ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

RISK FACTORS OF MORTALITY AMONG HIV/TB CO-INFFECTED PATIENTS UNDER ART: A CASE STUDY IN CHAGNI HEALTH CENTER, NORTHWEST ETHIOPIA.

Belete Mulatu

A THESIS SUBMITTED TO

THE DEPARTMENT OF STATISTICS

PRESENTED TO INPARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

Addis Ababa University

Addis Ababa, Ethiopia

September, 2013

This is to certify that the thesis prepared by Belete Mulatu, entitled: *Risk factors of mortality among HIV/TB co-infected patients under ART in Chagni Health Center, northwest Ethiopia* and submitted in partial fulfillment of the requirements for the degree of master of science in Statistics complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Advisor_____Signature_____Date_____

Examiner _____Signature_____ Date_____

Examiner_____Signature_____ Date_____

_____

Chair of Department or Graduate Program Coordinator

Abstract

Risk factors of mortality among HIV/TB co-infected patients under ART in Chagni Health Center, northwest Ethiopia.

Belete Mulatu

Addis Ababa University, 2013

HIV infection is a major risk factor for the development of tuberculosis (TB). Tuberculosis is a cause of significant morbidity and mortality in patients with AIDS. The objective of this study was to identify risk factors of mortality among HIV/TB co-infected patients in Chagni Health Center, Awi zone in Amhara Regional State, northwest Ethiopia. This is a retrospective study based on 510 cases of HIV-infected TB patients of age 15 years and above who received anti-TB treatment between September 1, 2010 and April 30, 2013 and were followed until July 31, 2013. The Kaplan-Meier method was used to estimate survival time and Cox's regression model was employed to identify covariates that have statistical significant effects on the survival of HIV/TB co-infected patients. The descriptive analysis showed that out of 510 patients, 200 (39%) patients died during ART and anti-TB treatment period, of these 113 (56.5%) were females. 303 (59.4%) patients had $CD4\ count \leq 200\ cells/mm^3$, 272 (53.3%) patients were in the age group $(15-30)$, 265 (52.2%) patients were under-weight ($BMI \leq 18.5$) and 234 (45.9%) patients started treatment on D4T-based drug regimens. The results of the multivariate PH Cox regression analysis showed that having $CD4\ count \leq 200\ cells/mm^3$ at the start of ART, presence of extra-pulmonary TB, being bedridden and WHO stage IV were associated with higher risk of mortality. On the other hand, those patients who took AZT-based drug regimens and having normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) experienced lower risk of mortality.

Acknowledgement

I would like to extend my sincere gratitude to my advisor Prof. Eshetu Wencheko for his unreserved guidance and comments from the stage of the research proposal up to thesis write-up.

My sincere acknowledgment also goes to my friends and colleagues who were there for me every step of the way especially I want to say thank you to, Sisay Awoke, and my wife.

I would also like to thank the Chagni Health Center manager Werkit Snishew for allowing me to obtain data and the manager of Chagni finance office for allowing me to use generator. I would also like to extend my sincere gratitude to the department of statistics for allowing me to print from the stage of the research proposal up to thesis write-up.

Finally I would like to thank Marta (ART team manager), Yeruk, Rahma (TB test manager) and Dereje (data clerk) of Chagni Health Center ART clinic for helping me to collect the data for this study.

'Thank you all!!

List of Abbreviations (Acronyms)

AIDS	Acquired Immune Deficiency Syndrome
aHR	adjusted Hazard Ratio
ART	Antiretroviral therapy
ARV	Antiretroviral
AZT	Zidovudine
BMI	Body Mass Index
CI	Confidence Interval
D4T	Stavudine
DOTS	Directly Observed Treatment, Short-Course
EPTB	Extra-pulmonary TB
FMOH	Federal Ministry of Health
HAART	Highly Active Antiretroviral Therapy
HCT	HIV counseling and testing
HIV	Human Immunodeficiency Virus
IQR	Inter Quartile Range
K-M	Kaplan-Meier
LR	Likelihood Ratio
M.TB	Mycobacterium Tuberculosis
MLE	Maximum Likelihood Estimate/Estimator
MOH	Ministry of Health
OI	Opportunistic Infections
OR	Odds Ratio
PH	Proportional Hazard
PIHCT	Provider-initiated HIV Counseling and Testing
PL	Partial Likelihood
PLWHA	People Living With HIV/AIDS
PTB	Pulmonary TB
TB	Tuberculosis
TDF	Tenofovir
UNAIDS	Joint United Nations Program on HIV/AIDS
WHO	World Health Organization
XDR	Extreme Drug Resistance

Table of Contents

List of Tables

List of figures

CHAPTER ONE

INTRODUCTION

1.1 Background

Tuberculosis (TB) is a deadly infectious disease caused by various species of Mycobacteria, mainly Mycobacterium tuberculosis (MTB). TB is the most common cause of death among people living with HIV/AIDS (PLWHA). The World Health Organization (WHO) recommends that all TB patients should be offered HIV testing and all PLWHA should be screened for TB. To increase uptake of HIV testing, WHO also recommends provider-initiated HIV testing and counseling (PIHTC) (WHO, 2004).

The human immunodeficiency virus (HIV) pandemic presents a massive challenge to the control of tuberculosis at all levels (TB/HIV guideline). HIV infection is a major risk factor for the development of TB and TB seems to make HIV infection worse. The increase in reported cases of TB since the mid-1980s is attributed, in part, to TB occurring in persons infected with HIV, the virus that causes AIDS. HIV robs the body off its natural ability to fight infection, making people with AIDS more likely to develop TB. HIV-infected person have weakened immune systems, and therefore have a much greater chance of developing active TB disease either by activation of latent infection or by being newly infected (WHO, 2004). As the immune system becomes increasingly damaged by HIV it becomes susceptible to opportunistic infections. Tuberculosis is the most common opportunistic infection complicating HIV infection especially in developing countries, and may occur at any stage in the course of immunodeficiency. TB/HIV co-infection significantly changes the natural history of both diseases and it is recognized as a major setback of both HIV and TB

infection control program (Godfrey-Faussett *et al*., 2002). HIV is the most important factor fuelling TB epidemic in populations with high HIV prevalence (WHO, 2004). This is because of the impaired cell-mediated immune responses associated with HIV, which results in enhanced susceptibility to TB and rapid progression of latent TB to active disease. In addition, HIV infection may also have an indirect effect on the incidence of TB by increasing transmission rates of MTB, with negative consequences for both HIV-negative and HIV-positive persons (Odhiambo *et al*., 1999; Sonnenberg *et al.*, 2004). Thus, in communities with high HIV rates, TB rates have been rising dramatically even where effective TB control strategies like Directly Observed Treatment Short-course (DOTS) is available. This underscored need to improve TB control wherever HIV co-infection is common. In response to this, WHO has developed a strategic framework with the goal of integrating TB and HIV/AIDS control programmes. One of the elements of this strategic framework is the widespread use of Highly Active Antiretroviral Therapy (HAART) as effective tool to TB control at the community level in countries with either an overlapping TB or HIV epidemic (WHO, 2003a). People whose immune systems have deteriorated due to advanced HIV disease were at greatly increased risk of developing active TB disease from a previously contained latent infection. Some immunocompromised people may not be able to contain a new TB infection and will immediately progress to active disease upon exposure to TB. Additionally people with HIV are more likely to develop extra-pulmonary TB (TB outside of the lungs) that may involve multiple organs and was harder to diagnose (Zvandasara et al., 2006).

1.2 Tuberculosis and HIV co-epidemics

Tuberculosis and acquired immunodeficiency syndrome are two completely different diseases. TB is caused by the tubercle bacillus, mycobacterium tuberculosis and spread through air. TB attacks the lungs, but sometimes other parts of the body too. AIDS is caused by human immunodeficiency virus. HIV can be spread by unprotected sexual relations, contaminated needles, breast milk, and transmission from an infected mother to her baby at birth. HIV attacks the immune system. As the two leading causes of infectious disease-associated mortality worldwide, TB and HIV diseases have been bound together from the early years of HIV/AIDS epidemic (Friedland et al., 2007). It has been known that there is an interaction of TB and HIV infection. HIV infection is the greatest known risk factor for progression of active TB disease (Daley et al., 1992) and reactivation of latent TB infection (Murray and Gursed, 1990). People infected with TB have only 10% chance of ever getting active TB in their lifetime. However, people infected with both HIV and TB have 50% chance of getting active TB in their lifetime (The White House Office of National AIDS Policy, 1995). HIV infection also increases TB case fatality (Ackah et al., 1995). On the other hand, TB is a leading cause of morbidity and mortality in population with high HIV prevalence. It has been shown that active TB promotes HIV replication, increases virus load, and so accelerates HIV disease progression and mortality (Toossi et al., 2001). For these reasons, the prevention and treatment of TB in HIV infected people is an important concern.

Globally, TB morbidity and mortality is partly attributed to the co-infection with HIV. TB/HIV co-infection significantly changes the natural history of both diseases. HIV related TB remains a serious challenge for health sector response to HIV. Thus, worldwide TB is one of the most common opportunistic infections and leading cause

of morbidity and mortality among HIV/AIDS patients accounting for about 30% of all death of HIV/AIDS patients (Raviglione, 1992 and Sharma, 2005). All countries with a generalized HIV epidemic state should aim to ensure that HIV testing is actively promoted and offered to all TB patients. The data available from these initiatives can form the basis of a reliable surveillance system that achieves high coverage (> 80%) of testing among TB patients (WHO, 2004). In some countries in Sub-Saharan Africa, up to 80% of people with TB are also living with HIV (UNAIDS update, 2009). Thus, the high rate of HIV infection and poorly functioning TB control programs in many parts of Africa are likely to contribute to the emergence of extreme drug resistant (XDR) tuberculosis, which could destabilize the control of TB in these areas (Dheda et al., 2010). TB is the leading cause of death among people who were HIV-positive, accounting for 13% of AIDS deaths worldwide. TB continues to claim more lives in Africa where the TB epidemic is still driven by the spread of HIV, which increases the likelihood of dying from the disease (Dheda et al., 2010). HIV-infection is now the most important single predictor of TB incidence in Sub-Saharan Africa. The region accounts for 70 % of the world's 14 million people who are co-infected (UNAIDS update, 2009). In some countries of Sub-Saharan Africa, up to 70% of patients with smear positive pulmonary TB is HIV-positive (Zachariah, 2010).

1.3 Global and Sub-Saharan African burden of HIV/TB co-infection

HIV/TB co-infection presents special challenges to the expansion and effectiveness of Directly Observed Therapy Short-course (DOTS) programs and the Stop TB Strategy. TB accounts for one-third of acquired immune-deficiency syndrome (AIDS) deaths worldwide and is one of the most common causes of morbidity in people living with HIV/AIDS (PLWHA) (WHO, 2009; Tsiouris et al., 2007). The global impact of the converging dual epidemics of TB and HIV remains as one of the major public health

problem or challenges of our time. Since 1990, the incidence of TB globally has risen dramatically, particularly in parts of the world where there was a significant overlap between HIV and TB epidemics. HIV-related TB remains a serious challenge for health-sector response and for people living with HIV. Currently, of the estimated 34 million people living with HIV, about one-third were estimated to have concomitant latent infection with mycobacterium tuberculosis. In 2010, of 8.8 million incident TB cases worldwide, 1.1 million (range, 0.9–1.2 million) were among people living with HIV and globally, there were an estimated 0.35 million (range, 0.32–0.39 million) deaths of HIV-associated TB. Sub-Saharan Africa continues to account for the global majority of the people living with HIV and TB, with an estimated 82%. In 2007, about 79% of the estimated total people living with HIV and TB were in this region posing other burdens like risk of drug resistant forms of TB due to difficulty and delays in diagnosis, increased mortality and greatly reduced survival time (UNAIDS update, 2009). HIV was the strongest risk factor for developing active TB disease, and in some countries up to 82% of people with TB have HIV (WHO, 2011). The dual epidemics were also of growing concern in Asia, where two-thirds of TB-infected people live and where TB now accounted for 40 percent of AIDS deaths. South-east Asia, mainly India, accounts for 13% of the remaining cases. However, in 2010, only 34% of TB patients globally knew their HIV status and only 5% of people living with HIV were screened for TB (WHO, 2011). Eastern Europe and the former Soviet Union have the fastest growing HIV epidemic in the world, and HIV co-infection is a factor that could exacerbate problems with multidrug-resistant TB (MDR-TB) epidemic in these regions (WHO, 2009). The overlap of TB/HIV co-infection with MDR-TB and extensively drug-resistant TB presents a tremendous challenge and threatens progress in controlling both TB and HIV/AIDS (WHO, 2009). Globally

about 11% of new adult cases of tuberculosis were HIV/AIDS co-infected where in Sub-Saharan Africa about 31% of new tuberculosis cases were HIV/AIDS co-infected (Corbett *et al.*, 2003). TB is the initial clinical manifestation of HIV disease for many people in Sub-Saharan Africa, and very high proportion of TB patients in this region are co-infected with HIV. HIV counseling and testing of TB patients is therefore an important method for identifying HIV-infected individuals, all of whom require HIV care and many of whom are indicated for antiretroviral therapy (ART) (Zachariah, 2010).

1.4 National burden of HIV/TB co-infection

Sub-Saharan African countries continued taking the leading position in HIV/TB morbidity and mortality rate, where TB epidemic was primarily driven by HIV infection. Ethiopia is one among these countries most heavily affected by HIV and TB co-infection. The global burden of death and disease caused by TB is concentrated particularly in low-income countries (WHO Report, 2011). Sub-Saharan Africa, including Ethiopia, is the area of highest prevalence of TB infection and in 2008, WHO report showed that Ethiopia ranks 7th among the world's 22 countries with a high tuberculosis burden with TB in an estimated annual incidence of 379 cases and prevalence of 643 cases per 100,000 populations (WHO, 2008) and the prevalence of HIV among TB patients is up to 41% (Demissie *et al.*, 2000; Yassin *et al.*, 2004). The high burden of TB in Ethiopia might in part be attributed to the rapid increase of HIV infection, because available data indicated that HIV/AIDS accounted for an estimated 32% or 141,000 total TB cases in 2005 (FMOH, 2007). On the other hand, WHO global report of 2008 estimated that in Ethiopia 40% of TB patients tested for HIV were HIV-positive. In Ethiopia routine data from 44 sites in the year 2005/6 showed 41% of TB patients were HIV-positive. In addition, another routine data collected in

2006/07 estimated that 31% of TB patients were HIV-positive (FMOH, 2007). TB was the cause of 76,000 deaths in Ethiopia, out of which 30% were among HIV-positive patients (WHO, 2009).

However, WHO recommended different collaborative activities for HIV/TB co-infections where one is initiation of antiretroviral therapy (ART). ART is an essential treatment for HIV infection in order to reduce the risks of death and HIV-related morbidities, or to improve survival time of PLWHIV. In addition, several observational studies, mostly done in resource-rich settings, have documented that HIV-infected TB patients receiving ART have higher rates of survival during and after TB treatment compared with patients who do not receive ART. Currently, the widespread use of highly active antiretroviral therapy (HAART) in developing countries has reduced the risk of treated HIV-infected persons developing TB by 70%–90%, compared with untreated individuals (Girardi *et al.*, 2000; Badri *et al.*, 2002).

1.5 Statement of the problem

HIV counseling and testing is fundamental to both HIV/AIDS prevention and TB treatment. Patients need to know their sero-status to benefit from available care and treatment options. Therefore, multi-focused counseling and testing strategies needed be instituted in order to reach risk groups (WHO, 2003a). The National TB and HIV guideline in Ethiopia recommends HIV counseling and testing as a routine care for TB patients. Chagni Health Center was implementing Provider Initiated HIV Counseling and Testing (PIHCT) in response to the high HIV prevalence among TB patients, and in an attempt to increase the uptake of HIV testing and ART. It is a well known fact that Ethiopia is one of the most countries hit by HIV/TB co-epidemics. Ever since the first infections were identified in the country, a lot has been done both

by the government and other stake holders to curb the effect of the diseases specially to suppress the risk of transmissions (FMOH, 2007). Numerous researches have also been conducted in an attempt to address many of the issues that arise in connection with HIV/TB co-epidemics. However, many of these research works mainly focused on the assessment of the prevalence and the study of numerous prevention measures that should be undertaken to stop or reduce the spread of the epidemics. And in most cases these studies were limited to urban areas mainly Addis Ababa. Given this as a drop back, this thesis focused on the consideration of some of the possible factors/variables that may highly influence the survival status of patients who are under ART in Chagni Health Center, northwest Ethiopia. Hence, this study deals with the risk factors of mortality among HIV/TB co-infected patients during ART and anti-TB treatment period.

1.6 Objective of the study

The general objective of the study was to identify the risk factors of mortality among HIV/TB co-infected patients treated under ART in Chagni Health Center, northwest Ethiopia.

The specific objectives of the study are:

- To compare the survival experience of HIV/TB co-infected patients in various categories of covariates.
- to identify the levels of risk factors that affect survival times of HIV/TB co-infected patients.
- To fit an appropriate model.

1.7 Significance of the study

The outcome of this study may provide information about high risk factors for mortality or the most influential covariates that have significant impact on survival of HIV patients during TB treatment by analyzing the impact of different covariates on survival of HIV/TB co-infected patients.

Specifically:

1. The results are expected to give some knowledge about risk factors of mortality among HIV/TB co-infected patients.
2. The results of this study could be used as input for other studies related to HIV/TB co-infected deaths.
3. This study could provide information to government and other concerned bodies in setting policies, strategies and further investigation for reducing HIV/TB mortalities.

1.8 Limitation of the study

- The study is conducted based on secondary data which might have incomplete and biased information.
- The study was restricted to adults in the age groups > 15 years, and results might not be applicable to infants and children.
- Parts of information on individuals are missed because of censored observations.
- The data on TB site do not identify those patients infected with both pulmonary TB and Extra-pulmonary TB.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. The Effect of ART on HIV/TB co-epidemics

The convergent human immunodeficiency virus (HIV) and tuberculosis (TB) pandemics continue to collectively exact significant morbidity and mortality worldwide. Highly Active Antiretroviral Therapy (HAART) has been a critical component in combating the scourge of these two conditions as both a preemptive and therapeutic modality. However, concomitant administration of antiretroviral and anti tuberculosis therapies poses significant challenges, including cumulative drug toxicities, drug-drug interactions, high pill burden, and the immune reconstitution inflammatory syndrome (IRIS), thus complicating the management of co-infected individuals.

Furthermore, the human immunodeficiency virus and HIV-associated tuberculosis (TB-HIV) epidemics remain uncontrolled in many resource-limited regions, especially in Sub-Saharan Africa. Ideally, the TB-HIV epidemic should be tackled through the widespread scale-up of preventive measures to enhance the WHO DOTS strategy for TB control. Frequent widespread HIV testing with early HIV diagnosis would permit implementation of the 'Three Is' strategy (intensified case finding, isoniazid preventive treatment and infection control) used in combination with a package of HIV care that includes timely initiation of ART. These interventions would reduce TB-HIV incidence rates (WHO, 2008; Lawn et al., 2009 and Harris et al., 2010). The optimal time to start treatment for HIV/AIDS has been a contentious issue since the introduction of HAART. In the revised 2009 guidelines, it is

recommended that HAART is initiated on all HIV patients with CD4 counts < 350 cells/µl, regardless of symptoms (Johansson et al., 2010).

The following literature search was used to further understanding of the risk factors affecting HIV/TB co-infected mortality and to select the covariates that could be important to include in this research. It is hoped that this review will serve as a standard for comparison for the intended analysis.

Collins et al (2010) conducted a retrospective cohort study of HIV patients who developed TB (excluding neurological disease) from January 1995 to December 2000 in a Peruvian HIV reference centre, Peru, and the patients were tracked for 24 months after TB diagnosis. They evaluated the survival and the effect of ART on mortality on these patients. The objective of the study was to assess the effect of antiretroviral therapy on survival of HIV-Infected tuberculosis patients. Survival was determined by Kaplan-Meier method and effect of ART on survival was evaluated using the Cox proportional hazards model. One hundred patients were studied (mean age 36 years, 79% male, median CD4 count 59 cell/µl). Fifty three received some form of ART: 31 mono/dual ART and 22 highly active antiretroviral therapy (HAART). 77% (36/47) of patients died in the no ART group, 65% (20/31) in the mono/dual ART group, and 27% (6/22) in the HAART group (p = 0.0001). The median survival time was 14 months (95% CI: 12–23) in the overall cohort, and 9 months (95% CI: 7–15) in the no ART group. The result of the study from the multivariable analysis revealed that a reduction of 70% in the risk of death (aHR = 0.3; 95% CI: 0.1–0.6) among mono/dual ART group patients, and 90% (aHR = 0.1; 95% CI: 0.06–0.4) among HAART patients, compared to those in the no ART control group. The study also suggested that patients with TB and HIV who did not receive ART had a short

survival time, and this was significantly improved with the addition of ART. The risk of death was reduced with both mono/dual ART and HAART.

Stockdale et al (2013) conducted an observational cohort study of all HIV-infected TB patients at a rural hospital in Kenya between 2005 and 2009. They analyzed the association between timing of initiation of ART and mortality, using Cox regression survival analysis, adjusted for measured confounders. The objective of the study was to provide information on the effect of timing of antiretroviral therapy (ART) initiation on outcomes of TB infection in real-life, non-clinical trial, rural settings in sub-Saharan Africa. A total of 404 antiretroviral-naive HIV/TB co-infected patients were included in the study. The result of the study revealed that initiation of ART during the first 8 weeks of TB treatment (early group) was not associated with changes in mortality at 1 year compared with initiation of ART after 8 weeks (late group) (aHR = 0.74; CI: 0.33–1.64). In patients with baseline CD4 counts ≤ 50 cells/μl, there was a significant reduction in mortality in the early group compared with the late group (aHR = 0.20; 95% CI: 0.042–0.99). In patients with a CD4 count > 50 cells/μl, there was no significant difference between early and late groups (aHR = 1.79; 95% CI: 0.64–5.03). Moreover, the study suggested that in HIV/TB co-infected patients in rural Kenya, early ART initiation (within 8 weeks) was associated with reduced mortality in those with CD4 counts ≤ 50 cells/μl. In patients with CD4 counts > 50 cells/μl, there was no association seen between timing of ART and mortality.

In another study conducted in Thailand, the impact of ART on survival of HIV-infected TB patients was measured using propensity score analysis that adjusted for factors associated with receiving ART. The aim of the study was to document the impact of ART on HIV-infected TB patients in public health programs in resource-limited settings. The study consisted of 626 TB patients who started ART and 643 who did not start ART. The result showed that out of 626 HIV-infected TB patients who started ART during TB treatment, 68 (11%) patients died and out of 643 HIV-infected TB patients who did not receive ART during TB treatment, 295 (46%) patients died. The univariate analysis showed that the estimated relative risk of dying for patients those who started ART during TB treatment as compared to those who did not receive ART during TB treatment was 0.24 (95% CI: $0.19 - 0.30$) and in patients with very low CD4 count (< 10 cells/mm^3), 12 patients out of 56 (21%) who received ART died compared with 35 out of 43 (81%) who did not receive ART ($aHR = 0.26$; 95% CI: $0.16 - 0.44$). After controlling for propensity to receive ART, the hazard ratio of death among patients treated with ART was 0.17 (95% CI: $0.12 - 0.24$) (Natpatou *et al.*, 2008).

Another study conducted in Thailand showed that the survival rates at 1, 2, and 3 years after TB diagnosis were 96%, 94%, and 88% respectively for ART+ group and 44%, 19%, and 9% respectively for ART- group (log-rank test, $P < 0.001$). Among patients in ART+ group, the patients who delayed ART ≥ 6 months after TB diagnosis had a higher mortality rate than those who initiated ART < 6 months after TB diagnosis ($aHR = 2.651$; 95% CI: 1.152 – 6.102) (Manosuthi et al., 2006).

2.2 Predictors of Mortality/Survival of HIV/TB co-infected Patients

To assess the incidence of mortality, clinical characteristics and outcome of co-infection with HIV and TB, Xueyan et al (2008**)** conducted a cross-sectional study of co-infected cases reported from nine domestic hospitals in mainland China. Treatments for TB and HIV were provided to 241 patients and mortality attributable to co-infection was reported for 15.8% of the cases. Immune function among most patients was suppressed based on reduced CD4 cell counts. HIV/TB co-infection was related to high mortality even when HAART and/or drug therapy for TB was provided. It was found that fever, fatigue, and weight loss were experienced by the majority of HIV/TB patients (87.6%, 61.0% and 60.2%, respectively), making these the leading clinical symptoms among co-infected patients.

A study conducted in Baltimore USA also revealed that, accounting for potential confounders, including CD4 cell count and viral load, the hazard of AIDS-related death was 2.4 times more for the person-time with TB compared to the person-time without TB **(**Lo'pez-Gatell et al., 2008**)**.

Andreychyn and Zhyvytsia (2013) conducted a prospective cohort study among HIV patients who developed TB from January 2005 to December 2006 in a Zaporizhzhya AIDS Center, Ukraine. The aim of the study was to introduce the effect of HAART on survival of HIV/TB co-infected patients and to evaluate the survival and the effect of HAART on mortality on these patients. Survival was determined by Kaplan-Meier method and the effect of HAART on survival was evaluated using Cox proportional hazards model. The study identified 80 patients who met the inclusion criteria. In multivariate analysis, patients with a CD4 cell count < 100 cell/µl had a 5-fold higher risk of mortality (aHR =5.2; 95% CI: 1.4–19.4) and those patients who had extra-

pulmonary tuberculosis had a 2-fold higher risk of mortality (aHR = 2.2; 95% CI: 1.1–8.3). The study also revealed that HAART significantly increased the probability of survival and reduced the risk of death for HIV/TB-infected patients in Ukraine.

Kassa et al (2012) conducted a study about the incidence of TB and early mortality in a large cohort of HIV infected patients receiving antiretroviral therapy at Zewditu Memorial Hospital in Addis Ababa, Ethiopia. The aim of the study was to measure the rate of TB, TB mortality, and associated risk factors following commencement of HAART in a cohort of patients attending HIV care. Survival analysis was done based on the Kaplan-Meier estimator and Cox proportional hazards model. The study revealed that at baseline, risk factors for mortality of patients associated with TB were WHO clinical stage III HIV infection (aHR = 2.53; 95% CI: 1.70–3.70), WHO clinical stage IV HIV infection (aHR = 3.86; 95% CI: 2.54–5.86), and patients who were being bedridden (aHR = 1.52; 95% CI: 1.13–2.05). The rate of mortality was 6.9% (incidence 2.8 per 100 person-years) and 57% of deaths occurred in the first six months of HAART initiation. WHO clinical stage IV HIV infection, CD4+ cell count < 50 cells/μl, being bedridden and TB after HAART initiation as baseline were independent predictors of mortality.

Sileshi et al (2013) conducted an institution-based retrospective cohort study among 422 TB-HIV co-infected patients being treated for tuberculosis between August, 2011 to January, 2012 in governmental health institutions in Bahir Dar, Northwest Ethiopia. The aim of this study was to assess predictors of mortality among TB-HIV co-infected patients being treated for TB. Kaplan-Meier test was used to estimate the probability of death after TB diagnosis. Cox proportional hazard model was used to determine factors associated with death after TB diagnosis. The study revealed that

independent predictors of mortality during TB treatment are: receiving ART (aHR = 0.35; 95% CI: 0.19–0.64); not having initiated cotrimoxazole prophylactic therapy (CPT) (aHR = 3.03; 95% CI: 1.58–5.79); being ambulatory (aHR = 2.10; 95% CI: 1.22–3.62); CD4 count between 0–75 cells/micro liter, 75–150 cells/micro liter, or 150–250 cells/micro liter (aHR = 4.83; 95% CI: 1.98–11.77, 3.57; 95% CI: 1.48–8.61, and 3.07; 95% CI: 1.33–7.07, respectively); and treatment in a hospital (aHR = 2.64; 95% CI: 1.51–4.62).

Lorent et al (2011) conducted a prospective observational cohort study of adults treated for TB at the internal medicine department of the Kigali University Hospital, Rwanda from May 2008 through August 2009. The aim of the study was to determine incidence, causes of, and risk factors for serious adverse events among patients on first-line anti-TB treatment, as well as the impact on anti-TB treatment outcome. Of 263 patients enrolled, 253 were retained for analysis. Overall, 40% had pulmonary TB, 43% extra-pulmonary TB and 17% a mixed form. The result of the study obtained from Cox model showed that for TB/HIV co-infected patients, extra-pulmonary TB (aHR = 2; 95% CI: 1.1–3.9) and CD4 count < 100 cells/mm^3 at TB diagnosis (aHR = 1.7; 95% CI: 1.0–2.9) were independent predictors of mortality.

Elliott et al (1995) examined the impact of HIV on mortality of patients treated for TB in a prospective study in Lusaka, Zambia. Patients with sputum smear-positive, miliary or meningeal tuberculosis were prescribed 2 months' daily treatment with streptomycin, thiacetazone, isoniazid, rifampicin, and pyrazinamide followed by 6 months with thiacetazone and isoniazid. Other patients were treated for 2 months with streptomycin, thiacetazone and isoniazid followed by 10 months with thiacetazone and isoniazid. 239 patients (65 HIV-negative and 174 HIV-positive) were followed to

2 years from start of treatment. The result of the study obtained from Cox model showed that the risk factors for death in HIV-TB patients were history of prolonged diarrhoea or fever, oral thrush, splenomegaly, allergy to tuberculin, low weight, anaemia or lymphopenia, and poor compliance with regimens containing rifampicin and pyrazinamide. Tuberculosis, even treated, was a major cause of death in patients with HIV infection.

Naidoo et al (2013) conducted a cross-sectional survey about predictors of TB and antiretroviral (ARV) medication non-adherence on 3107 patients older than 18 years in three study districts (14 primary health care facilities in each), in South Africa. The aim of the study was to identify factors associated with adherence to anti-TB and HIV treatment drugs. Associations between TB medication and ART non-adherence were identified using logistic regression analysis. Of the total, 1729 (56%) were HIV-positive patients. Among these 757 (44%) patients were on ART. The result of the study obtained from multivariate logistic model showed that the significant predictive factors for non-adherence to dual therapy (i.e. anti-TB treatment and ART) among HIV-TB co-infected patients were medium poverty (OR = 2.6; 95% CI: 1.46–4.65), high levels of poverty (OR = 3.89; 95% CI: 1.87–8.12), having one chronic condition (OR = 2.73; 95% CI: 1.31–5.65), having three or more chronic conditions (OR = 5.33; 95% CI: 2.27–12.55), high risk for alcohol misuse (OR = 13.1; 95% CI: 2.96–57.99) and having a partner who is HIV positive (OR = 3.12; 95% CI: 1.84–5.29). An additional predictor for non-adherence to anti-TB drugs was tobacco use.

Ismail and Bulgiba (2013) conducted a study by reviewing medical records at the time of TB diagnosis and subsequent follow-up of all newly registered TB patients with HIV co-infection at TB clinics in the Institute of Respiratory Medicine and three public hospitals in the Klang Valley, Malaysia between January 2010 and September 2010. Kaplan-Meier and multivariate Cox proportional hazards analysis were used to describe and identify predictors of death during TB treatment among TB/HIV co-infected patients. Of the 227 patients included in the study, 53 (23.3%) died at the end of the study. The study revealed that after adjusting for other factors, death in TB/HIV co-infected patients was associated with being Malay (aHR = 4.48; 95% CI: 1.73–11.64), CD4 T-lymphocytes count < 200 cells/µl (aHR = 3.89; 95% CI: 1.20–12.63), three or more opportunistic infections (aHR = 3.61; 95% CI: 1.04–12.55), not receiving antiretroviral therapy (aHR = 3.21; 95% CI: 1.76–5.85) and increase per 10^3 total white blood cell count per micro liter (aHR = 1.12; 95% CI: 1.05–1.20).

Hermans et al (2013) conducted a study about the risk of tuberculosis after antiretroviral treatment initiation from 2005 to 2009 in a large, urban HIV clinic in Uganda on 5,797 patients. The objective of the study was to evaluate the association of efavirenz (EFV) compared to nevirapine (NVP) with post-ART TB among patients initiated on first-line ART. Of the total, 3,484 (60%) were initiated on NVP and 2,313 (40%) were initiated on EFV. The result of the study obtained from multivariate Cox model showed that use of EFV compared to NVP was independently associated with higher TB incidence in patients with a baseline CD4 (+) T-cell count < 100 cells/mm^3 (aHR = 2.05; 95% CI: 1.29–3.27), but not at higher CD4 (+) T-cell counts (aHR = 0.71; 95% CI: 0.39–1.31).

Braitstein et al (2009) conducted a retrospective observational study to describe the observed incidence of and risk factors for TB diagnosis among HIV-infected children enrolled in a large network of HIV clinics in western Kenya. The study included 6,535 HIV-infected children aged between 0 and 13 years at AMPATH clinic. Survival data were analyzed using the Kaplan-Meier estimator and Cox model. Of the total, 50% were female and 234 (3.6%) were diagnosed with TB at enrollment. The study revealed that being severely immune-suppressed at enrollment (aHR = 4.4; 95% CI: 3.6-5.4), having ever attended school (aHR = 2.65; 95% CI: 2.2–3.3), being an orphan (aHR = 1.6; 95% CI: 1.3–1.9), being severely low weight-for-height at enrollment (aHR = 1.5; 95% CI: 1.3–1.6), and attending an urban clinic (aHR = 1.39; 95% CI: 1.16–1.67) were found to be risk factors for the incident of TB diagnosis. Children receiving combinations of antiretroviral treatment were less likely to be diagnosed with incident of TB (aHR = 0.15; 95% CI: 0.12–0.20).

Varma et al (2009) conducted a multi-center observational study of HIV-infected patients newly diagnosed with TB from 32 public TB treatment facilities from May 2005 through September 2006 in Thailand. The objective of the study was to identify treatment factors associated with improved survival during TB treatment. Survival was estimated using Kaplan-Meier method and multivariable proportional hazards analysis was conducted to identify factors associated with death. Of 667 HIV-infected TB patients enrolled, 450 (68%) were smear and/or culture positive. Death during TB treatment occurred in 112 (17%) cases. The result of the study revealed that in proportional hazards analysis, factors strongly associated with reduced risk of death were: ART use (HR = 0.16; 95% CI: 0.07–0.36), fluconazole use (aHR = 0.34; 95% CI: 0.18–0.64), and co-trimoxazole use (aHR = 0.41; 95% CI: 0.20–0.83). Among 126 patients that initiated ART after TB diagnosis, the risk of death increased the

longer that ART was delayed during TB treatment. Efavirenz- and nevirapine-containing ART regimens were associated with similar rates of adverse events and death.

Getahun et al (2011) conducted a retrospective cohort study about mortality and associated risk factors in a cohort of TB patients treated under DOTS programme from June 2004 to July 2009 in three randomly selected health centers in Addis Ababa, Ethiopia. The objective of the study was to determine the magnitude and identify risk factors associated with time to death among TB patients treated under DOTS programme. A step-wise multivariable Cox's regression model and Kaplan-Meier curves were used to model the outcome of interest. From a total of 6,450 registered TB patients 236 (3.7%) died at the end of the study. The death rate of pulmonary-positive, pulmonary-negative and extra-pulmonary TB patients were 2.7%, 3.6%, and 4.3%, respectively. The result of the study revealed that body weight at initiation of anti-TB treatment (< 35 kg), patient category, year of enrollment and treatment center were independent predictors for time to death.

Maruza et al (2012) conducted a retrospective cohort study between June 2007 and December 2009 with HIV-infected patients who started anti-tuberculosis treatment in the state of Pernambuco, Brazil. The aim of the study was to estimate the probability of survival and to identify risk factors for death in cohort of PLWHIV who started TB treatment. Survival data were analyzed using the Kaplan-Meier estimator, the log-rank test and Cox model. Out of a cohort of 2310 HIV-positive individuals, the data on 333 patients who commenced treatment for TB were analyzed. Cox model showed that females, age above 30 years, having anemia, not using HAART during treatment

for TB and a CD4+ lymphocyte count below 200 cells/mm^3 and treatment for TB having started in an out-patient clinic were associated with high risks of death.

Umphonsathien and Sungkanuparph (2011) conducted a retrospective cohort study among 184 patients with HIV/TB co-infection who started ART between January 2000 and December 2009 at Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. The aim of the study was to determine the association between the timing of ART initiation in HIV/TB co-infection and the occurrence of IRIS. Risk factors for TBIRIS were determined using Kaplan-Meier analysis and Cox proportional hazard model. Overall, 41.8%, 33.2%, and 25% had pulmonary, extra pulmonary, and disseminated TB, respectively. From Cox analysis, the risk of early ART within 4 months of TB treatment (aHR =2.972; 95% CI: 1.04–8.49) and disseminated TB (aHR = 4.489; 95% CI: 1.012–20.11) were significant risk factors for TB-IRIS.

Worodria et al (2011) conducted a study about the incidence and predictors of mortality and the effect of tuberculosis immune reconstitution inflammatory syndrome in a cohort of TB/HIV patients commencing ART on 618 patients from December 17, 2007 to December 31, 2009 at Mulago National Tuberculosis and Leprosy Program clinic in Kampala, Uganda. Predictors of mortality were assessed using Cox proportional hazards analysis. Of the total, only 302 patients fulfilled eligibility criteria. Overall, 53 patients died, 36 (68%) of these died within the first 6 months of TB diagnosis. Male sex (aHR = 2.19; 95% CI: 1.19–4.03), allergy to tuberculin skin test (aHR = 2.59; 95% CI: 1.10–6.12), a positive serum cryptococcal antigen result at enrollment (aHR = 4.27; 95% CI: 1.50–12.13) and no ART use (aHR =4.63; 95% CI: 2. 37–9.03) were independent predictors of mortality by multivariate analysis.

Deribe et al (2013) conducted a case–control study about predictors of mortality among HIV/TB co-infected persons on 919 patients enrolled in ART from June 8, 2003 to August 14, 2009 at Jimma University Specialized Hospital in Southwest Ethiopia. Bivariate analysis was done to determine associations between explanatory variables and the outcome variable (death). Logistic regression models were performed to identify predictors. Of the total, 69 (7.5%) died and among these deaths 43 (62.3%) had pulmonary TB, 14 (20.3%) had extra-pulmonary TB, 3 (4.3%) had disseminated TB and for 9 (13.0%) the type of TB was not recorded. The multivariate result revealed that male sex (OR = 2.04; 95% CI: 1.04–4.02), being bedridden at enrollment (OR = 2.84; 95% CI: 1.17–6.89), and cough of more than 2 weeks during initiation of ART (OR = 4.75; 95% CI: 2.14–10.56) were found to be the predictors of mortality among HIV/TB co-infected patients.

A study in North Carolina utilized surveillance data of HIV/TB patients on 543 cases that were reported between January 1, 1993 to December 31, 2003 from TB Control Program. TB cases that did not die prior to December 31, 2004 were right-censored as of that date. The study employed Kaplan-Meier method to estimate long-term survival, log-rank test to compare categories within group in survival and Cox proportional hazards method to compute hazard ratios. The result of the study showed that, in multivariable survival analysis, age greater than 45 years (aHR = 1.44, 95% CI: 0.96–2.15), baseline CD4 cell count (aHR = 0.76 per 100 cell increase, 95% CI: 0.65–0.88), and having HAART started during TB treatment (aHR = 0.43, 95% CI: 0.26–0.72) were associated with long-term survival after TB diagnosis. (Gadkowski et al., 2009).

CHAPTER THREE

DATA AND METHODOLOGY

3.1. Data

The data for this study were taken from the secondary source of ART clinic based on persons who were older than 15 years those who started anti-TB treatment between September 1, 2010 and April 30, 2013 and were followed until July 31, 2013 from Chagni Health Center, northwest Ethiopia. The Health Center served the population of Guanga Woreda and other adjacent Health Center's of Benshagul Gumze Regional States in North-West Ethiopia.

The Health Center started giving free ART service in 2008. It has a separate ART clinic and units for ART follow up for PLWHIV where the units in the clinic have nurses, a pharmacy, data clerks, patient adherence rooms, equipments and etc. Data for this study were extracted from the available standard national medical registers, which have been adopted by the Ministry of Health (MOH). The registers include the pre-ART register (register of patients at their first visit), the ART register (registration after ART initiation), and the follow-up patient form. The Health Center also has a separate clinic and units for TB treatment and follow-ups. In this case data regarding TB treatment and follow-ups were recorded on standard register given by MOH. For this study, both ART and TB registers have been looked through in order to record all persons who have been infected with HIV and started ART and have been under an anti-TB treatment.

This study reviewed patient’s intake forms and follow-up cards of HIV-TB patients by identifying documented clinical and laboratory variables and followed for the outcomes of either the event (death) or censored (dropped out, lost to follow up,

transferred out to other centers or on follow up at the end of study time) occurred. The total number of cases included in this study was 510 patients.

3.2. Variables included in the study

The response (dependent) variable: The response/outcome variable was survival time measured (in months) of HIV-TB co-infected patients, that is the length of time from anti-TB start date until the date of death/censored occurred.

Predictor (independent) variables:

The predictor (independent) variables included in the model were:

Covariates	Categories
Sex	$0 = female, \quad 1 = male$
Marital status	$1 = never\ married, \quad 2 = married, \quad 3 = others$
Level of education	$0 = no - education, \quad 1 = primary\ and\ above$
Functional status	$1 = working, \quad 2 = ambulatory, \quad 3 = bedridden$
TB type	$0 = pulmonary\ TB, \quad 1 = extra - pulmonary\ TB$
Substance use	$0 = no, \quad 1 = yes$
WHO clinical stage	$0 = stage\ I, 1 = stage\ II, 2 = stage\ III, 3 = stage\ IV$
Baseline CD4 count (cells/mm^3)	$0 = CD4 \leq 200, 1 = 200 < CD4 \leq 350, \ 2 = CD4 > 350$
Baseline weight	$0 = 30 - 40, 1 = 41 - 50, 2 = 51 - 60, \ 3 = > 60$
Baseline body mass index (BMI)	$0 = BMI \leq 18.5, 1 = 18.5 < BMI < 25, 2 = BMI \geq 25$
Baseline drug regimen type	$1 = AZT - based, 2 = D4T - based, 3 = TDF - based$
Baseline age in completed years	$0 = 15 - 30, 1 = 31 - 45, 2 = 46 - 60, 3 = > 60$

Note: All predictor variables were taken at baseline values. $BMI = \frac{weight}{height^2}$

We would like to provide the following brief information about the variables we categorized in this study

• **Marital Status:**- this categorization is based on the usual, never married, married, divorced and widowed. However, the data show that the number of patients whose

marital status divorced and widowed is too small to be categorized separately. Therefore, such patients are included in the “others” category.

• We categorized CD4 count and weight based on Federal Ministry Of Health guidelines for clinical and programmatic management of TB/HIV collaborative in Ethiopia. But BMI is categorized by using standard or metric measures and age is categorized based on previous researches and different journals.

Ethical consideration

An institutional ethical approval was obtained from the College of Natural and Computational Sciences, Addis Ababa University Ethics Review Committee. Following the approval, official letter of cooperation was written to the concerned bodies: Chagni Health Center by the College of Natural and Computational Science, Addis Ababa University. As the study was conducted through review of medical records, the individual patients were not subjected to any harm as long as the confidentiality is kept. In addition to that, no name or personal identifications were used on data collection form. The recorded data were not accessed by a third person, except the principal investigators, and was kept confidentially. Permission to conduct the study and to access the data was obtained from the Chagni Health Center authorities.

3.3. Methodology: Survival Analysis

Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event (Kleinbaum and Klein, 1996). Hence, survival analysis is also referred to as "time-to-event analysis". It is used in a number of applied fields, such as medicine, public health, the social sciences and engineering.

Survival analysis (lifetime analysis) was developed by biostatisticians modeling human life times in the medical and biological sciences. The methodological developments in survival analysis were largely achieved in the latter half of the 20^{th} century. One of the oldest and most straightforward non-parametric methods for analyzing survival data was given in Berkson and Gage (1950). One important development in non-parametric analysis methods was (Kaplan and Meir, 1958).

The response variable in survival analysis is survival time and is no longer limited to only time to death. It is a non-negative random variable used loosely for the time period from a starting time point to the occurrence of a any event. In the context of this study, survival time is the length of lifetime of HIV-TB co-infected patients from the date of TB treatment to the death of a patient.

3.3.1 The Nature of Survival Data: Censoring

The use of survival analysis, as opposed to the use of other statistical method, is most important when some subjects are lost to follow up or when the period of observation is finite certain patients may not experience the event of interest over the study period. In this latter case one cannot have complete information for such individuals. The incomplete observations are referred to as being censored. Most survival analysis consider a key analytical problem of censoring. In essence, censoring occurs when we

have some information about individual survival time, but we do not know the survival time exactly.

In reality such event can occur due to the following reasons:

1. An Individual's do not experience the event before the study ends.
2. An Individual's are lost to follow-up during the study period and
3. An Individual's withdraw from the study for unknown/known reasons.

There are three categories of censoring.

a) **Right censoring**: Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study.

b) **Left censoring**: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study; this is not common in survival studies.

c) **Interval censoring**: Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known.

3.3.2 Describing The Distribution of Time to An Event

Let the random variable T denote "time-to-event" of our interest. Of course, T is a positive random variable which has to be unambiguously defined. That is, we must be very specific about the start and end with the length of the time period in-between corresponding to T.

In routine data analysis, we may first present some summary statistics such as mean, standard error for the mean, etc. In analyzing survival data, however, because of possible censoring, the summary statistics may not have the desired statistical properties, such as unbiasedness. For example, the sample mean is no longer an

unbiased estimator of the population mean (of survival time). So we need to use other methods to present survival data. One way is to estimate the underlying true distribution. When this distribution is estimated (either parametrically or non-parametrically), then we can estimate other quantities of interest such as mean, median, etc. of the survival time.

The distribution of the random variable T can be described in a number of equivalent ways. There is of course the usual (cumulative) distribution function (Klein and Moeschberger, 1997)

$$F(t) = P(T \leq t), \quad t \geq 0.$$

which is right continuous, *i.e.*, $\lim_{u \to t^+} F(u) = F(t)$.

When T is a survival time, $F(t)$ is the probability that a randomly selected subject from the population will die before time t. If T is a continuous random variable, then it has a density function $f(t)$, which is related to $F(t)$ through the following equations

$$f(t) = \frac{dF(t)}{dt} \quad \text{so} \quad , \quad F(t) = \int_0^t f(u)du$$

3.3.3. Descriptive Methods For Survival Data

3.3.3.1 The Survival Function $S(t)$

Let T be a continuous random variable associated with the survival times, t be the realization of the random variable T and $f(t)$ be the underlying probability density function of the survival time t. The cumulative distribution function $F(t)$ represents the probability that a subject selected at random will have a survival time less than or equal to some stated value t,

$$F(t) = P(T \leq t) = \int_0^t f(u)\, du \qquad (3.1)$$

The survivor function, $S(t)$, is defined to be the probability that the survival time of a randomly selected subject is greater than some specified time t and so (Klein and Moeschberger, 1997)

$$S(t) = P(T > t) = 1 - F(t) = \int_t^{\infty} f(u)du\,, \quad t \geq 0. \tag{3.2}$$

The survivor function can, therefore, be used to represent the probability that, a randomly selected subject survives from the time origin to some specified time beyond t .

The survival function $S(t)$ is a non-increasing function over time taking on the value 1 at t = 0, *i.e.*, $S(0) = 1$. For a proper random variable T, $S(\infty) = 0$, which means that everyone will eventually experience the event. However, we will also allow the possibility that $S(\infty) > 0$. This corresponds to a situation where there is a positive probability of not "dying" or not experiencing the event. For example, if the event of interest is the time from response until disease relapse and the disease has a cure for some proportion of individuals in the population, then we have $S(\infty) > 0$, where $S(\infty)$ corresponds to the proportion of cured individuals (Tsiatis and Zhang, 2005).

From the equations 3.1 and 3.2 the relationship between $f(t)$ and $S(t)$ can be derived as

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}\big(1 - S(t)\big) = -\frac{d}{dt}S(t)\,, \quad t \geq 0. \tag{3.3}$$

3.3.3.2. The hazard function $h(t)$

The hazard rate is a useful way of describing the distribution of "time-to-event" data because it has a natural interpretation that relates to the aging of a population. We motivate the definition of hazard rate by first defining mortality rate which is a discrete version of the hazard rate.

The mortality rate at time t, where t is generally taken to be an integer in terms of some unit of time (e.g., years, months, days, weeks, etc), is the proportion of the population who fail (die) between times t and t + 1 among individuals alive at time t, i.e. (Tsiatis and Zhang, 2005)

$$\mathrm{m}(t) = P(t \leq T < t + 1 / T \geq t) \qquad (3.4)$$

The hazard function is widely used to express the risk of hazard of death at time t, and is obtained from the probability that an individual dies at time t, given that the individual has survived up to time t. It gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t. The hazard rate is not a probability, it is a probability rate. Therefore it is possible that a hazard rate can exceed one in the same fashion as a density function $f(t)$ may exceed one (Klein and Moeschberger, 1997).

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

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

$$h(t) = \lim_{h \to 0} \frac{P(t \leq T \leq t+h / T \geq t)}{h} \qquad (3.5)$$

By applying the theory of conditional probability and the relationship in equation (3.3), the hazard function can be expressed in terms of the underlying probability density function and the survivor function as follows

$h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = \frac{-d}{dt} \ln\{S(t)\}$. Thus, $h(t) = \frac{-d}{dt} \ln\{S(t)\}$.

From this, we can integrate both sides to get $\mathrm{H}(t) = \int_0^t h(u)\, du = -\ln\{S(t)\}$

where $\mathrm{H}(t)$ is referred to as the cumulative hazard function. Here we used the fact that $S(0) = 1$. Hence, $S(t) = e^{-\mathrm{H}(t)} = e^{-\int_0^t h(u)du}$ (Klein and Moeschberger, 1997).

3.4. Non-parametric methods

In survival analysis, it is always a good idea to present numerical or graphical summaries of the survival times for the individuals. In general, survival data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time (Collett, 2003). The Kaplan-Meier, Nelson-Aalen and Life Tables are the most widely used to estimate the survival and hazard functions. These methods are said to be non parametric or distribution-free since they do not require specific assumption to be made about the underlying distribution of the survival times (Klein and Moeschberger, 1997).

3.4.1 Life-table Estimator

The life table in Berkson and Gage (1950) is the earliest statistical method to study human mortality rigorously, but its importance has been reduced by the modern methods, like the Kaplan-Meier method in Kaplan and Meir (1958).

If censoring is occurred during the interval, then the life table estimate would be computed as (Tsiatis and Zhang, 2005)

$$\hat{S}^{LT}(t) = \prod_{x \le t}(1 - \hat{m}(x)) \quad \text{, and} \quad \hat{m}(x) = \frac{d(x)}{n(x) - \frac{w(x)}{2}}. \tag{3.6}$$

where:

- ✓ $d(x)$ denote the number of deaths at time *x*,
- ✓ $w(x)$ denote the number of censored individuals at time *x*,
- ✓ $\hat{m}(x)$ is the mortality rate at time x and
- ✓ $n(x)$ denote the number of individuals at risk just prior to time *x*, *i.e.*, a number of individuals in the sample who neither died nor were censored prior to time *x*.

The variance estimate of $\hat{S}^{LT}(\mathrm{t})$ is given by

$$\widehat{var}\{\hat{S}^{LT}(t)\} = \{\hat{S}^{LT}(t)\}^2 \sum_{x=1}^{t} \frac{d(x)}{\left\{n(x)-\frac{w(x)}{2}\right\}\left\{n(x)-\frac{w(x)}{2}-d(x)\right\}} . \quad (3.7)$$

3.4.2 Kaplan-Meier (KM) Estimator

The KM (product-limit) estimator is the life-table estimator with the time intervals taken so small that only at most one observation occurs in an interval (Kaplan and Meir, 1958). By convention, the KM estimator is a right-continuous step function that takes jumps at death times. 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. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included.

The Kaplan-Meier estimate can be expressed as (Tsiatis and Zhang, 2005)

$$\hat{S}^{KM}(t) = \prod_{x \le t}(1 - \hat{m}(x)) = \prod_{x \le t}(1 - \frac{d(x)}{n(x)}) . \quad (3.8)$$

where: $d(x), n(x)$ and $\hat{m}(x) = \frac{d(x)}{n(x)}$ are defined as in equation 3.6.

Generally $d(x)$ is either zero or one, but we allow the possibility of tied survival times in which case $d(x)$ may be greater than one.

3.3.3 Nelson-Aalen Estimator

The estimator $\sum_{x \le t} \frac{d(x)}{n(x)}$ is known as the Nelson-Aalen estimator of the cumulative hazard function $H(t) = \int_0^t h(x)dx$.

That is $\hat{H}^{NA}(t) = \int_0^t \frac{d(x)}{n(x)} = \sum_{x \le t} \frac{d(x)}{n(x)}$ and $\hat{S}^{NA}(t) = \exp(-\hat{H}^{NA}(t))$. (3.9)

where: $d(x)$ and $n(x)$ are defined as in equation 3.6.

Although the Nelson-Aalen estimate of the survivor function has been shown to perform better than the KM estimate in small samples, in many circumstances, the

two estimates will be very similar, particularly at the earlier survival times. Moreover, the KM estimate is regarded as an approximation to the Nelson-Aalen estimate.

The estimated variance of the Nelson-Aalen estimator is given by

$$\widehat{var}\{\hat{H}^{NA}(t)\} = \sum_{x \le t} \frac{\frac{d(x)}{n((x)} \left\{\frac{n(x)-d(x)}{n(x)}\right\}}{n(x)-1} \qquad (3.10)$$

and in the case of no ties $\widehat{var}\{\hat{H}^{NA}(t)\} = \int_0^t \frac{d(x)}{\{n(x)\}^2} = \sum_{x \le t} \frac{d(x)}{\{n(x)\}^2}$.

3.5. Non-parametric comparison of survival distributions

After obtaining statistics which provide a description of the overall survival experience, the survival and hazard functions, we usually turn our attention to a comparison of the survivorship experience in key subgroups in the data. These groups might be defined by the values of a covariate which are thought to be related to survival times.

The simplest way of comparing the survival times obtained from two groups of individuals is to plot the corresponding estimates of the two survivor functions on the same axes. In general, the pattern of one survivorship function lying above another means, the group defined by the upper curve lived longer or had a more favorable survival experience than the group defined by the lower curve. However, there are two possible explanations for the observed differences between two estimated survivor functions. One explanation is that there is a real difference between the survival times of the two groups of individuals, so that those in one group have a different survival experience from those in the other. An alternative explanation is that there is no real difference between the survival times of the two groups of individuals, and that the difference that has been observed is merely the result of chance variation. To help distinguish between the two explanations, with any degree of confidence, statistical test procedures should be used. These tests, which have the

same generalized form or algebraic presentation, include the Log Rank test, Generalized Wilcoxon test, Tarone and Ware test, Peto-Peto-Prentice test and Harrington-Fleming test.

The general form of these test statistics for the comparison of survival functions between two groups can be defined as follows (Tsiatis and Zhang, 2005):

$$T = \frac{\sum_x \; w(x) \left[dN_1(x) - \frac{dN(x)\; Y_1(x)}{Y(x)}\right]}{\left\{\sum_x w^2(x) \left[\frac{Y_1(x)\; Y_0(x)\; dN(x)\; [Y(x)-dN(x)]}{Y^2(x)\; [Y(x)-1]}\right]\right\}^{1/2}} \qquad (3.11)$$

where:

- ✓ $dN_0(x)$ and $dN_1(x)$ are the number of deaths observed at time x from treatments 0 (control group) and treatment1 (new treatment), respectively.
- ✓ $Y_0(x)$ and $Y_1(x)$ are the number of individuals at risk at time x from treatments 0 and 1, respectively.
- ✓ $Y(x) = Y_0(x) + Y_1(x)$ are the total numbers at risk at time x.
- ✓ $dN(x) = dN_0(x) + dN_1(x)$ are the total number of deaths at time x and
- ✓ $w(x)$ is the weight function.

The weights to be chosen depends on the type of alternative difference we wish to detect. If the weights $w(x)$ are stochastic (functions of data), then they need to be a function of the censoring and survival information prior to time *x*. The weight function *w*(*x*) can be used to emphasize differences in the hazard rates over time according to their relative values. For example, if the weight early in time is larger and later becomes smaller, then Log-rank test statistic would emphasize early differences in the survival curves (Tsiatis and Zhang, 2005).

3.5.1 The Log-rank test

The log-rank test, developed by Mantel and Haenszel, is a non-parametric test for comparing two or more independent survival curves. Since it is a non-parametric test, no assumption about the distributional form of the data is required.

The log rank test is the most commonly used test where $w(x) = 1$ for all x. The log-rank test statistic for testing $H_0: S_0(t) = S_1(t)$ is (Tsiatis and Zhang, 2005)

$$T = \frac{\sum_x \left[dN_1(x) - \frac{dN(x)\,Y_1(x)}{Y(x)}\right]}{\left\{\sum_x \left[\frac{Y_1(x)\;Y_0(x)\;dN(x)\;[Y(x)-dN(x)]}{Y^2(x)\;[Y(x)-1]}\right]\right\}^{1/2}} \qquad (3.12)$$

Under the null hypothesis that is 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, then $T \overset{a}{\sim} N(0,1)$ and T^2 follows a chi-square distribution with one degree of freedom.

3.6. Modeling of survival data for Cox regression model

Often one is interested in comparing two or more groups of survival times. If the groups are similar (homogeneous) except for the treatment under study then the nonparametric methods discussed above may be used directly. However, a problem frequently encountered in analyzing survival data is that of adjusting the survival function to account for concomitant information (sometimes referred to as covariates, explanatory variables or independent variables). Populations which exhibit heterogeneity are prevalent whether the study involves a clinical trial, a cohort study, or an observational study. After adjustment for these potential explanatory variables, the comparison of survival times between groups will be less biased and more precise than a simple comparison (Tsiatis and Zhang, 2005).

In the analysis of survival data, interest centers on the risk of hazard of failure at any time after the time origin of the study. As a consequence the hazard function is modeled directly in survival analysis. There are two broad reasons to model survival data. One objective of the modeling process is to determine which combinations of potential explanatory variables affect the form of the hazard function. In particular, the effect that the treatment has on the hazard of failure can be studied, as can the extent to which other explanatory variables affect the hazard function. Another reason for modeling the hazard function is to obtain an estimate of the hazard function itself for an individual from a set of explanatory variables. This may be of interest in its own right, but in addition, from the relationship between the survivor function and hazard function an estimate of the survivor function can be found (Collett, 2003).

3.7. Cox Regression Model

While non-parametric methods work well for homogeneous samples, they do not determine whether or not certain variables are related to the survival times. This need leads to the application of regression methods for analyzing survival data. The standard multiple linear regression model is not well suited to survival data for several reasons. Firstly, survival times are rarely normally distributed. Secondly, censored data result in missing values for the dependent variable (survival time) (Kleinbaum and Klein, 1996).

The Cox proportional hazards (PH) model is now the most widely used for the analysis of survival data in the presence of covariates or prognostic factors. This is the most popular model for survival analysis because of its simplicity, and not being based on any assumptions about the survival distribution. The model assumes that the

underlying hazard rate is a function of the independent covariates, but no assumptions are made about the nature or shape of the hazard function.

The Cox proportional hazards model proposed by Cox (1972) has become the standard in biomedical studies, particularly for settings in which the estimation of covariate effects is the primary objective.

The Cox Proportional Hazards model is given by (Tsiatis and Zhang, 2005)

$$h(t|x) = h_0(t)\exp(\beta' x) = h_0(t)\exp(\sum_{k=1}^{p}\beta_k x_k) \quad (3.13)$$

where:

- ✓ $h_0(t) = h(t|x = 0)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero and can take any shape as a function of t. The only requirement is that $h_0(t) > 0$.
- ✓ $x = (x_1, x_2, \dots, x_p)'$ is a $px1$ vector of explanatory variables for a particular individual, and
- ✓ $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is a $px1$ vector of regression coefficients.

Because of $h_0(t)$ is the non parametric part of the model and $\beta' x$ is the parametric part of the model the Cox's proportional hazards model is a semi-parametric model. From the interpretation of the model, it is obvious that β characterizes the "effect" of x. So β should be the focus of our inference while, $h_0(t)$ is a nuisance "parameter".

The survival function of the subpopulation with covariate x is:

$$S(t|x) = [S_0(t)]^{\exp(\beta' x)} \quad (3.14)$$

where: $S_0(t) = e^{-\int_0^t h_0(u)du}$ is the survival function of baseline population ($x = 0$).

The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good

estimate for regression coefficients β, hazard ratio, and adjusted hazard curves. The measure of effect is called hazard ratio. The hazard ratio for any two sets of covariates x and x^* is

$$HR = \frac{h(t|x)}{h(t|x^*)} = \frac{h_0(t)\exp(\beta' x)}{h_0(t)\exp(\beta' x^*)} = \exp(\beta'(x - x^*)), \text{ for all } t \geq 0, \quad (3.15)$$

which is a constant over time (so the name of proportional hazards model).

Equivalently, $\log\left(\frac{h(t|x)}{h(t|x^*)}\right) = \beta'(x - x^*)),$ for all $t \geq 0,$

In PH model, if there is a one unit increase in covariate x_k while other covariate values being held fixed, then

$$\log\left(\frac{h(t|x_k+1)}{h(t|x_k)}\right) = \log\{h(t|x_k + 1)\} - \log\{h(t|x_k)\} = \beta_k \quad (3.16)$$

Therefore, β_k is the increase in log hazard (*i.e.*, log hazard-ratio) at any time with a unit increase in the k^{th} covariate x_k. Equivalently,

$$\frac{h(t|x_k+1)}{h(t|x_k)} = e^{\beta_k}, \text{ for all } t \geq 0. \quad (3.17)$$

Thus, e^{β_k} is the hazard ratio associated with one unit increase in covariate x_k. So that $e^{\beta_k} - 1$ can be interpreted as the percentage change (increase or decrease) in hazard with one unit increase in x_k while adjusting for other covariates.

3.7.1 Partial likelihood estimate for Cox proportional hazards model

Suppose the survival data based on *n* independent observations are denoted by the triplet (t_i, x_i, δ_i) , *i=1,2,...,n.*

where: $t_{i,}$ is a censored/failure time for subject *i* and t_i be the ordered-event times for the m ≤n distinct events (deaths).

$\delta_i = \begin{cases} 0, for\ censored\ observations \\ 1, for\ failed\ observations \end{cases}$ *is event indicator.*

$x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})'$ is the value of the covariate vector for subject *i*.

Then 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. $f(t|x_i)$. 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. $S(t|x_i)$ **(**Hosmer and Lemeshow, 1999**)**.

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

$$((f(t|x_i))^{\delta_i} \ (S(t|x_i))^{1-\delta_i}. \qquad (3.18)$$

Consequently, the likelihood function which holds for any censored survival data with generalized hazard function $h(t|x_i)$, which may not assume PH, will be

$$L(\beta) = \prod_{i=1}^{n} (f(t|x_i))^{\delta_i} (S(t|x_i))^{1-\delta_i} = \prod_{i=1}^{n} (h(t|x_i))^{\delta_i} S(t|x_i) \qquad (3.19)$$

However, Cox showed that the relevant likelihood function which considers the baseline hazard rate as a nuisance parameter, he called it a partial likelihood function.

Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters ($h_0(t)$ in the Cox PH model. 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 $t_{(i)}|$ one event at time $t_{(i)}$)

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

Thus, for the proportional hazards model assuming no tied survival times the partial likelihood function is defined over all failure time $t_{(i)}$ that i = 1, 2, ..., m is gives as

$$\mathrm{PL}(\beta) = \prod_{i=1}^{m} \frac{\exp(\beta' x_i)}{\sum_{j \varepsilon R_{t(i)}} \exp(\beta' x_j)} = \prod_{i=1}^{n} \left[\frac{\exp(\beta' x_i)}{\sum_{j \varepsilon R_{t(i)}} \exp(\beta' x_j)} \right]^{\delta_i} \quad (3.20)$$

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 importance of using the partial likelihood is that this function depends only on β, the parameter of interest, and is free of the baseline hazard, $h_0(t)$, which is infinite dimensional nuisance function.

The log partial likelihood function is

$$\ell(\beta) = \sum_{i=1}^{m} \left[\beta' x_i - ln \left(\sum_{j \varepsilon R_{t(i)}} \exp(\beta' x_j) \right) \right] \quad (3.21)$$

We obtain the maximum partial likelihood estimate of the coefficients and their estimated variances by differentiating the right hand side of equation (3.21) with respect to the component of β, setting the derivative equal to zero and solving for the unknown parameters, and use the minus of the second derivative of the log partial likelihood as the information matrix respectively. Thus,

The score function is

$\mathrm{U}(\beta) = \frac{\partial \ell(\beta)}{\partial \beta}$ and the second derivative is $\frac{\partial^2 \ell(\beta)}{\partial \beta^2}$.

Consequently, $\frac{\partial^2 \ell(\beta)}{\partial \beta^2} < 0$ (See the details in Tsiatis and Zhang, 2005).

Therefore, $\ell(\beta)$ has a unique maximizer and can be obtained uniquely by solving the following partial likelihood equation:

$\mathrm{U}(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = 0$. This maximizer $\hat{\beta}$ defines the MPLE of β.

The quantity $-\frac{\partial^2 \ell(\beta)}{\partial \beta^2}$ is defined as the partial likelihood observed information and is denoted by $J(\beta)$. From the above the vector of efficient scores and the observed information matrix can be found, respectively, as follows:

$$U(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \left(\frac{\partial \ell(\beta)}{\partial \beta_1}, \frac{\partial \ell(\beta)}{\partial \beta_2}, \ldots, \frac{\partial \ell(\beta)}{\partial \beta_p}\right)' \text{ and} \tag{3.22}$$

$$J(\beta)_{pxp} = -\frac{\partial^2 \ell(\beta)}{\partial \beta^2} = \begin{bmatrix} -\frac{\partial^2 \ell(\beta)}{\partial \beta_1^2} & -\frac{\partial^2 \ell(\beta)}{\partial \beta_1 \partial \beta_2} & \cdots & -\frac{\partial^2 \ell(\beta)}{\partial \beta_1 \partial \beta_p} \\ -\frac{\partial^2 \ell(\beta)}{\partial \beta_2 \partial \beta_1} & -\frac{\partial^2 \ell(\beta)}{\partial \beta_2^2} & \cdots & -\frac{\partial^2 \ell(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & & \vdots \\ -\frac{\partial^2 \ell(\beta)}{\partial \beta_p \partial \beta_1} & -\frac{\partial^2 \ell(\beta)}{\partial \beta_p \partial \beta_2} & \cdots & -\frac{\partial^2 \ell(\beta)}{\partial \beta_p^2} \end{bmatrix} \tag{3.23}$$

Although the proportional hazards model for survival data assumes that the hazard function is continuous and, under this assumption, tied survival times are not possible, tied survival times can arise as a result of rounding process. Alternate methods have been suggested by Cox (modification of 1972) and approximations due to Breslow (1974) and Efron (1977) when there are ties. All of them reduce to the original partial likelihood suggested by Cox in the absence of ties. Despite its very complicated form, the appropriate likelihood function when there are ties is given by Kalbfleisch and Prentice (1980). (See the details in Klein and Moeschberger, 1997 and Collett, 2003).

3.8. Assessment of regression coefficients for Cox model

Naturally the next steps following the fit of a regression model are the assessment of the regression coefficients and the formation of confidence intervals for the parameters and related quantities. Under the assumption of proportional hazards, there are three different tests for model assessment (the significance of the coefficients): the partial likelihood ratio test, the Wald test and the score test. These tests are presented below as discussed in Klein and Moeschberger (1997) and Hosmer and Lemeshow (1999).

The Partial Likelihood Ratio Test (LR) is not only the easiest test to compute, but is also the best of the three tests for testing the significance of a subset of q explanatory variables from p explanatory variables. It is defined as

$$Q_{LR} = -2\,\ell(\hat{\beta}_{p-q}) - (-2\,\ell(\hat{\beta}_p))$$

Under the null hypothesis $H_0: \beta_q = (0,0,\ldots,0)'$ and for large sample the statistic $Q_{LR} \sim \chi^2(q)$, at a significance level α.

where: $\ell(\hat{\beta}_{p-q})$ and $\ell(\hat{\beta}_p)$, respectively, are the log partial likelihoods of the models which contains the p-q explanatory variables only (the restricted model) and all explanatory variables (the full model).

The Wald Test: To test $H_0: \beta_q = (0,0,\ldots,0)'$, we use the multivariable Wald statistic

$$Q_w = \hat{\beta}_q{}'\left[J_q(\hat{\beta})\right]^{-1}\hat{\beta}_q$$

where $\hat{\beta}_q$ *and* $J_q(\hat{\beta})$ are the corresponding estimates of β_q and sub matrix of the inverse of the observed information matrix from the full model. Under H_0 and for large samples the statistic $Q_w \sim \chi^2(q)$ at α level of significance.

The Wald test can also be used to test the significance of individual variables. The test statistic then becomes

$$Z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Under the null hypothesis : $H_0: \beta_j = 0$ the statistic $Z \sim N(0,1)$ at a significance level α, Consequently the 100(1-α)% Wald statistic based confidence interval for β_j is

$$\hat{\beta}_j \pm Z_{\alpha/2}se(\hat{\beta}_j)$$

where: $Z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution.

The Score Test: The score test statistic, to test $H_0: \beta_q = (0,0,\ldots,0)'$ is defined as:

$$Q_s = (\mathrm{U}(\beta_q, \hat{\beta}_{p-q}))'\left[J_q(\beta_q, \hat{\beta}_{p-q})\right]^{-1}(\mathrm{U}(\beta_q, \hat{\beta}_{p-q})$$

where $(\mathrm{U}(\beta_q, \hat{\beta}_{p-q})$ and $\left[J_q(\beta_q, \hat{\beta}_{p-q})\right]^{-1}$ are the score vectors and inverse of the observed information matrix evaluated at the hypothesized value of β_q and the

restricted partial maximum likelihood estimator of β_{p-q}. Under the above null hypothesis and for large sample $Q_s \sim \chi^2(q)$, at α level of significance.

3.9. 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).

Variable Selection Procedures

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).

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.

3.10. Model diagnostics for COX PH model

Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid, the fitted model must provide an adequate summary of the data upon which it is based. A complete and thorough examination of a model's fit and adherence to model assumptions is just as important as careful model development. Because methods used in assessing the adequacy of survival models have to cope with the occurrence of censored survival times, they are a little more complicated than the corresponding methods used in linear regression modeling.

Many model checking procedures are based on quantities known as residuals. Residuals are values that can be calculated for each observation and have the feature that their behavior is known, at least approximately, when the fitted model is satisfactory. The following residuals have been proposed for use by different authors in connection with the Cox regression model (Collett, 2003).

Martingale residuals are modified Cox-Snell residuals and, defined as $rm_i = c_i - r_i$, where c_i is censoring indicator and r_i is the Cox-Snell residual for the i^{th} individual given as $r_{i=\hat{H}_i(t)=-\ln \hat{S}_i(t)}$, where $\hat{H}_i(t)$ *and* $\hat{S}_i(t)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} individual at time *t*.

It can also be shown that these residuals sum to zero. Moreover, they are uncorrelated and their expected value is 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: The martingale, deviance and Cox-Snell residuals have the disadvantages that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. Schoenfeld proposed residuals that overcome these disadvantages (Schoenfeld, 1982). These residuals are calculated for each individual and for each covariate. Thus, the Schoenfeld residual for the i^{th} observation in the k^{th} covariate, the i^{th} component in the k^{th} score vector, is given by

$$rs_{ik} = \delta_i \left\{ x_{ik} - \frac{\sum_{j \varepsilon R_{t(i)}} x_{jk} \exp(\beta' x_j)}{\sum_{j \varepsilon R_{t(i)}} \exp(\beta' x_j)} \right\}$$

where, x_j is a vector of p fixed covariates for the j^{th} individual, x_{jk} is the value of k^{th} covariate on the j^{th} individual.

The sum of these residuals is zero and they have a large sample property that, their expected value is zero and they are uncorrelated with one another. The vector of these residuals for the i^{th} observation can be written as $rs_i = \left(rs_{i1}, rs_{i2}, \ldots, rs_{ip}\right)'$ and the convention is that rs_{ik} is set to be missing for censored observations.

Scaling a vector of Schoenfeld residuals by an estimator of its variance is more effective in detecting departures from the assumed model (Grambsch and Therneau, 1994). The vector of the scaled Schoenfeld residuals is then given by

$$rs_i^* = \{var(rs_i)\}^{-1} rs_i \approx mvar(\hat{\beta}) rs_i$$

where, m is the number of events (deaths).

The following are model diagnostics that are required to assess the model adequacy in this study.

3.10.1. Checking for Cox proportional hazard assumption

The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model. The hazards are not proportional means that the linear component of the fitted model varies with time in some manner. So that in order to use the Cox PH model, it has to be checked that the assumption of whether the effects of covariates on hazard ratio remain constant over time. Several procedures of graphical techniques and tests are proposed to investigate the proportionality assumptions in fitting the Cox model (Cox, 1972).

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 $\beta_i(t) = \beta_i + \delta_i g_i(t)$.

where $\beta_i(t)$ is time varying coefficient, β_i is constant, $g_i(t)$ is some specified function 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

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

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

This 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: $H_0: \delta = 0$ versus $H_1: \delta \neq 0$ and we use likelihood-based tests like Wald test. If $\delta = 0$ is not rejected, $\beta_i' s$ are not time varying coefficients and hence the proportional hazards assumption is satisfied. If $\delta = 0$ 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).

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).

3.10.2. 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 estimate 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, $LD_i = 2(l_p(\beta) - l_p(\beta_{-i})$, 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).

3.10.3. Testing for linearity of covariates

After identifying a particular set of explanatory variables on which the hazard function depends, it is important 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. An improvement for the fit of a model may be obtained by using some transformation of the values of a variable instead of the original one. LOESS smoothed curve can be superimposed on the scatter plots to give interpretation. If the functional form suggested (observed) that 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).

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1. Introduction

This study is a retrospective study based on 510 cases of HIV-infected TB patients of age 15 years and above who received anti-TB treatment between September 1, 2010 and April 30, 2013 and were followed until July 31, 2013.

In this section we present results of data analysis, interpretation and discussion. The first part presents summary statistics of factors considered in this study. The descriptive statistics shows four categories of WHO clinical stages. However, the clinical evidence shows that WHO clinical stage I and II are not relevant where HIV/TB co-infection study is done. Therefore, WHO stage I and II are not considered in model analysis. This has the consequence that the number of patients had to be reduced from 510 to 273 patients. The second part compares the survival time in different groups.

The third part of this section is about fitting the model. P values ≤ 0.05 and confidence level of 95% were considered to indicate statistical significance. Cox proportional hazard regression model was used to determine the hazard ratio (HR) of death for each main baseline predictor. To assess the association between baseline variables and mortality, two strategies were used. First, each baseline variable was entered into a single univariate Cox proportional hazards model. Second, a multivariate Cox proportional hazards model was fitted with the predictors that have P ≤ 0.25 in the univariate model. Eventually only those variables that remained significant at $p \leq 0.05$ in the final model were retained as independent predictors of survival/death of HIV/TB co-infected patients. The adequacy of the multivariate

model was also investigated. Finally, the results were interpreted and discussed. The statistical packages SPSS, SAS and STATA were employed to analyze the data.

4.2. Summary Statistics

The study included 510 HIV/TB co-infected patients, who started ART and anti-TB treatment in Chagni Health Center between the years 2010 and 2013, about whom complete information related to the study covariates was available. Of the total of 510 patients included, 323 (63.3%) were females, 81 (15.9%) were never married, and 251 (49.2%) were illiterate patients. Among the total, 303 (59.4%) patients had $CD4\ count \leq 200\ cells/mm^3$, 272 (53.3%) patients were in the age group $(15-30)\ years$, 265 (52.2%) patients were under-weight ($BMI \leq 18.5\ kg/m^2$) and 234 (45.9%) patients started treatment on D4T-based drug regimens. Of these 140 (42.2%), 101 (37.1%), 118 (44.5%) and 102 (43.2%) patients died during ART and anti-TB treatment period, respectively. When ART and TB treatment was initiated, 89 (17.5%), 148 (29%), 228 (44.7%) and 45 (8.8%) patients had clinical AIDS (WHO stage I, II, III and IV), respectively. 206 (40.4%) patients had extra-pulmonary TB and 11 (2.2%) patients had bedridden functional status due to the severe progression of the disease. Substance-use and opportunistic infections (OI) were prevalent among 314 (61.6%) and 249 (48.8%) patients, respectively at baseline (Table 4.1.1).

Table 4.1.1: Demographic and health factors of categorical covariates for HIV/TB co-infected patients under ART in Chagni Health Center.

Demographic and Health Factors		Status of censoring or event N = 510, dead = 200, censored = 310.			
		Total (in %)	Dead	Censored	Event/Death percentage
Sex	Female	323 (63.3)	113	210	35.0
	Male	187 (36.7)	87	100	46.5
Marital	Never married	81 (15.9)	36	45	44.4

status	Married	231 (45.3)	79	152	34.2
	Others	198 (38.8)	85	113	42.9
Level of education	No-education	251 (49.2)	94	157	37.5
	Primary and above	259 (50.8)	106	153	40.9
Substance use	No	196 (38.4)	64	132	32.7
	Yes	314 (61.6)	136	178	43.3
TB site	Pulmonary TB	304 (59.6)	107	197	35.4
	Extra-pulmonary TB	206 (40.4)	93	113	45.1
Functional status	Working	398 (78)	146	252	36.7
	Ambulatory	101 (19.8)	47	54	46.5
	Bedridden	11 (2.2)	7	4	63.6
WHO stage	Stage I	89 (17.5)	25	64	28.1
	Stage II	148 (29)	62	86	41.9
	Stage III	228 (44.7)	89	139	39.0
	Stage IV	45 (8.8)	24	21	53.3
Drug regimen type	D4T- based	234 (45.9)	102	132	43.6
	AZT-based	145 (28.4)	56	89	38.6
	TDF-based	131 (25.7)	42	89	32.1
Past OI	No	261 (51.2)	111	150	42.5
	Yes	249 (48.8)	89	160	35.7
CD4 count category	$CD4\ count \leq 200$	303 (59.4)	140	160	46.2
	$200 < CD4 \leq 350$	118 (23.1)	39	79	33.1
	$CD4\ count > 350$	89 (17.5)	21	68	23.6
BMI category	$BMI \leq 18.5$	265 (52.0)	118	147	44.5
	$18.5 < BMI < 25$	232 (45.5)	80	152	34.5
	$BMI \geq 25$	13 (2.5)	2	4	15.4
Age category	$15 - 30$	272 (53.3)	101	171	37.1
	$31 - 45$	200 (39.2)	79	121	39.5
	$46 - 60$	35 (6.9)	18	17	51.4
	> 60	3 (0.6)	2	1	66.7
Weight category	$30 - 40$	80 (15.7)	37	43	46.2
	$41 - 50$	249 (48.9)	91	152	39.5
	$51 - 60$	153 (30.1)	64	89	41.8
	> 60	27 (5.3)	2	25	7.4

Table 4.1.2: Descriptive statistics of the response variable included in the study.

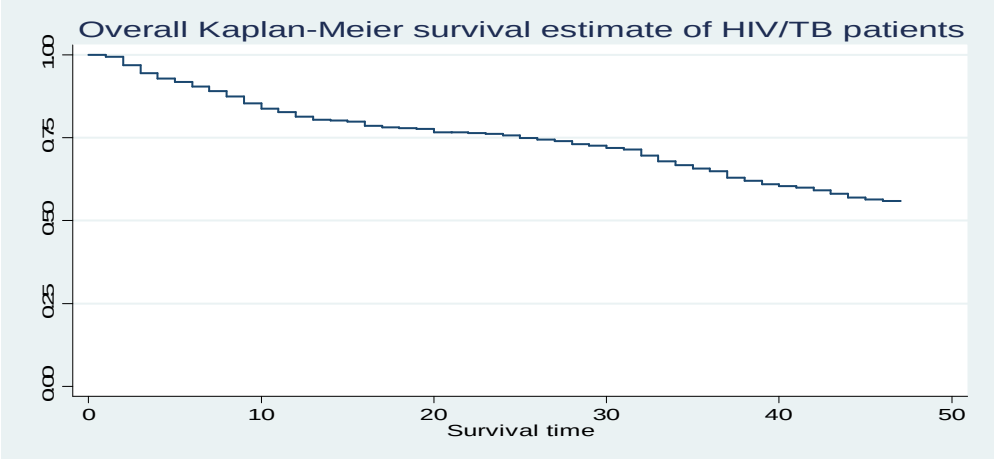
Patient status	Continuous variable	Mean	Standard deviation	Min.	Max.	Median	Q_1	Q_2
Dead	Survival time	19.49	14.23	1	46	15.5	7	33
Censored	Survival time	38.47	12.88	1	47	46	33	46
Overall	Survival time	31.03	16.31	1	47	37	14	46

4.3. Descriptive survival analysis

The patients were followed up for a median period of 37 months. The minimum follow up time was 1 month and the maximum was 47 months. 200 patients (39.2%) died during the follow up time, of which 13 (6.5%), 12 (6%), 10 (5%) and 7 (3.5%) deaths occurred within two months, three months, nine months and twelve months of TB treatment initiation with survival probability of 0.969, 0.945, 0.854 and 0.813, respectively. The overall mean estimated survival time of patients under the study was 35.5 (95% CI: $34.071-36.936$) months. Patients with higher CD4 count ($CD4 > 350\ cells/mm^3$) have relatively higher mean survival time (40.6 months) than lower CD4 count ($CD4 \leq 200\ cells/mm^3$) (32.3 months) patients. Patients with under-weight ($BMI \leq 18.5\ kg/m^2$) had survived for about 33.7 months while the mean survival time for normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) patients was 36.6 months. Table 4.3.1 Appendix presents the mean survival time of patients based on different socio-demographic and clinical characteristics.

We investigate whether the observed differences in data summary among different factors are statistically significant or not with the help of Kaplan-Meier survival method and the log-rank test. The log-rank test was performed to investigate the significance of the observed difference in the Kaplan-Meier estimates of the survivor functions among different categories of the factors.

Figure 4.2.1: The plot of the overall estimate of Kaplan-Meier survivor function of HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

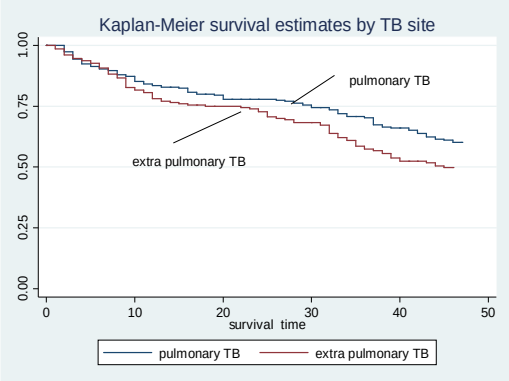


As can be seen from the graph of the estimate of overall Kaplan-Meier survivor function most of the deaths occurred in the earlier months of ART initiation and TB treatment, and it declined in the later months of follow up time.

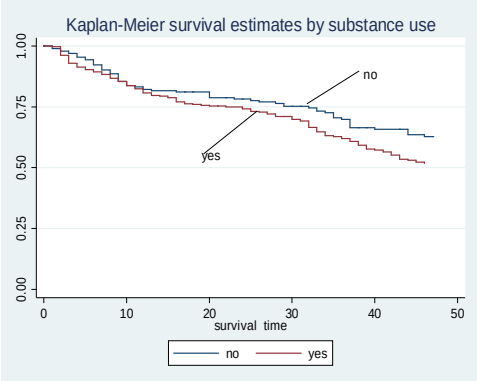
Separate graphs of the estimates of the Kaplan-Meier survivor functions, for different factors, have also been constructed in order to assess whether there were difference in survival experience between different groups of individuals. In general, the pattern that one survivorship function lying above another means that the group defined by the upper curve had a better survival than the group defined by the lower curve. The graphs of Kaplan-Meier survival estimates based on different categories of factors are presented in Figure 4.2.2 below and Figure 4.2.3 in the Appendix.

Figure 4.2.2: Plots of Kaplan-Meier survivor functions, based on different factors, for HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

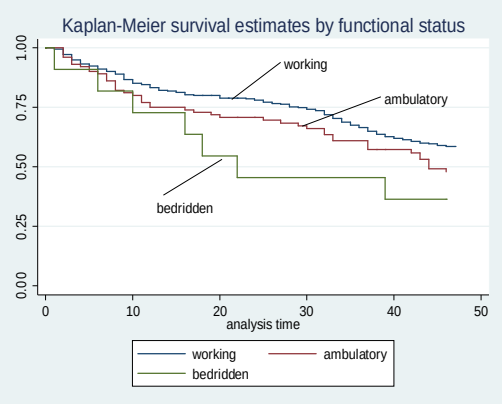
a)



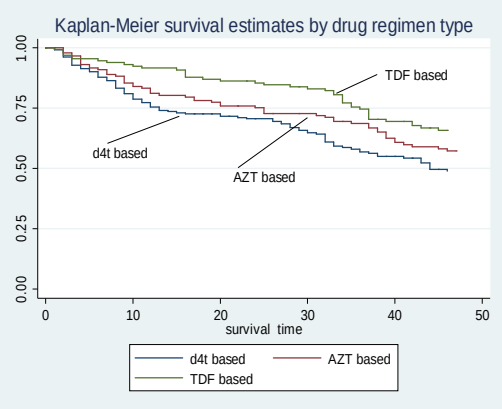
b)



c)



d)



As can be seen from Figure 4.2.2 above and Figure 4.2.3 in the Appendix, relatively larger gaps are observed in covariates such as substance use, sex, drug regimen type, functional status, BMI category, CD4 count category and TB site. However, most of the graphs did not show differences between different categories of marital status, past history of OI, level of education, weight category, age category and WHO clinical stages.

Moreover, according to the log-rank test result in Table 4.2.2 (Appendix), there was no significant difference in survival experience between the various categories of marital status, level of education, age category and past history of OI. However, as can be seen from the log-rank test result in Table 4.2.2, the P-value results confirmed that the survival experience of patients in the different categories of sex, substance-use, functional status, WHO clinical stage, drug regimen type, weight category, BMI category, CD4 count category and TB site differ significantly. Thus, those patients who were not using substances, had working functional status, had WHO clinical stage I, had pulmonary TB, those who were taking TDF-based drug regimens, had normal-weight (BMI between 18.5 kg/m^2 and 25 kg/m^2), weight greater than 60 kg, CD4 count greater than 350 cell/mm^3 and females had better survival experience.

4.3. Results of the Cox proportional hazards model

In statistical modeling, when the number of variables are relatively large, it could be computationally expensive to fit all possible models. A multivariate model containing the variables that were significant at modest ($P \leq 0.25$) levels of significance in the univariate analysis was fitted. As a result, twelve univariate Cox PH models, each containing one explanatory variable were fitted. Then, the multivariate Cox PH model was constructed by those factors which were significant in univariate Cox PH models.

Table 4.3.1: Results of the univariate Cox PH regression model for HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.						
Covariates	df	$\hat{\beta}$	s.e ($\hat{\beta}$)	Chi-square	Pr> $\chi 2$	unadjusted HR
Sex (female)	1	0.051	0.2164	0.0559	0.8131	1.052
Ref. group (male)						1
Age category	3			2.1492	0.5420	
31 – 45	1	0.214	0.2036	1.1021	0.2938	1.238

46 – 60	1	-0.091	0.4313	0.0441	0.8337	0.913
> 60	1	1.012	1.0118	1.0006	0.3172	2.751
Ref. group (15–30)						1
Marital status	2			0.1670	0.9199	
Married	1	-0.107	0.2663	0.1625	0.6869	0.898
Others	1	-0.059	0.2696	0.0473	0.8278	0.943
Ref. group (never-married)						1
Level of education (1^0 & above)	1	-0.019	0.1974	0.0098	0.921	0.981
Ref. group (no-education)						1
Past OI (yes)	1	0.075	0.1973	0.1445	0.7038	1.078
Ref. group (no)						1
TB site (EPTB)	1	0.424	0.1982	4.5862	0.0322	1.529
Ref. group (PTB)						1
Drug regimen type	2			10.784	0.0046	
AZT-based	1	-0.814	0.2577	9.9877	0.0016	0.443
TDF-based	1	-0.416	0.2488	2.7955	0.0945	0.660
Ref. group (D4T-based)						1
CD4 count category	2			11.685	0.0029	
$CD4\ count \leq 200$	1	1.423	0.5126	7.7077	0.0055	4.150
$200 < CD4\ count \leq 350$	1	0.866	0.5456	2.5212	0.1123	2.378
Ref. group ($CD4\ count > 350$)						1
BMI category	2			16.615	0.0002	
$Normal - weight$ ($18.5 < BMI < 25$)	1	-0.951	0.2366	16.151	<.0001	0.386
$Over - weight$ ($BMI \geq 25$)	1	-0.898	1.0066	0.0027	0.9583	0.407
Ref. group ($BMI \leq 18.5$)						1
Functional status	2			7.9590	0.0187	
Ambulatory	1	0.268	0.2224	1.4542	0.2279	1.308
Bedridden	1	1.020	0.3743	7.4299	0.0064	2.774
Ref. group (working)						1
WHO stage IV	1	0.489	0.2240	4.7624	0.0291	1.308
Ref. group (stage III)						1
Substance use (yes)	1	0.019	0.1974	0.0091	0.9240	2.774
Ref. group (no)						1

The univariate analysis of the data on survival times of HIV/TB co-infected patients suggested that not all of the 12 explanatory variables were needed in a proportional hazards model. As a result, the most appropriate subset of these covariates to be included in the multivariate model has to be selected based on their contribution to the maximized log partial likelihood of the model (-2LL). Moreover, using the Wald chi-

square test, the predictors that were found to be significant in the univariate analysis (i.e. P-values ≤ 0.25) will be considered for the next multivariate analysis.

According to the univariate analysis the covariates such as TB type, functional status, WHO clinical stage, baseline CD4 count category, baseline BMI category and drug regimen type were found significant factors to be entered in the multivariate model. The covariates: past history of OI, level of education, baseline age category, sex, substance-use and marital status did not bring significant changes in the value of -2LL($\hat{\beta}$). So, these variables became the first to be removed from the multivariate model. Therefore, we proceed fitting the initial multiple Cox proportional model by including the six covariates that were significant at univariate analysis. Then, another multivariate Cox proportional regression model will be fitted by eliminating those covariates that were not significant at 5% significance level step by step starting with a variable which have a largest Wald P-value.

Fortunately, from the total of six covariates that were significant at univariate analysis in Table 4.3.1, all of the covariates become significant at 5% level of significant in the initial multiple Cox proportional model. Therefore, there were no covariates that could be eliminated from the initial multiple Cox PH model. Hence, at this point we have a multivariate model which includes only the covariates given in Table 4.3.2.

Table 4.3.2: Results of the multivariate Cox PH regression model containing the variables significant at 20- 25% level in the univariate Cox PH regression model.

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	df	Chi-square	P-value
TB site	1	5.3746	0.0204
Drug regimen	2	6.4909	0.0390
CD4 count category	2	15.9939	0.0003
BMI category	2	16.0604	0.0003
WHO stage	1	5.4470	0.0196
Functional status	2	10.0055	0.0067

In the next step, the importance of the variables which were not found to be significant in the univariate analysis as predictors or useful confounder of survival experience of patients was assessed. Thus, these variables were included one at a time in the model containing the covariates in Table 4.3.2 and the improvement in -2LL($\hat{\beta}$) of the significant variables was examined (See Table 4.3.4 (a−e) Appendix).

However, the Wald test P-values in Table 4.3.4 (a−e) (Appendix) confirmed that none of them were found to be significant at 5% level of significance. Then we could conclude that none of those variables were found to be important to be included in the model. Therefore, the most appropriate main effects model was the one containing the covariates in Table 4.3.2.

The final step in model development strategy was consideration of interaction terms that may be useful in the improvement of the model. All possible two way interaction of the variables of the preliminary multivariate model were formed to see if the interaction effects could increase or decrease the survival time of HIV/TB co-infected patients. The significance of adding each of the interactions in the main effects model, one at a time, was checked using the Wald test.

However, the Wald test P-values in Table 4.3.5 Appendix indicated that none of the interactions was found to be significant at 5% level of significance. Thus, the final model would be the one containing only the main effects in Table 4.3.2. The parameter estimates and hazard ratios of the covariates are shown in Table 4.3.3.

Table 4.3.3: Estimated values of the coefficients, adjusted hazard ratios, 95% CI for the adjusted hazard ratio and P-values of the final Cox PH model for HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.							
Covariate	df	$\hat{\beta}$	s.e ($\hat{\beta}$)	Chi-square	Pr> $\chi 2$	aHR	CI for aHR
Functional status	2			10.005	0.0067		
Ambulatory	1	0.021	0.2322	0.0079	0.9292	1.021	0.647−1.609
Bedridden	1	1.210	0.3870	9.7675	0.0018	3.352	1.570−7.158
Ref. group (working)						1	
TB site (EPTB)	1	0.475	0.2049	5.3746	0.0204	1.608	1.076−2.403
Ref. group (PTB)						1	
Drug regimen type	2			6.4909	0.0390		
AZT-based	1	-0.670	0.2650	6.3838	0.0115	0.512	0.304−0.860
TDF-based	1	-0.251	0.2528	0.9894	0.3199	0.778	0.474−1.276
Ref. group (D4T-based)						1	
CD4 count category	2			15.994	0.0003		
$CD4\ count \leq 200$	1	1.548	0.5220	8.7933	0.0030	4.702	1.690−13.08
$200 < CD4\ count \leq 350$	1	0.784	0.5559	1.9884	0.1585	2.190	0.737−6.511
Ref. group ($CD4 > 350$)						1	
BMI category	2			16.060	0.0003		
$18.5 < BMI < 25$	1	-0.962	0.2404	16.025	<.0001	0.382	0.238−0.612
$BMI \geq 25$	1	-0.342	1.0177	0.3575	0.5499	0.710	0.097−5.219
Ref.group ($BMI \leq 18.5$)						1	
WHO stage IV	1	0.545	0.2337	5.4470	0.0196	1.725	1.091−2.728
Ref. group (stage III)						1	

4.4. Model Diagnostics

After a model had been fitted to an observed set of survival data the adequacy of the fitted model needs to be assessed. However, in order to use the parameter estimates, assessment of the important assumptions associated with the proportional hazards Cox regression model must be met. Indeed, the use of diagnostic procedures for model checking was an essential part of the model in process (Collett, 2003). In this study, three kinds of diagnostics namely, violation of the assumption of proportional

hazards, effect of influential observations and non-linearity in the relationship between the log hazard and the covariates are assessed.

4.4.1 Assessment for linearity of covariates in the model

In the process of model development, it is important to check whether the correct functional form of the continuous covariate holds in the model proposed to describe the data. A number of techniques which are designed to determine whether the data support the hypothesis that the effect of the covariate is linear in the log hazard. The resulting plots with a LOESS smoothed curve superimposed to aid in their interpretation. If the martingale residuals manifest a random pattern implying no systematic pattern and smoothed plots (LOESS) are approximately horizontal lines then the continuous variables had an approximated linear relationship with the survival time. But in this study there is no need for linearity checking since we used the continuous variables as categorical variable.

4.4.2 Checking for influential observations

Observations that have an undue effect on model-based inference are said to be influential. Therefore, in the assessment of model adequacy, it is important to determine whether any particular observation, if any, has an undue impact on inferences made on the basis of model fitted to an observed set of survival data. It was therefore of particular interest to examine the influence of each particular observation on these estimates.

The most direct measure of influence is $\hat{\beta}_j - \hat{\beta}_{j(-i)}$, where $\hat{\beta}_j$ is the jth parameter, $j = 1,2,\dots,p$ in the fitted Cox PH model and $\hat{\beta}_{j(-i)}$ is obtained by fitting the model after omitting observation i. In this way, we had to fit the $n + 1$ Cox models, one with the complete data and n with each observation eliminated. This procedure involves a

significant amount of computation if the sample size is large so that an approximation value was used for $\hat{\beta}_j - \hat{\beta}_{j(-i)}$ from the j[th] component of the vector $rs_{(i)}{}' V(\hat{\beta})$, where $rs_{(i)}$ is the $px1$ vector of score residuals for the i[th] observation which are modifications of Schoenfeld residuals and are defined for all observations, and $V(\hat{\beta})$ is the variance-covariance matrix of the vector of parameter estimates in the fitted Cox PH model.

The j[th] element of this vector is called delta-beta statistic for the j[th] explanatory variable, i.e., $\Delta_i\hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(-i)}$; it shows how much each coefficient will change by the removal of a single observation. This was done by examining the extent to which the estimated parameters and the maximized likelihood in the fitted model would be affected by omitting in turn the data record for each individual in the study.

Thus, the DFBETA and the likelihood displacement statistics were used to examine the untoward effect of each observation on the j^{th} parameter estimate and the maximized log partial likelihood, respectively in the fitted Cox regression model. A DFBETA value in excess of $2/\sqrt{n}$ merits further investigation. In this study, we would be concerned about absolute values in excess of $2/\sqrt{273}$ or 0.12.

The three largest changes in the parameter estimates are presented in Table 4.4.2.1 in the Appendix. As can be seen from Table 4.4.2.1 both absolute DFBETA values are less than 0.12. Thus, the change in parameter estimates by deleting these patients can be considered as insignificant. Thus, at this point we can conclude that neither the estimates for each of the parameters nor the set of parameter estimates are affected by any of the observations in the dataset.

4.4.3 Assessment of the proportional hazards assumption

The validity of Cox's regression analysis relies heavily on the assumption of proportionality of the hazard rates of individuals with distinct values of a covariate. If this assumption is not valid, then one may be appreciably misled by the results of the analysis. In this study, extended Cox model was used to test this assumption.

A graphical check was also used to provide any additional insight into any departure from proportionality. Thus, all interactions of covariates with the logarithm of survival times were modeled together with the main effects, and then formal tests were done about the coefficients of those interaction terms using the Wald test.

The result of the test was given in Table 4.4.3.1 (Appendix). The table shows that the Wald chi-square value and the corresponding P-values for each time dependent covariate. As can be seen from Table 4.4.3.1, the Wald P-values support that none of the coefficients of the interaction terms were significant at 5% level. Therefore, there was no evidence against the null hypothesis that the coefficients of the time varying variables (interaction terms) were zero ascertaining the validity of the proportional hazards assumption for the data.

The assumption of proportionality was also assessed graphically by plotting the scaled Schoenfeld residuals of each covariate against survival time. Figures 4.4.3.1 (a−f) in the Appendix shows the plots of scaled Schoenfeld residuals of each covariate in the model against survival times along with the LOESS smoothed curves. The residuals supported randomness to some extent and LOESS smoothed curves are approximately horizontal showing no evidence of non-proportionality.

Results of the likelihood ratio, score and Wald tests for model goodness of fit displayed in Table 4.3.6 of the Appendix suggested that the model fit is good (i.e. significant at 5% level of significance).

4.5. Interpretation and Discussion of Results

4.5.1. Interpretation of the Results

When multiple proportional hazards Cox model is used in the analysis of survival data, the coefficient of a categorical explanatory variable in the model could be interpreted as the logarithm of the ratio of the hazard of death to the baseline (reference group) hazard. On the other hand, the coefficient for a continuous explanatory variable is the estimated 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 interpretation of those variables that were significant in the final proportional hazards model of HIV/TB co-infected patients treated with ART in Chagni Health Center is as follows.

WHO stage, functional status, drug regimen type, CD4 count category, BMI category and TB site are the six categorical variables that were found to be strongly associated with the survival/death of HIV/TB co-infected patients in the multivariate Cox PH regression model.

Therefore, from the output obtained from the final multivariate Cox proportional hazards model (Table 4.3.3), we could conclude that the estimated relative risk (hazard ratio) of dying for patients who had extra-pulmonary TB compared to those who had pulmonary TB was 1.608 (95% CI: 1.076−2.403). This means that the rate of dying among patients with extra-pulmonary TB was about 1.6 times higher than among those patients who had pulmonary TB patients controlling for other variables

in the model. Thus, those patients who had extra-pulmonary TB were 61% more likely to die than those patients who had pulmonary TB controlling for other variables in the model. Since the confidence interval does not contain 1, an individual who had extra-pulmonary TB had a significantly higher hazard rate, at any given time, than patients who had pulmonary TB.

The estimated relative risk of mortality for patients those who took AZT-based regimen type and TDF-based regimen type compared to those who took D4T-based regimen type were 0.512 (95% CI: 0.304−0.860) and 0.778 (95% CI: 0.474−1.276), respectively. These results reveal that in case of patients who took AZT-based regimen type, they were more likely to have lower mortality by about 49% compared to those patients who took D4T-based regimen type. On the other hand, the p-value for TDF-based regimen type is greater than 0.05 and the confidence interval contains 1, this indicates that it has no statistically significant effect in the survival of HIV/TB co-infected patients compared to those patients who took D4T-based regimens.

The estimated relative risk of mortality for patients those who were being in WHO stage IV compared to those who were being in WHO stage III were 1.725 (95% CI: 1.091−2.728). This means that the rate of dying among patients with WHO stage IV was about 1.7 times higher than among those patients who were being in WHO stage III patients controlling for other variables in the model. Thus, those patients who were in WHO stage IV were 73% more likely to die than those patients who were in WHO stage III controlling for other variables in the model. Since the confidence interval does not contain 1, WHO stage IV patients had significantly higher hazard rate, at any given time, than WHO stage III patients.

Functional status was also associated with the survival/death of HIV/TB co-infected patients. The estimated relative risk of ambulatory and bedridden patients compared to those who were being in working patients were 1.021 (95% CI: 0.647−1.609) and 3.352 (95% CI: 1.570−7.158), respectively. This means that the rate of dying for bedridden patients were about 3.4 times higher than working patients controlling for other variables in the model. Thus, those patients being in bedridden were 235% more likely to die than those patients being in working functional status controlling for other variables in the model. On the other hand, the p-value for ambulatory patients is greater than 0.05 and the confidence interval contains 1, showing that there is no statistically significant effect in the survival of HIV/TB co-infected patients compared to those who were being working patients.

Baseline BMI category was associated with the survival/death of HIV/TB co-infected patients. The estimated relative risk of normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) and over-weight ($BMI \geq 25\ kg/m^2$) patients compared to those who had under-weight ($BMI \leq 18.5\ kg/m^2$) patients were 0.382 (95% CI: 0.238−0.612) and 0.710 (95% CI: $0.097-5.219$) respectively. These results reveal that in case of patients who had normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$), they were more likely to have lower mortality by about 62% compared to those patients who had under-weight ($BMI \leq 18.5\ kg/m^2$). On the other hand, the p-value for over-weight ($BMI \geq 25\ kg/m^2$) patients are greater than 0.05 and the confidence interval contains 1, implying that it has no statistically significant effect in the survival of HIV/TB co-infected patients compared to patients with under-weight ($BMI \leq 18.5\ kg/m^2$).

Baseline CD4 count category was also associated with the survival/death of HIV/TB co-infected patients. The estimated relative risk of $CD4\ count \leq 200\ cells/mm^3$ and $200\ cells/mm^3 < CD4\ count \leq 350\ cells/mm^3$ patients compared to those who had $CD4\ count > 350\ cells/mm^3$ patients were 4.702 (95% CI: 1.690−13.08) and 2.190 (95% CI: 0.737−6.511) respectively. This means that the rate of dying for $CD4\ count \leq 200\ cells/mm^3$ patients were about 4.7 times higher than $CD4\ count > 350\ cells/mm^3$ patients controlling for other variables in the model. Thus, those patients having $CD4\ count \leq 200\ cells/mm^3$ had a 370% higher risk rate of mortality than those patients having $CD4\ count > 350\ cells/mm^3$. On the other hand, the p-value for patients with $200\ cells/mm^3 < CD4\ count \leq 350\ cells/mm^3$ are greater than 0.05 and the confidence interval contains 1, showing that it has no statistically significant effect in the survival of HIV/TB co-infected patients compared to those patients with $CD4\ count > 350\ cells/mm^3$.

4.5.2 Discussion of the Results

This study identified covariates that are significantly associated with increased risk of mortality. Identifying patients at a higher risk of death has the advantage that due attention will be given to the risk group during their follow up time to minimize the risk of mortality while they are taking ART and TB treatment. This historical (retrospective) cohort study found that the significant predictors of lower/higher chance of survival in patients living with HIV/TB after initiation of ART and anti-TB treatment were: $CD4\ count \leq 200\ cells/mm^3$at baseline, normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) at baseline, being in advanced WHO stage IV, being in bedridden, taking AZT-based drug regimens and presence of extra-pulmonary TB.

The impact of CD4 cell count on survival rate has been assessed by several studies indicating that depletion of CD4 cell count is associated with high risk of death for HIV/TB co-infected patients. A prospective cohort study by Andreychyn and Zhyvytsia (2013) conducted at Zaporizhzhya AIDS Center, Ukraine revealed that patients with a CD4+ T-cell count < 100 cell/µl had a 5-fold higher risk of mortality (aHR = 5.2; 95% CI: 1.4−19.4). A similar study conducted in Addis Ababa, Ethiopia by Kassa et al (2012) showed that CD4+ cell count < 50 cells/µl as baseline was an independent predictor of mortality for HIV/TB co-infected patients. Another study conducted in the state of Pernambuco, Brazil by Maruza *et al* (2012) suggested that CD4+ T-lymphocyte count below 200 cells/mm^3 was a risk factor of death for HIV/TB co-infected patients. A study conducted at the Internal Medicine Department of the Kigali University Hospital, Rwanda by Lorent et al (2011) revealed that patients with a baseline CD4 count < 100 cells/mm^3 at TB diagnosis (aHR = 1.7; 95% CI: 1.0–2.9) was independent predictor of mortality. The current study showed that having $CD4\ count \leq 200\ cells/mm^3$ is associated with increased risk of death for HIV/TB co-infected patients (aHR = 4.702; 95% CI: 1.690 − 13.08).

Several studies showed that pulmonary TB was more frequent in PLWHIV than extra-pulmonary TB. Lorent et al (2011) conducted a prospective observational cohort study at the Internal Medicine Department of the Kigali University Hospital, Rwanda. Their study showed that patients who had an extra-pulmonary tuberculosis had a 2-fold risk of mortality (aHR = 2.0; 95% CI: 1.1−3.9). A similar study done by Andreychyn and Zhyvytsia (2013) showed that patients who had an extra-pulmonary tuberculosis had a 2-fold higher risk of mortality (aHR = 2.2; 95% CI: 1.1−8.3) compared to those patients who had pulmonary TB. Lo'pez-Gatell et al (2008**)** also revealed that the hazard of AIDS-related death was 2.4 times more for the person-time

with TB compared to the person-time without TB in Baltimore, USA. The current study also identified that the presence of extra-pulmonary TB significantly reduced the probability of survival and increased the risk of death for HIV/TB co-infected patients ($aHR = 1.608$; 95% CI: 1.076−2.403).

A study conducted in Addis Ababa, Ethiopia by Kassa et al (2012) showed that the advanced WHO clinical stage III HIV infection ($aHR = 2.53$; 95% CI: 1.70−3.70) and WHO clinical stage IV HIV infection ($aHR = 3.86$; 95% CI: 2.54−5.86) were risk factors for mortality of patients associated with TB. The current study showed that an advanced WHO stage IV is associated with increased risk of death for HIV/TB co-infected patients ($aHR = 1.725$; 95% CI: 1.091−2.728).

BMI category is another significant predictor of mortality of patients. Elliott et al (1995) conducted a prospective study in Lusaka showed that low weight was a major cause of death in patients with HIV infection. A similar study by Braitstein et al (2009) in a large network of HIV clinics in western Kenya showed that being severely low weight-for-height at enrollment ($aHR = 1.46$; 95% CI: 1.32−1.62) was an independent risk factor for having an incident TB diagnosis. Getahun et al (2011) conducted a retrospective cohort study in three randomly selected health centers in Addis Ababa, Ethiopia. The result revealed that body weight at initiation of anti-TB treatment (< 35 kg) was an independent predictor for time to death. The current study showed that having normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) is associated with decreased risk of death for HIV/TB co-infected patient ($aHR = 0.382; 95\%\ CI: 0.238 - 0.612$).

Drug regimen type is another significant predictor of mortality of patients. Elliott et al (1995) showed that poor compliance with regimens containing rifampicin and

pyrazinamide led to death in HIV/TB co-infected patients. Varma et al (2009) in a multi-center observational study in 32 public TB treatment facilities in Thailand revealed that fluconazole use (aHR = 0.34; 95% CI: 0.18–0.64) and co-trimoxazole use (aHR = 0.41; 95% CI: 0.20−0.83) were associated with low risk of death. They also suggested that efavirenz- and nevirapine-containing ART regimens were associated with similar rates of adverse events and death. Hermans et al (2013) found that in urban HIV clinics in Uganda the use of EFV compared to NVP was associated with higher TB incidence in patients with a baseline CD4 (+) T-cell count < 100 cells/mm^3 (aHR = 2.05; 95% CI: 1.29−3.27). The current study also identified that AZT-drug regimen type is associated with reduced risk of death for HIV/TB co-infected patients (aHR = 0.512; 95% CI: 0.304−0.860).

Deribe et al (2013) conducted a case–control study at Jimma University Specialized Hospital, southwest Ethiopia. The study showed that bedridden patients experienced a higher risk of death than working patients (OR = 2.84; 95% CI: 1.17−6.89) among HIV/TB co-infected patients. A study conducted in governmental health institutions in Bahir Dar, northwest Ethiopia by Sileshi et al (2013) also showed that ambulatory patients had twice the risk of death than working patients (aHR = 2.10). The current study found that bedridden patients had higher risk of death for HIV/TB co-infected patients (aHR = 3.352; 95% CI: 1.570−7.158).

CHAPTER FIVE

CONCLUSIONS and RECOMENDATIONS

5.1 Conclusions

In this study, the survival of HIV/TB co-infected patients who received TB treatment in Chagni Health Center from September 1, 2010 to July 31, 2013 was studied. The objective was to identify risk factors associated with high risk of mortality for HIV/TB co-infected patients using the methods of survival analysis.

The descriptive analysis showed that out of 510 patients, 200 (39%) patients died during ART and anti-TB treatment period, of these 113 (56.5%) were females. 303 (59.4%) patients had $CD4\ count \leq 200\ cells/mm^3$, 272 (53.3%) patients were in the age group $(15-30)\ years$, 265 (52.2%) patients were under-weight ($BMI \leq 18.5\ kg/m^2$) and 234 (45.9%) patients started treatment on D4T-based drug regimens. Of these 140 (42.2%), 101 (37.1%), 118 (44.5%) and 102 (43.2%) died during ART and anti-TB treatment period, respectively. The log rank test result showed that those patients who were females, who were not using substances, had working functional status, being in advanced WHO stage I, had pulmonary TB and those who were taking TDF based drug regimens had better survival experience.

The results of the multivariate proportional hazards Cox regression analysis showed that having $CD4\ count \leq 200\ cells/mm^3$ at the start of ART, the presence of extra-pulmonary TB, being bedridden and WHO stage IV contributed to high risk of mortality. On the other hand, AZT-based drug regimens and having normal-weight ($18.5\ kg/m^2 < BMI < 25\ kg/m^2$) reduced the risk of mortality.

5.2 Recommendations

- TB test must be done for people living with HIV in order to mitigate further deterioration of their health.
- Health workers should be cautious when a patient had lower baseline CD4 cell count ($CD4\ count \leq 200\ cells/mm^3$), and a patient is being bedridden and is being in WHO clinical stage IV.
- HIV/TB co-infected people should have awareness about the hazard of the risk factors identified in this study.

References

1. Ackah AA, Coulibaly D, Digbeu H, et al. Response to treatment, mortality, and CD4 lymphocyte counts in HIV-infected persons with tuberculosis in Abidjan, Cote d'ivoire. Lancet 345 (1995), 607-610.
2. Andreychyn M and Zhyvytsia D. Effect of antiretroviral therapy on survival of HIV/TB-infected patients in Ukraine. *Our Dermatol Online. 2013; 4(2):149-152*.
3. Badri M, Wilson D, Wood R. Effect of highly active antiretroviral therapy on incidence of tuberculosis in South Africa: a cohort study. Lancet 2002; 359: 2059-64.
4. Berkson, J., and Gage, R. P. Calculation of survival rates for cancer. Proceedings of the sta meetings of the Mayo clinic 25 (1950), 270-286.
5. Braitstein P, Nyandiko W, Vreeman R, Wools-Kaloustian K, Sang E, Musick B, Sidle J, Yiannoutsos C, Ayaya S, Carter EJ. The clinical burden of tuberculosis among human immunodeficiency virus-infected children in Western Kenya and the impact of combination antiretroviral treatment. Pediatr Infect Dis J. 2009 Jul; 28(7): 626-32.
6. Breslow, N. (1974). Covariance analysis of censored survival data. Biometrics, 30, 89-100.
7. Collett, D. Modelling Survival Data in Medical Research. Chapman and Hall, London, 2003.
8. Collins J, Jorge O, Alarcon J, Moore D and Hernandez A. Effect of antiretroviral therapy on survival of HIV-Infected tuberculosis patients in Peru. Rev Panam Infectol 2010; 12(2): 37-43.

9. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC and Dye C: The growing burden of tuberculosis global trends and interactions with the HIV epidemic. Arch Intern Med 2003; 165: 1009-1021.
10. Cox, D. R. Regression models and life-tables. Journal of the Royal Statistical Society. Series B 34 (1972), 187-220.
11. Daley GL, Small PM, Schecter GF, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus: an analysis using restriction-fragment-length polymorphism. The New England Journal of Medicine 326 (1992), 231-235.
12. Demissie M, Lindtjorn B, and Tegbaru B. HIV infection in tuberculosis patients in Addis Ababa. Ethiop J Health Dev 2000; 14: 277-82.
13. Deribe K, Yami A, Deribew A, Mesfin N, Colebunders R, Van geertruyden J. P, Woldie M and Maja T. Predictors of mortality among tuberculosis-HIV co-infected persons in Southwest Ethiopia: A case–control study. *Journal of the International Association of Providers of AIDS Care (JIAPAC)* 21 August 2013; 00(0) 1-5.
14. Dheda K, Shean K, Zumla A, Badri M, Streicher EM, Willcox P, et al. Early treatment outcomes and HIV status of patients with extensively drug-resistant tuberculosis in South Africa: a retrospective cohort study. Lancet May 22, 2010;375:1798-807.
15. Efron B. 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.
16. Elliott A, Halwiindi B, Hayes R, Luo N, Mwinga A, Tembo G et al,. The impact of human immunodeficiency virus on mortality of patients treated for

tuberculosis in a cohort study in Zambia. Transactions of the royal society of tropical medicine and hygiene (1995) 89,78-82.

17. Federal Ministry of Health Annual performance report of HSDP-111EFY 1999(2006/07), MOH, Ethiopia October 2007.
18. Fleming, T. R., and Harrington, D. P. Counting Processes and Survival Analysis. Wiley, New York, 1991.
19. Friedland, G., Churchyard, G. J., and Nardell, E. Tuberculosis and HIV co-infection: Current state of knowledge and research priorities. Journal of Infectious Diseases 196 (2007), Suppl 1:S1-S3.
20. Gadkowski L. B, Hamilton C. D, Ellen M. A, Fortenberry R, Luffman J, Zeringue E and Stout J. E (2009). HIV-specific health care utilization and mortality among TB/HIV co-infected persons. *AIDS patient care and STDs; 23(10).*
21. Getahun B, Ameni G, Biadgilign S and Medhin G. Mortality and associated risk factors in a cohort of tuberculosis patients treated under DOTS programme in Addis Ababa, Ethiopia. BMC Infectious Diseases 2011, 11:127.
22. Girardi E, Antonucci G, Vanacore P, Libanore M, Errante I, Matteelli A, and Ippolito G. Impact of combination antiretroviral therapy on the risk of tuberculosis among persons with HIV infection. AIDS 2000; 14:1985-91.
23. Godfrey-Fausset P, Maher D, Mukadi YD, Nunn P, Perriens J and Raviglione M. How human immunodeficiency virus voluntary testing can contribute to tuberculosis control. Bull World Health Organ 2002, 80:939-945.
24. Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika; 81:515-526.

25. Harries A D, Zachariah R, Corbett E L, Lawn SD, Santos-Filho ET, Chimzizi R, Harrington M, Maher D, Williams BG, De Cock KM. Confronting the HIV-associated tuberculosis epidemic—when will we act? Lancet 2010; 375: 1906–1919.
26. Hermans SM, Manabe YC, Kiragga AN, Hoepelman AI, Lange JM, van Leth F. Risk of tuberculosis after antiretroviral treatment initiation: a comparison between efavirenz and nevirapine using inverse probability weighting. Antivir Ther. 2013;18(4):615-22.
27. Hosmer, D.W. and Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York.
28. Ismail I and Bulgiba A. Predictors of death during tuberculosis treatment in TB/HIV co-infected patients in Malaysia. PLoS One. 2013; 8(8): e73250.
29. Johansson K. A, Robberstad B and Norheim O. N. Further benefits by early start of HIV treatment in low income countries. Survival estimates of early versus deferred antiretroviral therapy. January 2010, 7:3.
30. Joint United Nations programme on HIV/AIDS (UNAIDS). AIDS Epidemic Update 2009. Geneva; UNAIDS, 2009.
31. Kalbfleisch J. D and Prentice R. L (1980). The Statistical Analysis of Failure Time Data. New York: Wiley.
32. Kaplan, E., and Meier, P. Nonparametric estimation from incomplete observations. Journal of American Statistical Association 53 (1958), 457-481.
33. Kassa A, Teka A, Shewaamare A, Jerene D. Incidence of tuberculosis and early mortality in a large cohort of HIV infected patients receiving antiretroviral therapy in a tertiary hospital in Addis Ababa, Ethiopia. Trans R Soc Trop Med Hyg. 2012 Jun; 106(6):363-70.

34. Klein, J. P., and Moeschberger, M. L. Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York, 1997.
35. Kleinbaum, D. G and Klein M. Survival Analysis: A Self learning text. Springer, New York, 1996.
36. Lawless, J. F. Statistical Models and Methods for Lifetime Data Analysis. Wiley, New York, 1982.
37. Lawn S D, Kranzer K, Wood R. Antiretroviral therapy for control of the HIV-associated tuberculosis epidemic in resourcelimited settings. Clin Chest Med 2009; 30: 685–699.
38. Lo´pez-Gatell H., Cole S.R., Margolick J.B., Witt M.D., Martinson J., Phair J.P. and Jacobson L.P. (2008). Effect of tuberculosis on the survival of HIV-infected men in a country with low tuberculosis incidence. AIDS, 22:1869–1873.
39. Lorent N, Osee Sebatunzi O, Mukeshimana G, Ende J, Joannes Clerinx J. Incidence and risk factors of serious adverse events during anti tuberculosis treatment in Rwanda: A prospective cohort study. PLoS one. May 2011, volume 6, issue 5, e19566.
40. Manosuthi W, Chottanapand S, Thongyen S, Chaovavanich A, Sungkanuparph Survival rate and risk factors of mortality among HIV/tuberculosis-co infected patients with and without ART. J Acquir Immune Defic Syndr. 2006 Sep; 43(1): 42-46.
41. Mazura, M; Albuquerque, M. F. P. M; Braga, M. C; Barbosa, M. T. S; Byington, R; Coimbra, I; Moura, L. V; Batista, J. D. L; Diniz, G. T. N; Miranda-Filho, D. B; Lacerda, H. R; Rodrigues, L. C; Ximenes, R. A. A. Survival of HIV-infected

patients after starting tuberculosis treatment: a prospective cohort study. International Journal of Tuberculosis and Lung Diseases 2012; 16(5): 618-624.

42. Murray, J. F., and Gursed, d. HIV infection and tuberculosis. Respiration 57 (1990), 210-220.
43. Naidoo P, Peltzer K, Louw J, Matseke G, Mchunu G and Tutshana B. Predictors of tuberculosis (TB) and antiretroviral (ARV) medication non-adherence in public primary care patients in South Africa: a cross sectional study. BMC Public Health 2013, 13:396.
44. Natpatou S, Kevin P, Patcharin S, Sriprapa N, Norio Y, Wanpen W, Surin S, Wanchai S, Samroui K, Sakon I and Jay K. Antiretroviral Therapy for HIV-Infected Tuberculosis Patients Saves Lives but Needs to Be Used More Frequently in Thailand. Journal of Acquired Immune Deficiency Syndromes 2008; 48(2):181–189.
45. Odhiambo JA, Borgdorff MW, Kiambih FM, Kibuga D, Kwamanga D, Nganga L, Agwanda R, Kalisvaart N, Misljenovic O, Nagelkerke N, and Bosman M. (1999) Tuberculosis and the HIV epidemic: increasing annual risk of tuberculosis infection in Kenya, 1986-19.96. J Public Health 89:1078-82.
46. Raviglione MC, Narain JP, Kochi A. HIV-associated tuberculosis in developing countries: clinical features, diagnosis, and treatment. Bull World Health Organ 1992; 70:515-26.
47. Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. Biometrica; 69: 239-241.
48. Sharma SK, Mohan A, Kadhiravan T. HIV-TB co-infection: epidemiology, diagnosis and management. Indian J Med Res 2005; 121:550-67.
49. Sileshi B, Deyessa N, Girma B, Melese M and Suarez P. Predictors of mortality among TB-HIV Co-infected patients being treated for tuberculosis in

Northwest Ethiopia: a retrospective cohort study. BMC Infectious Diseases 2013, 13:297.

50. Sonnenberg P, Glynn JR, Fielding K, Murray J, Godfrey-Faussett P, and Shearer S. (2004) HIV and pulmonary tuberculosis: the impact goes beyond those infected with HIV/AIDS 18:657-62.
51. Stockdale A, Nkuranga J, Torok M, Faragher B and Lalloo D. Initiation of antiretroviral therapy in HIV-infected tuberculosis patients in rural Kenya: an observational study. Tropical Medicine and International Health 2013.
52. The White House Office of National AIDS Policy. HIV and TB: Tuberculosis. Ottawa, 1995.
53. Toossi Z, Mayanja-Kizza H, Hirsch GS, et al. Impact of tuberculosis (TB) on HIV-1 activity in dually infected patients. Glin. Exp. Immunol. 123 (2001), 233-238.
54. Tsiatis. A and Zhang. D. Analysis of Survival Data Lecture Notes. Daowen Zhang , Department of Statistics North Carolina State University , 2005.
55. Tsiouris SJ, Gandhi NR, El-Sadr WM, Friedland G. Tuberculosis and HIV--Needed: A New Paradigm for the Control and Management of Linked Epidemics. *Medscape General Medicine.* 2007; 9(3):62.
56. Umphonsathien M and Sungkanuparph S. Early Initiation of Antiretroviral Therapy in HIV/Tuberculosis Co-infection and Immune Reconstitution Inflammatory Syndrome. *J Infect Dis Antimicrob Agents 2011;28:15-23.*
57. Varma J, Nateniyom S, Akksilp S, Mankatittham W et al. HIV care and treatment factors associated with improved survival during TB treatment in Thailand: an observational study. BMC Infect Dis. 2009; 9: 42.

58. WHO. (2003a) Interventions to control TB in high HIV prevalent populations, p. 19-23 Guidelines for Implementing Collaborative TB and HIV Programme Activities. WHO, Geneva.
59. *WHO Report 2011: Global TB Control.*
60. WHO. guidelines for HIV surveillance among tuberculosis patients. Second ed. 2004.
61. World Health Organization. Global tuberculosis control. A short update to the 2009 report. Geneva, Switzerland: WHO, 2009.
62. World Health Organization. WHO Three I's meeting. Report of a joint WHO HIV/AIDS and TB Department meeting. Geneva, Switzerland: WHO, 2008.
63. Worodria W, Massinga-Loembe M, Mazakpwe D, Luzinda K, Menten J, Van Leth F, MD, Mayanja-Kizza H, Kestens L, Mugerwa R, Reiss P and Colebunder R. Incidence and predictors of mortality and the effect of tuberculosis immune reconstitution inflammatory syndrome in a cohort of TB/HIV patients commencing antiretroviral therapy. J Acquir Immune Defic Syndr. 2011; 58(1): 32-37.
64. Xueyan J., Hongzhou L., Yuexin Z., Zengquan Z., Hanhui Y., Qingxia Z., Hui W.,Fuxiang W., Wangqian J., Hongqing S., Limin D., Yan S., Yun H., Wubuer M., Aibaidoula Y., Xiaofeng S., Yadong H., Wei T., Boping Z. and Cao Y.(2008). A Cross-Sectional Study of HIV and Tuberculosis Co infection Cases in Mainland China. Southern Medical Association; 101(9):914-917.
65. Yassin MA, Takele L, Gebresenbet S, Girma E, Lera M, Lendebo E, and Cuevas LE. HIV and tuberculosis co-infection in the southern region of Ethiopia: a prospective epidemiological study. Scand J Infect Dis. 2004, 36:670-673.

66. Zachariah R, Spielmann MP, Harries AD, Salaniponi FL. Voluntary counseling, HIV testing and sexual behavior among patients with tuberculosis in a rural district of Malawi. Int J Tuberc Lung Dis2003 Jan;7(1):65-71.

67. Zvandasara, P., J.Hargrove, R. Ntozini, H. Chidawanyika, K. Mutasa, P. Iliff, L. Moulton, F. Mzengeza. L.Malaba, B. Ward, K. Nathoo, L. Zijenah, M. Mbizvo, C. Zunguza, J. Humphrey and the ZVITAMBO Study Group. 2006. "Mortality and Morbitidy among Postpartum HIV-positive and HIV-negative Women in Zimbabwe:Risk Factors, Causes, and Impact of Single-dose Postpartum vitamin A Supplementation." *The International Journal of Tuberculosis and Lung Disease* 43: 107-116.

APPENDIX

Table 4.2.1: Kaplan-Meier analysis of mean survival times for patients according to important socio-demographic and clinical characteristics of HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

Covariate	Category	Mean estimate	Std. Error	95% confidence Interval for mean
Sex	Female	36.7	.908	34.882−38.440
	Male	33.0	1.185	30.687−35.330
Marital status	Never married	35.1	1.711	31.773−38.480
	Married	36.2	1.023	34.193−38.204
	Others	34.0	1.223	31.603−36.397
Level of education	No-education	35.5	1.047	33.416−37.518
	Primary and above	35.5	1.315	32.900−38.055
Substance use	No	37.0	1.149	34.730−39.233
	Yes	34.1	.919	32.285−35.888
TB site	Pulmonary TB	36.7	.927	34.893−38.525
	Extra-pulmonary TB	33.2	1.144	31.015−35.498
Functional status	Working	36.3	.807	34.715−37.881
	Ambulatory	32.9	1.692	29.563−36.195
	Bedridden	26.9	5.159	16.797−37.021
WHO stage	Stage I	39.5	1.284	37.003−42.039
	Stage II	33.3	1.366	30.643−35.999
	Stage III	35.3	1.129	33.070−37.497
	Stage IV	32.1	2.493	27.235−37.009
Drug regimen type	D4T- based	32.5	.135	30.289−34.738
	AZT-based	35.8	1.363	33.095−38.437
	TDF-based	38.8	1.130	36.594−41.024
Past OI	No	35.2	1.033	33.172−37.222
	Yes	35.3	1.010	33.271−37.230
CD4 count category	$CD4 \leq 200$	32.3	1.031	30.231−34.275
	$200 < CD4 \leq 350$	38.5	1.203	36.117−40.832
	$CD4 > 350$	40.6	1.244	38.162−43.039
BMI	$BMI \leq 18.5$	33.7	1.022	31.699−35.707
	$18.5 < BMI < 25$	36.6	1.049	34.515−38.628
	$BMI \geq 25$	40.4	3.672	33.187−47.582
Age	$15 - 30$	36.1	.994	34.128−38.023
	$31 - 45$	34.4	1.158	32.102−36.640
	$46 - 60$	33.6	2.665	28.351−38.797
	> 60	38.7	3.953	30.918−46.415
Weight	$30 - 40$	32.6	1.994	28.740−36.556
	$41 - 50$	35.1	1.061	33.033−37.193

	51 – 60	35.1	1.248	32.631−37.522
	> 60	44.1	1.398	41.328−46.806
Overall survival time		35.5	.731	34.071−36.936

Table 4.2.2: Results of the Log-rank test for the categorical variables of HIV/TB co-infected patients treated under ART in Chagni Health Center.

Log-rank Test for Equality of Survivor Function N = 510, dead = 200, censored = 310.			
Covariates	df	Wald Chi-Square	P-value
Sex	1	7.1869	0.0073
Marital status	2	3.3668	0.2109
Level of education	2	2.1213	0.3454
Past History of OI	1	0.9014	0.3424
TB site	1	4.9894	0.0255
Drug regimen type	2	4.2757	0.0036
Substance use	1	4.3256	0.0375
Functional status	2	6.058	0.0484
WHO stage	3	0.8408	0.02
CD4 count category	2	13.84	0.0010
BMI category	2	19.07	0.0001
Age category	3	2.32	0.5096
Weight category	3	39.79	0.0000

Table 4.3.4 (a-e): Result of multivariate Cox PH model containing variables in Table 4.3.2 including those not significant in univariate analysis one at a time.

a) When level of education is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	5.2922	0.0214
Drug regimen	2	6.7032	0.0350
CD4 count category	2	16.0976	0.0003
BMI category	1	5.0571	0.0245
WHO stage	2	16.0457	0.0003
Functional status	2	10.2850	0.0058
Level of education	1	0.4004	0.5269

b) When past OI is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	5.3327	0.0209
Drug regimen	2	6.4200	0.0404
CD4 count category	2	15.9429	0.0003
BMI category	1	5.4434	0.0196
WHO stage	2	15.9176	0.0003
Functional status	2	9.9459	0.0069
Past OI	1	0.0042	0.9486

c) When age is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	5.3198	0.0211
Drug regimen	2	6.4331	0.0401
CD4 count category	2	15.8734	0.0004
BMI category	1	5.6327	0.0176
WHO stage	2	15.9941	0.0003
Functional status	2	10.0909	0.0064
Age	3	0.3699	0.9464

d) When marital status is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	4.8880	0.0270
Drug regimen	2	6.6037	0.0368
CD4 count category	2	16.4999	0.0003
BMI category	1	5.6359	0.0176
WHO stage	2	15.7727	0.0004
Functional status	2	10.2775	0.0059
Marital status	2	0.4484	0.7992

e) When substance-use is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	5.6213	0.0177
Drug regimen	2	6.9410	0.0311
CD4 count category	2	17.0300	0.0002

BMI category	1	5.7206	0.0168
WHO stage	2	16.7981	0.0002
Functional status	2	9.6791	0.0079
Substance-use	1	1.5022	0.2203

f) When sex is included

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	5.4928	0.0191
Drug regimen	2	6.4645	0.0395
CD4 count category	2	16.6581	0.0002
BMI category	1	5.5838	0.0181
WHO stage	2	15.9409	0.0003
Functional status	2	9.9573	0.0069
Sex	1	0.7955	0.3724

Table 4.3.5: Wald chi-square statistics and the corresponding p-values of possible two way interaction terms, added one at a time, to the variables included in the model in Table 4.3.2.

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.				
Interaction b/n covariates		df	Chi-square	P-value
Functional status	TB site	1	0.0274	0.8685
	Drug regimen	1	2.0257	0.1547
	CD4 count	1	0.4173	0.518
	BMI	1	0.0085	0.9266
	WHO stage	1	0.2351	0.6278
TB site	drug regimen	1	2.4187	0.1199
	CD4 count	1	0.4881	0.4848
	BMI	1	0.0023	0.9617
	WHO stage	1	0.1210	0.7279
Drug regimen	CD4 count	1	0.2706	0.6029
	BMI	1	0.0215	0.8834
	WHO stage	1	0.1520	0.6966
CD4 count	BMI	1	3.3971	0.0653
	WHO stage	1	0.0017	0.9674
BMI	WHO stage	1	0.2466	0.6195

Table 4.3.6: Results of the Likelihood ratio, Score and Wald tests for testing the global null hypothesis: BETA = 0.

Test	df	$\chi 2$	P-value
Likelihood Ratio	10	62.0022	< 0.0001
Score	10	60.2155	< 0.0001
Wald	10	56.4009	< 0.0001

Table 4.4.3.1: Results of the multivariate Cox PH regression model containing the variables in Table 4.3.2 and their interaction with log time (in months).

Analysis of Maximum Likelihood Estimates n = 273, dead = 103, censored = 170.			
Covariates	Df	Chi-square	P-value
TB site	1	0.3864	0.5342
Drug regimen	2	6.0092	0.0496
CD4 count category	2	7.3536	0.0253
BMI category	2	5.6844	0.0583
WHO stage	1	0.1412	0.7071
Functional status	2	5.6679	0.0588
TB site	1	2.4311	0.1189
Drug regimen	1	0.9383	0.3327
CD4 count category	1	3.8090	0.0510
BMI category	1	0.5444	0.4606
WHO stage	1	0.2654	0.6064
Functional status	1	0.6378	0.4245

N.B. Variables in the second group are interacted with the natural logarithm of survival times (in months).

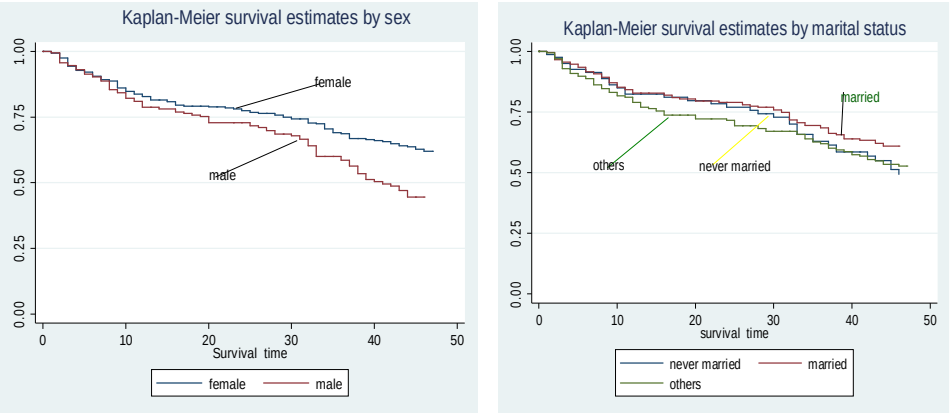
Table 4.4.2.1: The three highest differences in the parameter estimates of the variables included in the model in Table 4.3.2, when the data value for each patient is in turn deleted from the model.

Covariates	Deleted observation (i)	$\Delta_i\hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(-i)}$	$\lvert\Delta_i\hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(-i)}\rvert$
BMI $18.5<BMI<25$	84	0.0506086	0.0506086
	112	0.0496122	0.0496122
	218	-0.0538039	0.0538039
$BMI \geq 25$	81	0.07199035	0.07199035
	142	-0.02213151	0.02213151
	253	-0.03227359	0.03227359
CD4 count	133	0.0556519	0.0556519
	170	0.0564179	0.0564179

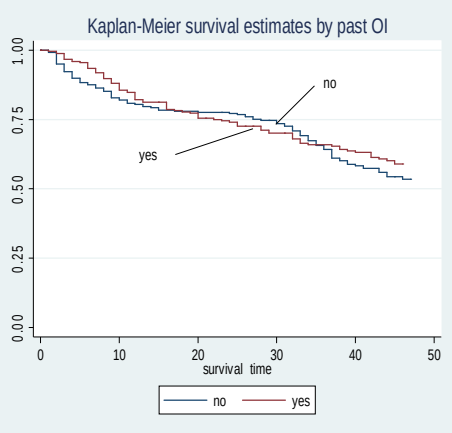
$CD4 \leq 200$	245	-0.0507357	0.0507357
$200 < CD4 \leq 350$	84	0.02490247	0.02490247
	112	0.0264319	0.0264319
	249	0.01827565	0.01827565
TB site EPTB	189	-0.0312713	0.0312713
	221	-0.033154	0.033154
	260	-0.038219	0.038219
Drug regimen AZT-based	29	0.0522285	0.0522285
	33	0.0585722	0.0585722
	53	0.054532	0.054532
TDF-based	43	0.0513594	0.0513594
	196	-0.0773575	0.0773575
	207	-0.067205	0.067205
WHO Stage IV	112	0.0511223	0.0511223
	207	-0.0585189	0.0585189
	241	-0.0647614	0.0647614
Functional status ambulatory	196	-0.05397	0.05397
	221	-0.0605718	0.0605718
	241	-0.0548348	0.0548348
Bedridden	49	0.0857192	0.0857192
	112	-0.0490256	0.0490256
	218	-0.01889965	0.01889965

Figure 4.2.3: Plots of Kaplan-Meier survivor functions, based on different factors, for HIV/TB co-infected patients treated under ART in Chagni Health Center, Chagni, 2013.

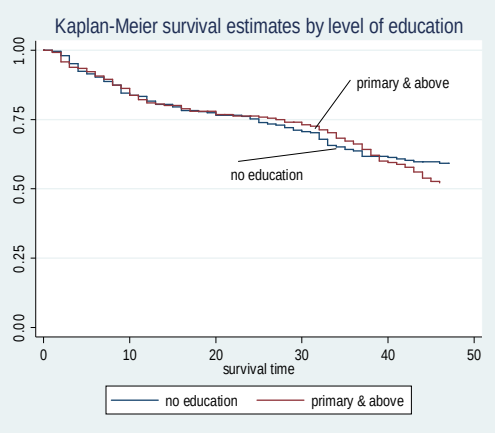
a) b)



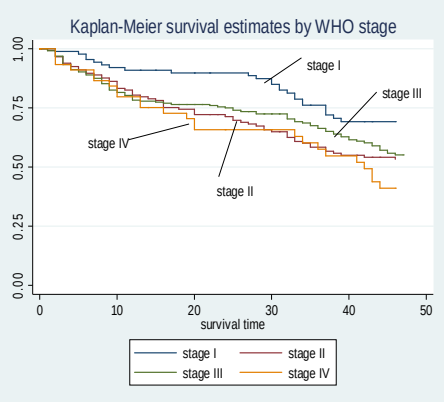
c)


Kaplan-Meier survival estimates by past OI
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
no
yes
no
yes

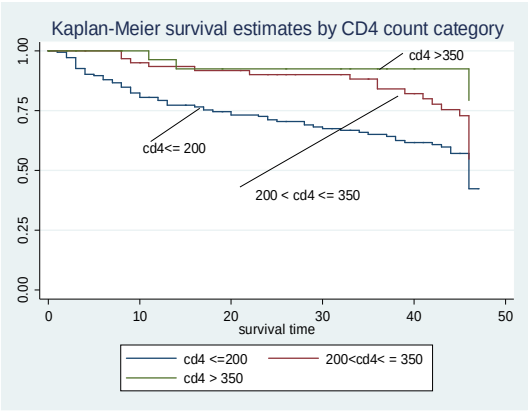
d)


Kaplan-Meier survival estimates by level of education
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
primary & above
no education
no education
primary & above

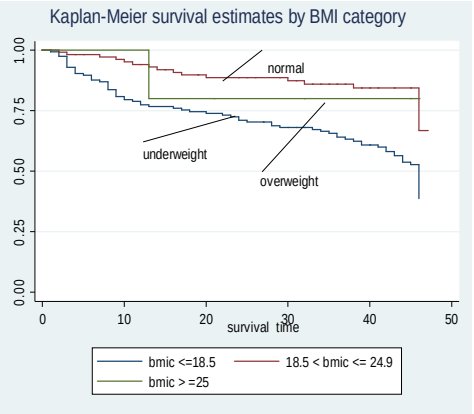
e)


Kaplan-Meier survival estimates by WHO stage
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
stage I
stage III
stage IV
stage II
stage I
stage II
stage III
stage IV

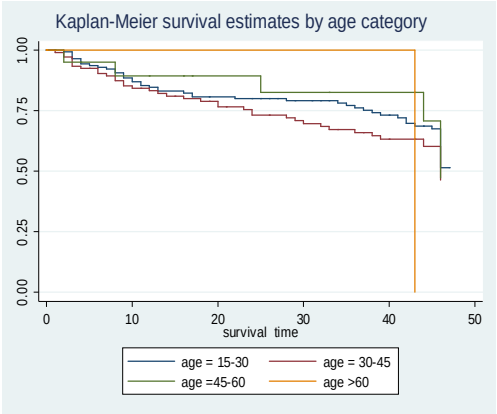
f)


Kaplan-Meier survival estimates by CD4 count category
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
cd4 >350
cd4<= 200
200 < cd4 <= 350
cd4 <=200
200<cd4< = 350
cd4 > 350

g)


Kaplan-Meier survival estimates by BMI category
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
normal
underweight
overweight
bmic <=18.5
18.5 < bmic <= 24.9
bmic > =25

h)


Kaplan-Meier survival estimates by age category
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
survival time
age = 15-30
age = 30-45
age =45-60
age >60

i)

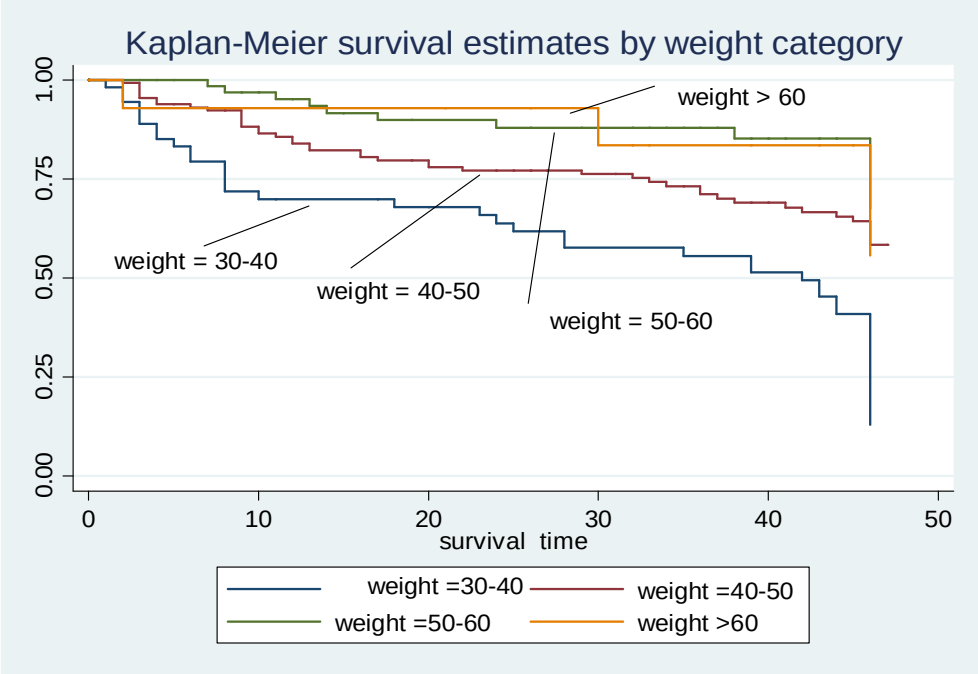
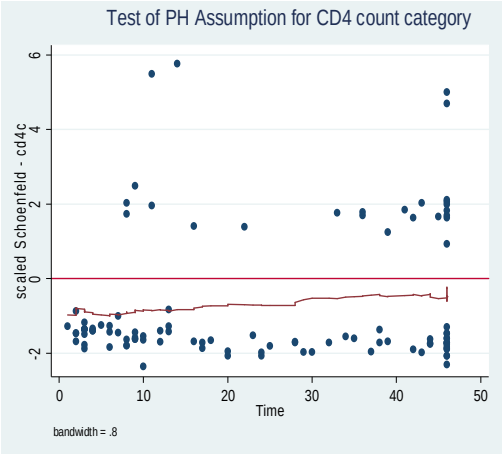
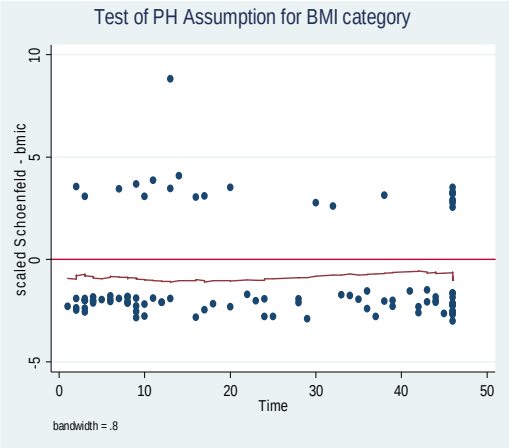


Figure 4.4.3.1(a – f): Graphs of the scaled Schoenfeld residuals and their LOESS smooth curves for the variables obtained from the multivariable PH Cox model in Table 4.3.2. The line that passes through zero is the reference line.

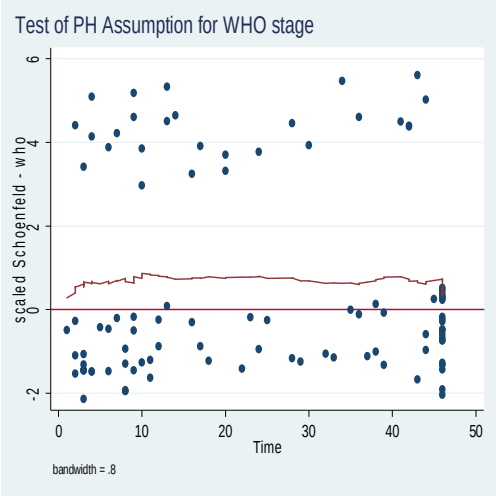
a)



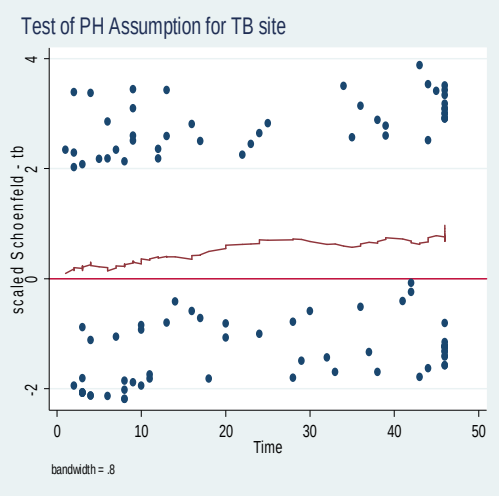
b)



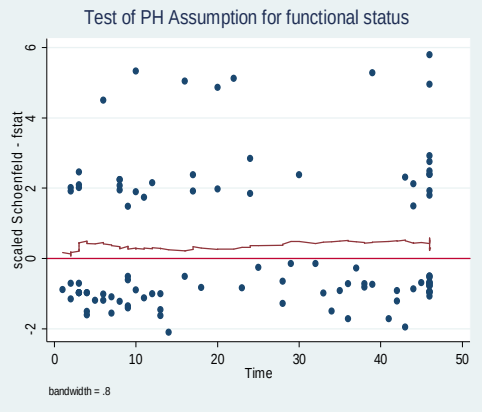
c)


Test of PH Assumption for WHO stage
scaled Schoenfeld - who
-2
0
2
4
6
0
10
20
30
40
50
Time
bandwidth = .8

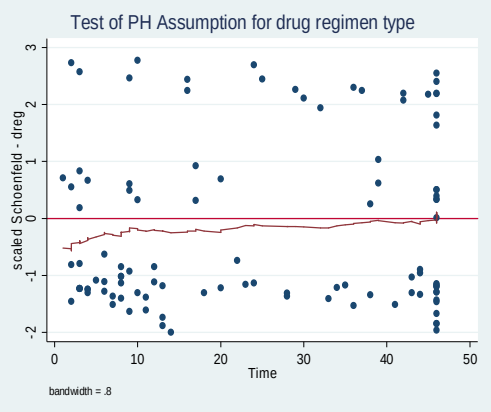
d)


Test of PH Assumption for TB site
scaled Schoenfeld - tb
-2
0
2
4
0
10
20
30
40
50
Time
bandwidth = .8

e)


Test of PH Assumption for functional status
scaled Schoenfeld - fstat
-2
0
2
4
6
0
10
20
30
40
50
Time
bandwidth = .8

f)


Test of PH Assumption for drug regimen type
scaled Schoenfeld - dreg
-2
-1
0
1
2
3
0
10
20
30
40
50
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: Belete Mulatu

Signature: _____

Date: _____

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

Name: Professor Eshetu Wencheko

Signature: _____

Date: _____