ADDIS ABABA UNIVERSITY

FACULTY OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

GRADUATE PROGRAMME

SURVIVAL OF TUBERCULOSIS PATIENTS TREATED UNDER DIRECTLY OBSERVED TREATMENT, SHORT COURSE IN ADDIS ABABA, ETHIOPIA

By **Abrham Keraleme**

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics

July, 2009

ADDIS ABEBA UNIVERSITY

FACULTY OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

SURVIVAL OF TUBERCULOSIS PATIENTS TREATED UNDER DIRECTLY OBSERVED TREATMENT, SHORT COURSE IN ADDIS ABABA, ETHIOPIA

By **Abrham Keraleme**

Approval by Board of Examiners

Ato Sileshi Fanta	_____
Dep't Head	**Signature**
_____	_____
Examiner	**Signature**
_____	_____
Examiner	**Signature**

Acknowledgement

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

I would like to express my gratitude to Dr. Fentaw Abegaz who helped me in adjusting the research topic and encouraged me to carry out this research work.

I would also like to express the great debt I owed to Seble Kebede and my family for providing financial, technical and moral support.

Furthermore, I want to thank Ato Taddele Cherinet, Ato Dejen Tesfaw, Ato Getahun Alem, Ato Zebene Ayele and Ato Yabebal Ayalew who have provided me help and moral support in one way or another to accomplish this study.

Finally, I would like to thank the Addis Ababa Health Bureau for allowing me to obtain data.

TABLE OF CONTENTS

List of Tables

List of Figures

Abbreviations

1. **DOTS** = Directly Observed Treatment, Short-course
2. **FMOH** = Federal Ministry of Health
3. **HIV** = Human Immunodeficiency Virus
4. **IP** = Intensive Phase
5. **LOWESS** = Locally Weighted Polynomial Regression
6. **MDR** = Multi-Drug Resistant
7. **NTLCP** = National Tuberculosis and Leprosy Control Program
8. **TB** = Tuberculosis
9. **WHO** = World Health Organization

Abstract

The objective of this study is to estimate survival probabilities and identify the risk factors for the death of tuberculosis patients in Addis Ababa during the treatment period. The data for this study are obtained from TB patients registered during September 2007 to January 2008 under DOTS at the Health Centers in Addis Ababa. The analytical methodologies used were the Kaplan-Meier to estimate the survival time and Cox's regression model was employed to identify the covariates that have a statistical significant effect on the survival longevity of TB patients. The estimation of the model parameters was done by partial maximum likelihood procedures. From the Cox regression model the factors Age, Initial weight, Category, HIV and interaction of Age by HIV were the risk factors for the death of TB patients. Furthermore it was found that the survival probabilities of TB patients with low initial weight, older age and HIV positive were low.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

1.1.1 Etiology and mode of transmission of tuberculosis

One of the most dangerous and killer diseases is Tuberculosis (TB). Tuberculosis is a bacterial disease caused mainly by Mycobacterium tuberculosis (M. tuberculosis). The mode of transmission of Mycobacterium species from person to person is well established. Virtually new infections with M. tuberculosis are acquired via airborne transmission. The sources of infections are persons with tuberculosis of the lung who are coughing and sneezing. Coughing and sneezing produces air droplets containing bacilli. Persons in the same household, or who are in frequent contact with an infectious patient have the greatest risk of being exposed to the bacilli (Murray and Lopez, 1996).

Once an individual is infected with the bacilli, he or she remains infected for many years, perhaps for life (Bass et al., 1990). A few infected persons develop clinically apparent disease within a few weeks after infection. Most, however, remain asymptomatic. However, infected persons can develop the disease at any time if they come under physical or emotional stress or if they become immune-compromised as occurs with human immunodeficiency virus (HIV) infection (Murray and Lopez, 1996).

1.1.2 Site of infection and laboratory diagnosis

Tuberculosis can affect virtually any tissue or organ but most commonly affects the lungs. As a result, tuberculosis is a disease with diverse clinical manifestations. Accordingly, the infections are broadly classified into pulmonary tuberculosis that accounts 85% of all tuberculosis cases and extra-pulmonary tuberculosis that compromises the remaining 15% of all tuberculosis cases. Pulmonary tuberculosis is further classified into smear-positive pulmonary tuberculosis that accounts for 75-80% of all pulmonary tuberculosis and smear-

negative pulmonary tuberculosis that accounts for about 20-25% of all pulmonary tuberculosis cases (FMOH, 2008).

Detection of the bacterium by direct microscope from sputum has been the main way for the diagnosis of tuberculosis. However smear positive rates among HIV-positive tuberculosis patients have markedly decreased and thus the most expensive and time taking culture technique is used for the diagnosis of tuberculosis (FMOH, 2002).

1.1.3 Target population (risk group)

Tuberculosis affects individuals of all ages of both sexes within every socio-economic group amongst the population. There are, however, groups which are more vulnerable to develop the disease. Poverty, malnutrition and crowded living conditions have been known for decades to increase the risk of developing the disease. Studies have indicated that poverty is associated with tuberculosis. The association with socio-economic status and tuberculosis arises in a variety of ways. Exposure is associated with crowding and quality of housing, which in turn may be associated with socio-economic status. Moreover, social mixing is associated with socio-economic status, perpetuating unequal disease distribution. Progression from infection to disease may depend on nutritional status and thus on poverty. The duration of infectiousness of sources of cases depends on access to adequate health care, which depends, in part, on socio-economic status.

Ethiopia is attacked by recurrent famine and widespread poverty that leads to severe malnutrition. Moreover, the health infrastructure is very poor. These social and economic factors are believed to aggravate the problem of tuberculosis in Ethiopia.

More recently, HIV infection has been identified as a major risk factor for developing tuberculosis. This is because infection with HIV destroys the immune defense mechanisms of the body and is, therefore, and important risk factor for the development of tuberculosis. It is estimated that 50 to 60% of HIV infected people will develop TB disease in their life time in contrast with HIV negative persons, whose life time risk is only 10% (FMOH, 2008).

The Human Immunodeficiency Virus (HIV) pandemic presents a massive challenge to the control of tuberculosis. The synergy between TB and HIV/ADIS is strong in high HIV prevalence populations. TB is a leading cause of morbidity and mortality, and HIV is fuelling the tuberculosis epidemic in Ethiopia. This unprecedented scale of the epidemic of HIV related tuberculosis demands concerted and urgent action.

HIV increases susceptibility to infection with M. tuberculosis, the risk of progression to TB disease, and the incidence and prevalence of TB. It also increases the likelihood of re-infections and relapses of TB (FMOH, 2008).

1.1.4 Global and National Morbidity and Mortality Burden of Tuberculosis

1.1.4.1 Global Burden of Tuberculosis

Recent evidences tend to demonstrate that TB prevalence and TB death rates are globally decreasing after having reached a peak. Since 2005, the TB incidence rate is in decline in all six WHO regions (FMOH, 2008).

In developing countries like Ethiopia tuberculosis continues to be a major health problem. This is because unlike in developed countries, HIV infection will result in a considerable increase of tuberculosis cases in those developing countries where both tuberculosis and HIV infection is prevalent (see below). The global burden of cases has been rising by 0.4 percent per year, but this overall global trend hides much faster increase in Sub Sahara Africa (Grimard and Harling, 2003). Tuberculosis remains, therefore, one of the top leading infectious diseases resulting in high morbidity and mortality.

Over 95% new tuberculosis cases and deaths occur in developing countries. The highest incidence and number of deaths occur in Asia and Sub-Saharan (Dye et al. 1999). Tuberculosis is the leading cause of death among persons with HIV infections, accounting for a third of AIDS related deaths worldwide (WHO, 1998).

1.1.4.2 National Burden of tuberculosis

According to the 2007 WHO estimates, the incidence of TB of all forms and smear positive TB stand at 341 and 152 per 100,000 population, respectively. The prevalence and mortality of TB of all forms is estimated to be 546 and 73 per 100,000 populations, respectively. In the year 2006/7 Ethiopia registered 129,743 cases of TB. According to latest estimates, Ethiopia stands 7th in the list of high burden countries for TB (FMOH, 2008).

Ethiopia is one of the highly affected countries by the TB/HIV co-epidemic. The WHO global report 2008 estimates that in Ethiopia 40% of TB patients tested for HIV are HIV positive, while routine data from 2006/07 estimates that 31% of TB patients are HIV positive (FMOH, 2008).

In 1992 a standardized TB prevention and control program, incorporating Directly Observed Treatment Short-course (DOTS), was started as a pilot survey in Arsi and Bale Zones, Oromia Region. The DOTS strategy has been subsequently scaled up in the country and implemented at national level. Currently the DOTS geographic coverage reaches 90% whereas the health facility coverage is 75% (FMOH, 2008).

1.2 Statement of the problem

Tuberculosis is a major public health problem throughout the world. According to the WHO Global Report 2007, one-third of the world's population is estimated to be infected with tubercle bacilli and hence at risk of developing active disease.

Currently, approximately one third of the world's population (1.9 billion) is infected with M. tuberculosis; the bacterium that causes tuberculosis. There are an estimated 8.8 million new tuberculosis cases and almost 1.6 million deaths from tuberculosis, including 195,000 patients infected with HIV. The global case fatality rate is 23 percent, but exceeds 50% in some African countries with HIV (Dye et al.1999). Among women of the reproductive age tuberculosis is the leading cause of death, surpassing all cases of maternal mortality (WHO, 1999).

According to the MOH hospital statistics data, tuberculosis is the leading cause of morbidity, the third cause of hospital admission (after deliveries and malaria), and the second cause of death in Ethiopia, after malaria. So, what are the factors that facilitate the death of TB patients? The most probable explanation might be HIV, localization of tuberculosis (pulmonary and extra pulmonary), history of previous treatment, Age, Marital Status, Category of patients (**Category I** is defined as all new cases that are sputum smear positive or seriously ill patients with smear negative or extra-pulmonary diseases, **Category II** patients are cases who have had previous anti-tuberculosis treatment and **Category III r**egime is prescribed to new patients who are smear- negative or extra-pulmonary TB and are not seriously ill) etc. It is, therefore, important to examine the risk factors for the survival of TB patients. This study attempts to identify the major risk factors.

1.3 Objectives

The general and specific objectives of this research paper are as follows:

1.3.1 General objectives

Based on data collected from Addis Ababa:

- To estimate survival time probabilities of TB patients during treatment period.
- To identify risk factors for death of TB patients during treatment period.
- To develop a statistical model that predicts the survival probability of TB patients.
- To investigate factors that jointly determines the death of TB patients.
- To estimate survival time probabilities of TB and HIV/ADIS patients during treatment period.

1.3.2 Specific objective

- To test the proportional hazard assumption in a survival model with log-minus log plots.
- To compare the survival probabilities of TB patients with respect to their categories.
- To compare the survival probabilities or TB patients with HIV (+ve and -ve) with respect to their categories.
- To assess the efficiency of DOTS strategy in TB controls.

1.4 Significance of the study

i. The results of this study might provide information to government and other concerned bodies in setting policies, strategies and further investigation for reducing death to TB patient.
ii. The results help donors and government to understand risk factors that influence the death of TB patients.
iii. The study could provide base-line data for detail and further studies in the future.

1.5 Limitation of the study

- The study has different limitations the major limitation of the study goes with the problems related to the use of secondary data. As different literature pointed out, there are different factors that are assumed to have impacts on the survival of TB patients such as alcoholism and smoking status. However, we did not get data on these variables to include in the analysis. This may make the study some what incomplete.

CHAPTER TWO

Literature Review

2.1 Literature related to the variables used in the study

A number of researches had been undertaken to identify the risk factors for death of tuberculosis patients. In the study of defaulting in three districts of Arsi Zone in Ethiopia Tekle et al. (2004) found that 11.3% of all tuberculosis patients in Ethiopia in 1997 - 99 were noncompliant from DOTS therapy. The adequate knowledge about the disease was found that minimizes defaulting from therapy in Ethiopia 2002. Toxicity of medication is also thought to be associated with non-compliance.

Tekle et al. had revealed in their study that medication side effects were significantly associated with defaulting. It was observed that defaulting rate in the case of those with adequate knowledge about the disease was less and as a result helps in reducing the toxicity. The second risk factor for noncompliance was the DOT providers.

In 1997 there were an estimated eight million new cases of tuberculosis world-wide and 16.2 million prevalent cases (Dye et al. 1999). Furthermore, tuberculosis is the largest single infectious cause of death among adults in the world, accounting for about two million deaths per year (Dye et al. 1999). Ninety five percent of cases and deaths occur in developing countries and tuberculosis accounts for approximately seven percent of all deaths in developing countries (Murray et al. 1990).

Erhabor et al. (2006) show that the age of patients in Osun State, Nigeria, ranged from 9 to79 years, the highest mortality was recorded between the 3rd and 4th decades of life with a male to female ratio 1.3: 1. Mortality was also high among married patients. The majority of patients died due to pulmonary TB. Human Immunodeficiency Virus (HIV) topped the list with 16.3% as compared to other factors.

A study conducted in Sao Paulo, Brazil showed that of all tuberculosis deaths identified, 78% had pulmonary form. The median age was 51 years; 52% had up to four years of schooling; 4% were probably living in the streets. Mortality rate increased with age; it was 5 cases in 10,000 for the entire city, ranging between 0 and 35 according to the district. Previous treatment was reported for 82 out of 232 patients, and of them, 41 defaulted treatments. Diabetes (16%), chronic obstructive pulmonary disease (19%), HIV infection (11%), smoking (71%), and alcohol abuse (64%) were also reported as a cause of death (Waldman et al. 2002).

In a study from the Netherlands survival probabilities for tuberculosis patients were estimated as 95% after six months of treatment and 6% died within one year while on treatment. The study identified the following significant factors for mortality of tuberculosis patients: male sex, age (>65 years), presence of malignancy, human immune-deficiency virus (HIV) infection, addiction to alcohol or drugs, localization of tuberculosis (pulmonary and extra pulmonary tuberculosis) and type of medical officer having made the diagnosis (specialist internal medicine) (Borgdorff et al. 1998).

A report from South India showed that 39 (6%) of 676 TB patients died during treatment period and higher death rates were independently associated with weight <35 kgs and history of previous treatment. The study recommended that possible role of nutritional interventions should be explored among underweight patients to reduce mortality (Santha et al. 2000).

A study conducted in Russia reported 183 (9.6%) deaths among 1916 TB patients. Old age, history of previous treatment for TB, multi-drug resistance TB and alcoholism have significance effect for death during treatment period (Mathew et al. 2006).

A study conducted in Chennai city, India, showed that 91% patients treated for TB survived the entire follow-up period of 600 days from the date of start of treatment and 9% patients died during the follow-up period. The risk factors identified for TB mortality were young age, male sex, smear positivity, treatment default, treatment failure and the combination of smoking and alcoholism (Kolappan et al. 2006).

In the study done by Vasantha et al. (2008) in Tiruvallur district, South India, the following results were obtained. Using multivariate analysis (Cox proportional hazard model), the

variables considered were sex, age, category, education, occupation, body weight at initiation of treatment (<35 kg), history of previous anti TB treatment, smoking and drinking habits, type of DOT providers, whether patient took treatment under supervision in intensive phase and continuation phase. Age (_≥45 years), previous history of treatment, alcoholism and body weight at initiation of treatment (<35 kg) were found to be risk factors for death during the treatment period. The other factors, namely sex, category, education, occupation, smoking, type of DOT providers and supervision under IP were not found to be significant for risk of death during treatment period.

Lefebvre and Falzon (2008) identify risk factors for death among TB patients in 15 European countries. Austria, Belgium, the Czech Republic, Denmark, Estonia, Germany, Ireland, Latvia, Lithuania, the Netherlands, Portugal, Slovakia, Slovenia, Sweden and the UK satisfied the selection criteria for the study. In the period 2002–2004, 82,314 TB cases were notified in the 15 countries, of whom 64% were smear-positive and 56% had drug susceptibility test results available; 39,566 (48%) cases with complete data for the variables of interest were included in the analysis. Patients with a previous history of TB were more likely to have MDR (Multi-Drug resistant). Overall, 3,085 (7.8%) cases died (ranging 5–12% between countries). The deaths were significantly associated with male sex, increasing age, European native, pulmonary localization and MDR, with a risk of death greater for secondary MDR than for primary MDR.

2.2 The model used in the study

The origin of survival analysis goes back to the time when mortality tables were introduced. Life tables are one of the oldest statistical techniques and are extensively used by medical statisticians and by actuaries. Yet relatively little has been written about their formal statistical theory. Kaplan and Meier (1958) gave a comprehensive review of earlier work and many new results. D. R. Cox (1972) was largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and more generally to the incorporation of regression-like arguments into life table analysis.

As the use of survival analysis grew, parametric models gave way to non-parametric and semi-parametric approaches for their appeal in dealing with the ever growing field of clinical trials in medical research.

Survival models have the capability of handling censored data. Cox (1972), Cox and Oakes (1984), Kalbflesch and Prentice (1980), Miller (1981) used survival analysis in modeling human lifetimes. Allison (1984), Blossfeld, Homerie, and Mayer (1989), Heckman and Singer (1985), Tuma and Hannan (1984) used survival analysis in economics and sociology. Several recent articles also described how survival models can be used to explore specific topics including social interaction (Allison and Liker (1982), Gardner and Griffin (1989)), organizational behavior (Fichman (1988) Morita, Lee, and Mowday (1989)), clinical trials (Greenhouse, Strangle, and Bromberg (1989)) and the life course (Johnson, (1988); Teachman, (1982)).

Fergusson, Horwood, and Shannon (1984) used hazard functions to study the time to marital breakdown after the birth of child. Hazard functions had been also used in studies of time to shift in attentions in classroom (Felmlee and Eder (1983); Felmlee, Eder and Tsui (1985)), time to change decision in the face of irrelevant information from a low-status partner (Hembroff and Myers(1984)), in study of relapse of mental illness (Lavori et al., (1984)), marital dissolutions (Morgan, Lye, and Condran (1988); Tuma, Hannan, and Groeneveld (1977)) and human lifetimes (Gross and Clark (1975)).

Proportional hazards modeling is the most frequently use type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (Stevens and Hollis (1989)), affective disorders (Shapiro et al. (1989)), childhood family breakdown (Fergusson, Horwood, and Dimond (1985); Fergusson, Horwood, and Shannon (1984)), interruptions in conversation (Dress (1986)), and employee turnover (Morita, et al. (1989), and in medical areas for identification of important covariates that have as significant impact on the response of the interested variables .

Vasantha et al. (2008) use proportional hazards model to examine risk factors for the survival of tuberculosis patients. Mathew et al. (2006) also use proportional hazards model and other non-parametric models (Kaplan-Meier) to examine causes of death during tuberculosis treatment in Tomsk Oblast, Russia.

Data in TB patients include censored data and therefore are not compatible with standard statistical models. This study uses the proportional hazards model to identify risk factors that are assumed to have influence on the survival longevity of tuberculosis patients.

CHAPTER THREE

Data and Methodology

3.1 Introduction

The data for this study are obtained from tuberculosis unit in Addis Ababa, capital city of Ethiopia which consists of 10 sub-cities divided into 99 kebeles. The DOTS strategy has been implemented in this area since 1994. There are 24 governmental health centers (HCs) participating in the program. During September 11, 2007- January 8, 2008, about 800 TB patients were registered for the treatment under DOTS at the HCs in Addis Ababa.

All the patients diagnosed with tuberculosis at one of these health centers (HCs) are given DOT in accordance with National Tuberculosis and Leprosy Control Programme (NTLCP) policies. Every dose of treatment is to be directly observed during the intensive phase (IP). This phase consists of three or more drugs for the first 8 weeks for new cases and 12 weeks for re-treatment cases and at least two drugs to be taken for 4-6 months during continuation phase. The drug must be collected every month and self administered by the patient, except for re-treatment cases.

3.2 Data

The data used in this study are collected from 637 patients. All patients registered for treatment within the indicated four months (except cases with missing values) in 24 Health Centers on socio-economic and demographic profiles, category, history of previous treatment and body weight at initiation of treatment from TB register maintained by health center TB unit were considered. Marital status, HIV and localization of tuberculosis were collected from Provider Initiated HIV Counseling and Testing Log Book.

The anti-tuberculosis regimens used for Category I and III patients were 2ERHZ/6EH and for Category II (Re-treatment regimen) patients was 2S(ERHZ) / 1(ERHZ) / 5E3$(RH)_3$ (H = isoniazid; R = rifampicin; Z = pyrazinamide; E = ethambutol; S = streptomycin. Numbers

before the acronyms indicate the duration of the treatment phase in months and numbers in subscript (for Category II only) indicate the number of times the drug is given each week whereas the drug is given daily for Category I and III patients). Treatment for category I and II patients was extended by another month if the sputum smear remained positive at the end of IP.

3.3 Definition of variables

The response (dependent) variable is continuous and describes the length of treatment time in days. The explanatory (independent) variables of interest in this analysis include socio-economic, demographic, and characteristics of disease and treatment profiles.

3.3.1 The Response Variable

The response variable for the i^{th} individual is represented by Y_i and it measures duration to event and it is defined by status variable (event or censoring variable). This is to say that the response variable is a censored survival time represented by variable time and event/death. Survival time is measures the follow-up of time from a defined starting point to the occurrence of a given event. This observation time has two components, the beginning point of the study time and the observation of time to the end. In survival analysis, the out come of interest (death in this study) is the duration of time until death occurs.

3.3.2 Predictor Variables

The predictor variables in survival data analysis are called covariates. These covariates can be categorical or continuous. In the development of the model we should also establish if the covariate is time dependent or not. Such cases affect how the covariates would be modeled in Cox proportional hazard procedure. The predictor

(covariate) variables which are assumed to influence the survival of TB patients included in the model are:

- Sex
- Age
- Initial Weight(kg)
- Localization of tuberculosis (pulmonary and extra pulmonary tuberculosis)
- Category
- Marital status
- HIV status
- History of Previous treatment

Table 3.1: Description of Variables included in the Analysis

❖ **Response Variable**

Variable	**Representation of variable**	**Category**
Length of treatment Time	Y	Continuous

❖ **Predictor Variables**

1. Demographic variables

Variables	**Representation of variable**	**Categories**
Gender (Sex)	X_1	0=Male, 1=Female
Age (Age)	X_2	continuous

2 Socio- economic Variables

Variables	**Representation of variable**	**Categories**
Marital Status (Mstat)	X_3	0=Single, 1=Others, 2=Married
Initial Weight (Intwt)	X_4	continuous
HIV Status (HIVs)	X_5	0= HIV Negative, 1= HIV Positive,

3 Treatment Variables

Variables	**Representation of variable**	**Categories**
Localization of TB (TBtype)	X_6	0=Extra Pulmonary, 1=Pulmonary
Category (Cat)	X_7	0=Category III, 1=Category II, 2=Category I
History of Previous treatment	X_8	0=New, 1=Others

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

- **Marital Status:**- this categorization is based on the usual, Single, 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.
- **History of Previous Treatment:-** This categorization is based on FMOH (2008) manual and different literature. Since the number of patients who defaulted or failed is too small to be categorized separately, such patients are included in the "Others" category.

3.4 Methods for data Analysis

3.4.1 Survival Data Analysis

Survival analysis is defined as a branch of statistics which deals with data related to time to an event. This topic is also called reliability analysis in engineering and duration analysis in economics or sociology.

The term survival analysis applies to techniques in which the data being analyzed represent the time it takes for a certain event to occur. The use of survival analysis, as opposed to the use of different statistical methods, is most important when there is no time-to-event record. In reality such situation can occur due to the following reasons:

- When an individual survive beyond the study period or the individual does not experience the event.
- Lost to follow-up, that is, an individual may drop out, transfer to other place, etc.
- Deaths due to other causes different from that/those specified in the study.

Therefore, survival data are almost always incomplete. The statistical terminology for such data is censoring. Censoring is common in survival analysis and it is considered as an important feature of survival data. Survival analysis is well suited to for such data which are very common in medical research since studies in medical areas have a special feature that follow-up studies could start at a certain observation time and could end before all experimental units had experienced an event. The most common encountered form of a censored observation is one in which observation begins at the defined time, say t=0, and terminates before the outcome of interest is observed. Since the incomplete nature of the observation occurs in the right tail of the time axis, such observations are said to be right censoring. The other mechanism that can lead to incomplete observation of time is truncation. A truncated observation is one which is incomplete due to a selection process inherent in the study design.

There are obviously many potential life models that overcome such incomplete observations. In some situations there may be reasons to select a particular family of models; the model my fit data on hand well, past experience may have shown the model to give a good description

of lifetime distribution from similar populations, there may be a knowledge of the underlying aging or failure process that suggests the validity of the model, and so on. In situations in which no family of models is singled out as being particular appropriate, the choice of the model is frequently made on the basis of considerations such as: the convenience of mathematically handling the model, the statistical methods available in connection with the model and the degree of complication of calculations involved in using the model.

Three additional points should be mentioned in connection with the choice of the model. Firstly, for any chosen particular model it has to fit the available data upon appropriate tests. Second, one should be aware of the consequences of departures from the assumed model on inferences made. Finally, although there is a lot of work based on lifetime distributions of parametric models, there are many situations in which it is desirable to avoid strong assumptions about the model. Nonparametric or distribution-free procedures are important in this case (Lawless, 1982).

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

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

3.4.1.1 Descriptive Methods for Survival data

This method is especially important if individuals are homogeneous at least within groups. In such situation it is appropriate to use the Kaplan-Meier survival estimator.

3.4.1.1.1 Kaplan-Meier estimator

The Kaplan-Meier estimator of the survivorship function [Kaplan and Meier (1958)], also called the product limit estimator, is probably the most popular approach because it :

i. is used by most software packages.
ii. incorporates information from all of the observations available, both uncensored and censored
iii. is based on individual observations, so it is more precise than the life table estimator

The Kaplan-Meier estimator of the survivorship function (or survival probability) $S(t) = P(T \geq t)$ is defined as:

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i} = \prod_{t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{3.1}$$

with the convention that

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

where

$t_{(1)},...,t_{(m)}$ is the set of m distinct death times observed in the sample

d_i is the number of deaths at $t_{(i)}$

n_i is the number of individuals "at risk" right before $t_{(i)}$.

3.4.1.2 Comparison of Survivorship Functions

When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. The figure in general shows if the pattern of one survivorship function lying above another which 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. Now the statistical question is whether the observed difference seen in the figure is significant. The general form of this test statistic is given by

$$Q = \frac{\left[\sum_{i=1}^{m} w_i \left(d_{1i} - \hat{e}_{1i}\right)\right]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \qquad (3.2)$$

In this expression

$$\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i} \quad \text{and} \qquad (3.3)$$

$$\hat{v}_{1i} = \frac{n_{1i} n_{oi} d_i \left(n_i - d_i\right)}{n_i^2 \left(n_i - 1\right)}$$

n_{oi} is the number at risk at observed survival time $t_{(i)}$ in group 0

n_{1i} is the number at risk at observed survival time $t_{(i)}$ in the group 1

d_{oi} is the number of observed deaths in group 0

d_{1i} is the number of observed deaths in group 1

n_i is the total number of individuals or risk before time $t_{(i)}$

d_i is the total number of deaths at $t_{(i)}$

The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times.

Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of group, and that the total number of observed events and the sum of the expected number of events is large, Q follows a chi-square distribution with one degree of freedom. We can also use the above test to compare k groups. In this study we use the log rank test and generalized Wilcoxon test which are special cases of Q.

3.4.1.2.1 The Cochran-Mantel-Haenszel Log rank test

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

Therefore, the log rank test statistic becomes

$$Q_{LR} = \frac{\left[\sum_{i=1}^{m}\left(d_{1i} - \hat{e}_{1i}\right)\right]^2}{\sum_{i=1}^{m}\hat{v}_{1i}} \qquad (3.4)$$

3.4.1.2.2 The Generalized Wilcoxon test

Gehan (1965) and Breslow (1970) generalized the Wilcoxon rank sum test to allow for censored data. This test uses weights equal to the number of subjects at risk at each survival time, i.e. $w_i = n_i$, and is called Wilcoxon or generalized Wilcoxon test in most software packages. Thus the Wilcoxon test can be defined as

$$Q_{GWt} = \frac{\left[\sum_{i=1}^{m} n_i\left(d_{1i} - \hat{e}_{1i}\right)\right]^2}{\sum_{i=1}^{m} n_i^2\hat{v}_{1i}} \qquad (3.5)$$

3.4.1.3 Regression Models for Survival Data

One of the most popular types of regression models used in survival analysis is the Cox proportional hazard model (Cox, 1972).

3.4.1.3.1 The Cox Proportional Hazards Regression Model

The Cox Proportional Hazard Model is a multiple regression method used to evaluate the effect of multiple covariates on the survival.

Cox (1972) proposed a semi-parametric model for the hazard function that allows the addition of covariates, while keeping the baseline hazards unspecified and can take only positive values. With this parameterization the Cox hazard function is

$$\lambda\left(t, X, \beta\right) = \lambda_o\left(t\right)e^{\beta' X} \qquad (3.6)$$

where

$\lambda_o(t)$ is the baseline hazard function that characterizes how the hazard function changes as a function of survival time,

$\lambda(t, X, \beta)$ represents the hazard function at time t with covariates X= $(X_1, ..., X_p)'$,

$\beta = (\beta_{1,...,}\beta_p)'$ is a column vector of p regression parameters,

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

t is the failure time.

The survival time of each member of the sample is assumed to follow its own hazard function. In such a case, the above model can equivalently be written as

$$\lambda_i(t, x_i, \beta) = \lambda_o(t)\exp(\beta_1 x_{i1} + ... + \beta_p x_{ip}) \qquad (3.7)$$

$i = 1,..., n,$ where n is total number of observations in the study.

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

The proportional hazards estimation method computes a coefficient for each predictor variable that indicates the direction and degree of flexing that the predictor has on survival.
The proportional hazard model is the most popular regression method for analysis of censored survival data. The popularity is because:

- It allows flexible choice of covariates (we can accommodate time varying, time independent, continuous and discrete covariates).
- It is fairly easy to fit.
- Standard software packages are programmed to handle proportional hazards model such as SPSS, SAS, STATA, S-PLUS etc.
- Does not make any assumption about the underlying survival distribution (does not require the knowledge of the shape of the survival distribution).
- Does not require estimation of the baseline hazards rate, λ_0 (t), to estimate the regression parameters.

The Cox proportional hazard model is formulated as the hazard function which measures the risk to death or rate of failure at time t.

3.4.1.3.1.1 Assumption of Cox proportional hazard model

(1) The baseline hazard , $\lambda_o(t)$ depends on *t*, but not on covariates $x_1,...,x_p$

(2) The hazard ratio, i.e., $e^{\beta' x}$, depends on the covariates $X = (x_1,...,x_p)'$, but not on time *t*.

(3) The covariates x_i do not depend on time *t*.

Assumption (2) is what led us to call this a proportional hazards model. To express this mathematically, consider two distinct values of the covariate X, say, x_1 and x_2.

$$\lambda(t, x, \beta) = \lambda_o(t) e^{\beta' X}$$

Then, the hazard ratio becomes:

$$\frac{\lambda(t, x_1, \beta)}{\lambda(t, x_2, \beta)} = \frac{\lambda_o(t) e^{\beta' x_1}}{\lambda_o(t) e^{\beta' x_2}}$$

$$= \frac{e^{\beta' x_1}}{e^{\beta' x_2}} \qquad (3.8)$$

$$= e^{\beta'(x_1 - x_2)}$$

is independent of time t.

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

3.4.1.3.1.2 Parameter Estimation

In Cox proportional hazards model we can estimate the vector of parameters β without having any assumptions about the baseline hazard $\lambda_o(t)$. As a consequence, this model is more flexible and an estimate of the parameters can be obtained easily.

Consider n independent individuals, the data that we need for the Cox proportional hazard model is represented by (t_i, δ_i, x_i) $i = 1,2,...,n$,

where t_i = the survival time for the i^{th} individual

δ_i = an indicator of censoring for the i^{th} individual given by 0 for censored and 1 for event/death

x_i = a vector of covariates for individual i $(x_{i1}, x_{i2},..., x_{ip})$.

The full likelihood for right censored data can be constructed as

$$L(\beta)=\prod_{i=1}^{n}\lambda(t_i,x_i,\beta\ )^{\delta_i}S(t_i,x_i,\beta\ ) \qquad (3.9)$$

where $\lambda(t_i,x_i,\beta\ )=\lambda_o(t_i)e^{\beta' x_i}$ is the hazard function for individual i.

$S(t_i,x_i,\beta\ )=(S_o(t_i))^{\exp(\beta' x_i)}$ is the survival function for individual i.

It follows that $L(\beta)=\prod_{i=1}^{n}(\lambda_o(t_i)\exp(\beta' x_i))^{\delta_i}(S_o(t_i))^{\exp(\beta' x_i)}$ (3.10)

The full maximum likelihood estimator of β can be obtained by differentiating the right hand side of equation (3.10) with respect to the components of β and the base line hazard $\lambda_o(t)$. This implies that unless we explicitly specify the base line hazard, $\lambda_o(t)$, we cannot obtain the maximum likelihood estimators for the full likelihood.

To avoid the specification of the base line hazard, Cox (1972) proposed a partial likelihood approach that treats the baseline hazard as a nuisance parameter and removes it from the estimating equation.

Partial likelihood

Instead of constructing a full likelihood, we consider the probability that an individual experiences an event at time t_i given that an event occurred at that time.

Let R_i denote the set of individuals at risk at time just prior to $t_{(i)}$. Assume that for the present case there is only one failure at time t_i, i.e, no ties. The probability that individual i with covariates x_i is the one who experience the event at time $t_{(i)}$.

=P(individual i has experiences an event at time $t_{(i)}$ | one event at time $t_{(i)}$)

$$\frac{\lambda(t,x_i)}{\sum_{j\in R_{t(i)}}\lambda(t,x_j)} \qquad (3.11)$$

And under the proportional hazards assumption on using equation (3.1), the ratio

$$\frac{\lambda_o(t)\exp\left(\beta' x_i\right)}{\sum_{j\varepsilon R_{t(i)}}\lambda_o(t)\exp\left(\beta' x_j\right)} \qquad (3.12)$$

shows the contribution to the partial likelihood at each death time $t_{(i)}$ by the individuals with covariate x_i in the risk set $R_{t(i)}$. Where $R_{t(i)}$ is the overall subjects in the risk set at time $t_{(i)}$.

By eliminating the base line hazards function, in the numerator and denominator, equation (3.12) becomes

$$\frac{\exp\left(\beta' x_i\right)}{\sum_{j\varepsilon R_{t(i)}}\exp\left(\beta' x_j\right)} \qquad (3.13)$$

Thus the partial likelihood is the product over all failure time $t_{(i)}$ for $i = 1,2,...,m$ of the conditional probability (3.13) to give the partial likelihood

$$L_p(\beta) = \prod_{i=1}^{m}\frac{\exp\left(\beta' x_i\right)}{\sum_{j\varepsilon R_{t(i)}}\exp\left(\beta' x_j\right)} \qquad (3.14)$$

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

$$l_p(\beta) = \sum_{i=1}^{m}\left[\beta' x_i - \ln\left(\sum_{j\varepsilon R_{t(i)}}\exp\left(\beta' x_j\right)\right)\right] \qquad (3.15)$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of (3.15) with respect to the component of β, setting the derivative equal to zero and solving for the unknown parameters.

The partial likelihood derived above is valid when there are no ties in the data set. But in most real situations tied survival times are more likely to occur. In addition to the possibility of more than one death at a time, there might also be more than one censored observations at a time of death.

To handle this real-world fact, partial likelihood algorithms have been adopted to handle ties. There are three approaches in common to estimate regression parameters when there are ties. The most popular and easy approach is Breslow's approximation.

The Breslow approximation

This approximation is proposed by Breslow and Peto by modifying the partial likelihood takes the following form

$$L_{\mathrm{B}}(\beta)=\prod_{i=1}^{m}\frac{\exp\left(\beta' s_i\right)}{\left[\sum_{l\varepsilon R_{t(i)}}\exp\left(\beta' x_l\right)\right]^{d_i}} \qquad (3.16)$$

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

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

Now the partial log likelihood of (3.16) is given as

$$l_{\mathrm{B}}(\beta)=\sum_{i=1}^{m}\left[\beta' s_i - d_i \ln\sum_{l\varepsilon R_{t(i)}}\exp\left(\beta' x_l\right)\right] \qquad (3.17)$$

We obtain the Breslow maximum partial likelihood estimator, adjusted for tied observation, by differentiating equation (3.17) with respect to the component of β and setting the derivative equal to zero and solving for the unknown parameters.

3.4.1.3.1.3 Interpretation of the coefficients of the Cox-regression model

The estimated coefficients for the predictor variables represent the slope or rate of change of a function of the outcome variable per unit of change in the predictor variable by keeping the remaining predictor variables fixed (Hosmer-Lemeshow, 1989). Thus interpretation involves two issues, determining the functional relationship between the outcome variable and the covariate and appropriately defining the unit of change for the predictor variable (Hosmer-Lemeshow, 1989).

The estimated regression coefficients $\hat{\beta}_i^{'}s$ reflect linear and non-linear relationships and they will be interpreted as the change in the log-hazards ratio for every unit increase/decrease, depending on the variable change in x_i, holding other predictors constant. For example, for a dichotomous covariate with value 1 and 0, the hazard ratios of being in the category of interest for the j[th] subject, becomes $\frac{\lambda_o(t)\exp(\hat{\beta}_i * 1)}{\lambda_o(t)\exp(\hat{\beta}_i * 0)} = \exp(\hat{\beta}_i)$, fixing the other covariates constant. It is interpreted as the hazard rate, or rate of death in our case, among subjects with i^{th} covariate value equals 1 is $\exp(\hat{\beta}_i)$ time higher than subjects with i^{th} covariate value equals zero, $i = 1,2,..., p$ and $j = 1,2,..., n$. For covariates having L levels (L>2), similarly interpretations can made by taking one of the L-levels as a reference category.

3.4.1.3.1.4 Model development

3.4.1.3.1.4.1 Selection of covariates

There are three kind of method selection of best subset of the covariates, i.e. purposeful selection, stepwise selection, best subset selection. But for this particular study purposeful selection is used.

- **Purposeful selection of covariates**

  Step1: it begins by including covariates that are statistically significant
  (p-value < 20-25% modest level of significant).
  Step2: include covariates that are considered more important.
  Step3: use these covariates (those selected out step1 and 2).
  Step4: select covariates that are statistically significant (p-value<0.05) using Wald test.
  Step5: retain some covariates that are “important” even though they are not significant and check also for potential confounders.
  Step6: in addition to these for deleted (remove) covariates check whether their contribution is statistically significant or not using particularly likelihood ratio test.
  This selection process provides preliminary subsets of covariates. These are considered as main effects.

3.4.1.3.1.4.2 Checking for linearity of continuous covariates

To check the linearity of continuous covariates we plot hazards against the midpoints of the class and also using plots of martingale residuals. But for this particular study the plot of martingale residuals is used.

- **Plot of martingale residual**

The martingale residual are defined as

$$\hat{M}_i = \delta_i - \hat{H}\left(t_i, x_i^{(-)}, \hat{\beta}^{(-)}\right) \qquad (3.18)$$

Where $x_i^{(-)}$ and $\hat{\beta}^{(-)}$ refer to the set of covariates and their corresponding coefficients after excluding the covariates for which we are checking the assumption of linearity.

δ_i is an indicator of censoring and

$\hat{H}\left(t_i, x_i^{(-)}, \hat{\beta}^{(-)}\right)$ is the cumulative hazard after excluding the covariate of interest.

Thus, to check for linearity we plot martingale residuals against the excluded covariates and see if the resulting plot is random i.e. no systematic pattern or if the resulting smooth plot is a straight line then it shows linearity.

3.4.1.3.1.5 Assessment of Model Adequacy

The methods for assessment of a flitted proportional hazards model are essentially the same as for other regression models. In general requirements for model assessment are

1. methods for testing the assumption of proportional hazards
2. subject-specific diagnostic statistics that extend the notations of leverage and influence to the proportional hazards model, and
3. overall summary measures of goodness of fit.

3.4.1.3.1.5.1 Checking for proportionality assumption

In order to use the Cox model, we must check the assumption of whether the effects of covariates on hazard ratio remain constant over time. This is a critical assumption of proportional hazards model and must be checked for each covariate.

Different studies suggest that several tests and graphical techniques can be used to assess proportionality assumptions in fitting the Cox model. The Grambsch-Therneau test of non-proportionality uses partial residuals for the test of proportional hazards assumption. In order to use this test for the i^{th} covariate Grambsch and Therneau (1994) propose a time-varying coefficient as

$$\beta_i(t) = \beta_i + \gamma_i g_i(t) \qquad (3.19)$$

where $\beta_i(t)$ is time varying coefficient, β_i is constant $g_i(t)$ is some specified function of time, usually $g_i(t) = \ln(t)$. The Cox proportional hazard model for time varying coefficient with $g_i(t) = \ln(t)$ becomes

$$\lambda(t, x_i, \beta_i(t)) = \lambda_o(t)\exp(\beta_i(t)x),$$

substitute $\beta_i(t) = \beta_i + \gamma_i g_i(t)$ gives

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

$$= \lambda_o(t)\exp(\beta_i x + \gamma_i (\ln t)x) \qquad (3.20)$$

This looks like the proportional hazards model where the interaction term, $x\ln(t)$ is included in the model in addition to the main effect x_i. To test the significance of the interaction term $x_i \ln(t)$, that is, $H_o : \gamma = 0$ against $H_1 : \gamma \neq 0$ we can use likelihood-based tests like Wald test. If $\gamma = 0$ is not rejected, $\beta_i^{'}s$ are not time varying coefficients and hence the proportional hazards assumption is satisfied. If $\gamma = 0$ is rejected then the proportional hazards assumption is not satisfied and we have to look for another model.

The Schoenfeld residuals graphical technique can be used to assess Cox model assumptions. 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 the time the i^{th} individual event (Schoenfeld, 1982). For greater diagnostic power the scaled schoenfeld residuals are considered, the scaling can be done on the variance of the i^{th} subject Schoenfeld residuals.

To check the proportionality assumption for each covariate, we plot the scaled Schoenfeld residuals on the Y-axis against log of survival time on the X-axis. 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.

3.4.1.3.1.5.2 Identification of influential and poorly fit subjects

Another important aspect of model evaluation is a thorough examination of regression diagnostic statistics to identify which, if any, observations:

1. have an unusual configuration of covariates,
2. exert an undue influence on the estimate of the parameters, and
3. have an undue influence on the fit of the model.

Statistics similar to those used in linear and logistic regression are available to perform these tasks with a fitted proportional hazards model. There are some differences in the types of statistics used in linear and logistic regression and proportional hazards regression, but the essential ideas are the same in all the three settings.

Leverage is a diagnostic statistic that measures how "unusual" the values of the covariates are for an individual. In linear and logistic regression leverage is the distance of the value of the covariates for a subject to the overall mean of the covariates. Leverage is not easily defined nor does it have the same nice properties in proportional hazards regression. This is due to the fact that subjects may appear in multiple risk sets and thus may be present in multiple terms in the partial likelihood.

The score process residual for the i^{th} subject on the k^{th} covariate may be expressed as

$$L_{ik} = \sum_{j=1}^{n} \left(x_{ik} - \bar{x}_{wjk}\right) dM_i\left(t_i\right) \qquad (3.21)$$

is a weighted average of the distance of the value, x_{ik}, to the risk set means, $\bar{x}_{wjk}$, where the weights are the change in martingale residual ($dM_i\left(t_j\right)$) defined as.

$$dM_i\left(t_j\right) = dN_i\left(t_j\right) - Y_i\left(t_j\right)\exp\left(x_i'\beta\right)\lambda_o\left(t_j\right) \qquad (3.22)$$

where $dN_i(t_j)$ is the change in the count function for the i^{th} subject at time t_j, always equal to zero for censored subjects and one for non-censored subjects, at actual observed survival time. The function $Y_i(t_j)$ is called the at risk process and defined as zero if $t_i \le t_j$ and one if $t_i \ge t_j$. $\lambda_o(t_j)$ is the value of $\frac{\delta_i}{\sum_{j \varepsilon R(t)} \exp(x_j' \beta)}$ evaluated at t_j

The net effect is that, for continuous covariates, the score residuals have the linear regression leverage property that the further the value is from the mean the larger the score residual is, but "large" may be either positive or negative. Thus, the score residuals are sometimes referred to as the leverage or partial leverage residuals. We plot score residuals against each continuous covariates to observe if there is individuals far away from the mean.

3.4.1.3.1.5.3 Overall goodness of fit

As in regression analysis, some measure analogous to R^2 may be of interest as a measure of model performance. There is not a single, simple, easy to calculate, useful, easy to interpret measure for a proportional hazards regression model. In particular, all measures depend on the proportion of values that are censored. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to a high percent of censored data. We use R^2 as it is the easiest and best one to use, and it is defined as

$$R_p{}^2 = 1 - \exp\left(\frac{2}{N}\left(LL_o - LL_{\hat{\beta}}\right)\right) \qquad (3.23)$$

where N is the total number of observation in the model.

LL_o is the Log partial likelihood for model zero.

$LL_{\hat{\beta}}$ is the Log partial likelihood for the fitted model with p covariates.

3.4.1.3.2 Extensions of the proportional Hazards model

We have used a proportional hazards model with a common unspecified baseline hazard function where all the study covariates had values that remained fixed over the follow-up period. Additionally, we have assumed that the observations of the time variable were continuous. In some settings one or more of these assumptions may not be appropriate.

Now to accommodate non-proportionality assumption one can apply stratified proportional hazards model in which the stratification in most cases is done by using a covariate fixed by design. Suppose we have $s = 1,2,...,S$ strata, and then allow the baseline unspecified hazard function to vary among the strata.

The hazard function for stratum, s is

$$\lambda_s(t, x, \beta) = \lambda_{so}(t)\exp(\beta' x) \tag{3.24}$$

The form of the partial likelihood for the s^{th} stratum is identical to the partial likelihood used in proportional hazards model, but it includes an additional subscript, s, indicating the stratum. The contribution to the partial likelihood for the s^{th} stratum is

$$L_{sp}(\beta) = \prod_{i=1}^{n_s}\left[\frac{\exp(\beta' x_{si})}{\sum_{j \varepsilon R(t_{si})}\exp(\beta' x_{si})}\right]^{\delta_{si}} \tag{3.25}$$

where n_{si}: the number of observations in the s^{th} stratum

t_{si}: the i^{th} observed value of time in s^{th} stratum

δ_{si}: the value of the censoring indicator associated with t_{si}

$R(t_{si})$: the risk set for subjects in stratum s at time t_{si}

x_{si}: the vector of p-covariates for subject i in stratum s.

The full stratified partial likelihood is obtained by multiplying the contributions to the likelihood, namely

$$L_{Sp}(\beta) = \prod_{s=1}^{S} L_{sp}(\beta) \tag{3.26}$$

The maximum stratified partial likelihood estimator of the parameter vector, β, is obtained by solving the p equations obtained by differentiating the $\log L_{Sp}(\beta)$ with respect to the p unknown parameters and setting the derivatives equal to zero. Finally model building and model assessment is the same as that of proportional hazards model.

CHAPTER FOUR

4 STATISTICAL DATA ANALYSIS AND DISCUSSION

4.1 Introduction

The results of the study are discussed in this chapter. The response variable, duration of treatment in days, is continuous. The censoring indicator (status) is 0 for censored observations and 1 for event, in our case death. In this study Cox survival regression model is used to see the relationship between the proposed independent variables and the response variable. We start our data analysis by giving the summary statistics for the categorical variables considered in the study; we then proceed to the bivariate analysis, checking assumptions and complete the final model in multivariate analysis.

4.2 Summary Statistics

The medical cards of 637 patients have been reviewed of which 15.54% (total 99) are death cases. A death proportion seems lower for females (14.24%) than for males (17.00%). The married group showed the highest percentage (16.87%) with respect to death proportions than the other two groups and HIV positive groups revealed the highest proportion of death (29.95%). Category I TB patients have the highest death proportion (23.08%) as compared to the other two groups while Category III TB patients show the lowest death rate. A patient who took the treatment for the first time, which is a new case, seems to have lower death proportion (14.81) than other groups (21.43) and Pulmonary TB patients' death proportion (19.32%) seems larger than Extra-Pulmonary TB patient. All the results have been summarized in Table 4.1 below.

Table 4.1: Demographic and Health factors by TB death

Summary of the Number of Event and Censored values						
Demographic and Health factors	Stratum	Value	Total	Event/ Death	Censored	Percent Death
Sex	1	0(Male)	300	51	249	17.00
	2	1(Female)	337	48	289	14.24
Marital Status	1	0(Single)	365	52	313	14.25
	2	1(Others)	29	6	23	10.69
	3	2(Married)	243	41	202	16.87
Localization of TB	1	0(Extra-pulm.)	254	25	229	9.84
	2	1(Pulmonary)	383	74	309	19.32
Categorization of patients	1	0(Category III)	416	50	366	12.02
	2	1(Category II)	26	4	22	15.38
	3	2(Category I)	195	45	150	23.08
Previous Treatment	1	0(New)	567	84	483	14.81
	2	1(Others)	70	15	55	21.43
HIV status	1	0(HIV-ve)	430	37	393	8.60
	2	1(HIV+ve)	207	62	145	29.95

4.3 Descriptive analysis

Before proceeding to more complicated models, we make a descriptive analysis that will use as initiation to our subsequent findings. Here we start with the test of equality of probabilities across the different groups of a categorical variable using Log Rank test. The null hypothesis to be tested is that there is no difference between the probabilities of an event occurring at any time point for each population. The SAS and SPSS results have been summarized in Table 4.2 below.

Table 4.2: Log rank test for equality of survival experience among the different groups of covariates

Test of Equality over Strata			
Variable	Chi-Square	DF	Pr > Chi-Square
Sex	1.0645	1	0.3022
Marital status	1.5193	2	0.4678
Localization of TB	9.3111	1	0.0023
Category	11.2690	2	0.0036
Previous treatment	2.4125	1	0.1204
HIV	46.9288	1	<.0001

The table shows that the different groups of Localization of TB, Category of TB patient and HIV status are statistically not equal in experiencing the death event, whereas levels of Sex, Marital status and history of previous treatment are statistically the same in experiencing the event death. The Log–Rank test results suggests that Localization of TB, Category of TB patient and HIV status are significant covariates whose different levels have an impact in the survival longevity of TB patients; while Sex, Marital status and History of Previous

Treatment does not have an impact. Plots of different groups of Category, Localization of TB and HIV status to compare the survival probability of TB patients are given below.

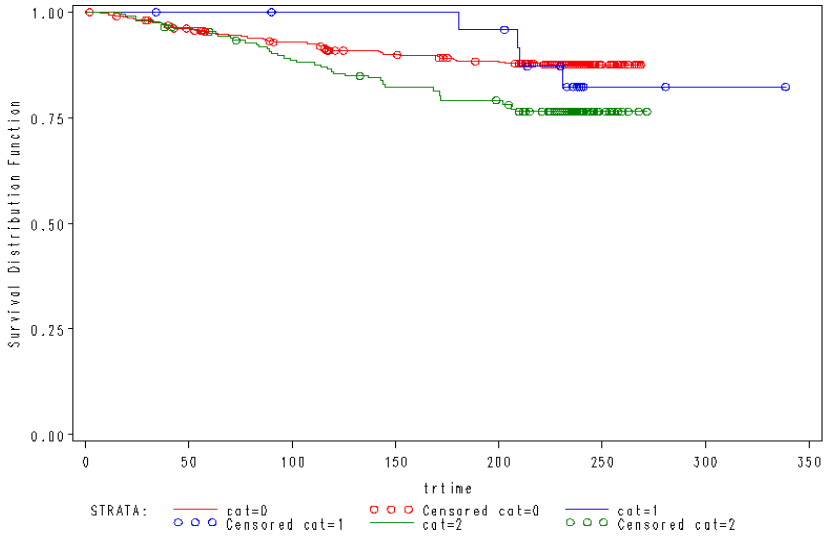


Figure 4.1: Survival curves of TB patients under three Categories

Figure 4.1 also supports that there is a difference in survival function between Categories and further identifies that, in general **Catigory II** patients survive longer than the other two where as **Catigory I** patients the lowest chance to survive.

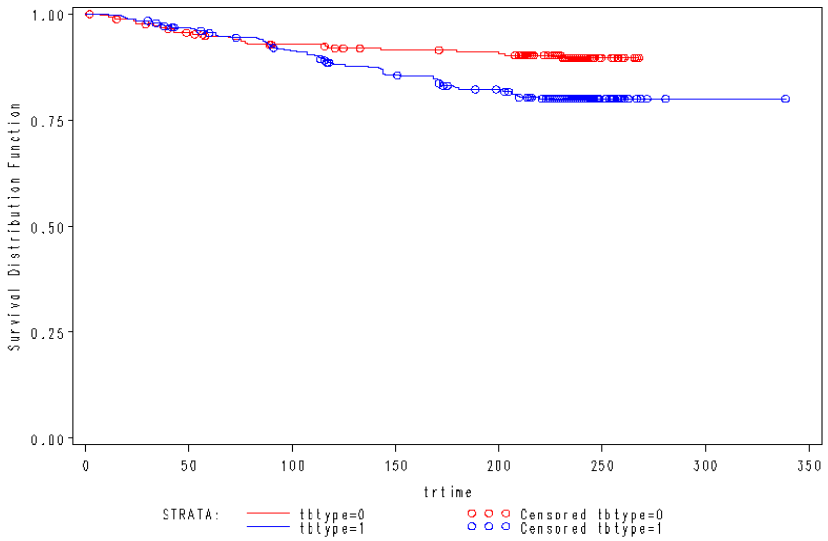


Figure 4.2: Survival curves of Extra-pulmonary and Pulmonary TB Patients

Again Figure 4.2 supports that there is a difference in survival function between Extra pulmonary and Pulmonary TB patients and further shows that Extra pulmonary patients

survive better than Pulmonary TB patients specially after the patient treated about three months.

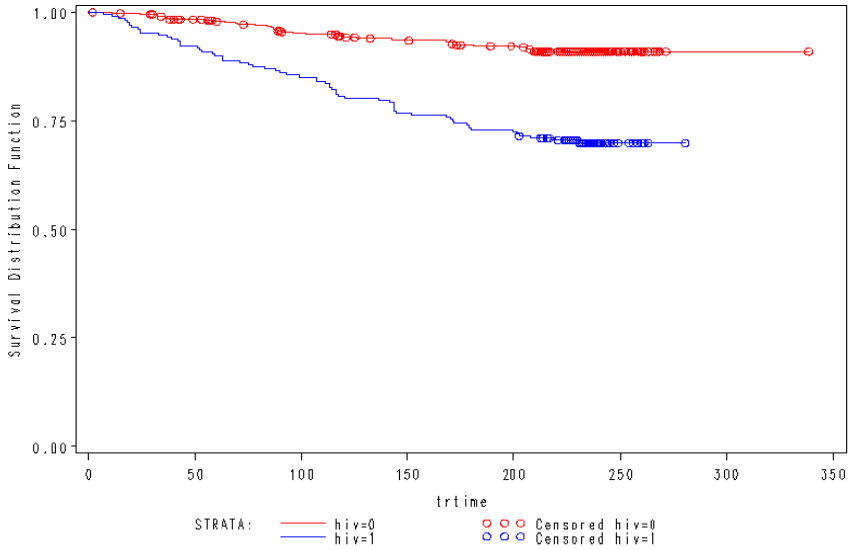


Figure 4.3: Survival curves of TB patients under HIV positive and HIV negative

Figure 4.3 shows that there is a difference in survival function between HIV negative TB patients and HIV positive TB patients.

4.3.1 Kaplan-Meier estimator

The Kaplan-Meier estimator of the survivorship function (or survival probability) of TB patients for each level of HIV status and Category of TB patients is calculated for every 15 days increase in treatment time. The following Table summarizes Minitab outputs.

Table 4.3: Estimates of Kaplan-Meier survival functions for HIV positive and HIV negative TB patients

	HIV negative				HIV positive			
Interval	n_i	d_i	c_i	$s(t)$	n_i	d_i	c_i	$s(t)$
15	430	1	2	0.9977	207	3	0	0.9855
30	427	1	3	0.9953	204	7	0	0.9517
45	423	5	5	0.9836	197	6	0	0.9227
60	413	2	6	0.9788	191	5	0	0.8986
75	405	3	1	0.9716	186	4	0	0.8792
90	401	6	3	0.9570	182	4	0	0.8599
105	392	3	1	0.9497	178	2	0	0.8502
120	388	3	6	0.9424	176	10	0	0.8019
135	379	1	3	0.9399	166	0	0	0.8019
150	375	2	1	0.9349	166	7	0	0.7681
165	372	0	0	0.9349	159	1	0	0.7633
180	372	4	4	0.9248	158	7	0	0.7295
195	364	1	1	0.9223	151	0	0	0.7295
210	362	5	0	0.9095	151	4	1	0.7101
225	357	0	0	0.9095	146	1	9	0.7053
240	357	0	357	0.9095	136	1	135	0.7001

From the above table we can say the probability of surviving of HIV (-) patients is 0.9095 for the entire treatment period, where as HIV (+) patients survival probability is 0.7001.

Table 4.4: Estimates of Kaplan-Meier survival functions for Categories I, II and III TB patients

	Category I				Category II				Category III			
Interval	n_i	d_i	c_i	$s(t)$	n_i	d_i	c_i	$s(t)$	n_i	d_i	c_i	$s(t)$
15	195	0	0	1.0000	24	0	0	1.0000	416	4	2	0.9904
30	195	4	0	0.9795	24	0	0	1.0000	410	4	3	0.9807
45	191	3	2	0.9641	24	0	0	1.0000	403	8	2	0.9613
60	186	2	1	0.9537	24	0	0	1.0000	393	5	5	0.9490
75	183	4	1	0.9329	24	0	0	1.0000	383	3	0	0.9416
90	178	6	0	0.9014	24	0	0	1.0000	380	4	2	0.9317
105	172	4	0	0.8805	24	0	0	1.0000	374	1	1	0.9292
120	168	5	0	0.8543	24	0	0	1.0000	372	8	8	0.9092
135	163	1	1	0.8490	24	0	0	1.0000	356	0	0	0.9092
150	161	5	0	0.8227	24	0	0	1.0000	356	4	0	0.8990
165	156	0	0	0.8227	24	0	0	1.0000	352	1	1	0.8964
180	156	6	0	0.7910	24	0	0	1.0000	350	5	5	0.8836
195	150	0	0	0.7910	24	1	0	0.9583	340	0	0	0.8836
210	150	5	0	0.7647	23	2	2	0.8750	340	2	5	0.8784
225	145	0	121	0.7647	19	0	0	0.8750	333	1	0	0.8758
240	24	0	24	0.7647	19	1	18	0.8289	332	0	332	0.8758

Table 4.6 shows that the probability of survival for Categories I, II, and III TB patients are 0.7647, 0.8289, and 0.8758, respectively, during treatment period.

4.4 Bivariate Analysis

For each covariate we will use a bivariate Cox proportional hazards model analysis that contains a single independent variable in order to have an idea about each covariate. Likelihood ratio chi-square test is used to test the significance of bivariate relationship.

In bivariate analysis, using likelihood ratio chi-square test, the variables that are found to be significant (p-value 0.2 is used as a criterion for significance) are Age of a patient(age), Initial weight of a patient(intwt), Localization of TB(tbtype), Category of a patient(cat), History of previous treatment(ptreat) and HIV status(hiv). Age of a patient, Initial weight, Localization of TB, Category and HIV status, with p-value less than 0.05(standard level of significance), have relatively strong associations to the death of TB patients. Table 4.7 summarizes the findings of Bivariate analysis (see Annex A).

Table 4.5: Bivariate analysis result for each covariate

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Sex	1	-0.20700	0.20110	1.0595	0.3033	0.813
Age	1	0.04013	0.00693	33.4934	<0.0001	1.041
Mstatus	2	0.10078	0.10351	1.518	0.468	1.106
Intwt	1	-0.04386	0.01225	12.8168	0.0003	0.957
Tbtype	1	0.69174	0.23136	8.9391	0.0028	1.997
Cat	2	0.33726	0.10306	11.257	0.004	1.401
Ptreat	1	0.43182	0.28033	2.3728	0.1235	1.540
Hiv	1	1.32399	0.20779	40.5992	<0.0001	3.758

4.4.1 Partial likelihood ratio test for the contribution of the interaction effect

From theoretical point of view the following possible interactions are expected. Moreover, we need to assess some realistic situations to see if two interaction effects can increase or decrease the survival time of TB patients. The partial likelihood ratio test is used to identify the significance of some reasonable and possible interactions (See Annex B). The hypothesis to be tested is

H_o: The model with only main effect fits the model equally well as the model having the main effect and their interaction as predictors.

H_1: H_o is not true

Decision: Reject H_o at $\alpha = 0.05$ level of significance if -2 LOG L_2- (-2 LOG L_1) $\geq$ $\chi_1^2(\alpha = 0.05)$ = 3.84, other wise do not reject H_o. This means we need to include the corresponding interaction in the multivariate analysis.

Table 4.6: Partial likelihood ratio test for checking interaction terms

Model Fit Statistics				
Variables	**-2 LOG L_2 With main effects only**	**-2 LOG L_1 With main effects and interaction**	**-2 LOG L_2- (-2 LOG L_1)**	Sig.
HIV, Age	**1191.142**	**1184.709**	**6.433**	**Reject**
HIV, Mstatus	**1212.298**	**1208.810**	**3.498**	**Do not reject**
HIV, Intwt	**1202.867**	**1199.719**	**3.148**	**Do not reject**
HIV, category	1202.041	1200.344	1.697	**Do not reject**
HIV, TB type	1201.647	1199.784	1.863	**Do not reject**
TB type, Category	1240.303	1239.808	0.495	**Do not reject**
TB type, Intwt	1233.131	1232.527	0.604	**Do not reject**

The table shows that only the interaction between HIV and age of patient (HIVage) is significant. And this is an indication that the interaction of HIV (+) and the age of the patient affects the survival time of the patient.

4.5 Multivariate Analysis

One problem with any bivariate analysis is that it ignores the possibility that a collection of covariates, each of which is weakly associated with the outcome, may have a significant effect when used together with other covariates in the model. If this is thought to be a possibility, then we should choose a significance level large enough to allow the suspected variables to become candidates for inclusion in the multivariate model. It is for this reason that we use p-value of 0.2 for selection of variables that are potentially candidates for the multivariate analysis from bivariate findings and those significant interactions on partial likelihood ratio test. To facilitate computation and interpretation, the coding scheme used in SAS, SPSS and STATA is given below in Table 4.9.

Table 4.7: Categorical Variable Coding

Variable			Frequency	Parameter Coding (1)	(2)
Sex	Male	0	300	0	
	Female	1	337	1	
Mstatus	Single	0	365	0	0
	Others	1	29	1	0
	Married	2	243	0	1
TBtype	Extra-Pulmonary	0	254	0	
	Pulmonary	1	383	1	
Cat	Category I	0	416	0	0
	Category II	1	26	1	0
	Category III	2	195	0	1
Ptreat	New	0	567	0	
	Others	1	70	1	
HIV	HIV-ve	0	430	0	
	HIV+ve	1	207	1	

The following table shows multivariate analysis done using the significant variables in the bivariate analysis and significant interaction terms based on the likelihood ratio test.

Table 4.8: Partial Likelihood Estimates for Fitted Proportional Hazards Model

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate(B)**	**Standard Error**	**Wald**	**Sig.**	**Exp(B)/Hazard ratio**
Age	1	.052	0.010	25.814	.000	1.054
Cat	2			5.120	.077	
cat(1)	1	0.034	.561	0.004	.952	1.066
cat(2)	1	0.504	.226	4.973	.026	1.594
Intwt	1	-.034	0.013	7.515	.006	.966
Tbtype	1	.181	0.256	.503	.478	1.199
Ptreat	1	.408	0.304	1.801	.180	1.504
Hiv	1	2.530	0.625	16.383	.000	12.558
HIVage	1	-0.036	0.016	5.465	0.019	0.964

The variables that are found to be insignificant at 10% level of significance in multivariate analysis are History of previous treatment and Localization of TB. We drop these variables for the next step and perform a multivariate analysis for the remaining five covariates. The following table shows the fitted Cox-proportional hazards model for covariates age, cat, intwt, hiv and interaction of hiv and age (HIVage).

Table 4.9: Partial Likelihood Estimates for Significant covariates

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate(B)**	**Standard Error**	**Wald**	**Sig.**	**Exp(B)/ Hazard ratio**
Age	1	.054	0.010	29.212	.000	1.056
Cat **cat(1)** **cat(2)**	2 1 1	 0.322 0.563	 .526 .209	7.268 0.374 7.253	.026 .541 .007	 1.379 1.757
Intwt	1	-.035	0.012	7.997	.005	.965
Hiv	1	2.548	0.620	16.903	.000	12.787
HIVage	1	-0.037	0.015	5.896	0.015	0.963

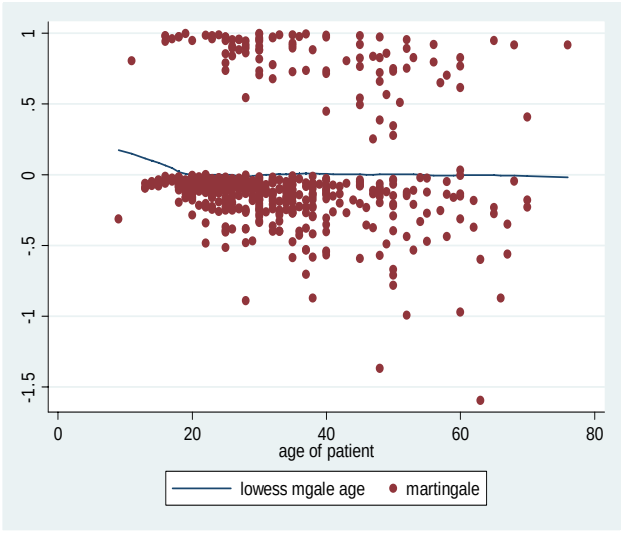
Here all covariates are significant at 5% level of significance. Thus, it becomes a preliminary final model after linearity of continuous covariates is checked.

4.5.1 Checking for the linearity of covariates in the model

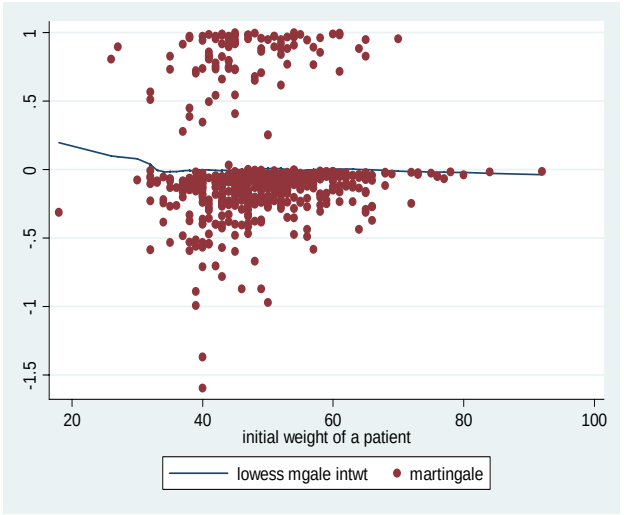
The next step is to examine the scale of continuous covariates in the preliminary main effects model. A number of techniques are available, all of which are designed to determine whether the data support the hypothesis that the effect of the covariate is linear in the log hazard and, if not, which transformation of the covariate is linear in the log hazard. Now we use the graphical method of testing linearity for continuous covariates.

As it was described in chapter three, the martingale residuals may be plotted against covariates to detect for the correctness of the functional form. The following STATA output shows the resulting graph.

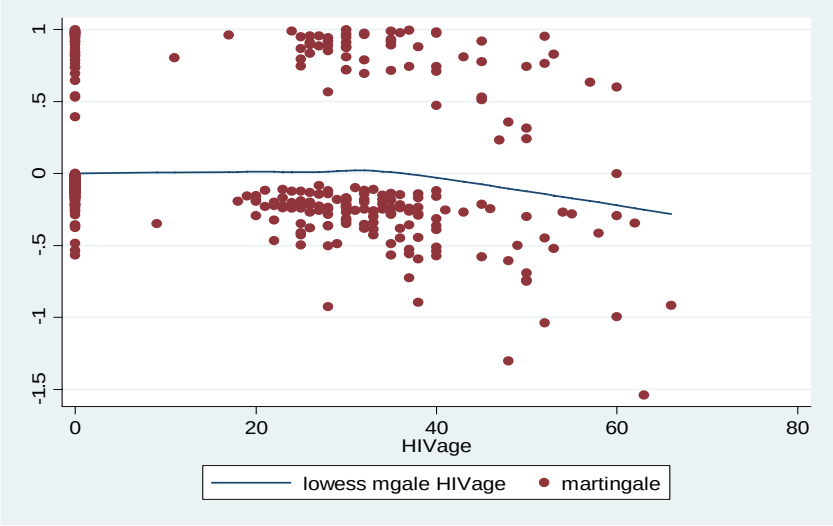
Graphical assessment of the functional form of the continuous covariate in the model



(a) Plot of Martingale Residuals for Age



(b) Plot Martingale Residuals for Initial Weight



(c) Plot of Martingale Residuals for HIV by Age interaction

Figure 4.4 Plots of Martingale residuals computed for continuous covariates (a) Age, (b) Initial Weight and (c) HIVage

The above graphs show the plot of martingale residual versus covariates. For each of the covariates, namely age, initial weight of TB patient and interaction of HIV status by Age, the plots do not show systematic patterns or trends and the resulting smoothed plots (LOWESS) are approximately straight line. Therefore the plots of martingale residual confirm that age and initial weight of a patient have an approximate linear relationship with the survival time (See annex C for Transformed covariates).

Table 4.10: Variables in the Preliminary Final Model

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate(B)**	**Standard Error**	**Wald**	**Sig.**	**Exp(B)/ Hazard ratio**
Age	1	.054	0.010	29.212	.000	1.056
Cat **cat(1)** **cat(2)**	2 1 1	 0.322 0.563	 .526 .209	7.268 0.374 7.253	.026 .541 .007	 1.379 1.757
Intwt	1	-.035	0.012	7.997	.005	.965
Hiv	1	2.548	0.620	16.903	.000	12.787
HIVage	1	-0.037	0.015	5.896	0.015	0.963

Table 4.12 shows computer output of the result of the fitted hazards model. Based on the result we look for predictors having statistical significant relationship with the hazards (death of TB patients). All the covariates namely age, category, initial weight, HIV and interaction of HIV and age are significant.

The values of the Wald statistics for individual β coefficients support that the estimated values ($\hat{\beta}_i^{'}s$) are significantly different from zero at α =0.05 level of significance for all the above five covariates. The remaining variables which were used in the bivariate analysis, including most interaction terms are found to be non-significant. We consider the model that contains these covariates as a preliminary final model and it could be the final model after we check proportionality assumptions.

4.5.2 Assessment of Model Adequacy

In our survival regression analysis assessment of model adequacy we must i) test the assumption of proportional hazards ii) check influence and poorly fit subjects and iii) overall summary measures of goodness of fit.

4.5.2.1 Test of the assumption of proportional hazards

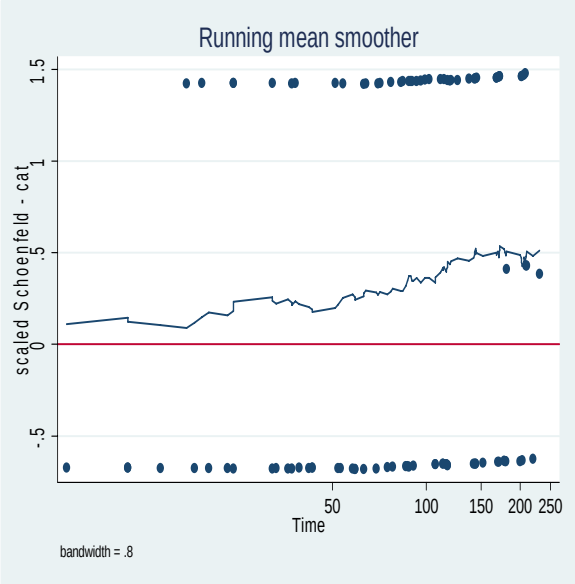
A proportional hazard is one of the very important assumptions in the Cox model. As it was described in chapter three, graphical diagnoses of scaled Schoenfeld residuals and likelihood based tests, like Wald test can be employed to assess the proportional hazard assumption to covariates that are significant in the multivariate analysis.

Under the assumption of proportionality of the proportional hazards model, the distribution of residuals over time is random and LOWESS smoothing line should be a straight line around zero.

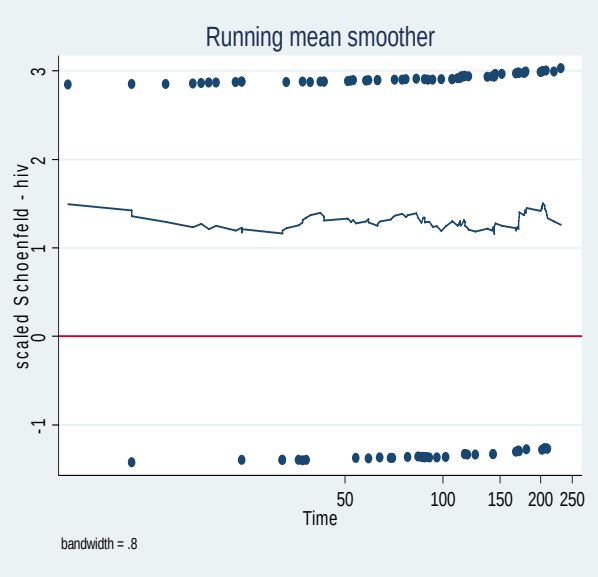
Graphical assessment of the proportional hazard assumptions

(a)Scaled Schoenfeld Residuals for Age

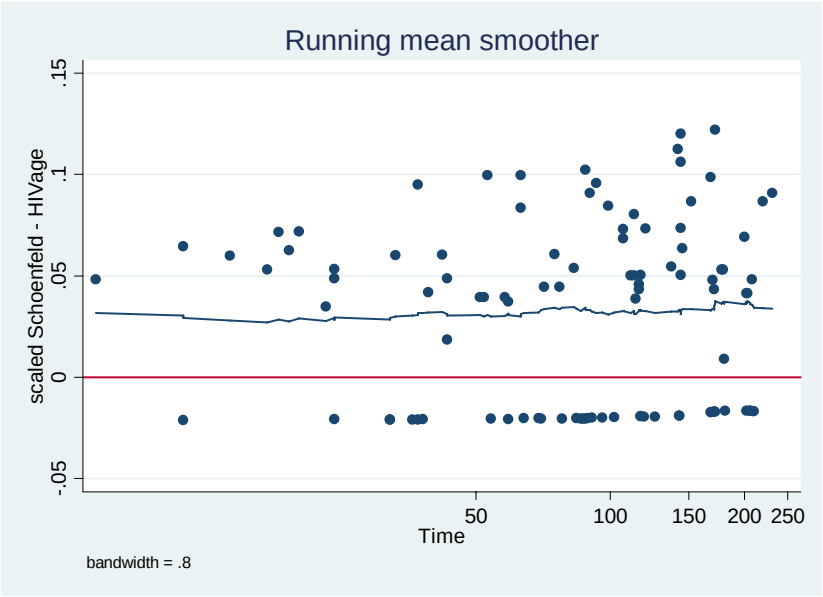
(b) Scaled Schoenfeld Residuals for Initial weight



(c) Scaled Schoenfeld Residuals for Category HIV



(d) Scaled Schoenfeld Residuals for



(e) Scaled Schoenfeld Residuals for HIV by Age interaction

Figure 4.5 plots of Scaled Schoenfeld residuals computed from the model in Table 4.12 for (a) Age, (b) Initial Weight, (c) Category, (d) HIV and (e) Age by HIV interaction.

The graphical display shows plots of the scaled Schonfield residuals against the survival time for each covariate namely Age, Initial Weight, Category, HIV status and the interaction of HIV and Age (HIVage) of a patient. In Figure 4.5 (a,b,d) and (e) plots of scaled Schoenfeld residuals show randomness. Moreover, the smoothed curve is an approximate horizontal line; so the above four covariates satisfied the assumption of proportional hazards. On the other hand Figure 4.5 (c) category of TB patients does not manifest randomness in the residuals over time and the smooth curve is not a horizontal straight line. This indicates that the covariate category does not satisfy the proportional hazards assumption.

The graphical test is not enough to be certain of the proportionality assumption of the model. The reason is that it is subject to the interpretation of different persons, that is, graphical method of assessing the proportionality assumption is more or less subjective. But we use it as a supportive argument for proportionality test.

One of the statistical tests for proportional hazards assumption is to generate time varying covariates by creating interactions of the predictors and a function of survival times, usually

covariate times log of time, and including these in the model. If any of the time dependent covariates are significant then those predictors do not show a proportional effect over the study period. That is the proportional hazard assumption fails to hold.

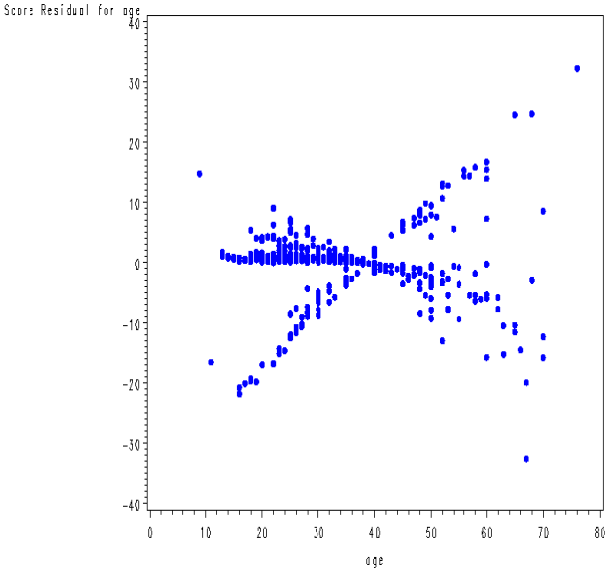
Table 4.11: SAS result of the assumption of proportionality test

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Age	1	-0.02205	0.06245	0.1246	0.7240	0.978
Intwt	1	0.02192	0.06866	0.1019	0.7495	1.022
Cat	1	-0.96531	0.64145	2.2647	0.1323	0.381
Hiv	1	2.91718	3.49231	0.6978	0.4035	18.489
HIVage	1	-0.02733	0.09146	0.0893	0.7651	0.973
Aget	1	0.01717	0.01362	1.5902	0.2073	1.017
intwtt	1	-0.01314	0.01552	0.7165	0.3973	0.987
Catt	1	0.28358	0.14238	3.9669	0.0464	1.328
Hivt	1	-0.09668	0.78217	0.0153	0.9016	0.908
HIVaget	1	-0.00188	0.02023	0.0086	0.9261	0.998

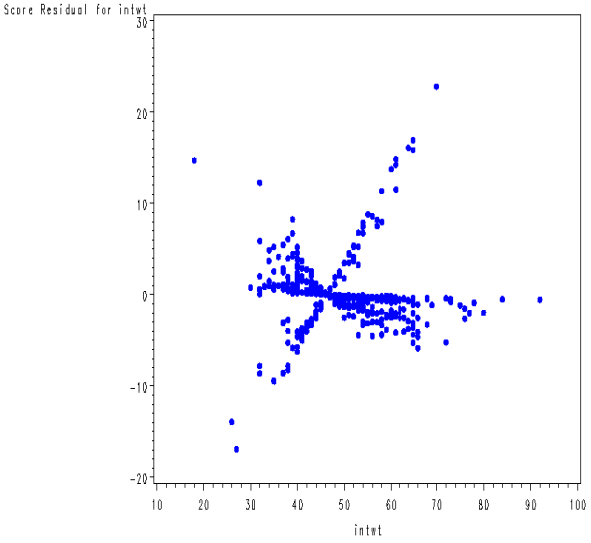
The table shows that the Wald chi-square value and corresponding p-values for each covariate. Since the p-value= 0.0464 of the likelihood ratio test is less than 0.05 for the covariate Category, we reject the proportional hazards assumption for the covariate Category. While the p-value for the other four covariates are greater than 0.05, implying that the proportionality assumption is satisfied.

4.5.2.2 Identification of influential and poorly fit subjects

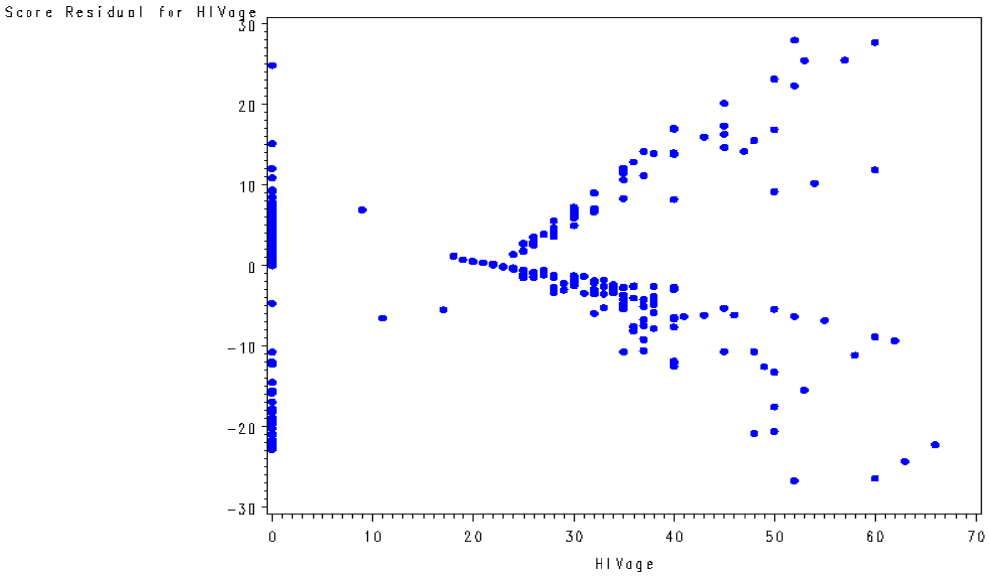
The graphs of the score residuals for covariates Age, Initial weight and the Age by HIV interaction obtained from the fitted model in Table 4.13 are shown in Figure 4.6. These three terms were chosen because they are continuous variables in the fitted model and are therefore most amenable to having their Score residuals examined graphically. The graphs for the categorical covariates are less interesting in that all the values fall on two or three vertical bands at zero, one and two covariate values.



(a) Score Residuals for Age



(b) Score Residuals for Initial Weight



(b) Score Residuals for Age by HIV interaction

Figure 4.6 plots of the Score residuals computed from the model for (a) Age, (b) Initial Weight and (c) Age by HIV interaction.

The purpose of giving the plot is to see whether there are subjects whose observations yield unexpectedly large values. This would be seen in the graph as a point lying further away from the others in the plot. In Figure 4.6(a) there is one point at the top right and another point at the bottom right that fall a bit further away from the rest of the points. However, the distance between these points and the others is not striking. Thus, we conclude that there are no high leverage values for age.

The Score residuals for the Initial weight are plotted in Figure 4.6(b). There is only one point at the top right that fall a bit further away from the rest of the points but the distance between the point and the others is not large. The same thing is true for Score residuals of HIV by age interaction. Here as we see from Figure 4.6(c) there are no strikingly large Score residuals. Thus, we conclude that there are no high leverage values for Initial weight and HIV by age interaction.

4.6 Extensions of the Proportional Hazards Model

As we attempted to show in the previous sections, the proportional hazards assumption is not satisfied for the covariate category in the data of TB patients. Hence, the best one can hope for is to apply stratified proportional hazards model in which the stratification is done by using the covariate category, as it is fixed by design. The only change in the model is the addition of strata statement together with the time dependent covariate in the SAS program and exclusion of that particular variable (category of TB patients) from the model.

The assumption is that we are fitting separate models for each level of a particular time-dependent covariate under the constraint that the coefficients are equal but the baseline hazard function are not equal. This is to say that each stratum can have a different baseline hazards function, while the coefficients of the remaining covariate are assumed to be constant across strata.

Table 4.12: Results for the final Proportional Hazard Model Stratified by Category of TB patients

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
age	1	0.05592	0.01012	30.5090	<.0001	1.058
intwt	1	-0.03541	0.01247	8.0656	0.0045	0.965
hiv	1	2.62815	0.62541	17.6592	<.0001	13.848
HIVage	1	-0.03929	0.01543	6.4832	0.0109	0.961

Log-likelihood=-500.183

4.7 Overall Goodness of Fit

We use R^2 as a measure of over all goodness of model fit. As it is defined in chapter three it is given as

$$R_p{}^2 = 1 - \exp\left(\frac{2}{N}\left(LL_o - LL_{\hat{\beta}}\right)\right)$$

where N=637 is the total number of observation in the model.

LL_o =-539.6465 is the Log partial likelihood for model zero.

$LL_{\hat{\beta}}$ =-500.183 is the Log partial likelihood for the fitted model with p covariates.

For the fitted model in Table 4.13, the value is

$$R_p{}^2 = 1 - \exp\left(\frac{2}{637}\left(-539.6465 - \left(-500.183\right)\right)\right)$$

$$= 0.116$$

The model displayed in Table 4.14 has passed all the tests for a good fitting model.

4.8 Interpretation and presentation of a final model

The model fit to the TB data, shown in Table 4.14 contains two continuous linear covariates (Age and Initial weight of a patient), one dichotomous covariate (HIV status) and an interaction between a continuous and a dichotomous covariate (Age and HIV status).

The estimated covariate coefficient $\hat{\beta}$ =-0.03541 for the continuous risk factor Initial weight implies that the hazards ratio is exp ($\hat{\beta}$) =0.965. This shows the change of hazard rate for every one kg increase in Initial weight. For a better interpretation we consider the change of the hazard rate for every five kg increase in Initial weight. The estimated hazard ratio for a five kg increase in Initial weight of TB patients is $0.84 = \exp(5 * -0.03541)$. The interpretation is that patients with five kg increase in initial weight, the probability of survival increased by about 14 percent, holding all others fixed.

Age and HIV status are present in the model, with both main effects and their interaction. Since HIV status is at two levels, we present hazard ratios for age at each HIV status rather

than for HIV status at each Age. The estimated log hazard as a function of the variables Age, HIV and Hivage, holding initial weight fixed is given as

$$Log\hat{H}(Age, HIVstatus, x) = \hat{\beta}_1 Age + \hat{\beta}_2 HIV + \hat{\beta}_3 HIV \times Age + \hat{\beta}' Z .$$

The next expression for the difference of an increase of say ten years of age holding HIV fixed is

$$Log\hat{H}R(Age+10, HIVstatus, x) = \begin{array}{l} \hat{\beta}_1(Age+10) + \hat{\beta}_2 HIV + \hat{\beta}_3 HIV \times Age + \hat{\beta}' Z \\ -\left[\hat{\beta}_1 Age + \hat{\beta}_2 HIV + \hat{\beta}_3 HIV \times Age + \hat{\beta}' x\right] \end{array}$$

$$= 10\hat{\beta}_1 + 10\hat{\beta}_3 HIV$$

Finally, the estimated hazard ratio for an increase of 10 years of Age at HIV=0 is

$$\hat{H}R(Age+10, Age, HIV = 0) = e^{10*0.05592}$$

$$=1.75$$

And at HIV=1 the hazard ratio is

$$\hat{H}R(Age+10, Age, HIV = 0) = e^{10*0.05592+10*(-0.03929)}$$

$$= 1.18$$

The interpretation is that being older by 10 years at HIV-ve increases the rate of death by about 75 percent. Where as at HIV+ve, the rate of death is increased only by 18 percent. This shows that death due to HIV is less dependent on Age.

Chapter Five

Conclusions and Recommendations

5.1 Conclusions

Our findings show that 84.5% of the patients were still alive at the end of eight months of anti-tuberculosis treatment. The Cox Proportional Hazards regression analysis was done to identify the effects of Sex, Age, Marital status, Initial weight, Category, Localization of TB, History of previous treatment and HIV status of TB patients on survival/death probability of TB patients.

The study suggests that Age, Category, Initial weight, HIV status and the interaction of HIV by Age have statistically significant effects on the survival longevity of TB patients. On the other hand Sex, Marital status, Localization of TB and History of previous treatment have no impact on the survival experience of TB patients.

The result of this study also indicated that survival/death status of TB patients does not show differences based on Sex, Marital status and history of previous treatment levels. However it depends on different groups of Category, Localization of TB and HIV status. Similarly, patients with small weight and older Age are less likely to survive. The study also showed that the highest death rate were associated with HIV positive patients.

5.2 Recommendations

Based on the result of the study different factors are identified for the death of TB Patients. And the following recommendations are made for health policy makers and clinicians:

- Since tuberculosis is a curable disease if properly treated, due attention must be given to DOTS strategy by government and non-government organizations.
- Health workers should be cautious when a patient has lower Initial weight and HIV-positive status. When this is the case appropriate clinical and non-clinical measures like medicine and support (can be home-based) should be provided.

- For those patients with low economic status, health workers/stakeholders need to find ways of supporting the patients with respect to improving their nutrition in particular and other assistance in general.
- Special attention should be given to HIV positive patients to receive ART to extend their lives.

Bibliography

[1]. Allison, P.D. and Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A coment on Gottman. Psychological Bulletin, 91, 393-403.

[2]. Allison, P.D. (1984). Event history analysis: Regression for longitudinal event data (Sage University paper series on quantitative applications in the social sciences, Number 07-046), Beverly Hills, CA: Sage.

[3]. Allison, P.D. (1995). Survival analysis using the SAS system: A practical guide, Cary NC: SAS Institute.

[4]. Anderson, P.K. (1982). Testing Goodness of Fit of Cox's Regression and life model. Biometrics, 38, 67-77.

[5]. Arjas, E. (1988). A graphical Method for Assessing Goodness of Fit in Cox's Proportional Hazard Model. Journal of the American statistical Association 83, 204-212.

[6]. Bass, J.B. Farer, Hopewell, P.C. O'Brien, R.Jacobs, R.F. Ruben, F.Snider, D.E.J. Thormton, G.(1990). "Treatment of tubeerculosis infections in adults and children", 149:1359-1374.

[7]. Bloessfeld, H.P., Hamerle, A. and Mayer, K.U. (1989). Event history analysis: Statistical theory and application in the social sciences. Hillsdale, NJ: Erlbaum.

[8]. Borgdorff M.W, Veen J, Kalisvaart N.A, Nagelkerke N. Mortality among tuberculosis patients in the Netherlands in the period 1993-1995. *Eur Respir J* 1998;**11**: 816-820.

[9]. Box-Steffensmeier, J.M and Zorn, J.W. (2001). Duration Models and Proportional hazards in political science. American Journal of Political Science. 45, 951-967.

[10]. Breslow, N. (1974). Covariance Analysis of Censored survival data, Biometrics 30,89-99.

[11]. Collette, D. (1994). Modeling survival data in medical research. London: Chapman and Hall.

[12]. Cox, D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society. Series B (Methodological), 34, No.2, 187-220.

[13]. Cox, D.R. (1975). Partial Likelihood Biometrika, 62, 269-276.

[14]. Cox, D.R. and Oakes, D. (1984). Analysis of survival data. London: Chapman and Hall.

[15]. Dye, C. Dolin, P. Scheele, S. and Athenian, V. (1999). Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. Journal of the American Medical Association, 282:677-686.

[16]. Dress, K.A. (1986). The effect of gender identify on conversation. Social Psychology Quarterly, 49, 294-301.

[17]. Erhabor, G.E., Adewole, 0.0. and Ogunlade 0. (2006). A Five-Year Review of Tuberculosis Mortality Amongst Hospitalised Patients in Ile-Ife. Department of Medicine, Obafemi Awolowo University Teaching Hospital Complex, Ile-Ife, Osun State, Nigeria

[18]. Federal Ministry of Health (2002). Manual Tuberculosis and Leprosy Prevention and control programs, Addis Ababa, Ethiopia.

[19]. Federal Ministry of Health (2008). Manual of Tuberculosis and Leprosy Prevention and control programs, Addis Ababa, Ethiopia.

[20]. Femalee, D. and Eder, D. (1983). Contextual effects in the classroom: the impact of ability groups on student attention. Sociology of Education, 56, 77-87.

[21]. Femalee, D., Eder, D. and Tsui, W.Y. (1985). Peer influence on classroom attention. Social Psychology Quarterly, 48, 215-226.

[22]. Fergusson, D.M., Horwood, L.J. and Diamond, M.E. (1985). A survival analysis of childhood family history. Journal of Marriage and the Family, 47, 287-295.

[23]. Fergusson, D.M., Horwood, L.J. and Shannon, F.T. (1984). Proportional hazards model of a family background. Journal of Marriage and the Family, 46, 539-549.

[24]. Fichman, M. (1988). Motivational consequences of absence and attendance: Proportional hazards estimation of a dynamic motivation model. Journal of Applied Psychology, 73, 119-134.

[25]. Fleming, T. and Harrington, D. (1991). Counting process and survival analysis. John Wiley and Sons, Inc., New York.

[26]. Gardner, W. and Griffin, W.A. (1989). Methods for the analysis of parallel streams of continuously reported social behaviors. Psychological Bulletin, 446-455.

[27]. Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnosis based on weighted residuals, Biometrika 81, 515-526.

[28]. Greenhouse, J.B., Stangle, D. and Bromberg, J. (1989). An introduction to the survival analysis method for the analysis of clinical trial data. Journal of Consulting and Clinical Psycology, 57, 536-544.

[29]. Grimard, F. and Harling, G(2003). The impact of Tuberculosis on Economic Growth. Department of Economics, McGill University, Montreal.

[30]. Gross, A.J. and Clark, V.A. (1975). Survival distribution. John Wiley and Sons Inc., New York.

[31]. Heckman, J. and Singer, B. (1985). Longitudinal analysis of labor market data. New York: Cambridge University press.

[32]. Hembroff, L.A. and Myers, D.E. (1984). Status characteristics: Degrees of task relevance and decision processes. Social Psychology Quarterly, 47, 337-346.

[33]. Johnson, D. (1988). Panel analysis in family studies. Journal of Marriage and Family, 50, 949-955.

[34]. Kalbfleisch, J.D. and Prentice, R.L.(1980). The Statistical Analysis of Failure Time Data. John Wiley and Sons, New York..

[35]. Khatri G.R, Frieden T.R. The status and prospectus of tuberculosis control in India. Int J Tuberc Lung Dis 2000;**4:** 193-200.

[36]. Kolappan C, Subramani R, Karunakaran K, Narayanan PR. Mortality of Tuberculosis patients in Chennai, India. Bull World Health Organ 2006; **84**:555-560.

[37]. Lavori, P.W., Keller, M.B. and Klerman, G.L. (1984). Relapses in affective disorders: A reanalysis of the literature using life-table methods. Journal of Psychiatric Research, 18, 13-25.

[38]. Lawless, J.F. (1982). Statistical models and methods for life time data. John Wiley and Sons, Inc., New York.

[39]. Lefebvre N and Falzon D. Risk factors for death among tuberculosis cases: analysis of European surveillance data. Eur Respir J 2008; 31: 1256-1260.

[40]. Mathew T A, Ovsyanikova T A, Shin S S, Gelmaova I, Balbuea D A, Atwood S, Peremitin G G, Strelis A K, Murray M B. Causes of death during tuberculosis Treatment in Tomsk Oblast, Russia. Int J Tuberc Lung Dis 2006; **10(8)**: 857-863.

[41]. Miller, R.G. (1981). Survival analysis. John Wiley and Sons, New York.

[42]. Moreau, T., O'Quigley, J. and Mesbsh, M. (1985). A global goodness of fit Statistics for the proportional hazards model. Applied statistics 34, 212-218.

[43]. Moreau, T., O'Quigley, J. and Mesbsh, M. (1986). On D. Schoenfeld's approach for testing the proportional hazards assumption. Biometrika 73, 513-515.

[44]. Morgan, S.P., Lye, D.N. and Condran, G.A. (1988). Sons, daughters and the risk of marital disruptions. American Journal of Sociology, 94, 110-129.

[45]. Morita, J.G., Lee, T.W. and Mowday, R.T. (1989). Introducing survival analysis to organizational researchers: A selective application to turnover research. Journal of Applied Psychology, 74, 280-292.

[46]. Murray, C.J.L. and Lopez, A.(1996). The global burden of tuberculosis. Cambridge, MA: Harvald University press.

[47]. Santha T, Garg R, Frieden T.R, Chandrasekaran V, Subramani R, Gopi P.G, Selvakumar N, Ganapathy S, Charles N, Rajamma J, Narayanan PR. Risk factors associated with default, failure and death among tuberculosis patients treated in a DOTS programme in Tiruvallur District, South India, 2000. Int J Tuberc Lung Dis 2002; **6(9):** 780-788.

[48]. Schoenfeld, D. (1982). Chi-squared goodness of fit tests for the proportional Hazards regression model. Biometrika 7, 145-153.

[49]. Shapiro, D.R., Quitkin, F.M. and Fliess, J.L. (1989). Response to maintenance therapy in bipolar illness: Effects of index episode. Archives of General Psychiatry, 46, 401-405.

[50]. Stevens, V.J. and Hollis, J.F. (1989). Preventing smoking relapse using an individually tailed skills training technique. Journal of Consulting and Clinical Psychology, 57, 420-424.

[51]. TB India 2006, RTCP Status Report, Central TB Division, Directorate General of Health Services, Ministry of Health and Family Welfare, Nirman Bhavan, New Delhi.

[51]. Teachman, J.D. (1982). Methodological issue in the analysis of family formulation and dissolution. Journal of Marriage and the Family, 44, 433-445.

[52]. Tekle B, Mariam DH, Ali A. Defaulting from DOTS and its determinants in three districts of Arsi Zone in Ethiopia. Int J Tuberc Lung Dis. 2002 6(7):573-9.

[53]. Therneau, T.R., Grambsch, P.M. and Fleming, T.R. (1990). Martingale based Residuals for Survival Models. Biometrika, 77, 147-160.

[54]. Thompson L.A., Chhikera, R.J. and Conkin, J. (2003). Cox Proportional Hazards Methods for Modeling the Time to Onset of Decomposition Sickness in Hypobaric Environments. NASA Technical Publication 2003-210791, Houston Johnson Space Center.

[55]. Tuma, N.B. and Hannan, M.T. (1984). Social dynamics: Models and Methods. San Diego, C.A. Academic Press.

[56]. Tuma, N.B., Hannan, M.T. and Groeneveld, L.P. (1977). Dynamic analysis of event histories. American Journal of Sociology, 84, 820-854.

[57]. Vasantha, M., Gopi, P.G. and Subramani, R. Survival of tuberculosis patients treated under DOTS in a rural tuberculosis unit (TU), South India.

[58]. Waldman EA, Ana Angelica Bulcao Portela Lindoso, Komatsu NK, de Figueiredo SM, Taniguchi M and Rodrigues LC. Profile of tuberculosis patients progressing to death, city of Sao Paulo, Brazil, 2002.

[59]. WHO (1998). Global tuberculosis program. "WHO report on tuberculosis epidemic." Geneva, Switzeland.

[60]. WHO(1999). Global tuberculosis control program report. World Health Organization. Geneva, Switzerland.

[61]. WHO (2007). Global tuberculosis control program report. World Health Organization. Geneva, Switzerland.

[62]. WHO (2008). Global tuberculosis control program report. World Health Organization. Geneva, Switzerland.

Annex A

Cox proportional hazard model containing individual covariates

Table A.1: Cox proportional hazard model containing only covariate Sex

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	1.0599	1	0.3032
Score	1.0633	1	0.3025
Wald	1.0595	1	0.3033

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
sex	1	-0.20700	0.20110	1.0595	0.3033	0.813

Table A.2: Cox proportional hazard model containing only covariate Age

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	29.5815	1	<.0001
Score	35.2854	1	<.0001
Wald	33.4934	1	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
Age	1	0.04013	0.00693	33.4934	<.0001	1.041

Table A.3: Cox proportional hazard model containing only covariate Marital status

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	0.9396	2	0.3324
Score	0.9509	2	0.3295
Wald	0.9479	2	0.3302

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
mstatus	2	0.10078	0.10351	0.9479	0.3302	1.106

Table A.4: Cox proportional hazard model containing only covariate Localization of TB

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.8792	1	0.0017
Score	9.3010	1	0.0023
Wald	8.9391	1	0.0028

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
tbtype	1	0.69174	0.23136	8.9391	0.0028	1.997

Table A.5: Cox proportional hazard model containing only covariate Initial weight

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	13.3924	1	0.0003
Score	12.5229	1	0.0004
Wald	12.8168	1	0.0003

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
intwt	1	-0.04386	0.01225	12.8168	0.0003	0.957

Table A.6: Cox proportional hazard model containing only covariate Category

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	10.4114	2	0.0013
Score	11.0893	2	0.0009
Wald	10.7089	2	0.0011

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
Cat	2	0.33726	0.10306	10.7089	0.0011	1.401

Table A.7: Cox proportional hazard model containing only covariate Previous treatment

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2.1443	1	0.1431
Score	2.4099	1	0.1206
Wald	2.3728	1	0.1235

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
ptreat	1	0.43182	0.28033	2.3728	0.1235	1.540

Table A.8: Cox proportional hazard model containing only covariate HIV status

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	42.2844	1	<.0001
Score	46.8780	1	<.0001
Wald	40.5992	1	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
Hiv	1	1.32399	0.20779	40.5992	<.0001	3.758

Annex B

Cox proportional hazard model containing two covariates and their interaction

Table B.1: Cox proportional hazard model containing HIV status, Initial weight and their interaction

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	54.9187	3	<.0001
Score	55.9223	3	<.0001
Wald	44.7859	3	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
HIVint	1	0.04594	0.02622	3.0695	0.0798	1.047
Hiv	1	-0.86694	1.22413	0.5016	0.4788	0.420
intwt	1	-0.06698	0.02150	9.7041	0.0018	0.935

Table B.2: Cox proportional hazard model containing only HIV status and Initial weight

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	51.7705	2	<.0001
Score	55.6959	2	<.0001
Wald	49.9317	2	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
hiv	1	1.26632	0.20868	36.8220	<.0001	3.548
intwt	1	-0.03669	0.01204	9.2809	0.0023	0.964

Table B.3: Cox proportional hazard model containing Category, Localization of TB and their interaction

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	14.8300	3	0.0020
Score	14.5504	3	0.0022
Wald	13.7088	3	0.0033

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
catTB	1	-0.21359	0.29390	0.5282	0.4674	0.808
tbtype	1	0.58580	0.28662	4.1771	0.0410	1.796
cat	1	0.41856	0.26782	2.4424	0.1181	1.520

Table B.4: Cox proportional hazard model containing Category and Localization of TB

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	14.3345	2	0.0008
Score	14.4345	2	0.0007
Wald	13.7706	2	0.0010

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
Cat	1	0.23773	0.11286	4.4374	0.0352	1.268
tbtype	1	0.49084	0.25305	3.7624	0.0524	1.634

Table B.5: Cox proportional hazard model containing Localization of TB, Initial weight and their interaction

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	22.1102	3	<.0001
Score	20.0987	3	0.0002
Wald	18.5143	3	0.0003

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
TBint	1	0.02175	0.02822	0.5940	0.4409	1.022
tbtype	1	-0.38297	1.32565	0.0835	0.7727	0.682
intwt	1	-0.05736	0.02416	5.6348	0.0176	0.944

Table B.6: Cox proportional hazard model containing Localization of TB and Initial weight

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	21.5067	2	<.0001
Score	19.7766	2	<.0001
Wald	19.8776	2	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
tbtype	1	0.63070	0.23181	7.4024	0.0065	1.879
intwt	1	-0.04156	0.01238	11.2769	0.0008	0.959

Table B.7: Cox proportional hazard model containing HIV status, Age and their interaction

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	69.9285	3	<.0001
Score	69.6426	3	<.0001
Wald	54.7517	3	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
HIVage	1	-0.03835	0.01523	6.3375	0.0118	0.962
Hiv	1	2.63873	0.61670	18.3080	<.0001	13.995
Age	1	0.05480	0.01003	29.8599	<.0001	1.056

Table B.8: Cox proportional hazard model containing HIV status and Age

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	63.4959	2	<.0001
Score	69.4792	2	<.0001
Wald	61.4315	2	<.0001

Analysis of Maximum Likelihood Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**
Hiv	1	1.18805	0.20853	32.4582	<.0001	3.281
Age	1	0.03687	0.00756	23.7943	<.0001	1.038

Annex C

Figures for checking Linearity for transformed continuous covariate

Figure C.1: Plot of Martingale Residuals For Age

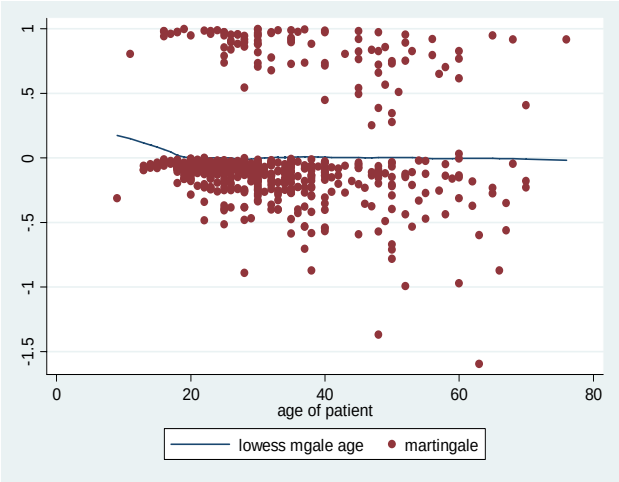
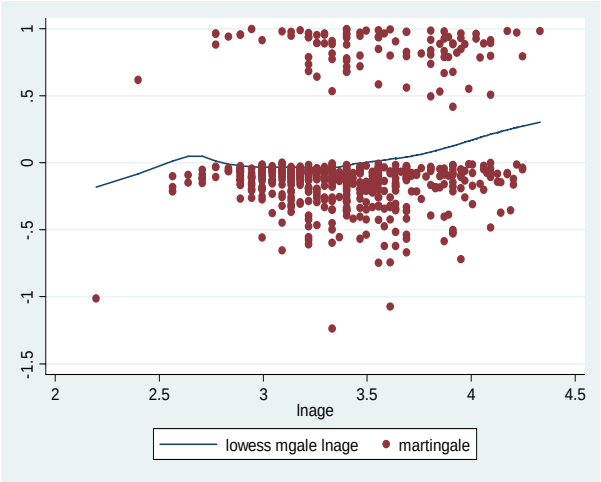


Figure C.2: Plot of Martingale residuals for Log of Age



FigureC.3: Plot of Martingale Residuals for Squareof Age

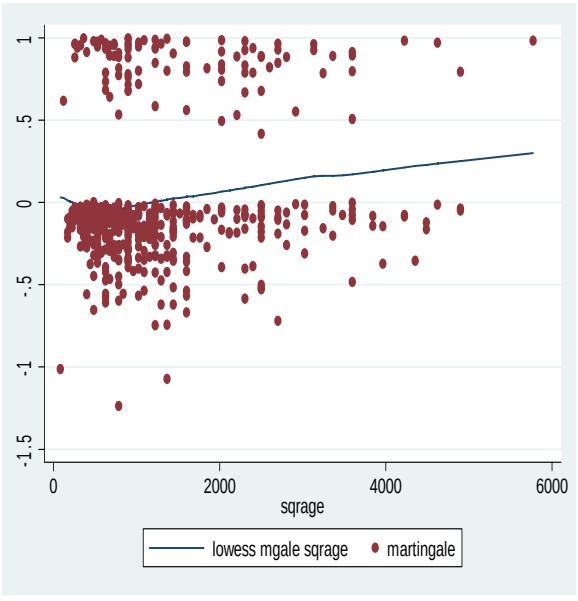


Figure C.4: Plot of Martingale Residuals for Initial weight

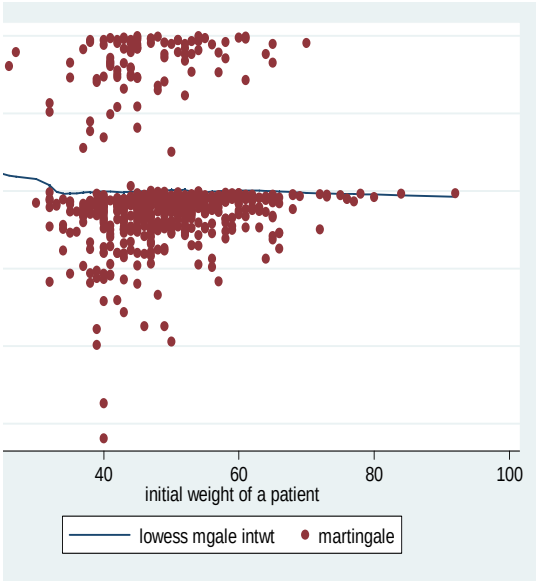


Figure C.5: Plot of Martingale Residuals For square of Initial weight

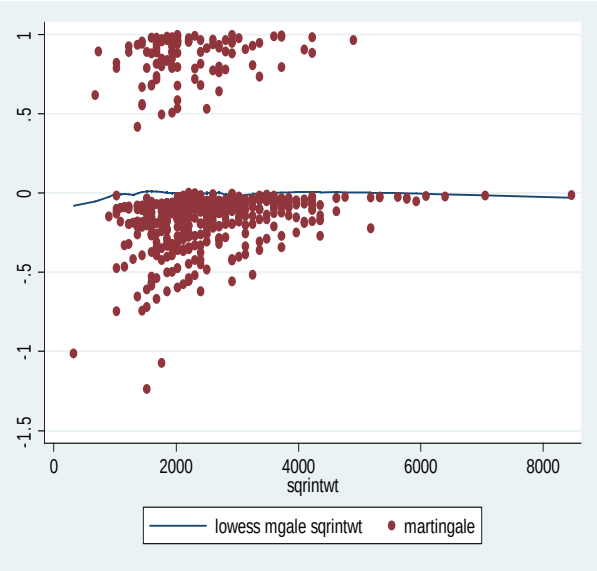
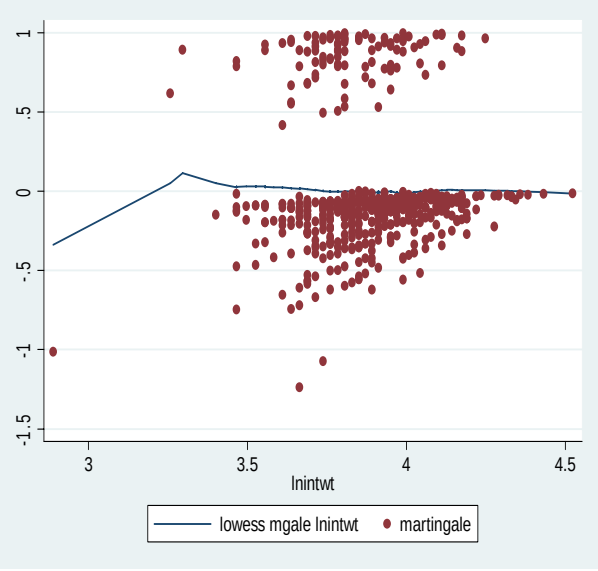


Figure C.6: Plot of Martingale Residuals for Log of Initial weight



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: Abrham Keraleme

Signature_____

Place of submission-DEPARTMENT OF STATISTICS,

SCIENCE FACULTY

ADDIS ABABA UNIVERSITY

Date of submission_____

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

Prof. Eshetu Wencheko

Signature_____