
አዲስ አበባ ዩኒቨርሲቲ
Addis Ababa University

# ADDIS ABABA UNIVERSITY
# OFFICE OF GRADUATE PROGRAMS
# COLLEGE OF NATURAL SCIENCES
# DEPARTMENT OF STATISTICS

RISK FACTORS OF MORTALITY AMONG PATIENTS WITH MULTI-DRUG RESISTANT TUBERCULOSIS IN ADDIS ABABA AND GONDAR, ETHIOPIA.

By

Solomon Molalign

Approved by the board of examiners:

Mekonnen Taddesse (Asst. Prof.) Signature_____Date_____
Examiner, Chairman, Department graduate committee

Shibru Temesgen (Dr.) Signature_____ Date_____
Examiner

Eshetu Wencheko (Prof.) Signature_____ Date_____
Advisor

Contents

**List of tables**

**List of Figures**

## Acknowledgement

I would like to express my sincere appreciation to my thesis advisor Prof. Eshetu Wencheko for his valuable suggestions, comments and patience during the entire time of the study.

My deep appreciation and thanks also goes to Dr. Mulugeta, from ALERT Hospital, in providing the data collected and my brothers Meseret Hailu and Tamirat Mollalign in providing other related materials to the study.

I am highly indebted to the Department of Statistics for financial support.

I would like to express my gratitude to the ALERT hospital and Gondar University Teaching and referral hospital administration office for allowing me to obtain data.

Finally, I acknowledge the help of Wollo University for sponsoring my MSc study.

## ACRONYMS

| | |
|---|---|
| ALERT | All African Leprosy Rehabilitation and Training Centre |
| AMK | Amikacin |
| ART | Antiretroviral therapy |
| CDC | Centres of Disease Control and Prevention |
| CI | Confidence Interval |
| CPM | Capreomycin |
| DOTS | Directly Observed Treatment, Short-Course |
| DST | Drug susceptibility test |
| E | Ethambutol |
| FMOH | Ministry of Health |
| GHC | Global Health Committee |
| HIV | Human Immunodeficiency Virus |
| HR | Hazard Ratio |
| INH | Isoniazed |
| KAN | Kanamycin |
| LR | Likelihood Ratio |
| MDR-TB | Multidrug resistance tuberculosis |
| MLE | Maximum LikelihoodEstimate/Estimator |
| NJMRC | National Jewish Medical and Research centre |
| OR | Odds ratio |
| RIF | Rifampicin |
| STM | Streptomysin |
| TB | Tuberculosis |
| UK | United Kingdom |
| UNICEF | United Nations Children`s Fund |
| UOGTRH | University of Gondar Teaching and Referral Hospital |
| WHO | World Health Organization |
| XDR-TB | Extensive drug resistance tuberculosis |

## Abstract

Risk factors of mortality among patients with Multidrug resistance Tuberculosis in Addis Ababa and Gondar, Ethiopia.

Solomon Molalign
Addis Ababa University, 2015

Multi-drug resistant TB (MDR-TB) occurs when the causative agent, *Mycobacterium tuberculosis*, becomes resistant to isonazid and rifampin, the two most effective drugs commonly used to treat TB. Despite high rates of MDR-TB in Ethiopia, little data exist on the prevalence of or risk factors for drug-resistant tuberculosis.

The objective of this study was to assess rates of and risk factors for MDR-TB in ALERT Hospital, Addis Ababa and Gondar University Teaching and Referral Hospital, Gondar, Ethiopia.
The study included 342 MDR-TB patients (142 from ALERT hospital and 200 from UOGTRH) under treatment from August, 2011 to August, 2014 for whom data for variables of interest were complete. Descriptive statistics, univariate and multivariate analyses were used as analytic methods. The Kaplan-Meier method was used to estimate the survival time and Cox's regression model was employed to identify the covariates that may have statistical significant effect on the survival of MDR-TB patients.

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 ($\widehat{HR}$ 4.7161; 95%; $CI$ 2.1861 – 10.1740), resistance to INH, RIF and at least one of other drugs (E, STM, KAN, AMK & CPM) ($\widehat{HR}$ 2.9771; 95%; $CI$ 1.3586 – 6.5238), smoking ($\widehat{HR}$ 3.17; 95%; $CI$ 1.32 – 7.64), weight ($\widehat{HR}$ 0.9093; 95%; $CI$ 0.8760 – 0.9440) and age ($\widehat{HR}$ 1.2199; 95%; $CI$ 1.0681 – 1.3933).

The mortality rate of patients was high at the earlier times of treatment. Laboratory and clinical factors were associated with increased risk of mortality. Thus, those patients with poor laboratory/clinical characteristics should be identified and treated before they progress to advanced disease stages.

# 1. INTRODUCTION

## 1.1. Background

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 (WHO, 2013).

When a person develops active TB (disease), the symptoms (cough, fever, night sweats, weight loss etc.) may be mild for many months. This can lead to delays in seeking care, and results in transmission of the bacteria to others. People ill with TB can infect up to 10-15 other people through close contact over the course of a year. Without proper treatment up to two-thirds of people ill with TB will die. Tuberculosis mostly affects young adults, in their most productive years. However, all age groups are at risk (WHO, 2013).

Over 95% of cases and deaths are in developing countries. According to the world Health Organization (WHO) Global TB Report 2012, in the year 2011 there were an estimated 8.7 million incident cases and 12 million prevalent cases of TB globally. Of these 1.1 million (13%) were people living with HIV. About 26% of TB cases occurred in Africa in 2011. The proportion of TB cases co-infected with HIV is highest in countries in the African region; overall, the African region accounted for 79% of TB cases among people living with HIV. In 2011, an estimated 990,000 deaths occurred among HIV negative cases of TB including 0.30 million deaths among women. This is equivalent to 14 deaths per 100,000 populations. In addition, there were an estimated 0.43 million deaths among TB cases that were HIV positive. Thus in total, approximately 1.4 million people died of TB in 2011 (FMOH, 2013).

Countries conduct surveillance of anti-TB drug resistance as component of any TB control programme with four main objectives: a) measure the burden of drug-resistant TB and accurately plan treatment programmes with second-line drugs; b) assess epidemiological trends as a reflection of the effectiveness of implemented drug-resistant TB prevention and control activities; c) design effective empirical, standardized regimens for the treatment of TB, particularly for patients who have already been treated for TB and return with the disease; and d) promptly identify local outbreaks of drug-resistant TB in order to respond in a timely way (WHO 2010).

Multi-drug resistant TB (MDR-TB) occurs when the causative agent, *Mycobacterium tuberculosis*, becomes resistant to isonazid and rifampin, the two most effective drugs commonly used to treat TB. MDR-TB results from either infection with organisms which are already drug-resistant or may develop in the course of a patient's treatment [WHO, 2013].

About 3.6% of new TB patients in the world have multidrug resistant strains (MDR-TB). Levels are much higher _about 20%_ among those previously treated for TB. The frequency of MDR-TB varies substantially between countries. About 10% of MDR-TB cases are also resistant to the two most important second-line drug classes, or extensively drug-resistant TB (XDR-TB). By September 2013, 92 countries had reported at least one XDR-TB case. About 480, 000 people developed MDR-TB. More than half of these cases were in India, China and the Russian Federation. It is estimated that about 9.0% of MDR-TB cases had XDR-TB. About 170,000 MDR-TB deaths are estimated to have occurred in 2012. Detection of MDR-TB patients is increasing. Almost 84,000 patients with MDR-TB were identified by WHO globally in 2012, up from 62,000 in 2011. The biggest increases were in India, South Africa and Ukraine. Enrolments in MDR-TB treatment in 2012 were equivalent to one in four of the MDR-TB cases estimated to occur among pulmonary TB patients. The ratio of enrolments to MDR-TB cases detected was about 92% but varied widely by country. About 48% of patients with MDR-TB enrolled on treatment in 2010 were reported to have been successfully treated (lower graphic). More patients have been monitored for outcomes over time but the proportion without outcome data remains high (WHO, 2013). Asia bears the burden of the epidemic as almost 50% of MDR-TB cases worldwide are estimated to occur in China and India (Ziglon, 2013)

MDR-TB patients respond poorly to short course chemotherapy and need to be treated intensively for up to 24 months with a regimen based on reserve anti-TB drugs (WHO, 2013).

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%. 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 (FMOH, 2013).

The Global Plan to Stop TB, 2011-2015 envisages that, in order to progress towards universal access, about one million MDR-TB patients need to be placed in treatment between 2011 and 2015. The Plan also aims to have at least 75% of MDR-TB patients completing their treatment successfully by 2015. In 2012, about 45% of cases targeted to be placed on treatment globally that year, were reported to have been enrolled. Among MDR cases that started treatment in 2010, the 75% treatment success threshold was achieved by 34 of the 107 countries reporting outcomes (WHO, 2013).

Of the 27 countries with a high burden of MDR-TB and extensively drug resistant TB (XDR-TB), 13 countries with data on treatment outcomes for MDR-TB cases reported a success of 25%-82% among patients that started treatment in 2007. However, it should be remembered that increases in the prevalence of resistance can be caused by poor or deteriorating TB control, the immigration of patients from areas of higher resistance, outbreaks of drug-resistant disease, and variations in surveillance methodologies (WHO, 2010).

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 (WHO, 2008). To

make things worse, in these TB and MDR-TB high burden countries patients stay in their communities for longer periods without being diagnosed or getting proper treatment. Even after diagnosis, because there are few diagnostic and treatment facilities and a lack of trained health professionals and drugs, patients do not start treatment immediately. This delay potentially allows easy spread of the disease to a large number of individuals within a short time (Hirpha et al, 2013).

Global Health Committee (GHC) reached a significant milestone in Ethiopia by initiating lifesaving therapy with 787 patients for drug-resistant TB since the program began in 2009. GHC initiated the countrywide program for drug-resistant TB in Addis Ababa at St Peter's Hospital in 2009 and in Gondar at the Gondar University Teaching and Referral Hospital in 2010 in partnership with the Ethiopian Ministry of Health, using approaches developed in Cambodia.

With funding from the Jolie-Pitt Foundation, the Lilly drug-resistant TB Partnership, and Annenberg Foundation, the first patients began with drug-resistant TB therapy under GHC in Ethiopia in 2009 in a converted isolation ward in St. Peter's Hospital in Addis Ababa. GHC has overcome numerous challenges along the way, including a year long delay in the availability of second-line TB drugs necessary to treat drug-resistant TB coupled with a year-long delay in the completion of the drug-resistant TB isolation ward at St. Peter's Hospital. But year-by-year progress has been steady. Patients receiving drug-resistant TB medicines and care are 74 in 2009 in Addis Ababa, 171, 342 and 600 patients in 2010, 2011 and 2012 in both Addis Ababa and Gondar. MDR-TB treatment was started in ALERT Hospital in 2011. This is the third MDR-TB centre in Ethiopia following St. Peter's TB Specialized Hospital in Addis Ababa and Gondar University Teaching and Referral Hospital.

Drug-resistant TB treatment requires the administration of five costly drugs with powerful side effects over a two-year period. This is compared with a six-month regimen of readily- available and easily tolerated drugs for "regular TB" where there is no drug resistance. It is particularly noteworthy in Ethiopia, which today ranks seventh among the 22 world's highest TB-burdened countries, and fifteenth among the 27 highest drug-resistant TB-burdened countries. Strikingly, WHO estimates that there are 6,000 new cases of drug-resistant TB each year in Ethiopia.

Outcomes of MDR-TB treatment have so far been not well described in Ethiopia. Therefore, examining cohorts who received a standardized second-line therapy and management of MDR-TB to determine the overall survival rate has a significant importance. Isoniazid, the most powerful mycobactericidal drug available, ensures early sputum conversion and helps in decreasing the transmission of TB. Rifampicin, by its mycobactericidal and sterilizing activities is crucial for preventing relapses. Thus, isoniazid and rifampicin are keystone drugs in the management of TB. While resistance to either isoniazid or rifampicin may be managed with other first-line drugs, resistant to both isoniazid and rifampicin (MDR-TB) demands treatment with second-line drugs. These drugs have limited sterilizing capacity and are not suitable for short course treatment. Thus, patients with MDR-TB require prolonged treatment with drugs that are less effective and more toxic (Sharma et al, 2005).

## 1.2. Statement of the problem

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. Ethiopia is one of the 27 high burden MDR-TB countries ranking 15$^{th}$ with more than 5000 estimated MDR-TB patients annually. Expanding access to MDR-TB therapy is urgently needed, yet poor implementation of such therapy can worsen the problem of XDR-TB. Understanding risk factors for death among MDR-TB patients is necessary to improve treatment outcomes (WHO, 2013). Hence, this study deals with survival and risk factors of mortality among MDR-TB patients.

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 elapsed 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, the research problems include: assessing and measuring survival/death status; identifying the risk factors of survival/death of MDR-TB patients; fitting appropriate survival model by including the significant risk factors on survival/death.

## 1.3. Objectives of the study

The general objective of the study is to identify predictors of all-cause mortality among patients with MDR-TB.

**Specific Objectives**

- ✓ To assess survival and predictors of mortality among patients under MDR-TB treatment in ALERT Hospital, Addis Ababa and University of Gondar Teaching and Referral Hospital, Gondar, Ethiopia.
- ✓ To compare the survival of MDR-TB patients with respect to their categories.

## 1.4. Significance of the study

- ➢ 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.
- ➢ The results would help donors and government to understand risk factors that influence the death of MDR-TB patients.
- ➢ The study could provide base-line data for detail and further studies in the future.

## 1.5. Limitation of the study

- ➢ The data were extracted from medical records of those already visited and registered at the hospitals.
- ➢ The study presumed that all deaths are caused by MDR-TB.
- ➢ The data did not provide information on other social determinants such as socioeconomic status, alcohol abuse, etc. and clinical factors such as MDR-TB category, adverse effects, mode of care, radiological findings, CD4 count etc. that would have been helpful in explaining MDR-TB rates.
- ➢ The study is based on baseline values of the variables of interest.

## 2. LITERATURE REVIEW

A number of researches had been undertaken to identify the risk factors for death of MDR-TB patients. Studies related to the survival of MDR-TB in Ethiopia are scarce. In this section, we review a few of them especially the ones that are closely related to the objectives of this study in addition to going through other studies.

Flament-Saillour et al (1999) conducted a study based on 51 cases of MDR-TB reported in France. The aim of the study was to analyze factors related to the persistence of MDR-TB sources of infection, the management, and the outcome of all MDR-TB cases that were reported in France during the year 1994. The data were retrospectively investigated in 1997. Fifty cases from metropolitan France and one from the French West Indies with MDR-TB were reported to the National Reference centre in 1994. The variables that were considered were age, sex, marital status, HIV status, CD4 count (HIV co-infected), prior treatment history, radiologic findings, microbiologic findings with drug susceptibility results and drug regimens. Categorical variables were compared using the chi-square. Odds ratios (OR) and their 95% confidence intervals ($CI_{95}$) were calculated using standard methods. The Mann-Whitney test was used for comparing continuous variables. Logistic regression models were used for multivariate analysis of factors associated with MDR-TB at the time of diagnosis. Survival analysis and identification of factors associated with poor survival in the MDR-TB group were done by the Kaplan-Meier method and Cox proportional hazard model, with death or treatment failure by December 1996 as the outcome variable. Factors significantly associated with MDR-TB in univariate analysis and factors clinically reasonable and not co-linear were included in the models. The Log-rank test was used to determine the level of statistical significance when comparing survival curves. The result of the study reveals that HIV co-infection, site of tuberculosis, and sputum smear results were not associated with MDR-TB in multivariate analysis.

A study aimed at describing the clinical, microbiological, molecular epidemiology and treatment of 90 MDR-TB cases was conducted in the UK (Drodniewski et al (2002)). All MDR-TB cases identified by health centres namely: the Public Health Laboratory Service Mycobacterium Reference Unit, the Scottish Mycobacteria Reference Laboratory, and PHLS Regional Centres for Mycobacteria from January 1, 1996 to June 30, 1997 were included. Data were sought on sex, age at diagnosis, ethnicity, country of birth, year of entry into the UK, history of prior TB,

and immunocompromised status. Clinical and radiological details were also sought. The date of death or whether the patient was alive on December 1, 1997 and December 1, 1998 was obtained in order to determine the length of survival from first diagnosis. Variables were assessed for their potential significance for survival using the log rank test. Relevant variables were included in a Cox proportional hazards model. Variables significant in the univariate analysis and to be included in Cox proportional hazards model were sex, immunocompromised status, HIV status, whether a fourth drug was given before MDR-TB diagnosis, whether appropriate three drug treatment was given based on in-vitro testing, whether there was resistance to pyrazinamide, ethambutol, prothionamide, amikacin, the number of drugs the infecting organism was resistant to, age in 10 year units, whether a bacterial culture was produced within 30 days, and whether a culture was produced and identified as MDR-TB within 60 days. It was found that immunocompromised status, the application of appropriate three drug treatment, whether or not Mycobacterium tuberculosis was cultured within 30 days, and age were associated with survival.

Chan et al (2004) conducted a study in the National Jewish Medical and Research centre (NJMRC). The study reviewed the records of 205 MDR-TB patients who were treated on the inpatient service and discharged between January 1, 1984 and December 31, 1998. Records were reviewed for demographics, previous number of drugs taken for TB and other potential risk factors for development of MDR-TB, susceptibilities to the anti-TB drugs and drugs used, drug toxicity, clinical complication, results of sputum cultures, and clinical outcomes. The study retrospectively analyzed the outcomes in 205 patients treated at NJMRC centre for MDR-TB. The study compared outcomes of patients treated; analyzed for the association of explanatory variables (age, sex, racial and ethnic distributions, number of drugs resistant, time with TB before first visit to NJMRC, users of flouroquinolones drug, extent of illness and surgery) with initial favourable response (survival time) using logistic regression. Estimate of the rates of survival was performed using Kaplan-Meier method. Survival analyses were performed based on different groupings of subjects. Within each analysis, comparisons between groups were made using the log-rank test. Survival rates were also estimated using Cox proportional hazards model in order that multiple predictors could be included in the same model. It was found that clinical complication and fluoroquinolone therapy were associated with improved microbiological and clinical outcomes in the 205 patients.

A prospective epidemiological case control study was conducted by Casal et al (2005). The aim of the study was to detect risk factors for multidrug resistance in patients with pulmonary tuberculosis in four European Union countries: France, Germany, Italy, and Spain. A total of 138 cases (resistant to both isoniazid (INH) and rifampicin (RIF)) and 276 controls (either not resistant to antituberculosis drugs or resistant to only one antituberculosis drug) were considered in the period from June 30, 1997 to June 30, 2000. For statistical and epidemiological analysis, Mantel–Haenszel's corrected chi square test was used to analyze the statistical significance of the difference in the amount of exposure in cases when compared to controls. A logistic regression model was used for multivariate analysis to identify, among a group of independent variables, those most significantly associated with MDR-TB. The predictors included were age, sex, birth place, TB contacts, previous tuberculosis, intravenous drug use, income, living situation, co-morbidity other than HIV, and HIV status. Risk factors associated with MDR-TB were age, known TB contacts, previous TB, income and intravenous drug use.

With the aim to retrospectively assess the burden, clinical characteristics, treatment outcomes, and long-term survival rate of patients with XDR-TB in a cohort of patients with HIV-negative MDR-TB, Kim et al (2008) conducted a study in South Korea. A total of 1,407 patients with culture-proven MDR-TB were enrolled from all national TB hospitals, all Korean National Tuberculosis Association chest clinics, and eight randomly selected university hospitals near Seoul. Medical records were reviewed for patients' demographics, TB treatment history, co-morbidities, acid-fast bacilli culture and drug susceptibility test (DST) results, chest radiographs, and treatment modalities and outcomes. The result of the study showed that age, BMI and DST results were statistically significant predictors. Mortality was approximately three to four times more likely in patients with XDR-TB than in patients with MDR-TB. The overall treatment success rate was less than 50%, and the treatment success rate in patients with XDR-TB was only 29%.

Holtz et al (2010) conducted a study to assess treatment outcome of MDR-TB/ XDR-TB in Latvia from 2000 to 2004. The study described demographic, clinical and treatment characteristics and compared differences between MDR-TB and XDR-Tb cases using chi-square test for dichotomous variables. Log rank test and the Kaplan-Meier survival curves were used to compare drug resistance groups with treatment outcomes. Covariates with clinical relevance for

poor outcome were retained for multivariable Cox proportional hazards analysis. HIV infection was not associated with all MDR-TB cases. Factors associated with risk of death were presence of XDR-Tb resistance, having bilateral cavitations and being retired or student. The interaction between age and previous treatment for MDR-TB was associated with poor outcome.

Balabanova (2011) conducted a retrospective national cohort study on MDR-TB/XDR-TB cases (n=1809) reported from 2002 to 2008 in Lithuania. The objective of the study was to identify risk factors influencing survival of patients with (MDR-TB/XDR-TB). The study used Kaplan-Meier survival curves and multivariable Cox regression to analyse time until death from any cause during a patient's treatment or follow-up, from the time of the first-recorded diagnosis of MDR-TB or XDR-TB. 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/XDR-TB diagnosis were associated with survival. There was no difference in survival of patients with primary MDR or XDRTB compared with those who developed drug resistance during treatment. There was no association of survival either with acquired TB before versus primary or with XDR-TB versus MDR-TB.

Farley et al (2011) conducted a prospective cohort study on MDR-TB patients with high HIV prevalence in South Africa. The study aimed to assess treatment outcomes MDR-TB patients from 2000 to 2004. The study used chi-square and Fisher exact test to compare demographic and clinical data. The study combined Completion and Cure into a single category and defined four mutually exclusive outcomes: Completion/Cure, Failure, Default, and Death to compare treatment outcomes. Cox proportional hazard model was used to examine multivariate factors associated with time to death. Factors associated with a significantly increased hazard of death were HIV infection, drug susceptibility, treatment regime, weight and therapeutic delay longer than two months.

A retrospective study on treatment outcome of MDR-TB was done by Anderson et al (2012) in the United Kingdom. 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, co-morbidities, 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. The study has also shown that receiving a fluoroquinolone or a bacteriostatic drug was more likely to have a successful treatment outcome compared to those who did not. Treatment with an injectable agent (Streptomycin, Amikacin, Capreomycin and Kanamycin) did not have a significant effect on treatment outcome. No significant interactions were detected and all factors significantly associated with treatment outcome.

Khan (2013) conducted a study to ascertain the social determinants of MDR-TB in the US between the years of 2005 and 2009 to better equip public health officials to deal with this growing threat. The study population was the US population, including all 50 states and the District of Columbia. It used the Centres for Disease Control and Prevention (CDC) Online Tuberculosis Information System as database. The variables that were examined were: whether or not the patient lived in a correctional facility at the time of diagnosis; HIV status; homelessness; whether or not the patient had an occupation; and place of birth (US-born or foreign born). Each variable was cross-tabulated with MDR-TB and then stratified by race/ethnicity, sex, and age. Mantel-Haenszel odds ratio was used to find associations between the above mentioned variables and MDR-TB. Living in a correctional facility and place of birth were found to be strongly related with MDR-TB.

Selamawit et al (2013) conducted a case control study between November 1, 2011 and February 28, 2012 in Addis Ababa. The aim of the study was to identify factors that determine the occurrence of MDR-TB among patients who had taken first line anti-TB treatment in Addis Ababa. Cases were selected from St. Peter Hospital and controls were selected from Addis Ketema Health Centre in Addis Ketema sub-city; Woreda 9 Health Centre in Kolfe Keranyo sub-city; Lideta Health Centre in Lideta sub-city; Kasanches Health Centre in Kirkos sub-city; and Woreda 19 and Nifas Silk Lafto health centres in Nifas Silk Lafto sub-city. The study included

134 cases and an equal number of controls. It was found that MDR-TB development is significantly associated with episodes of TB illness, interruption of first-line anti-TB treatment, education status, sex, the number of rooms in the patient's household, TB site, drug side effects during first-line treatment, lack of direct observation by health workers, and first-line anti-TB treatment. HIV status, history of smoking, experience of drug shortages, and family size were not significantly associated with MDR-TB development.

With the objective to assess the survival and predictors of survival time among patients under MDR-TB, Theodros et al (2013) conducted a 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.

## 3. DATA AND METHODOLOGY

### 3.1 Data

Data were obtained from a cohort of 342 MDR-TB patients enrolled in ALERT hospital in Addis Ababa, the capital city of Ethiopia 142 and in the University of Gondar Teaching Referral hospital 200 from august, 2011 to august, 2014. All patients who were diagnosed with a first MDR-TB episode and admitted to one of the MDR-TB treatment centres were included in the study. These were patients whose status has changed from TB to MDR-TB or were infected with MDR-TB without having had TB before.

Data were 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. Data were entered into a password protected computer to maintain confidentiality by the data clerk after checking for completeness and coding. Regular and daily supervision of the data collection process was done by the principal investigator. Health professionals who participated in the data collection were given orientation on how to appropriately extract the data.

### Study site

ALERT is a medical facility on the edge of Addis Ababa, specializing in Hansen's disease, also known as "leprosy". It was originally the All Africa Leprosy Rehabilitation and Training Centre (hence the acronym), but the official name is now expanded to include tuberculosis as All Africa Leprosy, Tuberculosis and Rehabilitation Training Centre. The Armauer Hansen Research Institute was founded in 1970, specializing in leprosy research. ALERT is the continuation and expansion of the leprosy hospital originally built by Dr. Thomas Lambie in 1922, which was later named the Princess Zänäbä Wärq Hospital. A memorandum to found ALERT was signed Dec. 11, 1965 by representatives of the Ministry of Health, Addis Ababa University, the International Society for the Rehabilitation of the Disabled, the Leprosy Mission, and Dr. Eugene Kellersberger of the American Leprosy Mission, who had the vision for establishing such a multifaceted centre and had been the main promoter of the project.

MDR-TB treatment was started in ALERT Hospital in 2011. This is the third MDR-TB centre in Ethiopia following St. Peter's TB Specialized Hospital in Addis Ababa and University of Gondar teaching referral Hospital in Amhara regional State. It has a capacity to provide service to over 300 patients annually. The center has 30 beds, counseling and monitoring facilities as well as surgical ward. Recently, 201 MDR-TB patients take treatment in this hospital, of which 40 patients take treatment for experiment purpose to reduce the time of follow up of MDR-TB treatment from two years to nine months. Therefore these 40 patients are not included in this study.

University of Gondar was established in 1954 as a Public Health College and Training Center (PHC & TC) in joint effort between the Imperial Ethiopian government, WHO, United States Operation Mission to Ethiopia, and UNICEF. University of Gondar Teaching Referral Hospital, which is the only referral hospital in North Gondar Administrative zone, is located in Gondar town, 741 km away North-West of Addis Ababa. The University of Gondar's MDR TB Program provides treatment services across north-west Ethiopia. The program began with three patients in August 2010.

## 3.2. Variables of the study

### 3.2.1. The Response Variable

The response or outcome variable in this study is the survival time measured (in months) from the date MDR-TB treatment's start until the date of the patient's death or censor.

### 3.2.2. Predictor Variables

On the basis of previous work, a set of variables is selected for the analysis. Considering the potential importance, the following socio-demographic factors and clinical factors have been considered.

- Sex( male, female)
- Age in completed years
- Baseline Weight in kilograms
- Marital status (single, Married, Others (widowed/divorced (separated)))

- Level of Education (No education, Primary, Secondary and above)
- HIV co-infection (Positive, Negative)
- Clinical complication(no, yes)
- Co-morbidities(no, yes)
- Drug susceptibility test results (INH & RIF only, INH, RIF and at least one of other drugs (E, STM, KAN, AMK & CPM)
- Therapeutic delay (less than one month, longer than one month)
- Smoking status (yes, No)

We do not have any prior knowledge of specific interactions that must be included. So, all possible bivariate interactions were considered.

## 3.3.Methodology

### 3.3.1. Survival Data Analysis

The term "survival analysis" pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study. In other words, survival analysis is an important statistical technique used to describe and model time–to–event data. 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. Time–to–event data can be found in many disciplines, for example in medicine:

- the time to death for patients having a certain disease (this explains the term survival analysis),
- the time to relapse of or cure from a certain disease, and
- the time to death of HIV patients after retroviral therapy.

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

- 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.
- 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.
- The multivariate Method uses Cox-proportional hazards model. It is considered as the most interesting survival modelling in the interest of examining the relationship between 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.

### 3.3.2. Descriptive methods for survival data

In any applied setting, a statistical analysis should begin with a thoughtful and thorough univariate description of the data. And this description includes life tables and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from all observations available.

#### 3.3.2.1. Survivor function S (t)

Survival data are not amenable to standard statistical procedures used in data analysis due to censoring. The survival time in days, weeks or months, whichever is the most appropriate, can then be calculated. The survivor function and hazard function are the two functions of central interest in summarizing survival data. The actual survival time of an individual, *t*, can be regarded as the value of a random variable *T*, which can take any non-negative value. The different values that *T* can take have a probability distribution, and we call *T* the random variable associated with the survival time. When the random variable *T* has a probability distribution with underlying probability density function f (*t*), the cumulative distribution function (cdf) of *T*, denoted $F_T(t)$, is then given by

$$F_T(t) = P_T\,(T \leq t) = \int_0^t f(u)du\,, t > 0 \qquad (3.1)$$

It is the probability that an individual will die before time *t*.

The basic quantity employed to describe time-to-event phenomena is the survival function, the probability of an individual surviving or being event-free beyond time *t* (experiencing the event after time *t*). It is defined as $S(t) = P(T > t)$. When $T$ is a continuous random variable, the survival function is the complement of the cumulative distribution function, that is

$$S(t) = 1 - F(t) \qquad (3.2)$$

Since S (*t*) is a probability, S (0) = 1 and as t approaches ∞, S (t) approaches 0.

**Median Survival Time**

Median survival time m is defined as that value for which S (m) = 0.5. Sometimes, it is denoted by $t_{0.5}$ or $t_{med.}$ If S (t) is not strictly decreasing, m is the smallest number such that S (m) ≤ 0.5 or $t_{mec} = S^{-1}(0.5)$.

### 3.3.2.2. Hazard function h (t)

The hazard function *h* (*t*) gives the instantaneous potential for failing at time *t*, given that the individual has survived up to time *t.* It is also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic process, the age specific failure rate in epidemiology, the inverse of the Mill's ratio in economics or simply the hazard rate. In contrast to the survivor function, which focuses on failing, the hazard function focuses on not failing, that is, on the event occurring. Thus, in some sense, the hazard function can be considered as giving the opposite side of the information given by the survivor function.

The hazard function $h(t) \geq 0$ is given as:

$$h(t) = lim_{\Delta t \to 0} \frac{P\{an\ individual\ fails\ in\ the\ time\ interval(t, t+\Delta t) given\ survives\ until\ time\ t\}}{\Delta t}$$

$$= lim_{\Delta t \to 0} \frac{P\{t \leq T \leq t + \Delta t \,|\, T \geq t\}}{\Delta t} \qquad (3.3)$$

By applying the theory of conditional probability and the relationship, the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d}{dt} lnS(t). \qquad (3.4)$$

The corresponding cumulative hazard function H(t) is defined as

$$H(t) = \int_0^t h(u)du = -lnS(t) \qquad (3.5)$$

Then, $S(t) = exp\,(-H(t))$ *and* $f\,(t) = h\,(t)S(t)$.

### 3.3.2.3. Estimation of survivorship function

In practice, when using actual data, we usually obtain estimated survivor function and obtain curves that are step functions, rather than smooth curves.

**Kaplan-Meier estimator**

The Kaplan-Meier (KM) estimator of the survivorship function [Kaplan and Meier (1958)], also called the product limit estimator, is mostly used to estimate the survivor and hazard functions. It 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. The Kaplan–Meier estimate of the survival function is an empirical or non-parametric method of estimating S (t) from non- or right-censored data. It is extremely popular as it requires only very weak assumptions and yet utilises 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*, $S(t) = P(T \geq t)$ is defined as:

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} = \prod_{t_{(i)} \leq t} 1 - \frac{d_i}{n_i} \qquad (3.6)$$

with the convention that

$\hat{S}(t) = 1$ for $t < t_{(1)}$.

where, $t_1, t_2, \ldots, t_n$ is the set of survival times 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

$$Var\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \sum \frac{d_i}{n_i(n_i - d_i)} \qquad (3.7)$$

### 3.3.2.4. Comparison of Survivorship Functions

After obtaining statistics which provide a description of the overall survival experience, the survival and hazard functions, one is expected to proceed with a comparison of the survivorship experience of subgroups in the data. These groups might be defined by the values of a covariate

which are thought to be related to survival times. When comparing groups of subjects, it is always preferable to begin with a graphical display of the data in each group. In studies of survival time, we should graph the Kaplan-Meier estimator of the survival function for each of the groups. That is, plotting the corresponding estimates of the two survivor functions on the same axis of Kaplan-Meier estimator. In general, if the plot shows the pattern of one survivorship function lying above another, this means the group defined by the upper curve lived longer, or had a more favourable survival experience than the group defined by the lower curve. But, the statistical question is whether the observed difference seen on the plot is significant. This needs to be answered using appropriate statistical test (Hosmer and Lemeshow, 1999).

The general form of test statistic that deal with this issue is given as

$$Q = \frac{[\sum_{i=1}^{m} w_i(d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^{m} w_i{}^2 \hat{v}_{1i}} \qquad (3.8)$$

where:

$m$ is the number of rank ordered event (death) times.

$d_{1i}$ is the observed number of events (death in group1at event time $t_i$ .

$\hat{e}_{1i} = \frac{n_{1i}-d_i}{n_i}$ is the expected number of events (deaths) corresponding to $d_{1i}$ .

$n_{1i}$ is the number of individuals at risk in group1 just prior to event (death) time $t_i$.

$n_{2i}$ is the number of individuals at risk in group 2 just prior to event (death) time $t_i$.

$\hat{v}_{1i} = \frac{n_{1i}n_{2i}d_i(n_i-d_i)}{n_i{}^2(n_i-1)}$ is the variance of the number of events $d_{1i}$ at time $t_i$.

$w_i$ is the weight for censor adjustment at failure time $t_i$

$n_i$ and $d_i$ are the number of individuals at risk and number of death in both groups ( i.e., group 1 and group 2) just prior to event time $t_i$ ,respectively.

The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times. Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, Q follows a chi-square distribution with one degree of freedom. We can also use the test based on Q above to compare more than two groups.

The log rank test which is a special case of Q is used in this study.

**Log rank test**

The log rank test, sometimes called the Cox-Mantel test, is the most well known and widely used test statistic. This test statistic is based on weights equal to one, i.e. $w_i = 1$.

Therefore, the log rank test statistic becomes

$$Q = \frac{[\sum_{i=1}^{m}(d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^{m}\hat{v}_{1i}} \qquad (3.9)$$

Under the null hypothesis that two survival functions are equal, the log rank test statistic $Q$ has an approximation of chi-square distribution with one degree of freedom for large samples. The null hypothesis of equality of survival functions will be rejected for large values of Q.

### 3.3.3. Modelling 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 modelling can be used.

Through a modelling 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 centres on the risk or hazard of death at any time after the time origin of the study. As a consequence, the hazard function is modelled 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 modelled. However, many of the principles and procedures used in a linear modelling carry over to the modelling survival data.

There are two broad reasons for modelling survival data. One objective of the modelling process is to determine which combination of potential explanatory variable affect the form of the hazard function. In particular, the effect that the treatment has on the hazard of death can be studied, as can the extent to which other explanatory variables affect the hazard function. Another reason for

modelling the hazard function is to obtain an estimate of the hazard function itself for an individual. This may be of the interest in its own right, but in addition, from the relationship between the survivor function and hazard function, an estimate of survival function can be found. 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 counselling 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.

#### 3.3.3.1. The Proportional Hazards Regression Model

Proportional hazards model was proposed by Cox (1972) and has also come to be known as Cox regression model. Although the model is based on the assumption of proportional hazards, no particular form of probability distribution is assumed for the survival times. The model is therefore referred to as a semi-parametric 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*** = $(x_1, x_2, \dots, x_p)'$. Let $h_0(t)$ be the hazard function for an individual for whom the values of all the explanatory variables that make up the vector ***x*** are zero. The function $h_0(t)$ is called the baseline hazard function. The hazard function of the $i^{th}$ individual can be written as

$$h_i(t) = h_0(t)e^{\beta' x_i} = h_0(t)\exp\left(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\right), i = 1,2, \dots, n \qquad (3.10)$$

where: $n$ is total number of observations in the study.

$x_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)'$ is a column vector of measured covariates for the $i^{th}$ individual (patient) which are assumed to affect the survival probability.

$\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is a column vector of $p$ regression parameters associated with explanatory variables.

$e^{\beta' x_i}$ characterizes how the hazard function changes as a function of subject covariates

$\beta' x_i$ is called the linear component of the model, also known as the risk score or prognostic index for the $i^{th}$ individual.

*t* is the failure time.

Since this model can be re-expressed in the form

$$\log\left\{\frac{h_i(t)}{h_0(t)}\right\} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \qquad (3.11)$$

the PH model may also be regarded as a linear model for the logarithm of the hazard ratio.

The Cox proportional hazard model is popular because it allows a flexible choice of covariates: time varying, time-independent, continuous and discrete. Two other issues that make it popular are that it does not make any assumption about the underlying survival distribution and also does not require estimation of the baseline hazard rate, $h_0(t)$ to estimate the regression parameters.

#### 3.3.3.2. Assumption of Cox proportional hazard model

1. The baseline hazard function $h_0(t)$ depends on *t*, but not on covariates $x_1, x_2, \ldots, x_p$.
2. The hazard ratio, $e^{\beta}$, depends on the covariates $X = (X_1, X_2, \ldots, X_p)'$ not on time.
3. The covariates in the Cox model $X_i$ s' are time-independent.

Because of assumption (2) it is called a proportional hazards model. To show this issue mathematically, consider two distinct values of a continuous covariate *X*, say, $x_{i1}$ and $x_{i2}$

$$h(t, X, \beta) = h_0(t) e^{\beta' X} \qquad (3.12)$$

Then, the hazard ratio becomes

$$\frac{h(t, X_1, \beta)}{h(t, X_2, \beta)} = \frac{h_0(t) e^{\beta_i X_{i1}}}{h_0(t) e^{\beta_i X_{i2}}} = e^{\beta_i (X_{i1} - X_{i2})} \qquad (3.13)$$

which is clearly independent of time.

This reveals that the ratio of the hazard functions for two individuals with different covariate values does not vary with time.

#### 3.3.3.3. Fitting the proportional hazards model

Fitting the proportional hazards model to an observed set of survival data entails estimating the unknown coefficients of the explanatory variables, $X_1$, $X_2$, . . . , $X_p$, in the linear component of

the model, $\beta_1, \beta_2, \ldots, \beta_p$. The baseline hazard function, $h_0(t)$, may also need to be estimated. It turns out that these two components of the model can be estimated separately. The $\beta$'s are estimated first and these estimates are then used to construct an estimate of the baseline hazard function. This is an important result, since it means that in order to make inferences about the effect of *p* explanatory variables $X_1, X_2, \ldots, X_p$, on the relative hazard, $h_1(t)/h_0(t)$, we do not need an estimate of $h_0(t)$. The regression coefficients in the proportional hazards Cox model, which are the unknown parameters in the model, can be estimated using the method of maximum likelihood.

Consider n independent individuals, the data that we need for the Cox proportional hazard model is represented by triplet$(t_i, \delta_i, x_i), i = 1,2, \ldots, n$, where

$t_i$ is the survival time for $i^{th}$ individual

$\delta_i$ an indicator of censoring for the $i^{th}$ individual given by 0 for censored and 1 for death

$x_i$ a vector of covariates for individual *i* ($x_{i1}, x_{i2}, \ldots, x_{ip}$)

The likelihood for survival data is constructed by considering the censored observations and the failed observations separately. In the case of failed observations the survival time is exactly $t_i$, thus the contribution to the likelihood is the probability that the subject fails at time $t_i$, i.e. $f(t_i, \beta, x_i)$. And, for censored observations the contribution for the likelihood is that the probability that a subject survives at least $t_i$ time units, i.e. $S(t_i, \beta, x_i)$(Hosmer and Lemeshow, 1999).

In general, a concise way to denote the contribution of each observation to the likelihood is the expression:

$$f(t_i, \beta, x_i)^{c_i} \, S(t_i, \beta, x_i)^{1-c_i}$$

The full likelihood for right censored data can be constructed as

$$L(\beta) = \prod_{i=1}^{n} h(t_i, x_i, \beta)^{\delta_i} \, S(t_i, x_i, \beta) \qquad (3.14)$$

where $h(t_i, x_i, \beta) = h_0(t_i) e^{\beta' x_i}$ is the hazard function for individual *i*

$S(t_i, x_i, \beta) = S_0(t_i)e^{\beta' x_i}$ is the survival function for individual *i*

Then, the full maximum likelihood becomes

$$L(\beta) = \prod_{i=1}^{n}(h_0(t_i)e^{\beta' x_i})^{\delta_i}(S_0(t_i)e^{\beta' x_i}) \quad (3.15)$$

Full maximum likelihood requires that we maximize (3.15) with respect to the unknown parameter of interest, β, and unspecified baseline hazard and survival functions. This indicates that unless we explicitly specify the baseline hazard, $h_0(t)$ we cannot obtain the maximum likelihood estimators for the full likelihood. But, Cox (1972) proposed using an expression he called "partial likelihood function" that depends only on the parameter of interest.

#### 3.3.3.4. Partial likelihood

Suppose $m$ failures occur out of $n$ subjects with $m \leq n$. Suppose that $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(m)}$ are the $m$ distinct ordered failure times observed, and let $R_i$ be the set of subjects at risk prior to time $t_{(i)}$. Assume that there is only a single failure at time $t_i$, i.e. no ties. Then $P(individual\ i\ has\ experienced\ an\ event\ at\ time\ t_{(i)} \mid one\ event\ at\ time t_{(i)})$

$$= \frac{h(t, x_i)}{\sum_{j\varepsilon R_{t(i)}} h(t, x_j)} = \frac{h_0(t)exp(\beta' x_i)}{\sum_{j\varepsilon R_{t(i)}} h_0(t)exp(\beta' x_j)} = \frac{exp(\beta' x_i)}{\sum_{j\varepsilon R_{t(i)}} exp(\beta' x_j)} \quad (3.16)$$

When that there are no tied times assumed, the partial likelihood is defined over all failure time $t_{(i)}$ that $i = 1,2,..,m$ and given as

$$L_p(\beta) = \prod_{i=1}^{m} \frac{exp(\beta' x_{(i)})}{\sum_{j\varepsilon R_{t(i)}} exp(\beta' x_j)} \quad (3.17)$$

where the product is over $m$ distinct ordered failure times and $x_{(i)}$ denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$. The log partial likelihood function is

$$l_p(\beta) = \sum_{i=1}^{m} \left[ \beta' x_{(i)} - \ln \sum_{j\varepsilon R_{t(i)}} exp(\beta' x_j) \right] \quad (3.18)$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of (3.17) with respect to the component of β, setting the derivative equal to zero and solving for the unknown parameters. However, this partial likelihood function methods are based on the assumption that there are no tied values among the observed survival times. But, in most real situations tied survival times are more likely to occur. In addition to the possibility of more than one death at a time, there might also be more than one censored observations at a time of death. A number of approaches to handle tied data have been suggested and, of these, three are used by software packages: an exact expression that is derived in Kalbfleisch and Prentice (2002) and approximations due to Breslow (1974) and Efron (1977). However, approximations derived by Breslow (1974) and Efron (1977) are designed to provide expressions that are more easily computed than the exact partial likelihood, yet that still account for the fact that ties are among the observed values of survival time. In many applied settings there is little or no practical difference between the estimators obtained from the two approximations. Because of this and since the Breslow approximation is more commonly available and popular it is used mostly (Hosmer and Lemeshow, 1999).

**The Breslow approximation**

This approximation is proposed by Breslow and Peto to modify the partial likelihood and has the form

$$L_B(\beta) = \prod_{i=1}^{m} \frac{exp(\beta' s_i)}{\left[\sum_{l \varepsilon R_{t(i)}} exp(\beta' x_l)\right]^{d_i}} \tag{3.19}$$

where $d_i$ the number of deaths occurred at time $t_{(i)}$

$s_i$ the sum of covariates over $d_i$ subjects at time $t_{(i)}$

Then, the partial log of (3.18) is given as

$$l_B(\beta) = \sum_{i=1}^{m} \left[\beta' s_i - d_i \ln \sum_{l \varepsilon R_{t(i)}} exp(\beta' x_l)\right] \tag{3.20}$$

Breslow maximum partial likelihood estimator, adjusted for tied observation is obtained, by differentiating equation (3.19) with respect to the components of β and setting the derivative equal to zero and solving for the unknown parameters.

### 3.3.4. 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 Hosmer and Lemeshow, 1999).

**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 univariable analysis of the association between survival time and all important covariates (Hosmer and Lemeshow, 1999). Recommendable procedure in selecting variables in the study Hosmer and Lemeshow (1999) and Collett (2003) recommended the following procedure in variable selection.

1. Include all variables that are significant in the univariable analysis at the 25 percent level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.
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 any of the main effects currently in the model can be considered for inclusion if the inclusion significantly modifies the model.

### 3.3.5. Model diagnostics for Cox PH model

Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid in any sense of the word, the fitted model must provide an adequate summary of the data upon which it is based. The methods for assessment of a fitted proportional hazards model are essentially the same as for other regression models (Hosmer and Lemeshow, 1999).

Residuals are used to investigate the lack of fit of a model and useful for examining different aspects of the model. The following residuals have been proposed for use by different authors in connection with the Cox regression model.

**Martingale residuals:** It is a slight modification of the Cox-Snell residuals and is defined as $\widehat{M}_i = \delta_i - r_i$ where, $\delta_i$ is the censoring indicator and $r_i$ is Cox-Snell residual for $i^{th}$ individual given as $r_i = \widehat{H}_i(t) = -\ln \hat{S}_i(t)$, where $\widehat{H}_i(t)$ and $\hat{S}_i(t)$ are the estimated values of the cumulative hazard and survivor functions of the $i^{th}$ individual at time *t*. The martingale residuals have the property that $\sum_{i=1}^{n} \widehat{M}_i = 0$. They are uncorrelated and with expected value zero in large samples. In this respect they have properties similar to those possessed by residuals encountered in linear regression analysis (Collett, 2003).

**Schoenfeld residuals:** This overcomes the problem that the above three residuals depend heavily on observed survival time and cumulative hazard function. They are computed for each individual and covariate. It follows that, the Schoenfeld residual for the $i^{th}$ individual and $k^{th}$ covariate is defined as:

$$\hat{s}_{ik} = \delta_i \left[ X_{ik} - \frac{\sum_{j \varepsilon R_{(t_i)}} X_{ik}\, exp(\hat{\beta}' X_j)}{\sum_{l \varepsilon R_{(t_i)}} exp(\hat{\beta}' X_j)} \right] \qquad (3.21)$$

where, $X_j$ is a vector of p fixed covariates for the $j^{th}$ individual, $X_{ik}$ is the value of $k^{th}$ covariate on the $j^{th}$ individual.

Because of that, Schoenfeld residuals are defined only for the uncensored observations in which case $\hat{s}_{ik} = \delta_i \left[ X_{ik} - \frac{\sum_{j \varepsilon R_{(t_i)}} X_{ik} \, exp(\hat{\beta}' X_j)}{\sum_{l \varepsilon R_{(t_i)}} exp(\hat{\beta}' X_j)} \right]$ and for each covariate it must sum to zero. In addition, they are uncorrelated and with expected value zero (Schoenfeld, 1982).

Most of the model diagnostics in survival data are based on the residuals stated above. The following are model diagnostics that are required to assess the model adequacy in this study.

#### 3.3.5.1. Testing for the Nonlinearity of covariates

After identifying a particular set of explanatory variables on which the hazard function depends, it is desirable to check that the correct functional form has been adopted for the continuous covariates. The Martingale residuals can be plotted against covariates to detect nonlinearity. Nonlinearity is not an issue for categorical variables, so we only examine plots of martingale residuals against continuous covariate. LOESS smoothed curve can be super imposed on the scatter plots to give interpretation. If the functional form observed in using the plots has some pattern, which is non linear, the covariate can be so transformed and the martingale residuals again should be plotted against the transformed covariate. A horizontal straight line which is drawn as a reference through zero would then confirm that the appropriate transformation has been used to the covariate. In addition, if the resulting smooth plot is a straight line compared to the reference line, then it shows linearity (Collett, 2003).

#### 3.3.5.2. Examining influential observations

Another important aspect of model evaluation is through diagnostic statistics in order to identify which subjects have an unusual configuration of covariates or observations that have influence on the estimates of the parameters or on the fit of the model. In other words a fitted model is particularly sensitive to one or more observations in the data set. Such observations can be termed as influential observations. Conclusions from survival analyses are often framed in terms of estimates of the relative hazard, which depends on the estimated values of the coefficients in the Cox regression model. Thus, it is desirable to examine the influence of each observation on

these estimates. The interest is about observations that influence estimate of hazard functions and the complete estimate of the model and identifications of these observations. This could be done by fitting the model to all *n* observations in the data set, and then fitting the same model to the sets of n-1 observations obtained by omitting each of the *n* observations in turn. The interest is to determine if the result would change when a particular observation is removed from the analysis (Collett, 2003).

Suppose that $l_p(\beta)$ is log partial likelihood and $\hat{\beta}_j$ is the corresponding $j^{th}$ parameter estimates of the model containing all the *n* observations and $l_p(\beta_{-i})$ be the log partial likelihood and $\hat{\beta}_{j(-i)}$ is the $j^{th}$ parameter estimate of the model containing only the $n-1$ observations after deleting the $i^{th}$ observation, respectively. Then the statistic, $\Delta_i\hat{\beta}_j = \hat{\beta}_j - \hat{\beta}_{j(-i)}$, which is known as DFBETA, can be used as a measure of how the $j^{th}$ parameter estimate would change if the $i^{th}$ observation was deleted from the data set. On the other hand, the statistic, $LD_i = 2l_p(\beta) - 2l_p(\beta_{-i})$, which is called the likelihood displacement statistic, can be used as a measure of how the maximized partial log likelihood changes if the $i^{th}$ observation was deleted from the data set. Observations that influence a particular parameter estimate have a large absolute value of DFBETA than other observations in the data set. Observations that do influence the overall fit of the model are those which have large values of likelihood displacement statistics than the other observations in the data set (Collett, 2003).

#### 3.3.5.3. Checking Cox Proportional Hazard Assumption

In order to use the Cox model, it has to be checked that the assumption of whether the effects of covariates on hazard ratio remain constant over time. This is a vital assumption of proportional hazards model and must be assessed for each covariate. Several procedures of graphical techniques and tests are proposed to investigate the proportionality assumptions in fitting the Cox model (Cox, 1972). The Schoenfeld residuals are employed to assess this assumption.

The Schoenfeld residuals graphical technique can be used to assess Cox model proportionality assumption. The technique is based on individual contributions to the log partial likelihood and measures the difference between the covariate for the $i^{th}$ individual and a weighted average of the covariate over the risk set at each event. To check the proportionality assumption for each

covariate, we plot the scaled Schoenfeld residuals against log of survival time. If the proportional hazards assumption is satisfied, the distribution of residuals over time is random, that is, does not show a particular trend, and the smoothed plot called Locally Weighted polynomial regression (Lowess) line summarizing the residuals should be a straight line and close to the horizontal reference line. Otherwise, a plot of scaled Schoenfeld residuals for a given covariate may reveal a violation of the proportional hazards assumption (Schoenfeld, 1982). Formal tests need to detect any time dependency in particular covariates, after allowing for the effects of explanatory variables that are known. Testing the dependency of covariates on time is equivalent to testing for a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time. A non-zero slope is an indication of a violation of the proportional hazard assumption. The Grambsch-Therneau test of non-proportionality uses partial residuals for the test of proportional hazards assumption. In order to use this test for the $i^{th}$ covariate, Grambsch and Therneau (1994) propose a time-varying coefficient as

$$\beta_i(t) = \beta_i + \gamma_i g_i(t) \tag{3.22}$$

where $\beta_i(t)$ is time varying coefficient, $\beta_i$ is constant, and $g_i(t)$ is some specified fuction of time, usually $g_i(t) = \ln t$;

Then, the Cox proportional hazard model for time varying coefficient with $g_i(t) = \ln(t)$ is defined as

$h(t, x_i, \beta_i(t)) = h_0(t)\exp(\beta_i(t)x)$, by substituting for $\beta_i(t)$ and $g_i(t)$ becomes

$$= h_0(t)\exp(\beta_i + \gamma_i \ln t)x$$

$$= h_0(t)\exp(\beta_i x + \gamma_i(\ln t)x \tag{3.23}$$

Equation (3.23) is the proportional hazards model with the interaction term, $xln(t)$ and main effect $x_i$. To test the significance of the interaction term $xln(t)$, we perform the test:

$H_0: \gamma = 0$ versus $H_1: \gamma \neq 0$ and we use likelihood-based tests like Wald test. If $\gamma = 0$ is not rejected, $\beta_i's$ are not time varying coefficients and hence the proportional hazards assumption is satisfied. If $\gamma = 0$ is rejected then the proportional hazards assumption is not satisfied, that leads to the need of other methods that cope with time-dependency (Schoenfeld, 1982).

# 4. STATISTICAL DATA ANALYSIS

## 4.1. Introduction

In this section we present results of data analysis. The first part presents summary statistics of factors considered in this study. The second part compares the survival time in different groups. Descriptive survival analysis was employed to compare the survival time in different groups. The third part is about fitting Cox proportional hazard model. Then, the adequacy of the model is investigated. The statistical packages SPSS and STATA were employed to analyze the data.

## 4.2. Summary Statistics

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).

The distributions of the socio-demographic characteristics of the cohort are summarized in Table 4.1.

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.

The overall median estimated survival time of patients under the study was 16 months. The minimum follow-up time was 1 month and the maximum was 42 months [Table 4.2]. Males had relatively lower survival time (14.55 months) than females (16.78 months). The median weight of patients is 50.5 Kg, with non-smokers (88.9 %). 8.5 % of patients had clinical complication due to Pneumothorax (7), pneumonia (3), hemoptysis (3), Cor pulmonale (8) and others (8). About 95% of the patients had no medical diagnosis other than tuberculosis; 5% had co-morbidities diabetes (3.5%) and others (Myocardial Infarction or Asthma (1.5%)) other than HIV. Drug susceptibility test showed that 269 (78.7%) of the patients were resistant only to rifampicin and isonizid anti-TB drugs; the rest were resistant to RIF, INH and at least one of other drugs (E, STM, KAN, AMK & CMP). The length of time from MDR-TB confirmation to

start of treatment (therapeutic delay) was less than one month for 239(69.9%) patients and longer than one month for 103(30.1%) [Table 4.1].

**Table 4.1: Demographic and Health factors of categorical covariates by MDR-TB in ALERT hospital and Gondar University Teaching and Referral Hospital.**

| Demographic and Health Factors | | Summary of the Number of Event and Censored values | | | |
|---|---|---|---|---|---|
| | | Total (%) | Death | Censored | Death % |
| Sex | Male | 195(57) | 21 | 174 | 10.8 |
| | Female | 147(43) | 16 | 131 | 10.9 |
| Marital Status | Single | 158(46.2) | 14 | 144 | 8.9 |
| | Married | 155(45.3) | 18 | 137 | 11.6 |
| | Others(widowed/divorced) | 29(8.5) | 5 | 24 | 17.2 |
| Educational Status | No Education | 123(36) | 16 | 107 | 13.0 |
| | Primary | 100(29.2) | 9 | 91 | 9.0 |
| | Secondary and above | 119(34.8) | 12 | 107 | 10.1 |
| HIV Status | Negative | 271(79.2) | 28 | 243 | 10.3 |
| | Positive | 71(20.8) | 9 | 62 | 12.7 |
| Co-morbidities | No | 325(95) | 34 | 291 | 10.5 |
| | Yes | 17(5.0) | 3 | 14 | 17.6 |
| Clinical Complication | No complication | 313(91.5) | 25 | 288 | 8.0 |
| | Yes | 29(8.5) | 12 | 17 | 41.4 |
| Drug Susceptibility test | INH & RIF only | 269(78.7) | 24 | 245 | 8.9 |
| | INH, RIF and at least one of other drugs (E, STM, KAN, AMK & CMP) | 73(21.3) | 13 | 60 | 17.8 |
| Therapeutic Delay | Less than one month | 239(69.9) | 23 | 216 | 9.6 |
| | More than One month | 103(30.1) | 14 | 89 | 13.6 |
| Smoking | Yes | 38(11.1) | 8 | 30 | 21.1 |
| | No | 304(88.9) | 29 | 275 | 9.5 |

**Table 4.2: Summary statistics of continuous variables included in the study**

| Patient Status | Continuous Variables | Mean | Standard deviation | Min. | Max. | Median | $\overline{Q^1}$ | $\overline{Q^3}$ |
|---|---|---|---|---|---|---|---|---|
| Censored | Time | 16.37 | 8.311 | 1 | 40 | 18 | 9 | 23 |
| | Age | 30.59 | 12.134 | 1 | 72 | 28 | 22 | 38 |
| | Weight | 50.71 | 9.710 | 7 | 88 | 51 | 46 | 56 |
| Death | Time | 8.38 | 9.13 | 1 | 42 | 5 | 2 | 12.5 |
| | Age | 36.19 | 12.789 | 16 | 76 | 36 | 26.5 | 43.5 |
| | Weight | 44.77 | 7.347 | 32 | 60 | 47 | 40 | 50.5 |
| Overall | Time | 15.51 | 8.750 | 1 | 42 | 16 | 8 | 23 |
| | Age | 31.19 | 12.311 | 1 | 76 | 28 | 22 | 40 |
| | Weight | 50.07 | 9.652 | 7 | 88 | 50.5 | 45 | 55 |

## 4.3. Descriptive survival analyses

The patients were followed up for a median period of 16 months. The minimum follow-up time was 1 month and the maximum was 42 months. A total of 37 subjects (10.8%) died during the maximum follow up time of 42 months; 11, 12 and 5 deaths occurred in the first, second and third three months of MDR-TB treatment initiation, respectively. That means 28 patients died in the first nine months of follow up.

Next, we investigate whether the observed differences in data summary among different factors are statistically significant or not with the help of Kaplan-Meier survival estimates and the log-rank test. The following graph of the estimate of overall Kaplan-Meier survivor function reveals that most of the deaths occurred in the earlier months of MDR-TB treatment initiation.

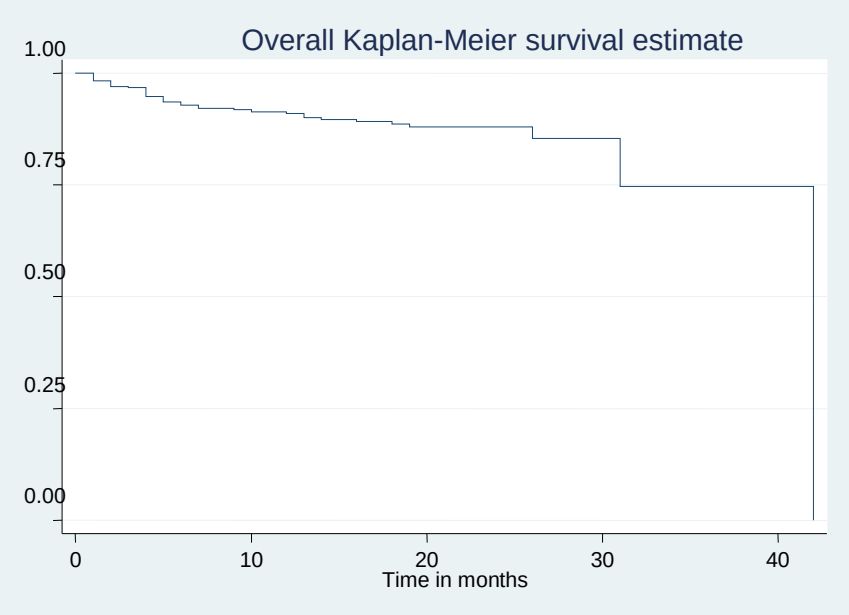



**Figure 4.1: The plot of the overall estimate of Kaplan-Meier survivor function of MDR-TB patients in ALERT and Gondar University Teaching and Referral hospital, Ethiopia.**

The Kaplan-Meier estimator survival curve gives the estimate of survivor function among different strata or groups of covariates to make comparisons. Separate graphs of the estimates of the Kaplan-Meier survivor functions are constructed for different categorical covariates; in so doing it is possible to see the existence of difference in survival experience between the indicated categories of individuals. The pattern that one survivorship function lies above another means the group defined by the upper curve has a better survival than the group defined by the lower curve.

Most of the graphs did not show differences between different categories. Relatively larger gaps are observed in covariates such as clinical complication and drug susceptibility test. The graphs of Kaplan-Meier survival estimates based on different categories of factors are presented in the displays of Figure 1(a-i) in Appendix D.

To check for significance differences among categories of factors that are shown using the Kaplan-Meier estimates of the survivor functions, we employ a log-rank statistical test.

**Table 4.3: Results of the Log-rank test for the categorical variables of MDR-TB patients in Addis Ababa and Gondar University Teaching and Referral hospital.**

| Log-rank test for equality of survivor functions of categories | | | |
|---|---|---|---|
| Covariates | Chi-Square | DF | p-value |
| Sex | 0.22 | 1 | 0.6372 |
| Marital Status | 2.12 | 2 | 0.3470 |
| Educational Level | 1.72 | 2 | 0.4236 |
| HIV Status | 0.86 | 1 | 0.3525 |
| Co-morbidities | 0.03 | 1 | 0.8548 |
| Clinical Complication | 42.56 | 1 | 0.0000 |
| DST | 1.31 | 1 | 0.2526 |
| Therapeutic Delay | 0.03 | 1 | 0.8694 |
| Smoking | 5.76 | 1 | 0.0164 |

The log-rank test indicates that statistically there is a significant difference of survival experience among groups of clinical complication and smoking status at 5% level of significance.

On the other hand, there are statistically no significant differences in survival/death experience among groups of the remaining categorical covariates: sex, marital status, educational level, HIV status, drug susceptibility test, co-morbidities and therapeutic delay.

## 4.4. Results of the Cox proportional hazards model

In order to study the relationship between survival time and covariates, a regression modelling approach to survival analysis using the Cox proportional hazards model can be employed for estimating the regression coefficients, making interpretation based on the hazard function, conducting statistical tests and constructing confidence intervals. Checking the adequacy of a model and its development precedes interpretation of results obtained from the fitted model.

The aim of model development is to obtain a model that satisfactorily describes the data at hand. This is done in order to identify statistically significant factors that influence the survival of MDR-TB patients using a multivariate Cox proportional hazard model.

The model will be constructed by first identifying factors which are significant at 25% level of significance in a univariate Cox proportional hazard analysis. The Wald chi-square test is used to test the significance. Accordingly, Table 4.4 is the summary of a univariate analysis used to select potential predictors for further analysis.

**Table 4.4: Results of a univariable proportional hazards Cox regression model of MDR-TB patients in ALERT and Gondar University Teaching and Referral hospital.**

| Analysis of Maximum Likelihood Estimates | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Covariates | DF | $\hat{\beta}$ | $se(\hat{\beta})$ | Chi-Square | -2LL | Sig. | HR | Lik. Rat. Sig | Score sig |
| Sex, Female | 1 | -0.160 | 0.341 | 0.22 | 392.957 | 0.6389 | 0.852 | 0.6376 | 0.639 |
| Age | 1 | 0.023 | 0.012 | 3.86 | 389.615 | 0.0494 | 1.024 | 0.0590 | 0.049 |
| Marital Status | 2 | | | 2.03 | 391.405 | 0.3632 | | 0.4118 | |
| Marital Status married | 1 | -0.937 | 0.757 | 0.19 | | 0.6661 | 1.169 | | 0.666 |
| Marital Status others | 1 | -0.782 | 0.750 | 2.01 | | 0.1559 | 2.099 | | 0.156 |
| Education | 2 | | | 1.68 | 391.534 | 0.4323 | | 0.4392 | |
| Primary | 1 | -0.386 | 0.417 | 0.85 | | 0.3552 | 0.680 | | 0.355 |
| Secondary and above | 1 | -0.468 | 0.393 | 1.42 | | 0.2337 | 0.626 | | 0.234 |
| HIV status, positive | 1 | 0.355 | 0.385 | 0.85 | 392.381 | 0.3570 | 1.426 | 0.3718 | 0.357 |
| Co-morbidities, yes | 1 | 0.133 | 0.730 | 0.03 | 393.147 | 0.8554 | 1.142 | 0.8582 | 0.855 |
| Clinical Complication, yes | 1 | 1.996 | 0.360 | 30.66 | 370.124 | 0.0000 | 7.359 | 0.0000 | 0.000 |
| DST | 1 | 0.403 | 0.356 | 1.28 | 391.958 | 0.2475 | 1.496 | 0.2491 | 0.247 |
| Therapeutic Delay | 1 | 0.058 | 0.354 | 0.03 | 393.153 | 0.8703 | 1.059 | 0.8703 | 0.870 |
| Smoking | 1 | -0.930 | 0.403 | 5.32 | 388.735 | 0.0211 | 0.395 | 0.0350 | 0.021 |
| Weight | 1 | -0.058 | 0.013 | 18.67 | 378.142 | 0.0000 | 0.944 | 0.0001 | 0.000 |
| The value of -2LL for the null model is 393.179 | | | | | | | | | |

Consequently, the most important subset of predictors to be included in the multivariable model will be selected based on their contribution to the maximized log partial likelihood of the model (-2LL). The summary above reveals that the highest reduction in $-2LL(\hat{\beta})$ is observed for clinical complication where the value for the null/empty model, was reduced from 393.179 to 370.124. The difference 23.055 is statistically significant (p-value <0.001). The next highest change is obtained for weight where the difference equal to 15.037 was found to be significant at p-value < 0.001.

Proceeding in this manner covariates will be eliminated in order of the magnitude in which they increase the $-2LL(\hat{\beta})$. Thus, using the Wald chi-square test, the predictors that are found to be significant at 25% level of significance, and will be considered for the next multivariate analysis were: Age, Clinical Complication, Drug Susceptibility test, Smoking Status and weight of the patients. Furthermore, Age, Clinical Complication, Smoking Status and weight have strong associations with survival of MDR-TB at p-value less than 0.05. Since keeping the covariates sex, marital status, education, co-morbidities and therapeutic delay in the model does not bring significant changes in the value of $-2LL(\hat{\beta})$, these variables become the first to be removed from the multivariate model. We proceed fitting the initial multiple Cox proportional model by including the five covariates that are significant at univariate analysis. Then, a multivariate Cox proportional regression model was fitted by eliminating those covariates that are not significant at 5% level of significance. Consequently, all the five covariates are found to be significant at 5% level [see Appendix A]. Hence, we have a multivariate model which includes the five covariates namely: Age, weight, Clinical complication, Drug Susceptibility test and Smoking Status.

The next step is assessing the importance of the variables which were not found to be significant in the univariate analysis as predictors or confounders of survival experience of patients and their effects. This can be achieved by adding non-significant variables in the univariate analysis (one at a time) to the model containing the five covariates. The result of the analysis revealed that, none of those single variables were found to be significant; therefore they cannot be retained in the model. Thus, variables that were not significant at univariate analysis are not confounders of the main factors in the preliminary model [see Appendix B, Tables: 1-6].

The final step in model development strategy is consideration of interaction terms that may be useful in the improvement of the model. We do not have any prior knowledge of specific interactions that we must include so we will consider all the possible bivariete interactions to see if the interaction effects can increase or decrease the survival time of MDR-TB patients. The Wald test is used to assess the significance of reasonable and possible interactions. The test confirmed the presence of a significant interaction between age and weight [See Appendix B, Table: 7].

Even though the covariate weight was not significant in the multivariate model, we cannot eliminate it since it has a significant interaction effect with covariate age. Now, the result ensures that the preliminary model of the study will contain the five covariates namely Age, weight, Clinical complication, Drug Susceptibility test and Smoking Status and the interactions between age and weight (Age*Weight). The parameter estimates and hazard ratios of the covariates are shown in Table 4.5 below.

**Table 4.5: Estimated values of the proportional hazards model to the data from MDR-TB patients in ALERT and university of Gondar Teaching and Referral hospital.**

| **Variables in the multivariate Cox regression model** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SE | | Df | Sig. | Est. HR | 95% r HR (Exp( | |
| | | | | | | | Lower | Upper |
| Weight | -0.0192 | 0.0451 | 0.18 | 1 | 0.6707 | 0.9810 | 0.8979 | 1.0717 |
| Age | 0.1987 | 0.0678 | 8.60 | 1 | 0.0034 | 1.2199 | 1.0681 | 1.3933 |
| Clinical complications | 1.5510 | 0.3923 | 15.63 | 1 | 0.0001 | 4.7161 | 2.1861 | 10.1740 |
| DST | 1.0910 | 0.4003 | 7.43 | 1 | 0.0064 | 2.9771 | 1.3586 | 6.5238 |
| Smoking | 1.1537 | 0.4486 | 6.61 | 1 | 0.0101 | 3.1697 | 1.3159 | 7.6354 |
| Age*Weight | -0.0037 | 0.0015 | 5.83 | 1 | 0.0157 | 0.9963 | 0.9933 | 0.9993 |

Even though the interaction of age and weight had a significant effect on the mortality of MDR-TB patients, the hazard ratio is not much different from unity. This means there is not much change on the model. So, we not consider the interaction. Consequently the final model of the study will contain the five main covariates namely age, weight, clinical complication, drug susceptibility test and Smoking Status.

**Table 4.6: Estimated values of the proportional hazards of the final model to the data from MDR-TB patients in ALERT Hospital and Gondar University Teaching and Referral Hospital.**

| **Variables in the multivariate Cox regression model** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SE | | Df | Sig. | Est. HR | 95% r HR (Exp( | |
| | | | | | | | Lower | Upper |
| Weight | -0.095 | 0.0190 | 24.82 | 1 | 0.0000 | 0.9093 | 0.8760 | 0.9440 |
| Age | 0.0366 | 0.0132 | 7.59 | 1 | 0.0059 | 1.0372 | 1.0106 | 1.0646 |
| Clinical complications | 1.6341 | 0.3901 | 17.55 | 1 | 0.0000 | 5.1251 | 2.3860 | 11.0090 |
| DST | 0.8447 | 0.3732 | 5.12 | 1 | 0.0236 | 2.3274 | 1.1199 | 4.8367 |
| Smoking | 0.8790 | 0.4328 | 4.13 | 1 | 0.0422 | 2.4084 | 1.0313 | 5.6251 |

## 4.5. Model Diagnostics

In model construction it is necessary to make diagnostics before concluding that the model with the five predictors. This includes: testing the assumption of proportional hazards, a check for influence and poorly fit subjects and overall summary measures of goodness of fit.

### 4.5.1. Test of the assumption of proportional hazards

One of the main assumptions of the Cox proportional hazard model is that the hazard ratios are constant overtime. That means, the risk of failure must be the same no matter how long subjects have been followed. In order to test this assumption, the extended Cox model is employed and a graphical display is used to substantiate the same. Thus, in this study, using a test based on the interaction of the covariates with the log of time is used to see if the assumption of proportionality is violated or not. Therefore, one of the statistical tests for proportional hazards assumption is to generate time varying covariates by creating interactions of the predictors and a function of survival times, usually covariate times log of time, and including these in the model. If a time-dependent covariate is significant, this indicates a violation of the proportionality assumption for that specific predictor. The result of the test is given in Appendix C (Table 1). The table shows the Wald chi-square value and corresponding p-values for each covariate. Since the p-value of the Wald test is greater than 0.05 for all covariates, there is no evidence against the proportionality of hazard assumption. All interactions of covariates with the logarithm of survival times are modelled together with the main effects. The Wald test is used to check the significance of the interaction terms at 5% level of significance. The result of the test indicates that none of the coefficients of interaction terms are significant at 5% level (i.e. higher p-value) (Appendix C: Table1). The result reveals non-significance of time-dependent covariates. On the other hand, there are no covariates which show a trend/pattern with time. This indicates that the hazard ratios will be constant over the study time showing that there is no sufficient evidence to reject the null hypothesis.

Another method of testing the proportionality assumption is by using the Schoenfeld and scaled Schoenfeld residuals. Since the test results in Table 2 (Appendix C) are not significant (p-values over 0.05) we cannot reject proportionality. In addition, the assumption of proportionality was also assessed graphically by plotting the scaled Schoenfeld residuals of each covariate against log time (Appendix D: Figure 2).

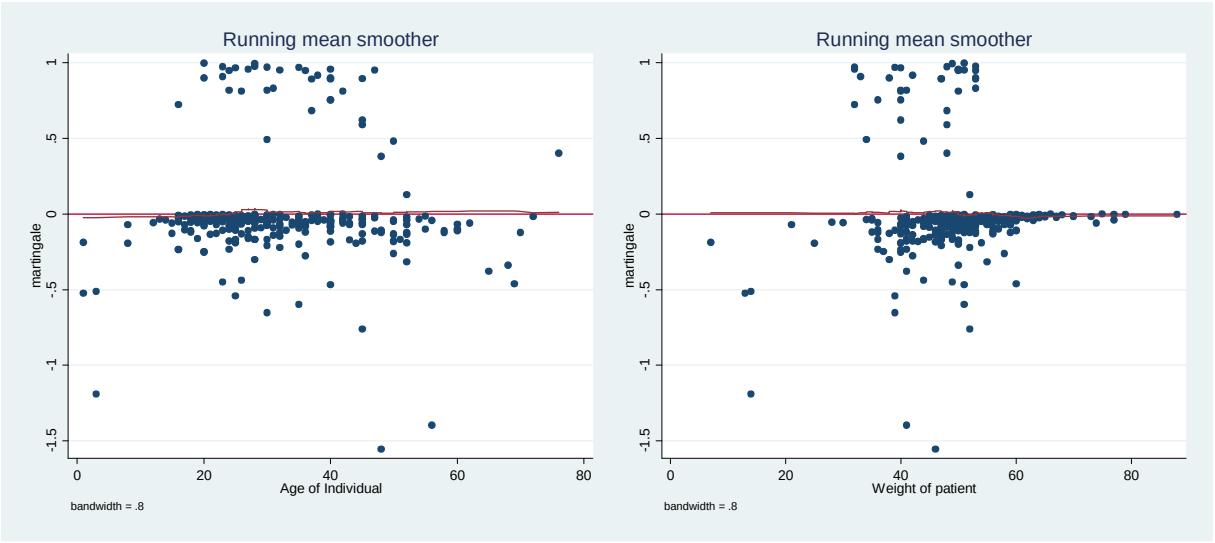
### 4.5.2. Assessment of Influential Observations

The next step we follow in regression diagnostics is to see if there are any observations that have undue influence on the estimates of the Cox regression parameters, or have an unexpected influence on the fit of the model. The DFBETA statistic for measuring the influence of the $i^{th}$ observation is defined as the one-step approximation to the difference in the MLE of the regression parameter vector with $i^{th}$and the MLE of the regression parameter vector without the $i^{th}$observation. As a result, the first five largest changes in parameter estimates are shown in Table 3 of Appendix C.

The largest difference for age occurs for patient 35. The change in the parameter estimate on omitting the data for this patient is -0.0043677. Therefore, omission of this patient decreases the hazard of death relative to the baseline hazard. The standard error of the parameter estimate for age in the full data set is 0.0133, and so the maximum amount by which this estimate changed when one observation is deleted is about 33% of the standard error (less than one standard error). Thus, the change in age effect by deleting this patient can be considered as insignificant.

Omitting the data from patient 82 from the dataset brought the largest changes in the parameter estimates for baseline weight. The maximum change in the parameter estimates for weight when this observation is omitted in turn is 0.0085845. The standard error of the parameter estimate for weight in the full data set is 0.0191. So the maximum amount by which this estimate changed when one observation is deleted is about 45% of the standard error. The effect of deleting these observations increases the relative hazard of death but this increment is not large. The differences in the parameter estimates for the levels of the categorical variables were also assessed. However, the largest differences are less than a quarter of the standard error of the corresponding estimate.

The largest difference for clinical complication occurs for patient 35. The change in the parameter estimate on omitting the data for this patient is -0.1068666. Therefore, omission of this patient decreases the hazard of death relative to the baseline hazard. The standard error of the parameter estimate for clinical complication in the full data set is 0.390. Hence the maximum amount by which this estimate changed when one observation is deleted is about 27% of the standard error (less than one standard error). Thus, the change in clinical complication effect by


Running mean smoother
martingale
1
.5
0
-.5
-1
-1.5
0
20
40
60
80
Age of Individual
bandwidth = .8
Running mean smoother
martingale
1
.5
0
-.5
-1
-1.5
0
20
40
60
80
Weight of patient
bandwidth = .8

### 4.5.4. Goodness of Fit of the Final Model

We can evaluate the fit of the model by using the Cox-Snell residuals. If the model fits the data well then the true cumulative hazard function conditional on the covariate vector has an exponential distribution with a hazard rate of one (see Figure 4.4 below).

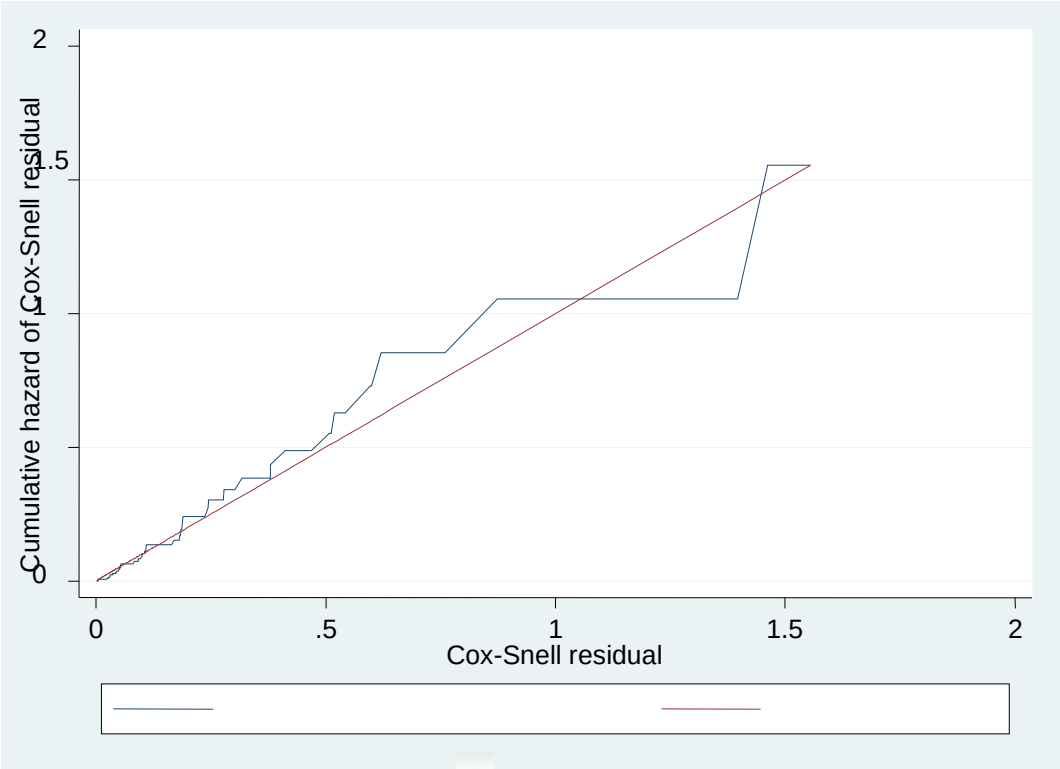



**Figure 4.4: Cumulative hazard plot of the Cox-Snell residuals of the proportional hazards Cox regression model in table 4. The $45^0$-straight line through the origin is drawn for reference.**

We see that the hazard function follows the 45 degree line very closely except for large values of time. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. Overall we would conclude that the final model fits the data very well.

## 4.6. Interpretation of the results

When the multiple proportional hazards Cox model is used in the analysis of survival data, the estimated coefficient of a categorical explanatory variable in the model can be interpreted as the logarithm of the ratio of the hazard of death to the baseline (reference group) hazard. On the other hand, the estimated coefficient for a continuous explanatory variable is the change in the

logarithm of the hazard ratio for a unit increase in the value of the explanatory variable after adjustment for the remaining variables in the model. Thus, the interpretations of those variables that were significant in the final proportional hazards model of MDR-TB patients treated in ALERT and UOGTRH are given in the following part of this section.

The model fit to the MDR-TB data, shown in Table 4.6 contains two continuous linear covariates (Age and baseline weight of a patient), three dichotomous covariates (clinical complication, drug susceptibility test and smoking status).

Suppose we take an increment of weight by kg to make comparisons. Then, the estimated hazard ratio for a five kg increase in initial weight will be 0.62 =exp (5*-0.095). This means that the hazard rate decreased by 38% for every 5kg increase in weight.

Age of MDR-TB patients is found to have a significant effect on the mortality of patients estimated HR=1.0372. Since the estimated hazard ratio is greater than unity, other things being equal, the higher the age of a patient, the greater the hazard of death at any given time. In particular an increase of one year in age of patients increases the risk of death by 3.7%. Since a one-year age increasing doesn't make sense the result based on a five-year increase of age is 20% increase in risk of death.

The hazard ratio (also called relative risk) indicates that patient with clinical complication experience approximately 5 times of the risk of death faced by patient with no clinical complication, while all other variables are held constant. Patient resistant to at least three anti TB drug including INH and RIF, while holding all other variables constant, are 2.3 times more likely to die than patient with only INH and RIF.

High mortality occurred among smokes ($\widehat{HR}$ 2.4084; 95%; $CI$ 1.0313 – 5.6251) than among non-smokers. The result reveals that the rate of dying among Smoker patients is 2.4 times higher than with non smoker patients, controlling other variables constant in the model.

# 5. DISCUSSION, CONCLUSIONS AND RECCOMENDATIONS

## 5.1. Discussion

This research was conducted with the objectives to study survival time and identify predictors of mortality in a cohort of MDR-TB patients. The analysis revealed that the proportional hazards assumption of the Cox regression model was not violated. Therefore, further stratification has not been made to estimate the hazard. We found higher hazard of death or equivalently lower survival rate in patients who were smokers, patients whose TB was found to be resistant to RIF, INH and at least one of other drugs and patients who had clinical complications during the treatment period.

The median survival of MDRTB patients was about 1.25 years which is less than the median survival for Lithuania 4 years (Balabanova et al (2011)) and UK 3.78 years (Drodniewski et al (2002)).

In the current study total death rate was 10.8% for the cohort and the rate is similar to that in another study where the death rate for pooled appropriate data from 21 countries was 11% (Johnston (2009)). Death rate of 7%, 15% and 23.4% had been reported for the Latvian cohort (Holtz et al (2010)), Uzbekistan cohort (Cox (2007)) and South African cohort (Farley et al (2011)), respectively.

A retrospectives study conducted by Balabanova et al (2011) in Lithuania, Oladimeji et al (2014) in Nigeria and Theodros et al (2013) in St. Peter TB specialized hospital, Addis Ababa, Ethiopia showed HIV seropositivity was a significant risk factor of death during MDR-TB treatment period. In our study HIV seropositivity was not significantly associated with death. A study conducted in Tanzania by Range et al (2012) showed HIV status was not found to be a risk factor of death among MDR-TB patients. A case control study conducted by Selamawit et al (2013) in Addis Ababa, Ethiopia showed that HIV status was not significantly associated with MDR-TB development. A similar finding was observed in the United Kingdom (Anderson et al (2012)) and four European Union countries (France, Germany, Italy and Spain) (Casal et al (2005)) in which HIV seropositivity was a not significant risk factor of death during MDR-TB treatment period. However, among patients who were immunecompromised risk of death was high.

The impact of clinical complication on survival rate has been assessed by several studies. Chan et al (2004) found that clinical complications due to pneumonectomy and lobectomy had been found to be associated with survival outcome of MDR-TB patients. A study conducted in Uzbekistan by Cox et al (2007) indicated that clinical complication is a risk factor of mortality and treatment failure. A retrospective cohort study conducted in St. Peter TB specialized hospital, Addis Ababa, Ethiopia by Theodros et al (2013) showed that patients with clinical complication had a higher hazard of mortality (HR 1.90, 95% CI 1.52 - 2.39. P < 0.001) than those who did not have any complication. The finding of the current study agrees with the above cited findings with the estimated hazard ratio ($\widehat{HR}: 5.1251; 95\%\ CI\ 2.3860 - 11.0090, P < 0.0001$).

A study by Dhingra et al (2008) suggested that co-morbidities did not influence the outcome of MDR-TB treatment. Theodros et al (2013) showed that co-morbidities did not have a significant influence on the risk of mortality associated with MDR-TB. The current study shows that co-morbidities were not risk factors of mortality. In contrast, a study by Khurram et al (2012) showed co-morbidities had influence on mortality among MDR-TB patients. A study of MDR-TB in Estonia, showed that co-morbidities (AOR, 2.62; 95% CI, 1.00-6.87) were independent risk factors for treatment failure (Kim et al (2007)).

A study conducted in New Delhi, India by Dhingra et al (2008) showed over 70% of patients were resistant to three or more antitubercular drugs. A study by Khurram et al (2012) conducted in Rawalpindi, Pakistan showed resistant to more than two drugs was a risk factors of mortality among MDR-TB patients. The finding of the current study also agrees with the above sited findings. In current study, 21% of the patients were resistant to at least one other first-line drug apart from INH and RIF.

A prospective study conducted by Farley et al (2011) in South Africa showed patients weight group (<45kg) and intermediate weight group at baseline (46-60kg) had high hazard of death $\left(\widehat{HR}: 2.52; 95\%\ CI\ 2.04 - 3.13, P < 0.0001\right)$ and ($\widehat{HR}: 1.48; 95\%\ CI\ 1.26 - 1.74, P < 0.0001$). A study by Dhingra et al (2008) conducted in New Delhi, India suggested that baseline weight predicts the poor outcome of MDR-TB patients. The current study also identifies low

weight as a risk factor with estimated hazard ratio ($\widehat{HR}: 0.9093; 95\%\ CI\ 0.8760 - 0.9440, P < 0.0001$).

A study conducted in Tanzania by Range et al (2012) showed that age was not a risk factor of poor outcome among MDR-TB patients. Oladimeji et al (2014) found that age was not a risk factor of mortality among MDR-TB patients in Nigeria. A prospective epidemiological case control study was conducted in four European Union countries (France, Germany, Italy, and Spain) by Casal et al (2005) showed that age is associated with MDR-TB (AOR=3.10). The current study found that age has a slight contribution to mortality of MDR-TB patients ($\widehat{HR}: 1.0372; 95\%\ CI\ 1.0106 - 1.0646, P < 0.006$).

A study conducted by Shin et al (2006) in Russia showed that alcohol use during treatment was significantly associated with poor treatment outcome. A study by Selamawit et al (2013) suggested that smoking status was not significantly associated with MDR-TB development. Theodros et al (2013) showed that smoking status has a significant influence on the risk of mortality associated with MDR-TB. The current study asserted that smoking is a risk factor of mortality among MDR-TB patients with estimated hazard ratio $\widehat{HR}: 2.4084[95\%\ CI\ 1.0313 - 5.6251, p < 0.043]$.

## 5.2. Conclusions

Mortality rate was high in the earlier months of MDR-TB treatment initiation and stabilized in later stage; 11 and 12 deaths occurred in the first and second three months of MDR-TB treatment initiation, respectively. Using Cox proportional hazard model covariates that are significantly influence the survival of MDR-TB patients are identified. Five main factors that are identified to affect survival of the patients significant at 5% level are age, baseline weight, clinical complication, DST, and smoking status. It is important to note that the conclusions have been reached based only on the factors/variables included in the study. This can be seen as a limitation.

## 5.3. Recommendations

- The strong associations of poor survival with social factor like smoking emphasize an urgent need for non-medical interventions on cessation of smoking to improve survival of patients.
- Health workers should be cautious when a patient has lower baseline weight and when MDR-TB patients are resistant to more TB drugs.
- Health workers and peer educators and data clerks, working with patients under MDR-TB, should be given special training to improve the quality of data records of patients. Moreover, attempt should be made to investigate the causes of deaths that occurred out of hospitals.

## References


1. Anderson L F, S Tamne, J P Watson, T Cohen, C Mitnick, T Brown, F Drobniewski, I Abubakar, Treatment outcome of multi-drug resistant tuberculosis in the United Kingdom: retrospective-prospective cohort study from 2004 to 2007. (Euro Surveill. 2013;18 (40):pii=20601).
2. Balabanova Y, Radiulyte B, Davidaviciene E,et al. Survival of drug resistant tuberculosis patients in Lithuania: retrospective national cohort study BMJ Open2011;1:e000351. doi:10.1136/bmjopen-2011-000351
3. Balabanova Y, Ruddy M, Hubb J et al. (2005) Multidrug-resistant tuberculosis in Russia: clinical characteristics, analysis of second-line drug resistance and development of standardized therapy. European Journal of Clinical Microbiology and Infectious Diseases 24, 136–139.
4. Breslow, N. (1974). Covariance Analysis of Censored survival data, Biometrics 30,89-99.
5. Casal M., M. Vaquero, H. Rinder, E. Tortoli, J. Grosset, S. Rusch-Gerdes, J. Gutierrez, and V. Jarlier. A Case-Control Study for Multidrug-Resistant Tuberculosis: Risk Factors in Four European Countries. Microb Drug Resist. 2005;1: 62-9.
6. Chan E. D., Valerie L, Matthew J. S, Julanie F. C, Mai-Lan N. H, Marian G, and Michael D. I. (2004). Treatment and outcome analysis of 205 patients with multidrug-resistant tuberculosis. American Journal of Respiratory and Critical Care Medicine 169, 1103–1109.
7. Collett, D. (1994). Modeling survival data in medical research. London: Chapman and Hall.
8. Cox HS, Kalon S, Allamuratova S, Sizaire V, Tigay Z. N, et al (2007) Multidrug-Resistant Tuberculosis Treatment Outcomes in Karakalpakstan, Uzbekistan: Treatment Complexity and XDR-TB among Treatment Failures. PLoS ONE 2(11): e1126. doi:10.1371/journal.pone.0001126
9. Cox, D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society. Series B (Methodological), 34, No.2, 187-220.
10. Dhingra V. K, Rajpal S., Mittal A and M. Hanif: Outcome of Multi-Drug Resistant Tuberculosis cases treated by individualized regimens at A Tertiary Level Clinic. Indian J Tuberc 2008; 55:15-21

11. Drobniewski F, I Eltringham, C Graham, J G Magee, E G Smith, B Watt, A national study of clinical and laboratory factors affecting the survival of patients with multiple drug resistant tuberculosis in the UK. Thorax2002; 57:810–816.
12. Efron Bradley; The Efficiency of Cox's Likelihood Function for Censored Data: Journal of the American Statistical Association; Vol. 72, No. 359 (Sep., 1977), pp. 557-565. Available at: http://www.jstor.org/stable/2286217 (accessed on May 21, 2014)
13. Farley JE, Ram M, Pan W, Waldman S, Cassell GH, et al. (2011) Outcomes of Multi-Drug Resistant Tuberculosis (MDR-TB) among a Cohort of South African Patients with High HIV Prevalence. PLoS ONE 6(7): e20436. doi:10.1371/journal.pone.0020436
14. Federal Ministry of Health (FMOH). Guidelines for clinical and programmatic management of TB, TB/HIV and leprosy in Ethiopia fifth edition march, 2013,Addis Ababa
15. Flament-Saillour M, Robert J, Jarlier V and Grosset J (1999) Outcome of multi-drug-resistant tuberculosis in France. A nationwide case–control study. American Journal of Respiratory and Critical Care Medicine 160, 587–593.
16. Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika; 81:515-526.
17. Holtz TH, V. Leimane, G. Dravniece, V. Riekstina, I. Sture, S. Kammerer, M.P. Chen, G. Skenders. (2010) Treatment outcome of multidrug/extensively drug-resistant tuberculosis in Latvia, 2000–2004. European Respiratory Journal: Print ISSN 0903-1936.
18. Hosmer, D.W. and Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York.
19. Johnston JC, Shahidi NC, Sadatsafavi M, Fitzgerald JM (2009) Treatment Outcomes of Multidrug-Resistant Tuberculosis: A Systematic Review and Meta-Analysis. PLoS ONE 4(9): e6914. doi:10.1371/journal.pone.0006914
20. Kalblfleisch J D and Prentice R L (2002) The Statistical analysis of Failure Time Data (second edition) a John Wiley & Sons, inc., publication., Hoboken, New Jersey
21. Kelly PM, Cumming RG, Kaldor JM (1999) HIV and tuberculosis in rural sub-Saharan Africa: a cohort study with two year follow-up. Trans R Soc Trop Med Hyg 93: 287-293.
22. Khan, R., "The Social Determinants of Multidrug Resistant Tuberculosis in the United States between 2005 and 2009" (2013). Public Health Thesis; Georgia State University.

23. Khurram M., Hamama Tul Bushra Khaar, Fahim M.: Multidrug-resistant tuberculosis in Rawalpindi, Pakistan. J Infect Dev Ctries 2012; 6(1):29-32.
24. Kim Doh Hyung, Hee Jin Kim, Seung-Kyu Park, Suck-Jun Kong, Young Sam Kim, Tae-Hyung Kim, Eun Kyung Kim, Ki Man Lee, Sung Soon Lee, Jae Seuk Park, Won-Jung Koh, Chang-Hoon Lee, Ji Yeon Kim, and Tae Sun Shim, Treatment Outcomes and Long-term Survival in Patients with Extensively Drug-resistant Tuberculosis, Am J Respir Crit Care Med Vol 178. pp 1075–1082, 2008
25. Kim HR, Hwang SS, Kim HJ, Lee SM, Yoo CG et al. (2007) Impact of extensive drug resistance on treatment outcomes in non-HIV-infected patients with multidrug-resistant tuberculosis. Clin Infect Dis, 45: 1290-1295. doi:10.1086/522537. PubMed: 17968823
26. Oladimeji O, Isaakidis P, Obasanya OJ, Eltayeb O, Khogali M, et al. (2014) Intensive-Phase Treatment Outcomes among Hospitalized Multidrug-Resistant Tuberculosis Patients: Results from a Nationwide Cohort in Nigeria. PLoS ONE 9(4): e94393. doi:10.1371/journal.pone.0094393
27. Range N., Friis H., Mfaume S., Magnussen P., Changalucha J., Kilale A., Mugomela A. and Andersen A. B.: Anti-tuberculosis drug resistance pattern among pulmonary tuberculosis patients with or without HIV infection in Mwanza, Tanzania: Tanzania Journal of Health Research.
28. Schoenfeld, D. (1982). Partialresiduals for the proportionalhazards regression model. Biometrica; 69: 239-241.
29. Selamawit Hirpa, Girmay Medhin, Belaineh Girma, Muluken Melese, Alemayehu Mekonen, Pedro Suarez and Gobena Ameni.: Determinants of multidrug-resistant tuberculosis in patients who underwent first-line treatment in Addis Ababa: a case control study.BMC Public Health201313:782.
30. Sharma SK, Mohan A, Kadhiravan T (2005) HIV-TB co-infection: Epidemiology, diagnosis and management. Indian J Med Res 121: 550-567.
31. Shin S. S., Pasechnikov A. D., Gelmanova I. Y., Peremitin G. G., Strelis A. K., Andreev Y. G., Golubchikova V. T., Tonkel T. P., Yanova G. V. Nikiforov M., Yedilbayev A., Mukherjee J. S., Furin J. J., Barry D. J., Farmer P. E., Rich M. L., and Keshavjee S.: Treatment outcomes in an integrated civilian and prison MDR-TB treatment program in Russia. INT J TUBERC LUNG DIS 10(4):402–408

32. Theodros Getachew, Alemayehu Bayray and Berhe Weldearegay: 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 January, 2013.
33. World Health Organization, Global Tuberculosis Control Report; 2010 tuberculosis.
34. World Health Organization, Global Tuberculosis Control Report; 2013 *tuberculosis global facts www.who.int/tb/data.* (accessed on March, 2015)
35. World Health Organization. Guidelines for the programmatic management of drug-resistance tuberculosis. Geneva, Switzerland; WHO; 2008.
36. Zignol M. Incidence of multidrug-resistant tuberculosis: 2003 global estimates; 2003. Available at: http://www.stoptb.org/wg/dots_plus/meetings.asp (accessed on June 5, 2014).

**APPENDIX A: Results of the multivariable proportional hazards Cox regression model Results of the multivariable proportional hazards Cox regression model containing the variables significant at 25% level in the univariable.**

| Testing Global Null Hypothesis (whether the model fit better than the null model) BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.48 | 5 | 0.0000 |

| Covariates/ Factors | Wald | Df | Sig. |
|---|---|---|---|
| Weight | 24.95 | 1 | 0.0000 |
| Age | 7.66 | 1 | 0.0057 |
| Clinical complications | 17.52 | 1 | 0.0000 |
| DST | 5.11 | 1 | 0.0238 |
| SmokingStatus | 4.10 | 1 | 0.0428 |

Note that the value of -2LL for the model containing the covariates in this table is 241.697

**APPENDIX B: The result of multivariable Cox hazard model those variables not significant in the univariate analysis by including one at a time.**

**Table 1: When sex is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.63 | 6 | 0.0000 |

| Covariates/ Factors | Wald | DF | Sig. |
|---|---|---|---|
| Weight | 24.21 | 1 | 0.0000 |
| Age | 7.59 | 1 | 0.0059 |
| Clinical complications | 17.05 | 1 | 0.0000 |
| DST | 5.15 | 1 | 0.0232 |
| SmokingStatus | 4.13 | 1 | 0.0422 |
| Sex | 0.15 | 1 | 0.6960 |

**Table 2: When Marital Status is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.55 | 7 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | DF | Sig. |
|---|---|---|---|
| Weight | 24.72 | 1 | 0.0000 |
| Age | 5.57 | 1 | 0.0183 |
| Clinical complications | 17.37 | 1 | 0.0000 |
| DST | 5.25 | 1 | 0.0219 |
| SmokingStatus | 3.54 | 1 | 0.0600 |
| Marital Status | 0.20 | 2 | 0.9045 |

**Table 3: When Educational level is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 52.08 | 7 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | DF | Sig. |
|---|---|---|---|
| Weight | 22.91 | 1 | 0.0000 |
| Age | 5.75 | 1 | 0.0165 |
| Clinical complications | 18.10 | 1 | 0.0000 |
| DST | 5.42 | 1 | 0.0199 |
| SmokingStatus | 4.43 | 1 | 0.0353 |
| Educational level | 0.65 | 2 | 0.7223 |

**Table 4: When Co-morbidities is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.44 | 6 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | DF | Sig. |
|---|---|---|---|
| Weight | 24.39 | 1 | 0.0000 |
| Age | 7.65 | 1 | 0.0057 |
| Clinical complications | 17.42 | 1 | 0.0000 |
| DST | 5.09 | 1 | 0.0241 |
| SmokingStatus | 3.94 | 1 | 0.0472 |
| Comorbidities | 0.01 | 1 | 0.9177 |

**Table 5: When Therapeutic delay is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.90 | 6 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | DF | Sig. |
|---|---|---|---|
| Weight | 24.97 | 1 | 0.0000 |
| Age | 6.91 | 1 | 0.0086 |
| Clinical complications | 17.46 | 1 | 0.0000 |
| DST | 4.68 | 1 | 0.0305 |
| SmokingStatus | 4.51 | 1 | 0.0337 |
| Therapeutic delay | 0.48 | 1 | 0.4901 |

**Table 6: When HIV Status is included**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 51.71 | 6 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | DF | Sig. |
|---|---|---|---|
| Weight | 24.79 | 1 | 0.0000 |
| Age | 7.37 | 1 | 0.0066 |
| Clinical complications | 17.47 | 1 | 0.0000 |
| DST | 4.74 | 2 | 0.0294 |
| SmokingStatus | 3.72 | 1 | 0.0537 |
| HIV Status | 0.29 | 1 | 0.5911 |

**Table 7: Wald statistics and corresponding p-values of possible interaction terms, added one at a time, to the variables included in the model in APPENDIX A.**

| Interactions | Wald $\chi^2$ | DF | Prob > chi2 |
|---|---|---|---|
| Age*Weight | 5.79 | 1 | 0.0161 |
| Age*ClinicalCompl | 1.73 | 1 | 0.1889 |
| Age*DST | 1.01 | 1 | 0.3143 |
| Age*SmokingStatus | 0.42 | 1 | 0.5158 |
| Weight*ClinicalComp | 0.00 | 1 | 0.9457 |
| Weight*DST | 0.09 | 1 | 0.7613 |
| Weight*SmokingStatus | 1.05 | 1 | 0.3060 |

| | | | |
|---|---|---|---|
| ClinicalComp*DST | 0.17 | 1 | 0.6789 |
| ClinicalComp*SmokingStatus | 1.73 | 1 | 0.1881 |
| Drugsusctest*SmokingStatus | 1.39 | 1 | 0.2381 |

**Table 8: Results of the multivariable proportional hazards Cox regression model containing an interaction**

| Testing Global Null Hypothesis | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 60.46 | 6 | 0.0000 |

| Covariates/ Factors | Wald $\chi^2$ | Df | Sig. |
|---|---|---|---|
| Weight | 0.19 | 1 | 0.6640 |
| Age | 8.56 | 1 | 0.0034 |
| Clinical complications | 15.66 | 1 | 0.0001 |
| DST | 7.42 | 1 | 0.0064 |
| SmokingStatus | 6.57 | 1 | 0.0104 |
| Age*Weight | 5.79 | 1 | 0.0161 |

**APPENDIX C: Results of Model diagnostics**

**Table 1: Result of test of proportional hazards assumption (test results based on the Wald test)**

| | Covariates | DF | Coef. | Std. Err. | Wald $\chi^2$ | p-value | Haz. Ratio |
|---|---|---|---|---|---|---|---|
| Main | Weight | 1 | -0.0569 | 0.03610 | 2.48 | 0.1150 | 0.9447 |
| | Age | 1 | 0.0019 | 0.02663 | 0.01 | 0.9407 | 1.0019 |
| | 2.Clinicalco~s | 1 | 2.5416 | 0.7228 | 12.39 | 0.0004 | 12.6995 |
| | 2.drugSustes~t | 1 | 0.4628 | 0.7286 | 0.40 | 0.5253 | 1.58856 |
| | 2.SmokingS~s | 1 | -0.3910 | 0.9322 | 0.18 | 0.6749 | 0.6764 |
| tvc (time-varying covariets ) | Weight | 1 | -0.0255 | 0.01977 | 1.67 | 0.1963 | 0.9748 |
| | Age | 1 | 0.0210 | 0.01364 | 2.38 | 0.1232 | 1.0212 |
| | Clinicalco~s | 1 | -0.6045 | 0.4418 | 1.87 | 0.1712 | 0.5463 |
| | drugSustes~t | 1 | 0.2845 | 0.3956 | 0.52 | 0.4720 | 1.3291 |
| | SmokingSta~s | 1 | 0.8776 | 0.5113 | 2.95 | 0.0861 | 2.4050 |

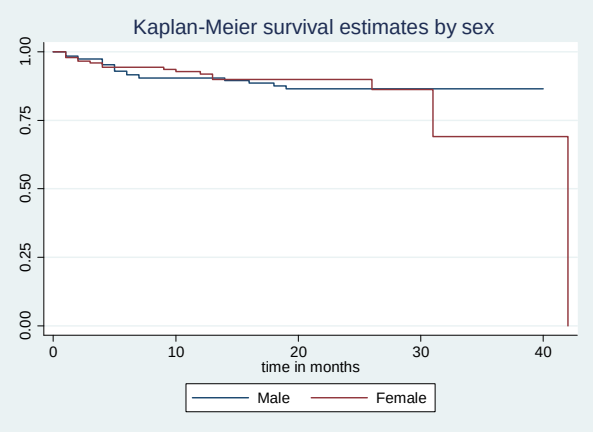
**Table 2: Test of proportional hazards assumption (test results based on Schoenfeld residuals)**

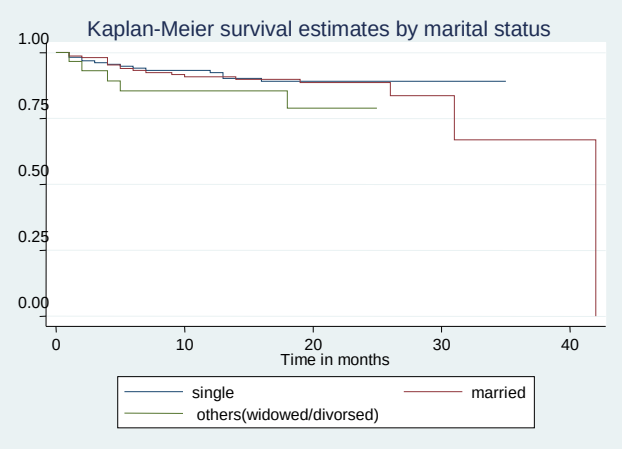
| | Rho | Chi Square | Df | Prob>Chi-square |
|---|---|---|---|---|
| Weight | -0.2109 | 1.19 | 1 | 0.2749 |
| Age | 0.1896 | 1.11 | 1 | 0.2913 |
| Clinical complication | -0.1566 | 0.96 | 1 | 0.3278 |
| DST | -0.0993 | 0.39 | 1 | 0.5304 |
| Smoking status | -0.1014 | 0.44 | 1 | 0.5057 |
| Global test | | 3.16 | 5 | 0.6755 |

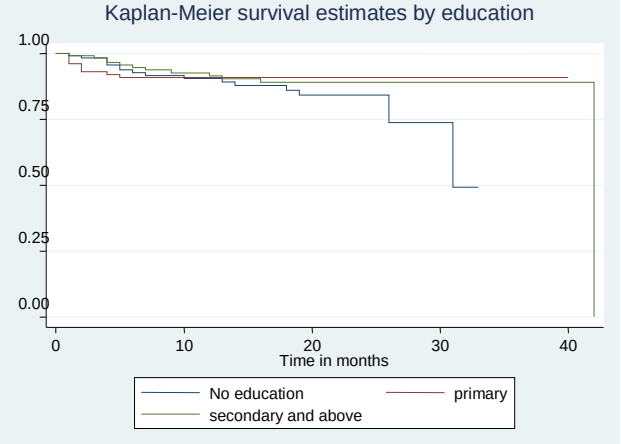
**Table 3: The five highest approximate differences in the parameter estimates of the variables included in the model in Appendix A**

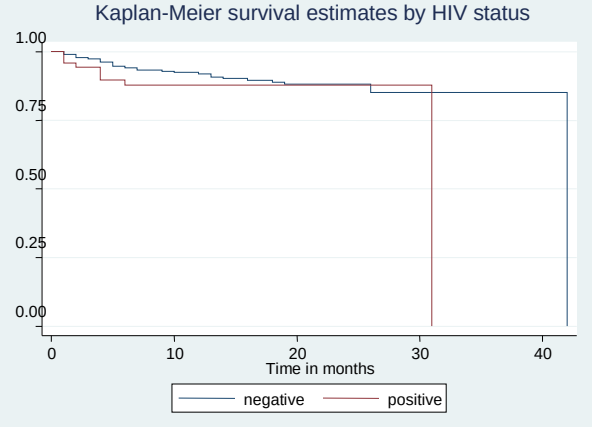
| Covariates | Deleted observations | $\Delta_{(i)} = \beta - \beta_{(i)}$ | $\lvert\beta - \beta_{(i)}\rvert$ |
|---|---|---|---|
| Age | 35 | --0.0043677 | 0.0043677 |
| | 240 | -0.0036268 | 0.0036268 |
| | 159 | -0.0028012 | 0.0028012 |
| | 235 | 0.0025676 | 0.0025676 |
| | 192 | -0.0025662 | 0.0025662 |
| Weight | 82 | 0.0085845 | 0.0085845 |
| | 159 | 0.0054742 | 0.0054742 |
| | 223 | 0.0046785 | 0.0046785 |
| | 35 | 0.0044252 | 0.0044252 |
| | 240 | 0.004178 | 0.004178 |
| Clinical complication, Yes | 35 | -0.1068666 | 0.1068666 |
| | 9 | 0.1063618 | 0.1063618 |

| | | | |
|---|---|---|---|
| | 240 | 0.0987577 | 0.0987577 |
| | 181 | 0.09711645 | 0.09711645 |
| | 100 | -0.0892519 | 0.0892519 |
| DST, INH, RIF and at least one of other drugs | 82 | -0.1070685 | 0.1070685 |
| | 33 | 0.084071 | 0.084071 |
| | 266 | 0.082851 | 0.082851 |
| | 38 | 0.0787478 | 0.0787478 |
| | 159 | -0.0768021 | 0.0768021 |
| Smoking status, Yes | 9 | -0.1764252 | 0.1764252 |
| | 192 | 0.1463362 | 0.1463362 |
| | 20 | 0.1387952 | 0.1387952 |
| | 33 | 0.1328053 | 0.1328053 |
| | 266 | 0.1258256 | 0.1258256 |

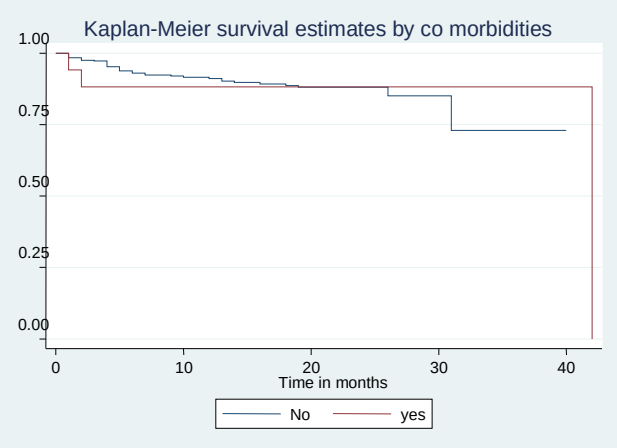


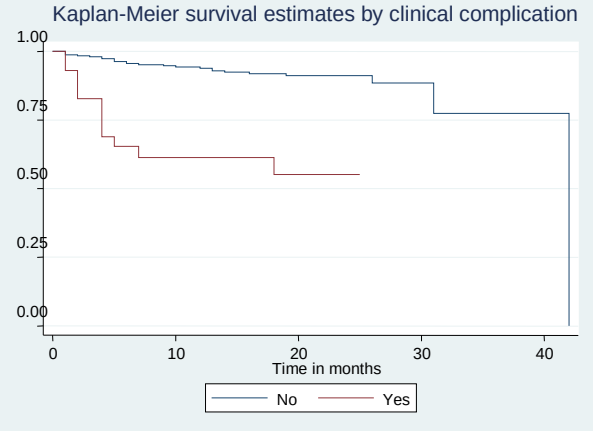



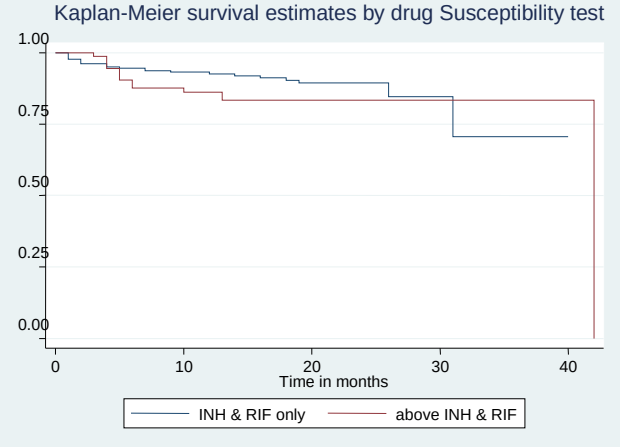
e.


Kaplan-Meier survival estimates by co morbidities
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Time in months
No
yes

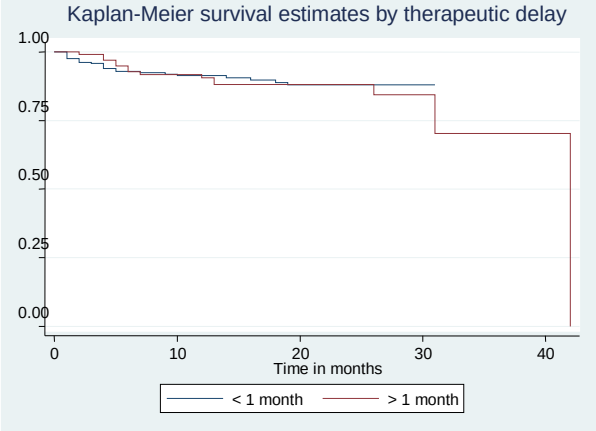
f.


Kaplan-Meier survival estimates by clinical complication
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Time in months
No
Yes

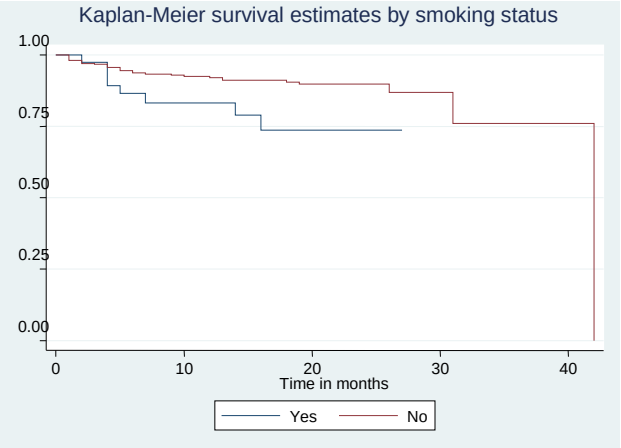
g.

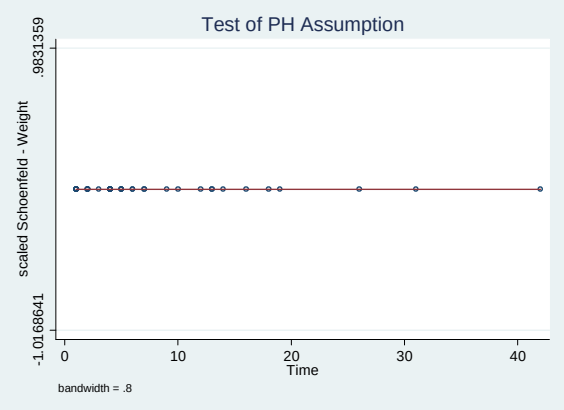

Kaplan-Meier survival estimates by drug Susceptibility test
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Time in months
INH & RIF only
above INH & RIF

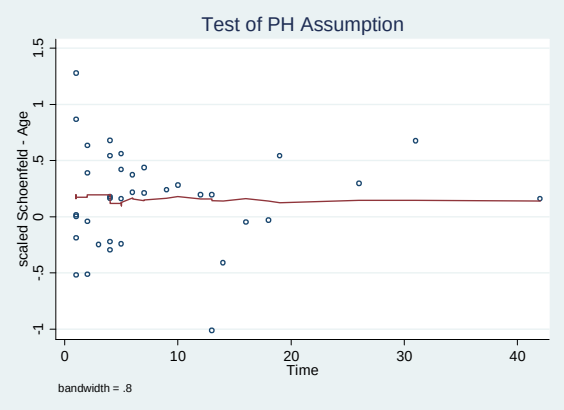
h.

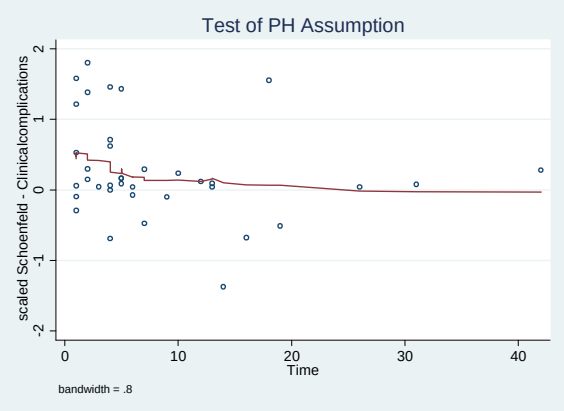

Kaplan-Meier survival estimates by therapeutic delay
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Time in months
< 1 month
> 1 month

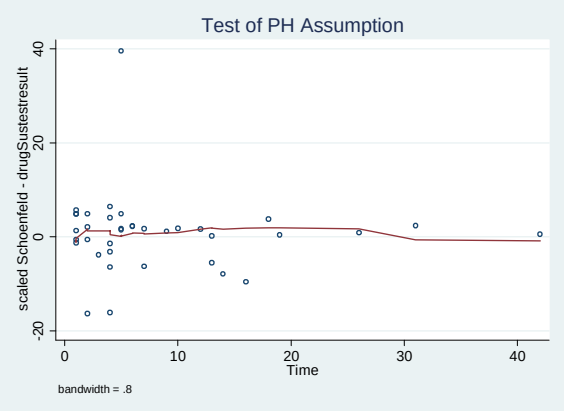
i.

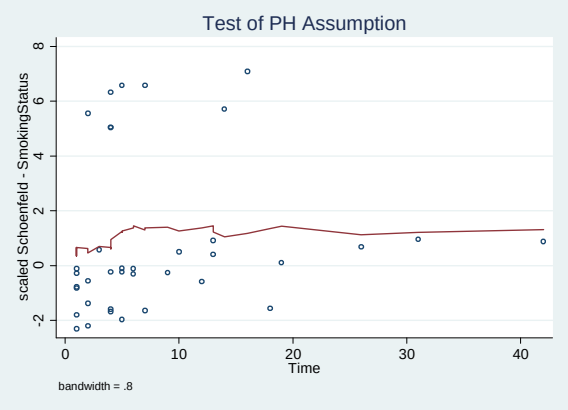

Kaplan-Meier survival estimates by smoking status
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Time in months
Yes
No


Test of PH Assumption
scaled Schoenfeld - Weight
.9831359
-1.016864
0
10
20
30
40
Time
bandwidth = .8


Test of PH Assumption
scaled Schoenfeld - Age
1.5
1
.5
0
-.5
-1
0
10
20
30
40
Time
bandwidth = .8


Test of PH Assumption
scaled Schoenfeld - Clinicalcomplications
2
1
0
-1
-2
0
10
20
30
40
Time
bandwidth = .8


Test of PH Assumption
scaled Schoenfeld - drugSustestresult
40
20
0
-20
0
10
20
30
40
Time
bandwidth = .8


Test of PH Assumption
scaled Schoenfeld - SmokingStatus
8
6
4
2
0
-2
0
10
20
30
40
Time
bandwidth = .8

# DECLARATION

I, the undersigned, declare that this thesis is my original work, has not been presented for degrees in any other University and all source materials used for the thesis have been duly acknowledged.

Name: Solomon Molalign

Signature: …….

Date: ……

Place: College of Natural Science, Addis Ababa University

This thesis has been submitted for examination with my approval as a University advisor.

Name: Professor Eshetu Wencheko

Signature: ……..

Date: ……..

Place: College of Natural Science, Addis Ababa University