2.131	0.033050 *	[0.8404	0.9927]
as.factor(HIV)1	0.516100	1.675481	0.255703	2.018	0.043554 *	[1.0150	2.7656]
as.factor(smok)1	0.854755	2.350798	0.274401	3.115	0.001840 **	[1.3729	4.0252]
as.factor(com)1	0.769042	2.157699	0.340802	2.257	0.024035 *	[1.1064	4.2081]
as.factor(TBs)1	-1.098333	0.333427	0.303683	-3.617	0.000298 ***	[0.1839	0.6046]
as.factor(TBt)1	-0.781259	0.457829	0.597953	-1.307	0.191363	[0.1418	1.4780]

Rsquare= 0.206 (max possible= 0.902) Likelihood ratio test= 73.82 on 7 df, p=2.489e-13

Wald test = 76.82 on 7 df, p=6.106e-14 Score (logrank) test = 85.49 on 7 df, p=9.992e-16

Source: Saint Peter's Specialized TB hospital, Ethiopia; f from January 3, 2014 to October 28, 2016

**P-value<.005 was statistically significant. Hazard ratio =HR,*

CI=Confidence interval , β=Parameter of estimate, SD =Standard deviation

Table 6: After variable TB site (TB type) removed

Covariate	β	HR	SD	Z-value	P-value	95% CLOF HR	
Age	0.044	1.045363	0.009235	4.804	1.56e-06 ***	[1.0266	1.0645]
Bmi	-0.095	0.909269	0.042528	-2.237	0.025318 *	[0.8366	0.9883]
as.factor(HIV)1	0.526	1.691695	0.254551	2.065	0.038892 *	[1.0272	2.7861]
as.factor(smok)1	0.853	2.346735	0.274904	3.103	0.001916 **	[1.3692	4.0222]
as.factor (com)1	0.822	2.276161	0.338912	2.427	0.015230 *	[1.1714	4.4227]
as.factor(TBs)1	-1.132	0.322404	0.302958	-3.736	0.000187 ***	[0.1780	0.5838]

R square= 0.201 (max possible= 0.902) Likelihood ratio test= 71.67 on 6 df, p=1.856e-13

Wald test = 76.08 on 6 df, p=2.298e-14 Score (logrank) test = 84.01 on 6 df, p=5.551e-16

Source: Saint Peter's Specialized TB hospital, Ethiopia; from Jan 3, 2014 to October 28, 2016

**P-value<.005 was statistically significant. Hazard ratio =HR, CI=Confidence interval, β=Parameter of estimate, SD =Standard deviation*

Table 7:Test of assumption of constant proportional hazard after smoking is stratified

Covariate	Rho	Chisq	p
Age	0.06421	0.3399	0.560
Body mass index	-0.05383	0.1549	0.694
HIV co-infection	0.03243	0.0799	0.777
Comorbidity	0.13077	1.2025	0.273
TB status	-0.00573	0.0023	0.962
GLOBAL	NA	1.8662	0.867

Appendix B: Kaplan-Meier survivor estimates for different categories or groups

Figure 1: Plots of Kaplan-Meier survivor estimates for different categories or groups

Figure 1.1 kaplan-meier survival estimate for sex categories.

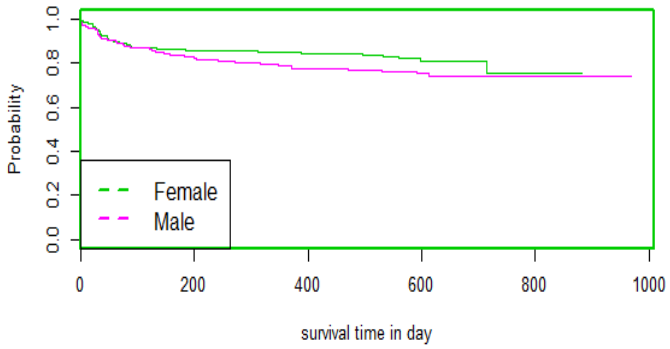


Figure 1.2 kaplan-meier survival estimate for TB type categories.

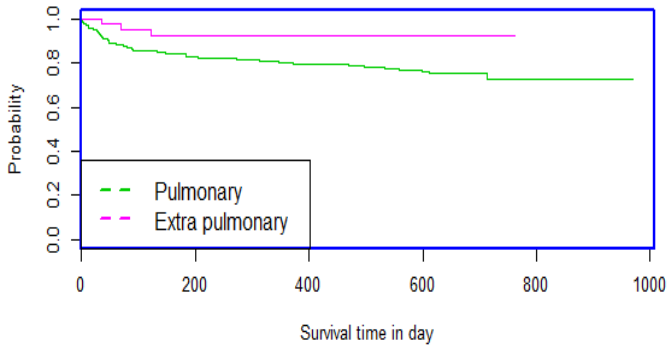


Figure 1.3 kaplan-meier survival estimate for TB status categories.

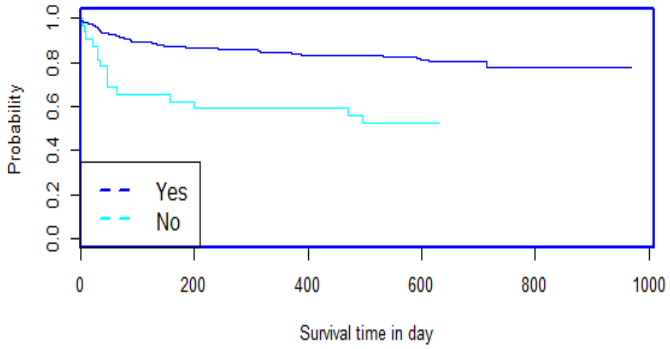


Figure 1.4 kaplan-meier survival estimate for smoking status categories

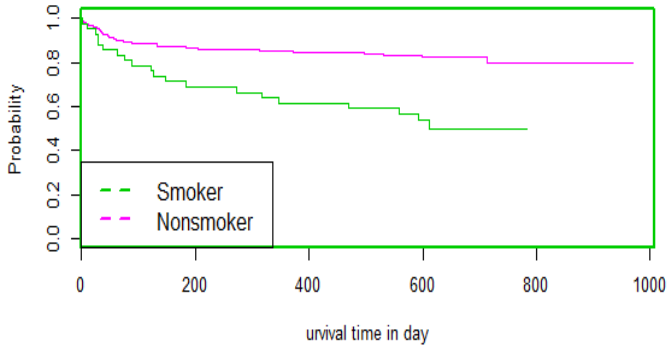


Figure 1.5 kaplan-meier survival estimate for alcohol use categories.

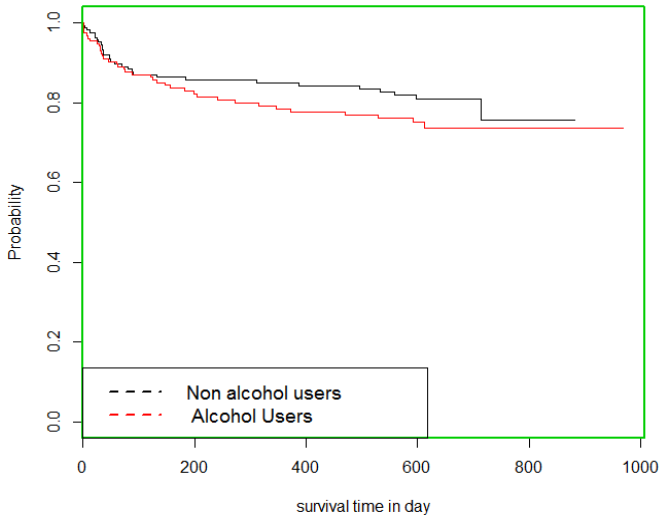


Figure 1.6 kaplan-meier survival estimate for HIV co-infection.

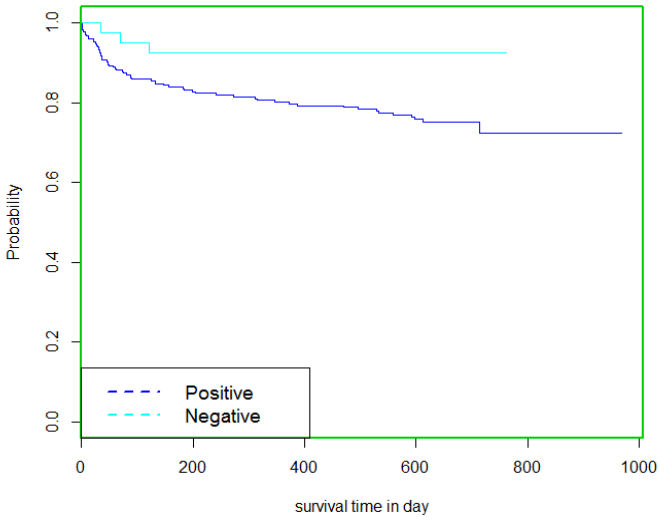


Figure 1.7 kaplan-meier survival estimate for co-morbidity categories.

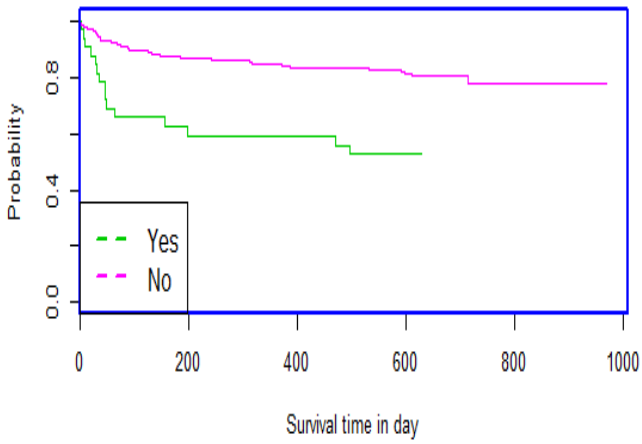


Figure 1.8 kaplan-meier survival estimate for level of education.

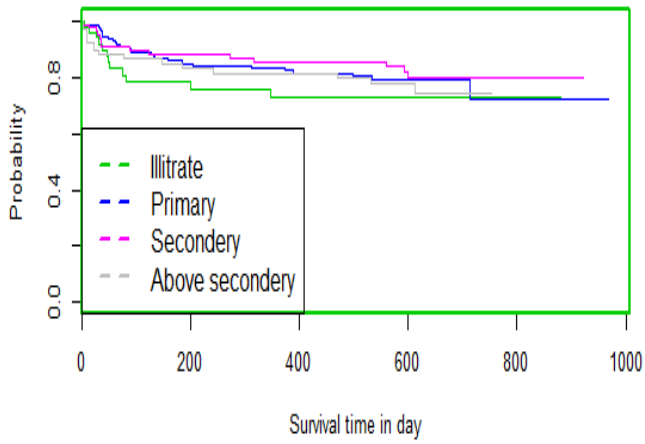


Figure 1.9 kaplan-meier survival estimate for region.

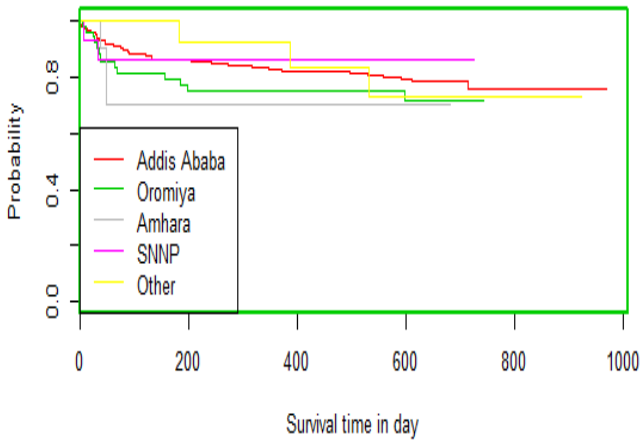
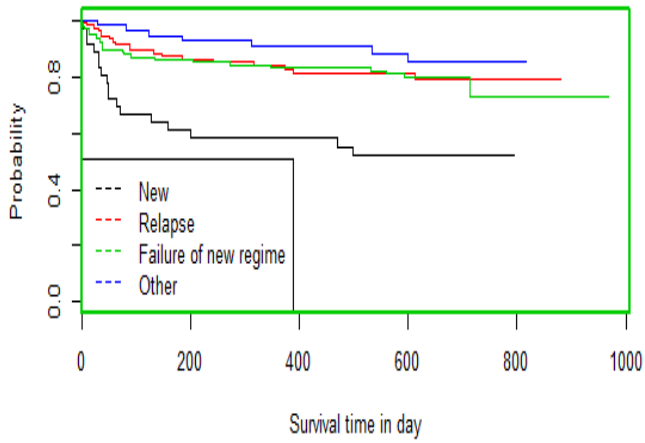


Figure 1.10 kaplan-meier survival estimate for registration group.



Appendix C: Cox PH model diagnosis plots

Figure 2: Plots of scaled Schoenfeld residuals against time for the covariates.

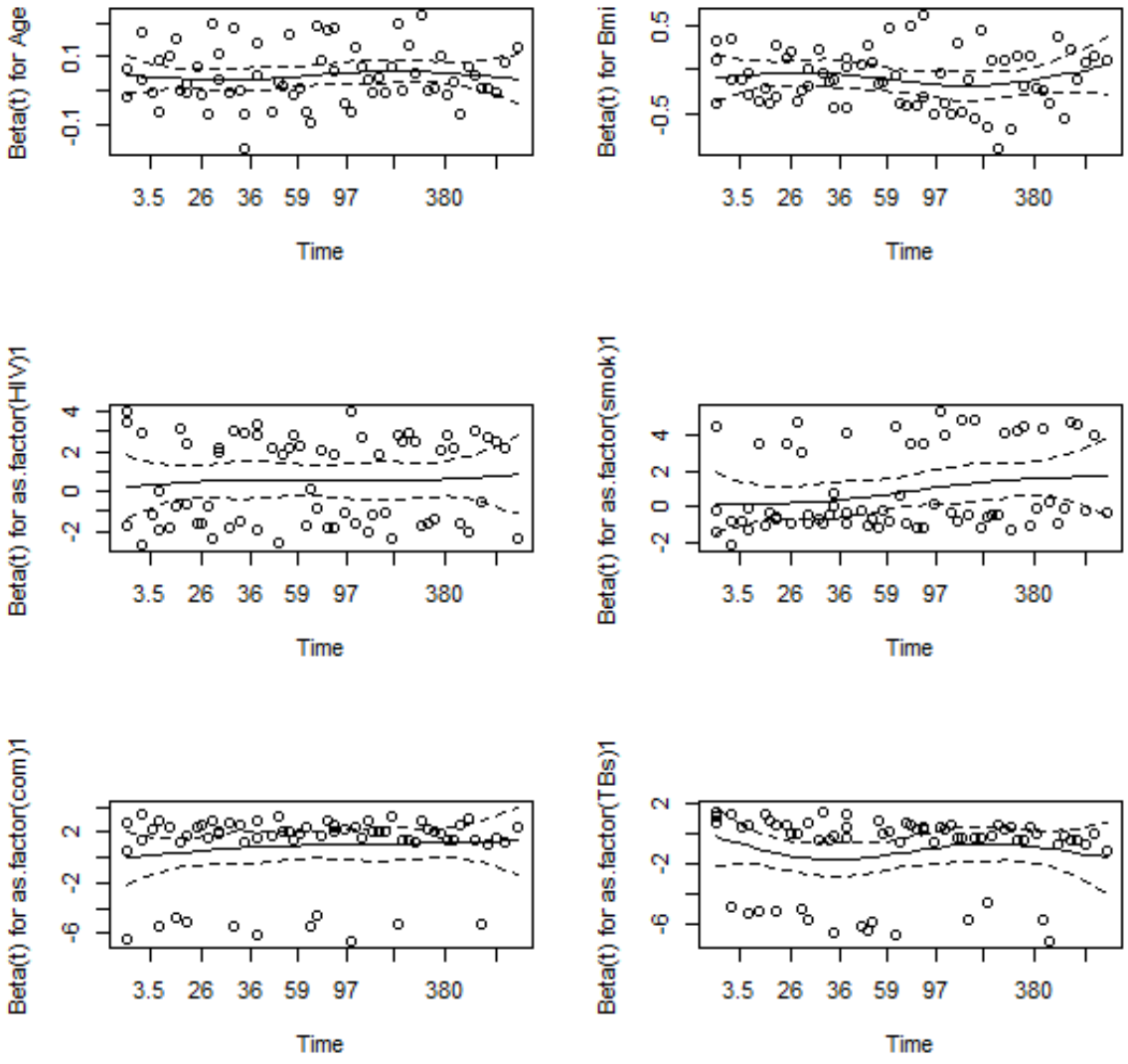


Figure 3: plot of observation verse scale change in coefficients

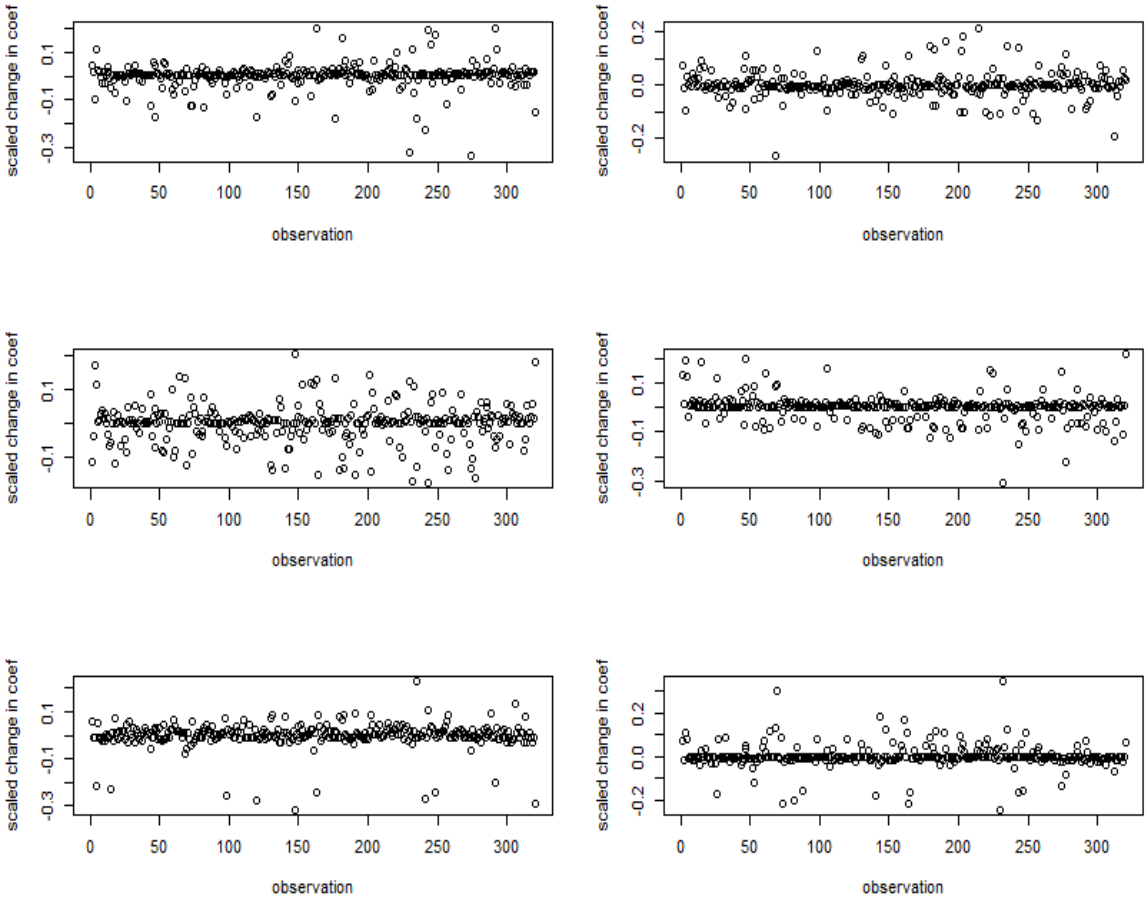
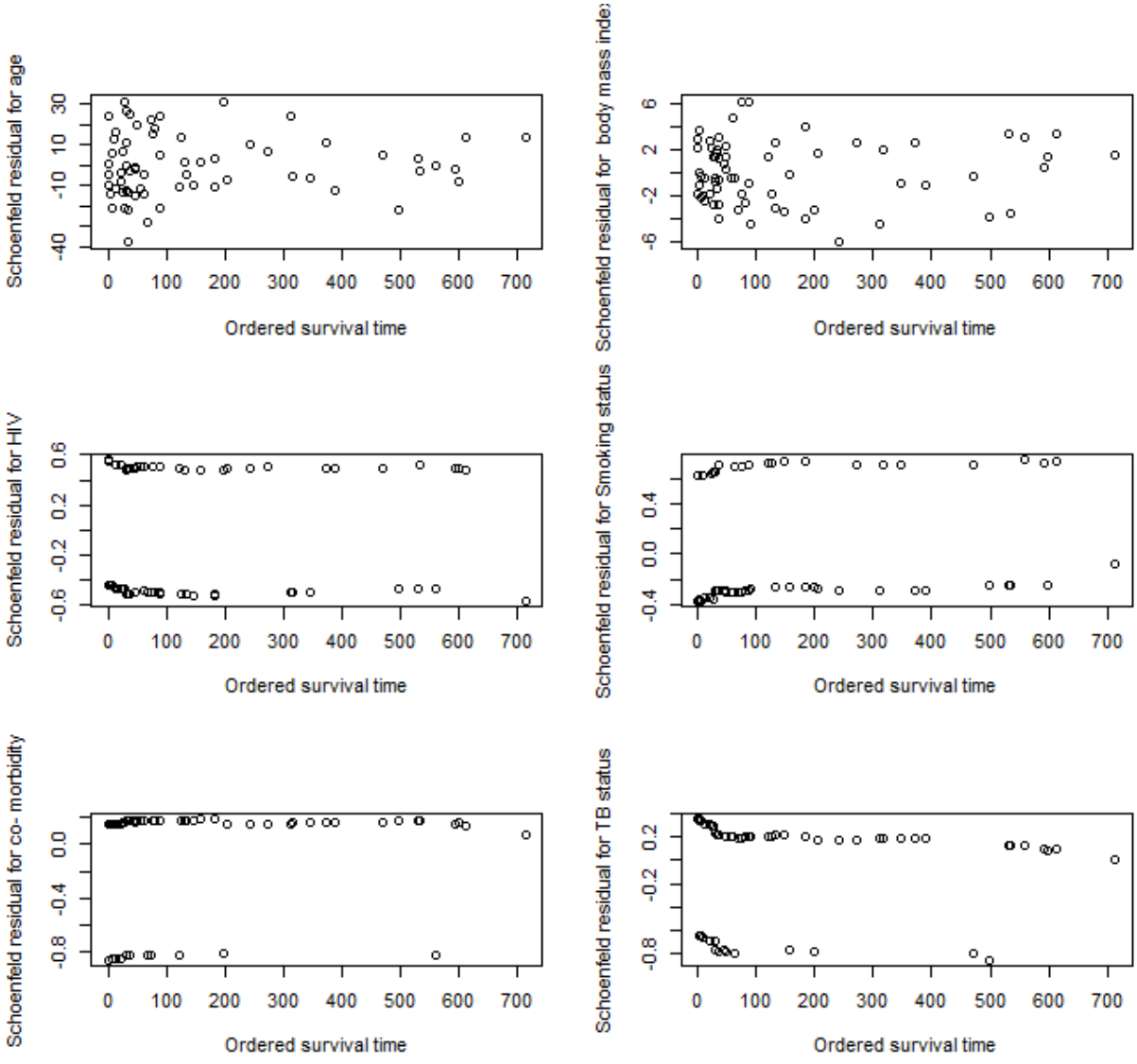


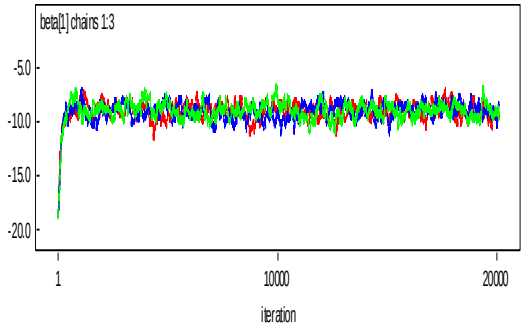
Figure 4: Plot of ordered survival time verses schoenfeld residual of covariate

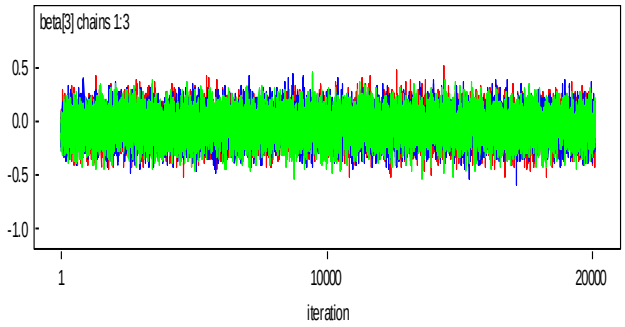


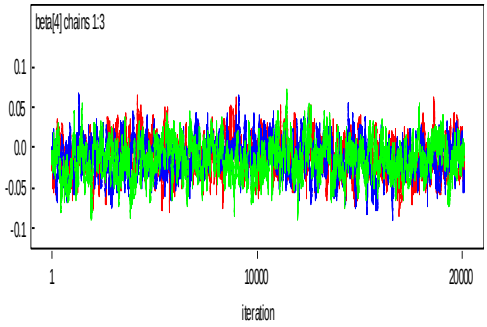
Appendix D: Bayesian parameter estimate and test of convergence

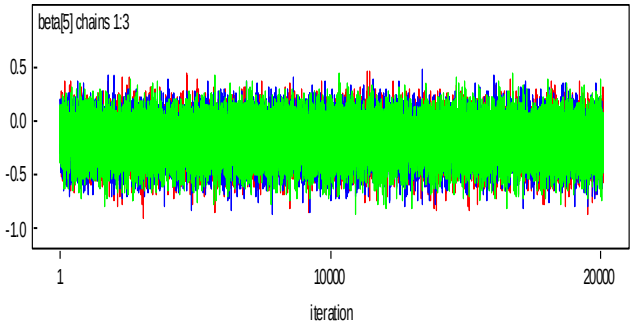
Table 8: Bayesian Parameter Estimate for Covariates with their Corresponding 95% credible Interval

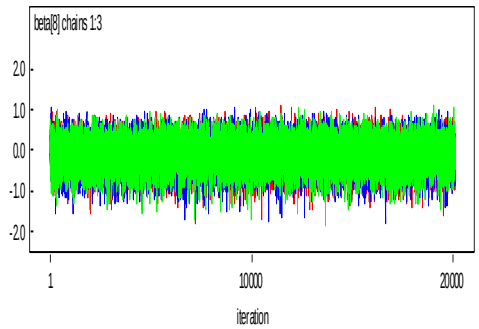
Covariate	Nodes	mean	Sd	MC error	2.5%	median	97.5%	sample
Constant	beta[1]	-9.081	0.8743	0.03722	-10.54	-9.026	-7.715	60300
Age	beta[2]	0.012	0.0055	9.544E-5	8.052E-4	0.0117	0.0225	60300
Sex Male	beta[3]	-0.047	0.1273	0.001318	-0.2967	-0.04776	0.2036	60300
Bmi	beta[4]	-0.013	0.0216	7.147E-4	-0.05555	-0.01264	0.0282	60300
TB status Pulmonary	beta[5]	-0.176	0.1719	8.937E-4	-0.5235	-0.1728	0.1499	60300
Tb site Yes	beta[6]	-0.549	0.5872	0.02344	-1.818	-0.5176	-0.5066	60300
Oromiya	beta[7]	0.065	0.1663	0.001725	-0.2681	0.06645	0.3815	60300
Amhara	beta[8]	-0.081	0.3505	0.002532	-0.8093	-0.06602	0.5601	60300
SNNP	beta[9]	-0.250	0.2926	0.002554	-0.8563	-0.2389	0.2973	60300
Other	beta[10]	0.301	0.263	0.003131	-0.2332	0.3085	0.8001	60300
HIV Positive	beta[11]	-0.012	0.1383	0.001037	-0.287	-0.01154	0.2562	60300
Smoker	beta[12]	0.155	0.2154	0.001541	0.0744	0.1575	0.5695	60300
Alcohol user	beta[13]	-0.069	0.2318	0.00189	-0.5321	-0.06693	0.3772	60300
Primary	beta[14]	-0.173	0.1776	0.002978	-0.513	-0.1754	0.1856	60300
Secondary	beta[15]	-0.328	0.1938	0.002736	-0.7051	-0.3294	0.0598	60300
Above secondary	beta[16]	-0.219	0.2108	0.003137	-0.6286	-0.2205	0.1992	60300
Comorbidity yes	beta[17]	0.0558	0.1365	0.001965	0.0106	0.05531	0.3233	60300
Relapse	beta[18]	0.1806	0.5665	0.02231	-0.8243	0.1408	1.431	60300
Failure of new regime	beta[19]	0.1268	0.5636	0.02227	-0.8684	0.08609	1.378	60300
Other	beta[20]	0.0431	0.569	0.02215	-0.9681	0.002915	1.294	60300
	sigma	1.524	0.1146	0.004822	1.368	1.517	1.688	60300

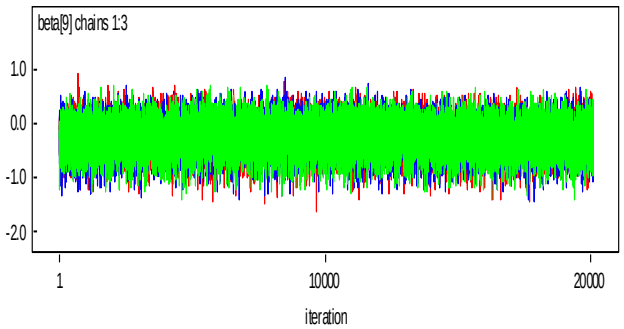

beta[1] chains 1:3
-5.0
-10.0
-15.0
-20.0
1
10000
20000
iteration

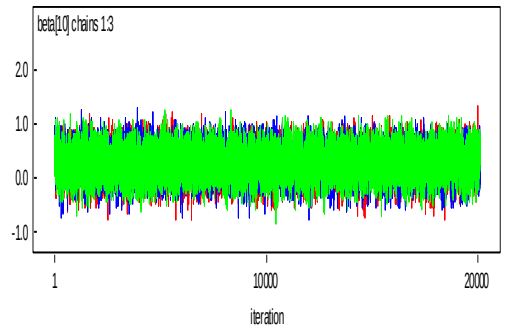

beta[3] chains 1:3
0.5
0.0
-0.5
-1.0
1
10000
20000
iteration

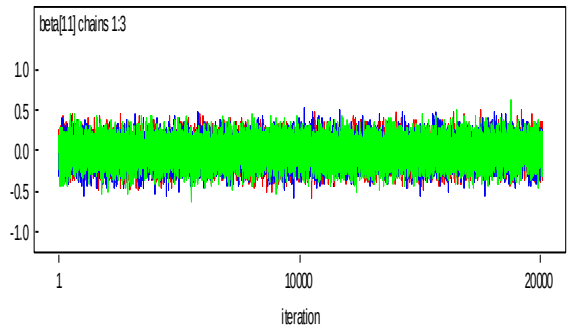

beta[4] chains 1:3
0.1
0.05
0.0
-0.05
-0.1
1
10000
20000
iteration


beta[5] chains 1:3
0.5
0.0
-0.5
-1.0
1
10000
20000
iteration


beta[8] chains 1:3
2.0
1.0
0.0
-1.0
-2.0
1
10000
20000
iteration


beta[9] chains 1:3
1.0
0.0
-1.0
-2.0
1
10000
20000
iteration


beta[10] chains 1:3
2.0
1.0
0.0
-1.0
1
10000
20000
iteration


beta[11] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
1
10000
20000
iteration

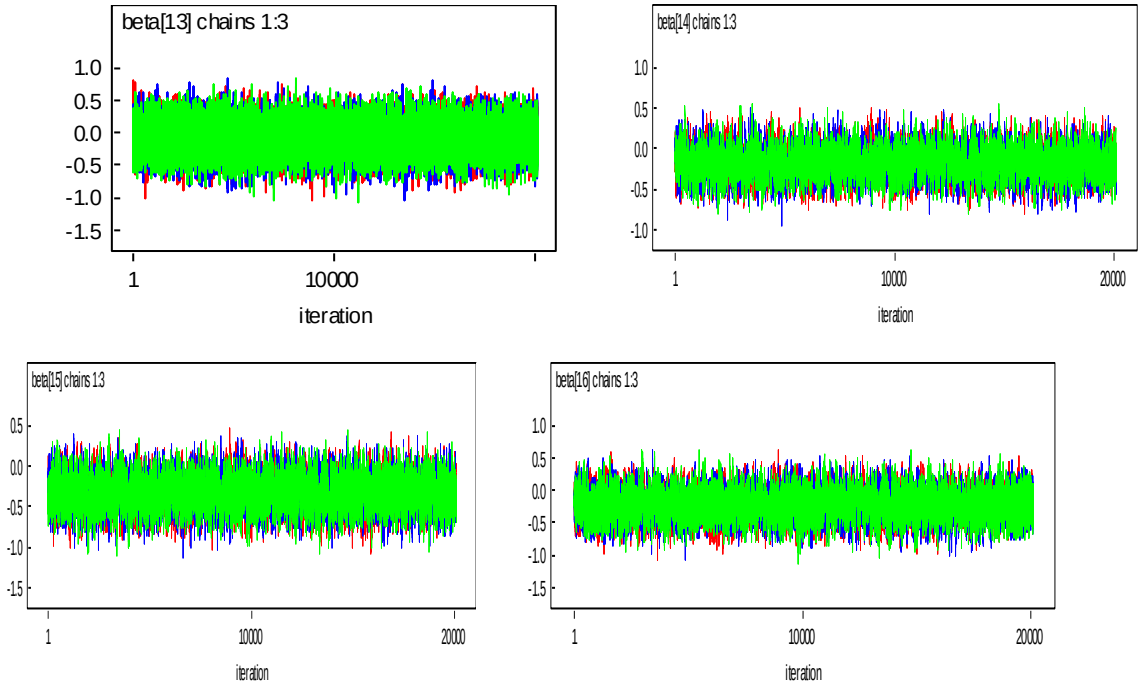
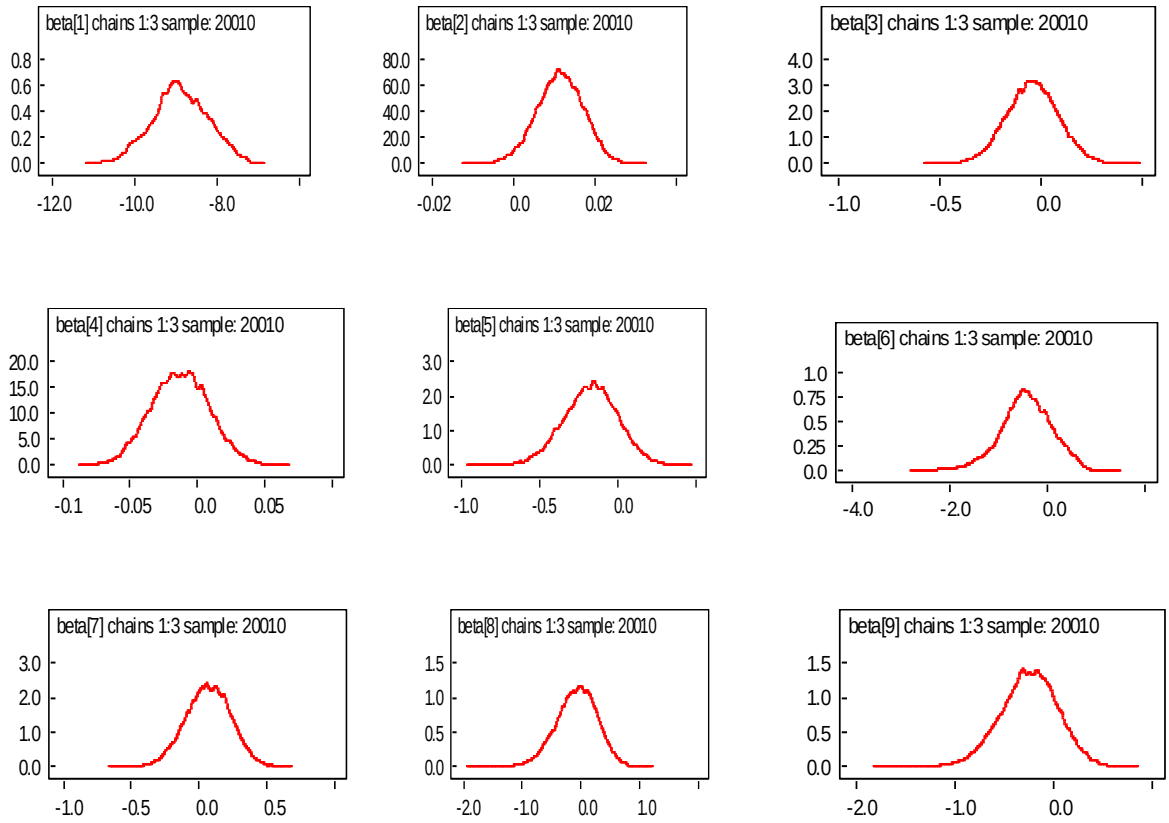


Figure 5:Time series (history) plot for all parameter to test convergence



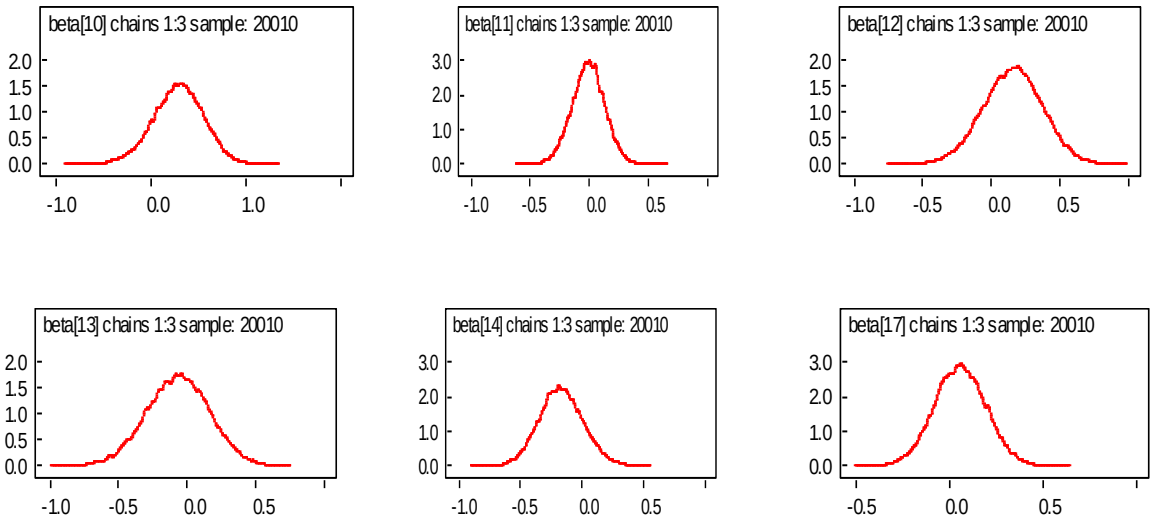
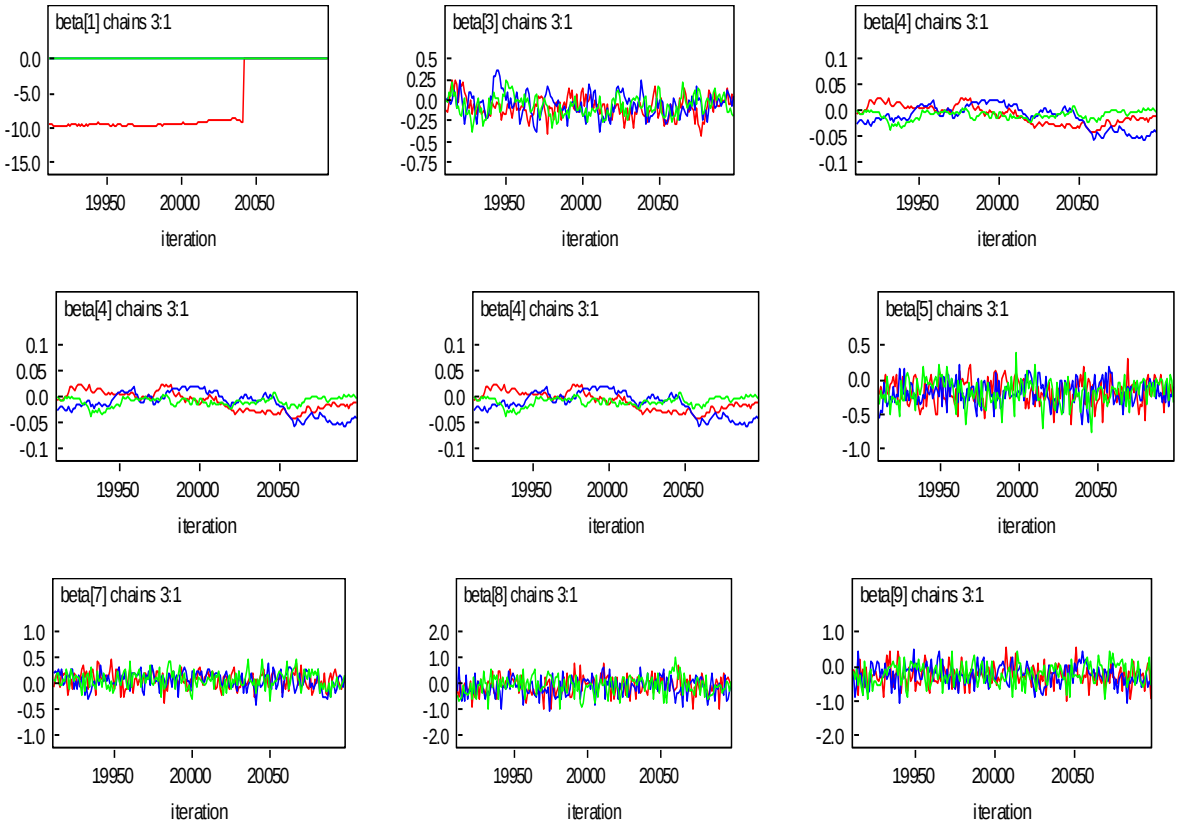


Figure 6: Convergence Analysis using Density for all parameters



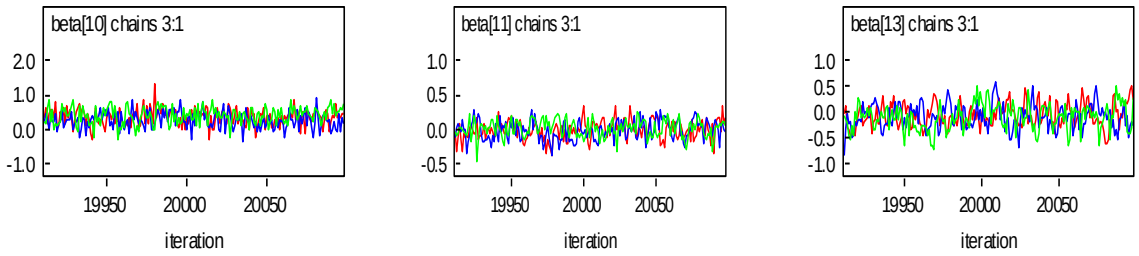


Figure 7: Convergence Analysis using Trace plot for parameters

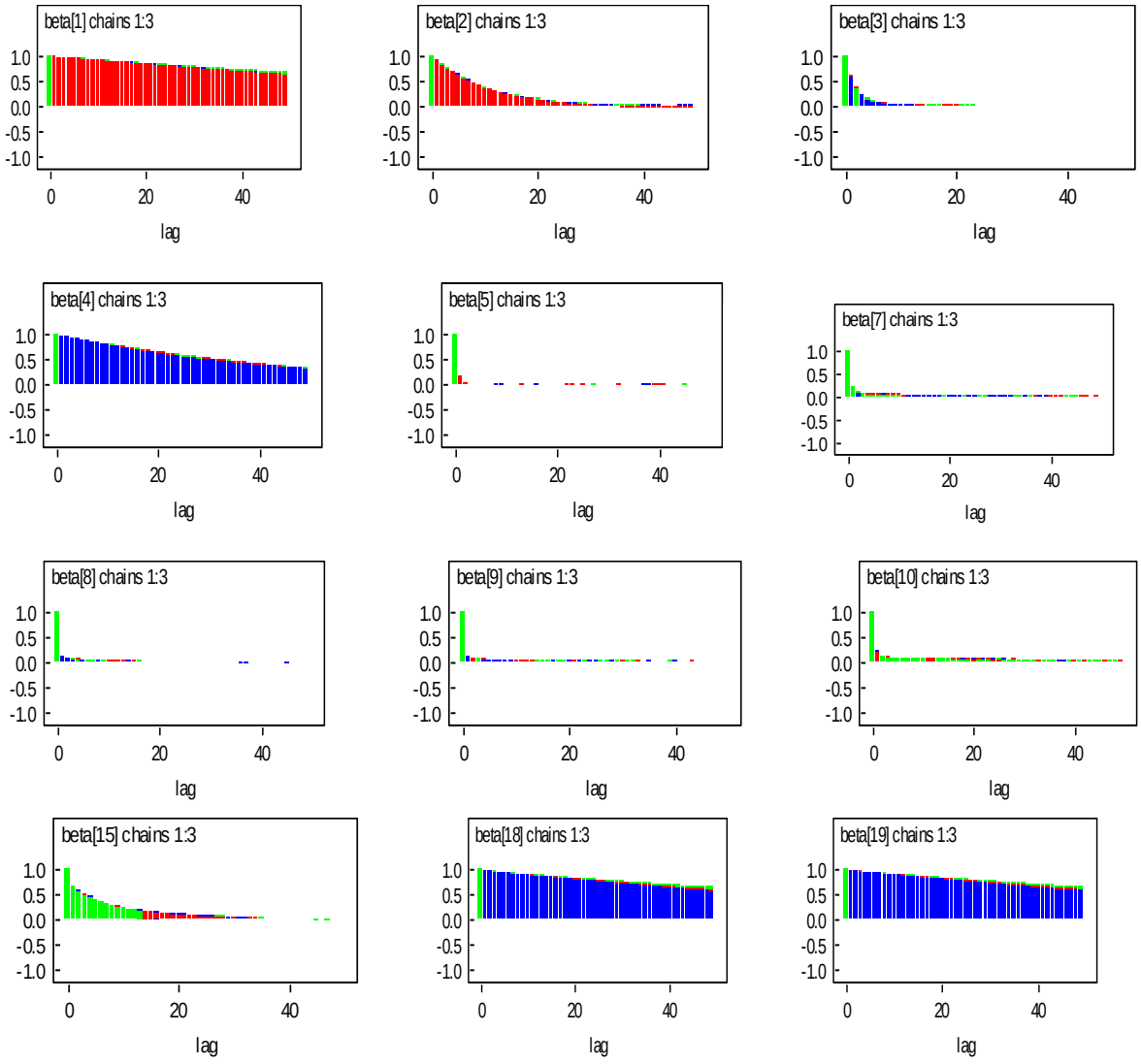


Figure 8: Autocorrelations plot for all parameters