ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAMME
FACULTY OF SCIENCE
STATISTICS DEPARTMENT

SURVIVAL TIME OF HIV-INFECTED CHILDREN UNDER 15 YEARS OF AGE AFTER INITIATION OF ANTIRETROVIRAL THERAPY:

A CASE STUDY IN BAHIR DAR

By
Habtamu Atnafu

A thesis submitted to the school of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Applied Statistics

November, 2011

TABLE OF CONTENTS

Pages

ACKNOWLEDGMENTS

Blessed God!!! I thank you for this and for that.

I am extremely grateful to my sincere advisor Professor Eshetu Wencheko, for his invaluable help, beyond-mind tolerance, un-annoyingly read all my trials and corrected and guided me in to proper track. Had I been alone without him, nothing would have been done.

I would like to thank also Dr. Lynn Kay and Sr. Jacqueline Bernhard for their concern to this study especially I want to say thank you Dr. Lynn for your will and effort to help me with the thesis.

My sincere gratitude goes to my special friend Abraham Kebede, who keeps on encouraging and supporting me in all rounds.

I would like to acknowledge the spiritual support I got from my wife (Rahel) and son (new baby) to the completion of this study.

List of tables **Page**

List of figure **Page**

List of ACRONYMS

3TC	Lamivudine
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
AZT	Zidovudine
D4T	Stavudine
EFV	Efavirenz
FHAPCO	Federal HIV/AIDS Prevention and Control Office
FMOH	Federal Ministry of Health
I-TECH	International Education and Training Center on HIV
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
NVP	Neverapine
NIAID	National Institute of Allergy and Infectious Diseases
TB	Tuberculosis
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

ABSTRACT

Even though the use of Antiretroviral treatment has made a marked influence on the survival time of HIV-infected children, all children who received ART drugs do not have the desired level of immune reconstitution. The objective of this study is to assess the survival time of HIV-infected children after ART initiation and to identify factors other than ART that influence the survival time of HIV-infected children who continuously followed ART at Felege-Hiwot Hospital in Bahir Dar. In this study survival data analysis is used to assess survival length of 255 HIV-infected children who followed ART at Felege-Hiwot Hospital in Bahir Dar from 2006 to 2009 for 30 months. To compare the survival time of different groups of children defined by a factor, Kaplan-Meier survival function and the Log-Rank test are used. Cox proportional hazards model is used to model the relationship between significant explanatory variables and survival time. Of the 255 children in the study 71(27.84%) died during the study period. The overall estimated mean survival time of children was 22.381 months with standard error 0.693. Age group, Hemoglobin group and WHO clinical stage of HIV- infected children have been found to be significant factors that affect survival time of children. Therefore, special attention should be given to younger children in ART, children should start ART treatment when they have higher hemoglobin values and children should begin treatment when they are at a lower clinical stage.

ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAMME
FACULTY OF SCIENCE
STATISTICS DEPARTMENT

SURVIVAL TIME OF HIV-INFECTED CHILDREN AFTER INITIATION OF ANTIRETROVIRAL THERAPY:

A CASE STUDY IN BAHIR DAR

By
Habtamu Atnafu

Approved by the board of examiners:

_____	_____
Chairman, Department graduate committee	Signature
_____	_____
Examiner	Signature
_____	_____
Examiner	Signature

Tel. 251-1-239465 P.O.Box 1176 Addis Ababa, Ethiopia

DECLARATION

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

Name: Habtamu Atnafu

Signature: _____

Place: Faculty of Science, Addis Ababa University.

Date: _____

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

Name: Professor Eshetu Wencheko

Signature: _____

Date: _____

Chapter One

1. Introduction

1.1. Background of the study

Acquired Immune Deficiency Syndrome (AIDS) is a disease caused by a retrovirus known as human immunodeficiency virus (HIV). Although important progress has been achieved in preventing new HIV infections and in lowering the annual number of AIDS related deaths, the number of people living with HIV continues to increase. AIDS-related illnesses remain one of the leading causes of death globally and are projected to continue as a significant global cause of premature mortality in the coming decades (UNAIDS 2008). The virus attacks the immune system and weakens the body's natural defense system to fight against infection.

According to UNAIDS, around 33.4 million people are living with HIV throughout the world, including the approximately 2.7 million newly infected in 2008 and over 2 million have lost their lives to the disease leaving behind orphaned children and ravaged communities in 2008. The greatest burden of the disease is concentrated in developing countries with the least ability to cope. The countries of sub-Saharan Africa are home to approximately 22.4 million people living with HIV/AIDS (Joint UNAIDS 2009).

The HIV pandemic created unprecedented burden on the economies and health care systems of affected countries, particularly in Sub-Saharan Africa, where prevalence is highest. The total number of children less than 15 years of age living with HIV/AIDS in 2008 worldwide is 2.1 million including 430,000 newly infected of which 390,000 new cases are in Sub-Saharan Africa during 2008 (UNAIDS 2009). In Ethiopia the total number of people who

have died due to HIV/AIDS in 2006 alone was 88,997 and in 2007 it was estimated that 71,902 people would die (FMOH, 2007). In 2010, AIDS related death is expected to decline to 28,073 which might be as a result of ART. Currently an estimated 1,216,908 people are living with HIV/AIDS of which 79,871 are children. It is estimated that 397,818 of the HIV positive cases are in need of ART out of which 26,053(6.5%) are children under 15 years of age. It is also estimated that the adult (15-49) HIV prevalence in Amhara region in 2010 is 2.9% with 155,694 male and 223,372 female cases who live with the virus, this may increase the number of HIV positive children in the region due to mother to child transmission at birth and/or during breast feeding (National AIDS resource center 2010).

Recognizing the urgent need for antiretroviral treatment, the government of Ethiopia issued the first antiretroviral guideline in 2003, which is the same year as the antiretroviral treatment has began. The first antiretroviral guideline was revised in 2005 to facilitate a rapid scale up of ART. Within two years patients on treatment at 117 hospitals and 108 health centers increased from 900 to 62,221 (FMOH 2007) and in 2008 a total of 400 health care facilities, 277 public health centers and 123 public and private hospitals, are rendering ART services in the country. The number of people who were able to access ART has substantially increased from 900 in 2003 to 180,447 in 2008 and in 2010 as part of the global issue the government of Ethiopia is working to provide universal access to HIV/AIDS treatment (Seyoum *et al.* 2009).

The use of highly active antiretroviral therapy (HAART) has resulted in a significant reduction in AIDS-related deaths and complications among adults and adolescents. However, the medical management of HIV-infected children remains challenging. Access to HIV treatment is limited and early treatment initiation can cause serious complications. Since

there is currently no cure for the disease HIV, a balance between treating the disease and maintaining quality of life must be weighed carefully. An evaluation to determine an appropriate time to initiate HAART is necessary to improve both the quality of life and survival of HIV infected children (NIAID 2008).

Comprehensive care and support for HIV-infected children should be provided in a care-and -treatment centre, preferably where the patients receive treatment. Close regular follow-up is important since these children are at risk of morbidity and mortality. Mortality estimates in Africa show that without treatment 35.2% of HIV-infected children will die in their first year and 52.5% by age two. This underscores the importance of timely antiretroviral treatment -care and -support (FMOH 2007).

In resource-rich setting HAART has changed the face of paediatric HIV. HIV-infected children now survive to adolescence and adulthood. However, the experience in paediatric HIV/AIDS care and treatment in Ethiopia is limited (FMOH, 2007). Major obstacles to scaling up paediatric care include: lack of human resources and scarcity of paediatric providers, no systematic effort to identify and follow HIV exposed infants, limited availability of virologic testing, lack of provider-initiated HIV testing, missed opportunity for testing children, insufficient advocacy and understanding that ART is efficacious in children, and limited experience with program implementation to provide paediatric HIV/AIDS care and treatment. Consequently too few children could start on ART in Ethiopia (FMOH 2007).

This research work is undertaken against the above background and explores the factors that have strong association with the survival experience of HIV-infected children who started ART in one of the government hospitals in the regional state of Amhara at Bahir Dar in Felege-Hiwot Hospital.

1.2. Statement of the problem

All children who received ART drugs do not have the desired level of immune reconstitution as shown by the increase in CD4 lymphocyte counts, viral load depletion, decreased risk of opportunistic infection, decreased risk of death and improved survival. Some patients show more rapid immune reconstitutions while others have slow and incomplete immune reconstitutions and remain with greater risk of AIDS-associated events and death. Moreover, patients may die with undetectable viral load and adequate CD4 count recovery. Therefore, this research paper attempts to get a closer insight if other factors that contribute to the above variation and to study empirically the effect of different ART drugs on the survival of children.

1.3. Objectives of the study

The main objective of this study is to assess the survival of children after they started ART at Felege-Hiwot hospital in Bahir Dar city, which is one of the cities in Ethiopia, where HIV has high prevalence (National Factsheet, 2010)

In the light of this major objective, the study intends to:

- identify factors other than ART that affect the survival time of children after initiation of ART
- examine the effect of different ART drug combinations on the survival time of children

1.4. Significance of the study

It is not a hidden fact that HIV/AIDS pandemic has caused a serious burden on the economic, social and political issues of the world in general, and in developing countries in particular. Currently, countries including Ethiopia are trying to provide ART drugs to people who are living with HIV/AIDS, in order to decrease opportunistic infectious diseases, HIV/AIDS-related death and to improve quality of life to those infected with HIV. However, very little attention has been given to look into how factors other than ART drugs may influence the survival of HIV infected children. This study is, therefore, undertaken with the aim to provide information about factors that influence the survival of HIV infected children.

Chapter Two

2. Literature Review

2.1. HIV/AIDS

The risk of HIV infection increases by number of sexual partners, intravenous drug use, any sex without condom, alcohol and other drug use, tattoos and body piercing with contaminated needles or instruments. Since AIDS was first identified, researchers and scientists have tried their best to find medicine and vaccine but the results everywhere were not successful. Consequently the lives of human beings have been observed to be threatened due to infection with the virus. On top of all, since the disease is affecting the most productive citizens, it is natural that it causes damage to national economy (Bollinger 1999; UNAIDS 2000).

2.2. HIV/AIDS in Ethiopia

The first evidence of HIV epidemic in Ethiopia was detected in 1984. Since then, AIDS has claimed the lives of millions and has left behind hundreds of thousands of orphans (FHAPCO, 2007). Ethiopia is one of the hardest hit Sub-Saharan Africa countries by the HIV pandemic. In 2009 a single point estimate of AIDS related deaths for 2010 was 44,751 of which 7,214 were children and there were about 84,189 HIV positive pregnancies from which 14,140 gave birth to HIV positive children and this resulted in an estimated increase in the number of HIV positive children from 72,945 in 2009 to 79,871 in 2010 (National AIDS resource center 2009).

According to the National Factsheet, 2010 of National AIDS resource center the total number of HIV positive people in 2010 is estimated to be 1,326,329 including 137,494 new HIV infections and excluding 28,073 AIDS related deaths during the year. It is also estimated that

a total of 90,610 HIV positive children under 15 years of age including 14,276 new infections and excluding 3,537 AIDS related deaths of children during the year.

According to the FHAPCO single point estimate for prevalence of HIV/AIDS in Ethiopia, the adult (15-49) HIV prevalence for 2007 is estimated at 2.1% of which 7.7% is urban and 0.9% is rural (EMOH 2007). In 2010, the FHAPCO estimates of the overall adult (15-49) HIV prevalence is 2.4%. Urban and rural HIV prevalence rates were 7.7% and 0.9%, respectively. In 2010, an estimated 28,073 Ethiopians died of AIDS scaling the number of children who have lost one or both parents to AIDS to 804,184 (National Factsheet 2010).

HIV/AIDS has been and still is the greatest challenges to the Ethiopian health system, as elsewhere in sub-Saharan African countries. It has remained among the major causes of deaths over the past two decades. Currently in 2010, more than one million people are estimated to be living with HIV in Ethiopia of whom nearly 397,818 need ART care and treatment (National Factsheet 2010).

2.3. Goals of Antiretroviral Therapy (ART)

Infection with HIV causes immunologic deficiency that results in depletion of CD4 cells and suppression of cell mediated immune defenses. HIV infects CD4 cells by interacting with their CD4 receptors, which allows the virus to gain entry into the cells. The invading HIV replicates within the CD4 cells, destroys them and spreads to other CD4 cells, depleting the CD4 cell population. Because CD4 cells direct and activate immune responses, many immune functions degrade as a result of HIV infection (Virco 2008).

Individuals with weakened immune defenses are susceptible to infections caused by opportunistic pathogens that do not normally cause disease for immune competent individuals. These infections are known as opportunistic infections. Once an opportunistic

infection has begun, it can rapidly spread throughout the body via the circulatory system, damages vital organs and becomes fatal (Virco 2008).

The goals of therapy in treating HIV/AIDS infected individuals are, therefore, in view of the following points:

- **Clinical goal** which aims to extend life expectancy and quality of life for patients infected with HIV,
- **Virological goal** is to reduce the HIV viral load to the lowest level possible in order to prevent disease progression and limit development of resistance to ARV drugs,
- **Immunological goal** is to preserve and restore immunologic functioning in the normal range. This involves the quantitative component of CD4 cell count in the normal range. It also involves the qualitative goal of resisting infections by opportunistic pathogens, and
- **Epidemiological goal** is to reduce transmission of HIV to others. (Virco 2008).

The clinical symptoms of HIV infection were evolved on the depletion of CD4 cells and the replication of HIV RNA. Therefore, CD4 cell count and HIV viral load are two of the most important predictors of the clinical prognosis of HIV-infected subjects.

Highly active antiretroviral therapy (HAART) and a single or combination of several drugs, have high activity to inhibit HIV RNA replication. HIV cocktail therapy is a combination reagents which inhibit the replication of HIV RNA at different stages of HIV life-cycle. The currently available HIV inhibition reagents can be categorized into Nucleotide reverse transcriptase inhibitors, Non-Nucleotide reverse transcriptase inhibitors, protease inhibitors, and integrated zinc-finger inhibitors (Zhang 2007).

2.4. Importance of ART in Ethiopia

In the face of competing demands such as malaria, TB, and famine some question whether an investment on ART in Ethiopia is justifiable. Given the impact of AIDS across society and the potential of ART to reduce the burden, the justification for pursuing this agenda is unarguable. HIV and AIDS are affecting every sector of Ethiopian society. At the macro level, the health, agriculture, education, business and industry sectors are all adversely impacted by the disease. Families and communities are likewise affected. The MOH estimate that the annual mortality rate for those in the 15-49 age range will increase from a projected 200,000 without factoring in AIDS to over 350,000 in 2004 (with the AIDS epidemic). Besides these numbers, the impact of AIDS increased absenteeism in the workplace, reduction of productivity, reduced family income, and increased family expenditure on health care and burial rituals. In a resource poor country such as Ethiopia, the economic impact of AIDS-related illness and death is severe (AIDS Resource Center *et al* 2005).

In contrast to Ethiopia**,** AIDS-related deaths and illnesses in countries where ART has been available since the mid 1990s have considerably declined (UNAIDS 2004). The experience of developed nations, as well as countries such as Brazil (from middle income countries which produces ART drugs), have proven that ART treatment reduces disease burden and dependence, and increases the function, well-being, and productivity of individuals. This in turn, can help offset some of the consequences of the HIV and AIDS pandemic (AIDS Resource Center *et al* 2005).

2.5. General Overview of the Use of Antiretroviral Therapy

During the year 1996 the advent of potent Highly Active Antiretroviral Therapy (HAART) that met the above mentioned goals led to a revolution in the care of patients with HIV/AIDS in the developed world. These treatments have dramatically reduced rates of mortality and

morbidity and have improved the quality of life of people with HIV/AIDS. Most importantly, through antiretroviral therapy, HIV/AIDS which was once an infectious disease with an almost universally fatal outcome has been transformed into a manageable chronic infectious disease (Palella 2003 as cited in Selamawit 2009).

A study which includes 285 HIV-positive children treated with first-line ART for at least 24 months to identify risk factors associated with treatment failure in two hospitals of Cambodia at Angkor Hospital for Children and Donkeo Referral Hospital, has shown that ART treatments have the desired effect to improve the survival of children, CD4 count progress, viral load suppression and weight for age Z-score increment, for children with viral load less than 1000 copies/ml. On the other hand for children associated with viral load greater than 1000 copies/ml only CD4 counts and CD4 percentages below the threshold is responsible for severe immunologic suppression at 24 month which predicts treatment failure at this time. However, orphan status and caregiver characteristics, including literacy, age and socio-economic status, were not associated with treatment failure after 24 months of ART (Isaakidis 2010).

A retrospective cohort study was designed to assess clinical factors associated with growth in HIV infected children on ART in Uganda. Height and weight measurements were taken pre-ART and post-ART initiation for at least 6 months from 749 children included in the study. Descriptive and logistic regression analyses were conducted to identify covariates associated with risk of either stunting or being underweight. Children in World Health Organization (WHO) clinical-stages II, III, and IV at baseline were 1.5 times more likely to become underweight from that of Clinical stage I, but Initiation of ART resulted in improvement in mean standardized weight-for-age. Weight-for-age z-score improved significantly after

initiation of ART. This pediatric population gained weight more rapidly than height after initiation of ART (Kabue 2008).

In a study to evaluate changes and risk factors for death among HIV-infected children in Paediatric AIDS Clinical Trials Group 219/219c in US among 3553 HIV-infected children was followed up for a median of 5.3 years. The study shows that increased risk of death was significantly associated with low CD4, pneumonia and AIDS-defining illness at entry. Whereas, decreased risks of mortality were identified for children timely began highly active antiretroviral therapy (Brady *et al.* 2010).

On the other hand a case control study conducted by Fontana and colleagues (1999), to see body composition in HIV infected children in relation with disease progression and survival using a total of 86 HIV infected and 113 HIV uninfected children showed that weight in HIV infected children was significantly less than in a control children with similar age; moreover weight were significantly associated with increased risk of death. Therefore the authors conclude that body weight for age is a good prognostic indicator (Fontana *et al.* 1999).

The department of paediatrics in New York University, New York, U.S.A. has conducted a study to measure the efficacy of highly active antiretroviral therapy in HIV infected children in resource poor setting. To measure the clinical, immunological, and virological effects of HHART on HIV-infected children in Mombasa, Kenya, data were taken and analyzed from 29 children. The result reveals that weight for age and CD4 cell have increased stepwise, while viral load decreased from a baseline by a factor more than 17 times (Song *et al.* 2007).

Another study at primary care clinics in Lusaka, Zambia, held to demonstrate the clinical and immunological outcomes on 2938 children enrolled in a paediatric treatment program who started ART have verified that mortality rate was significantly associated with: CD4 cell depletion, lower weight for age, younger age, and anemia. And the mean CD4 cell percentage at ART initiation for more than 53% of children who had at least one repeat measurement has shown an increasing trend as measured every six months after initiation of ART (Bolton-Moore *2*007).

A study conducted by George and colleagues (2007), on 236 HIV infected children taking ART in Port-au-Prince, Haiti between 1 May 2003 to 30 April 2006 to assess the outcome of paediatric ART using retrospective data, demonstrated that age below 18 months, CD4T cell percentage and weight for age Z score were significant baseline predictors of child mortality (George *et al*. 2007).

A study was conducted on children aged 12 months and older from 10 studies (nine African, one Brazilian) to evaluate the prognostic value of selected laboratory and growth markers on the short-term risk of mortality in a total of untreated HIV-infected children in resource-limited settings. The result of this study has shown that, CD4% and CD4 cell count were the strongest predictors of mortality, followed by weight-for-age and hemoglobin (Gibb *et al* 2008).

In another study by Walker and colleagues (2007) to see the impact of daily contrimoxazole prophylaxis and antiretroviral therapy on hospital admission of HIV infected Zambian children using randomized control trial, it was reported that mortality rate was high among children who had lowest current weight for age. The authors conclude that, low weight for age score are independent predicator of death.

Another study from Cape Town on children known to be vertically infected with HIV has revealed that the risk of death was significantly associated with age less than 6 months and severity of disease at time of diagnosis. The median survival for all the children from time of diagnosis was 32 months. Infants diagnosed before 6 months of age had significantly shorter median survival (10 months) compared with 36 months for those diagnosed at 7-12 months of age. For the children over the age of 12 months the cumulative proportion surviving 48 months was 78%. Children with severe disease (category C) had a median survival of 21 months, significantly lower than that in category B (32 months). For the children in category A the cumulative proportion surviving at 48 months was 66% (Hussey *et al* 1998).

The introduction of ART presents an enormous opportunity in terms of reducing morbidity and mortality due to AIDS worldwide. Ethiopia has been engaged in the scale up of ART access to its people since 2005. The free ART program was launched in July 2005. Despite the many challenges, ART scale up has recorded the greatest achievement over the last few years. The service has been expanded from only three health facilities in 2005 to 400 in 2008. The number of people who started on ART has also shown an unprecedented increase during the same period from 900 in 2005 to 180,447 by end of December 2008 (Seyoum *et al*. 2009).

A study by Selamawit Ejigu (2009) on 423 patients identified factors on those patients at a risk of treatment failure has shown that the mean survival time (without treatment failure) was 53 months. Females were found to have a higher survival time of 57 months and males have significantly higher risk of developing treatment failure. Those with two or more episodes of poor adherence during their follow-up have a significantly higher failure compared to those with no episode of poor adherence. Missed appointment is another

independent predictor of treatment failure. The study has shown that non-adherence to medication and clinic visits are associated with treatment failure. Following patients closely for their level of adherence and their trend of missing clinic visits can be used to help identify those at higher risk of treatment failure. Providing intense adherence counseling for these patients may prevent occurrence of failure.

Another study by Binyam Taye (2007) that has been undertaken to assess the impact of malnutrition in survival of HIV infected children after initiation of ART based on 475 HIV infected children has shown that CD4 count, hemoglobin value and weight for age were baseline predictors despite the obvious benefit of ART use on HIV related survival.

According to the studies above, the survival time of HIV infected patient after initiation of ART is a function of baseline variables like CD4 count, clinical stage of the disease, weight, age, drug adherence, types of treatment (mono-therapy, bi-therapy or triple therapy), viral load, nutrition, hemoglobin value and so on. This study is designed to 'identify additional' factors that affect the survival time of HIV infected children after initiation of antiretroviral treatment (ART) as a case study in Felege-Hiwot hospital ART unit at Bahir Dar city.

Chapter Three

3. Methodology

3.1. Background information of study Area

The study was conducted in Bahir Dar, the regional state of Amhara which is 560 Km away from the capital city of Ethiopia, Addis Ababa, in North West Ethiopia. Felege Hiwot Referral Hospital is a government hospital in Bahir Dar which is one of the three referral hospitals in Amhara Regional State of Ethiopia. The hospital provides clinical care for patients infected with HIV/AIDS including children since 2005 free of charge to patients.

In Bahir Dar the prevalence of HIV/AIDS is too high. The prevalence of the disease, particularly among children causes a major problem that needs priority. Among pregnant women at the anti-natal care unit of the hospital, for instance, the prevalence of HIV/AIDS is one of the highest compared with other cities in Ethiopia (National Factsheet, 2010).

3.2. Variables in the study

Variables to be considered in this study were selected based on another similar study (Binyam, 2008). Additionally other variables that were expected to be factors that affect the survival time of HIV-infected children after they have begun ART in Bahir Dar town at Felege-Hiwot referral hospital were included.

3.2.1. The response variable

The response variable in this study was the survival time of HIV-infected children after starting ART. This was measured as the time a child was followed up from the time the child

began to receive treatment, until the time to an event (death) or they were lost to follow up (for those right censored subjects).

3.2.2. Explanatory variables/factors

The independent variables included in the study were factors that were thought to influence the survival status. Even though it would have been good to include different socio-economic and virological variables, that can affect the survival time of HIV-infected children, because of limited resources the study focused only on clinical and immunological characteristics. However, as similar studies have shown the explanatory variables/factors included here were expected to show significant influence on the survival time of HIV-infected children.

Table 3.1. Explanatory variables/factors included in the study

Explanatory variables / factors	Description	Value labeled
X_1	Age	Ages less than 1.5 years = 0 Ages between 1.5-5 years = 1 Ages between 5 - 14 years = 2
X_2	Gender	Female = 0 Male = 1
X_3	WHO Clinical Stage	Clinical Stage- IV = 0 Clinical Stage- III = 1 Clinical Stage- II = 2 Clinical Stage- I = 3
X_4	CD4 count Cut off point for CD4 lymphocyte counts for different age groups were those that related to the category of severe	Above the threshold = 0 Below the threshold =1

	immunodeficiency: For those younger than one year : less than 1500 cell/µl For those between 1-3 years age :less than 750 cell/µl For those between 3-5 years age :less than 350 cell/µl For those older than 5 years age :less than 250 cell/µl are considered as below the threshold.	
X_5	Prophylaxis taken	No = 0 Yes = 1
X_6	Reported functional status Working: means actively participates in age appropriate activities. Ambulatory: means limited tolerance for activities Bedridden: means no tolerance for activities.	Working = 0 Ambulatory = 1 Bedridden = 2
X_7	Weight Normal weight for age: the weight of a child is within 2 standard deviations from the median weight of normal children of the same age. Low weight for age: the weight of a child is below 2 standard deviations from the median weight of normal children of the same age. Very low weight for age: the weight of a child is below 3 standard deviations from the median weight of normal children of the same age.	Normal weight = 0 Low weight for age = 1 Very low weight for age = 2
X_8	Hemoglobin	Less or equal to 7.00 gm/dl = 0 7.00-8.50 gm/ dl =1 8.50- 10.00 gm/dl =2 More than 10.00 gm/dl = 3

X_9	Reason for starting ART	CD4 count only = 1 Clinical only = 2 CD4 + clinical = 0
X_{10}	Type of ART drug	D4t-3TC-NVP = 0 D4t-3TC-EFV = 1 AZT-3TC-NVP = 2 AZT-3TC-EFV = 3
X_{11}	TB treatment at start	Yes = 0 No = 1

3.3. The data

The study was based on secondary data from hospital records on all children who had enrolled in Felege-Hiwot referral hospital pediatric ART unit from 2007 to early 2009. The hospital serves as a referral hospital for nearby lower level hospitals and health centers. Bahir Dar Felege-Hiwot hospital provides care to HIV-infected adults and children through the government free antiretroviral treatment program since 2005, with additional support from the nongovernmental organization such as I-TECH Ethiopia. Bahir Dar Felege-Hiwot hospital ART unit is well staffed with ART trained physicians, nurses, counselors and laboratory technicians working full time. The hospital is under the Amhara Administration Health Bureau and gets technical and financial support from I-TECH.

This study used data obtained from patients' follow up records of HIV cohort database which we filtered by age to include children of age below 15 years. Data we`re collected based on the child's identification number without any direct contact with a child so as to maintain the confidentiality of the child's record. The total number of children on ART included more than 819 children from the HIV cohort database. Demographic data, laboratory and clinical

information of all HIV-infected children who were being followed-up between the years 2007 to 2009 were recorded. However, only 255 patients with a full record of variables were included in the study.

3.4. Methods of Descriptive Statistical Data Analysis

Before beginning any statistical investigation, we should perform a thoughtful and thorough univariate and bivariate analysis of the data so as to obtain a clear sense of the distributional characteristics of the outcome variable and to identify explanatory variables/factors that explain the survival time of children. For survival data, however, some of the observations were incomplete (because of censoring). Conventional univariate statistics like mean, median, standard deviation are not appropriate in a study of survival time.

Hence, we should look for other approaches to overcome the problem of censoring. In this situation we must obtain functions of survival time, T, in order to obtain estimates of the parameters of interest.

In what follows we introduce the cumulative distribution function F(t), survival function S(t), the density functions f(t) and hazard function h(t) that are functions defined for survival time T.

The cumulative distribution function is a fundamental building block of survival data analysis when the data include censored observations. The cumulative distribution function is a probability that a subject selected at random will have a survival time less than some stated value, *t*.

Let *T* be a random variable representing survival time of subjects in the population, and *t* be the realization of *T*. Then the cumulative distribution function of *T* is given as

$$F(t) = \Pr(T < t).$$

The survival function of *T*, S(*t*) gives the proportion of children whose survival times are *t* or longer. S(*t*) is the proportion still alive at time *t*.

$$S(t) = Pr(T \geq t) = 1 - F(t).$$

The probability density function (pdf) of *T*, f(*t*) gives the probability of deaths occurring at time *t*.

This is the chance that a child chosen at random will die at time *t*.

$$f(t) = \frac{dF(t)}{dt} = lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P[t < T \leq t+\Delta t]}{\Delta t} = \frac{d\{1-S(t)\}}{dt} = \frac{-dS(t)}{dt}.$$

The other function associated with the random variable *T* is the hazard function. The hazard function h(*t*) gives the probability of death during a short interval after time *t*, for those children who are alive at time *t*. The hazard function h(*t*), is the conditional probability density function, which is the instantaneous rate of failure/death for a child at time *t* given that a child alive at time *t*.

$$h(t) = \frac{\text{probability of a death at time } t}{\text{proportion alive at time } t}$$

$$= lim_{\Delta t \to 0} \frac{P[t < T \leq t+\Delta t \mid T \geq t]}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P[t < T \leq t+\Delta t] / P[T \geq t]}{\Delta t}$$

$$= lim_{\Delta t \to 0} \frac{P[t < T \leq t+\Delta t] / \Delta t}{P[T \geq t]} = \frac{f(t)}{S(t)}$$

We estimate the survival experience of different groups of explanatory variables through a non-parametric method called Kaplan-Meier survival function estimator (Collett 2003).

3.4.1. The Kaplan-Meier estimator of the survivorship function

The Kaplan-Meier estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps defined by the observed survival and censored times. The steps are intervals defined by a rank ordering of survival times. Each interval begin at an observed time and ends just

before the next ordered time and is indexed by the rank order of the time point defining its beginning.

The Kaplan-Meier estimator at any point in time is obtained by multiplying a sequence of conditional survival probability estimators. Each conditional probability estimator is obtained from the observed number at risk of dying (n) and the observed number of deaths and is equal to $(n - d)/n$.

The Kaplan-Meier estimator of the survivorship function at time *t* is obtained from the equation:

$$\widehat{S_{KM}(t)} = \prod_{t(i)\leq t} \frac{ni-di}{ni} \quad \text{with} \quad \widehat{S(t)} = 1 \text{ if } t \leq t_{(1)}.$$

The Kaplan-Meier estimator of the survivorship functions has to be obtained separately for each group of study subjects.

Other estimator of survivorship function is $\hat{S}(t) = e^{-\hat{H}(t)}$ where $\hat{H}(t) = -\int_0^t h(u)du$.

The function $\hat{H}(t)$ is an important analytical tool in the analysis of survival time data. It plays a central role in regression modeling of survival time data and is called the cumulative hazard function. To see how survival function related with the cumulative hazard function as given above, let us begin at the hazard function h(t) as:

$$h(t) = \frac{f(t)}{S(t)}$$

$$h(t) = \frac{1}{S(t)} \cdot \frac{-dS(t)}{dt} = \frac{-dlnS(t)}{dt}$$

$$-h(t) = \frac{dlnS(t)}{dt}$$

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

$$S(t) = e^{-\int_0^t h(u)du}$$

$$S(t) = e^{-H(t)}$$ (Hosmor & Lemeshow 1998).

This expression suggest that if we can specify the hazard function, h(t), then it is relatively easy to obtain an expression for any of the other functions of survival time T. Hence, we need to specify a regression model type of the hazard function as a function of time and covariates.

3.4.2. Regression model for the hazard function

There are a number of statistical models that can be used to predict a response variable from a set of explanatory variables/factors. Statistically, the specification of a model requires choosing both the systematic and error components.

- The choice of the systematic component involves an assessment of the relationship between an 'average' of the outcome variable and the explanatory variables/factors. This may be determined by an exploratory analysis of the current data and/or past experience.
- The choice of an error component involves specifying the statistical distribution of what remains to be explained after the model is fit.

In some applied setting the model must describe both the basic underlying distribution of the response variable (error component) as well as characterize how that distribution changes as a function of covariates (systematic component) and in another setting a model that addresses the latter one is sufficient.

In this study we need to specify a model that characterizes how the distribution changes as a function of covariates. That is, the systematic component only is to be used, and therefore the assumptions required for the error component become unnecessary. Models used to describe the survival experience in a comparative sense are often called semi-parametric regression models (Hosmer & Lemeshow 1998).

In this setting the end products of the statistical analysis are estimated parameters that compare the survival experience of selected subgroups, and they can be handled by specifying the hazard function. By specifying a model through the hazard function, we may address specific questions such as how survival is related to the treatments under study and other subject characteristics. A regression model for the hazard function

$$h(t, \mathbf{X}, \boldsymbol{\beta}) = h_0(t)r(\mathbf{X}, \boldsymbol{\beta})$$

This hazard function is the product of two functions. The function, $h_0(t)$, characterizes how the hazard function changes as a function of survival time. $r(\mathbf{X}, \boldsymbol{\beta})$ characterizes how the hazard function changes as a function of subject covariates.

The function must be chosen such that $h(t, \mathbf{X}, \boldsymbol{\beta}) > 0$. Note that $h_0(t)$ is the hazard function when $r(\mathbf{X}, \boldsymbol{\beta}) = 1$. $h_0(t)$ is referred to as the baseline hazard function. Under the above model the ratio of the hazard functions for two groups of the study subjects with covariate values denoted X_1 and X_0 is:

$$HR(t,X_1,X_0) = \frac{h(t,X1,\beta)}{h(t,X0,\beta)} = \frac{h_0(t)\, r(x1,\beta)}{h_1(t)\, r(x0,\beta)}$$

$$= \frac{r(X1,\beta)}{r(X0,\beta)}$$

The hazard ratio (HR) depends only on the function r(X, β). When the ratio function $HR(t,X_1,X_0)$ is easily interpreted, the actual form of the baseline hazard function is of little importance. This leads to assume that different groups of the study subjects have similar shaped hazard functions. That is, the two groups have constant relative risk over time (Hosmer & Lemeshow 1998).

3.4.3. Cox-proportional hazards regression model

The Cox proportional hazards regression model is a model which assumes that the log of hazard rate is related additively with a function of covariates or hazard rate is related multiplicatively with a function of covariates. The hazard function h(t, **X**, **β**) is related to the covariates as a product of baseline hazard $h_0(t)$ and a function on covariates r(**X**, **β**).

Cox (1972) was the first to propose r(**X**, **β**) = exp(**X'β**) and the baseline hazard is left unspecified as nuisance parameter in the model. Therefore the Cox proportional hazard function is given as

$$h(t, \mathbf{X}, \boldsymbol{\beta}) = h_0(t) \exp(\mathbf{X'\beta})$$

Note $h_0(t)$ is the hazard function, where all values of the covariate are zero, i.e. r(X=0, β) = 1.

One way to specify the distribution of survival time is through the hazard function. Thus, the survival function given by the Cox proportional hazard function is

$$S(t, \mathbf{X}, \boldsymbol{\beta}) = [S_0(t)]^{exp(X'\beta)}.$$

where $[S_0(t)] = e^{-H_0(t)}$ is the baseline survivorship function.

3.4.4. Assumptions Concerning to Cox-proportional hazards regression model

The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model. Specifically the Cox proportional hazards model has a hazard function of the form h(t, **X**, **β**) = $h_0(t)$exp(**X**, **β**). This function has two parts, the baseline hazard function, $h_0(t)$, and the exponential predictor, exp(**X**, **β**).

Cox proportional hazards regression assumes that different groups have similar shaped hazard functions or assumes that any two groups have constant relative risk over time. That is,

1. Study group *j* has a hazard function h_j (*t*, **X**, **β**) that is a positive multiple of the baseline hazard as: $h_j(t, \mathbf{X}, \boldsymbol{\beta}) = h_0(t)\, exp(\mathbf{X}, \boldsymbol{\beta})$.
2. There is a baseline hazard function $h_0(t)$ common to all individuals in all the study groups.
3. Explanatory variables act only on the hazard ratios (relative risks) *exp*(**X**, **β**), and they do not affect the baseline hazard.

3.5. Comparison of survival functions

The very right place to begin comparing the survival experience between groups of study subjects is the plot of the survival functions of $\widehat{S_{KM}(t)}$ for each group of the study subjects. At any point in time the proportion of subjects estimated to be alive is greater for groups represented by the upper curve than the other group represented by the lower curve. The statistical question is whether the observed difference in the graph is significant or not. A test procedure must be developed for judging whether there is a statistically significant difference in survival experience between groups or not. The test statistic used is obtained by computing the difference between the observed and expected number of deaths in each group of the study subjects. The table below is used for testing the equality of survivorship functions in *K* groups at observed survival time t_i.

Group Event	1	2	3	…	*k*	….	*K*	Total
death	d_{1i}	d_{2i}	d_{3i}	…	d_{ki}	….	d_{Ki}	d_i
alive	$n_{1i}- d_{1i}$	$n_{2i}- d_{2i}$	$n_{3i}- 2d_{3i}$	…	$n_{ki}- d_{ki}$	….	$n_{Ki} - d_{Ki}$	n_i-d_i
at risk	n_{1i}	n_{2i}	n_{3i}	…	n_{ki}	….	n_{Ki}	n_i

Then the hypothesis to be tested becomes

H_0: $S_1(t) = S_2(t) = S_3(t) = \ldots\ldots = S_K(t)$

against H_1: Not H_0.

Of the various test statistics the Log-rank test statistic considered here is

$$Q_{LR} = \sum_{k=1}^{K} \frac{(d_{k+} - \hat{e}_{k+})^2}{\hat{e}_{k+}}.$$

$Q_{LR} \sim \chi^2_{(k-1)}$ under the null hypothesis.

In the above $d_{k+} = \sum_{i=1}^{m} d_{ki}$ and $\hat{e}_{k+} = \sum_{i=1}^{m} e_{ki}$.

The test rule used is

There is no statistically significant difference in survival experience between groups at α – level of significance if the computed value of $Q_{LR} < \chi^2_{(k-1)}$ i.e. do not reject H_0.

There is statistically significant difference in survival experience between groups at α –level of significance if the computed value of $Q_{LR} \geq \chi^2_{(k-1)}$ i.e. reject H_0.

3.5.1. Fitting the Cox-proportional hazards regression model

Consider the probability that a child experience the event at time t_i given that an event occurs at that time. Let R_i denote the set of children at risk at time just prior to t_i. Then the probability that child *i* with covariates $\mathbf{X_i}$ would experience the event at time t_i is given as:

$$Pr_i = \frac{Pr(individual\ i\ has\ an\ event\ at\ time\ t_i\ |Survival\ at\ time\ t_i)}{Pr(one\ event\ at\ time\ t_i\ |Survival\ at\ time\ t_i)}$$

$$Pr_i = \frac{Pr\ (t_i \leq T < t_i + \Delta t\ |T > t_i, Xi)}{Pr\ \left(\cup_{j=1}^{Rj} (t_j \leq T < t_j + \Delta t)\ \middle|\ T > t_j\right)}$$

$$Pr_i = \frac{Pr(t_i \leq T < t_i + \Delta t|T > t_i, Xi)}{\sum_{j=1}^{Ri} Pr[(t_j \leq T < t_j + \Delta t|T > t_j)]}$$

$$Pr_i \cong \frac{h(t, \mathbf{X}, \boldsymbol{\beta}) \times \Delta t}{\sum_{j=1}^{Ri} h(t|, \mathbf{Xj}, \boldsymbol{\beta}) \times \Delta t} \quad \text{because} \quad h(t_i) = \lim_{\Delta t \to 0} \frac{Pr(t_i \leq T < t_i + \Delta t|T > t_i, Xi)}{\Delta t}$$

$$Pr_i \cong \frac{h0(ti)\exp(X'\beta)}{\sum_{j=1}^{Ri} h0(tj)\exp(X'\beta)} \cong \frac{\exp(X'\beta)}{\sum_{j=1}^{Ri}\exp(X'\beta)}$$

$$Pr_i\,(T=t_i) = P \cong \frac{\exp(Xi'\beta)}{\sum_{j=1}^{Ri}\exp(Xj'\beta)}$$

$\ln(P)= \ln[\frac{\exp(Xi'\beta)}{\sum_{j=1}^{Ri}\exp(Xj'\beta)}] = \mathbf{Xi'\beta} - \ln[\sum_{j=1}^{Ri}\exp(\mathbf{Xj'\beta})]$ where **β** is a vector of coefficients and **X** is the vector of explanatory variables/factors.

For *n* children each with (t_i, c_i, $\mathbf{X_i}$), the contribution of an observation to the likelihood function is defined as the product $[f(t_i, \boldsymbol{X}, \boldsymbol{\beta})]^{ci}\,[S(t_i, \boldsymbol{X}, \boldsymbol{\beta})]^{1-ci}$. The likelihood function becomes

$$l(\beta) = \prod_{i=1}^{n}[f(t_i, \boldsymbol{X}, \boldsymbol{\beta})]^{ci}[S(t_i, \boldsymbol{X}, \boldsymbol{\beta})]^{1-ci}$$

where t_i is the length of time that the i^{th} subject is observed (survival time), c_i is an indicator of censoring for the i^{th} subject, $\underline{X}_i$ is a vector of covariates for individual *i*, for i=1,2,…,*n*.

The estimates of the parameter, β, are the value that maximizes the log-likelihood function.

$$L(\beta) = \sum_{i=1}^{n}\{ci\ ln[f(t, \boldsymbol{X}, \boldsymbol{\beta})] + (1-ci)\ln[S(t, \boldsymbol{X}, \boldsymbol{\beta})]\}$$

$$= \sum_{i=1}^{n}\{ci\ ln[h(t, \boldsymbol{X}, \boldsymbol{\beta}) \times S(t, \boldsymbol{X}, \boldsymbol{\beta})] + (1-ci)\ln[S(t, \boldsymbol{X}, \boldsymbol{\beta})]\}$$

$$= \sum_{i=1}^{n}\{ci\ lnh(t, \boldsymbol{X}, \boldsymbol{\beta}) + ci\ lnS(t, \boldsymbol{X}, \boldsymbol{\beta}) + lnS(t, \boldsymbol{X}, \boldsymbol{\beta}) - ci\ ln[S(t, \boldsymbol{X}, \boldsymbol{\beta})]\}$$

$$= \sum_{i=1}^{n}\{ci\ lnh(t, \boldsymbol{X}, \boldsymbol{\beta}) + lnS(t, \boldsymbol{X}, \boldsymbol{\beta})\}$$

$$= \sum_{i=1}^{n}\{ci\ ln[h_0(t_i)] + ci\boldsymbol{X_i'\beta} + e^{\mathbf{X_i\beta}}\ lnS_0(t_i)\}$$

However, since the log-likelihood function involves the unspecified baseline hazard and unspecified baseline survivorship functions, we are not going to use full maximum likelihood that requires maximizing the log-likelihood equation $L(\beta) = \sum_{i=1}^{n}\{ci\ ln[h_0(t_i)] + ci\boldsymbol{X_i'\beta} + e^{\mathbf{X_i\beta}}\ lnS_0(t_i)\}$ with respect to the unknown parameter of interest, β, as well as the unspecified baseline hazard and survivorship functions. Instead we use here the partial likelihood which is defined as:

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{Xi'\beta}}{\sum_{j \in R(ti)} e^{Xj'\beta}} \right]^{c_i}$$

where the summation in the denominator is over all subjects in the risk set at time t_i, denoted by $R(t_i)$. The expression above is modified to exclude terms when $c_i = 0$, becomes

$$l_p(\beta) = \prod_{i=1}^{m} \frac{e^{X(i)'\beta}}{\sum_{j \in R(t(i))} e^{X(j)'\beta}} \quad .$$

where the products is over the *m* distinct ordered survival times and $\mathbf{X_{(i)}}$ the vector of values of covariates for the subject with ordered survival time $t_{(i)}$. The log-partial likelihood function is

$$L_p(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left\{ \boldsymbol{Xi'\beta} - ln \left[\sum_{j \varepsilon R_{t(i)}} e^{Xj'\beta} \right] \right\}$$

Observe that the log-partial likelihood is only a function of **β**'s.

3.5.2. Estimation of the regression parameter **β**'s

Since **β** is a vector of *p* parameters of the model and $(X_1, X_2, \ldots, X_p)$ is the vector of explanatory variables, then there are *p*-equations to solve, one for each parameter when the partial derivative of the log-partial likelihood with respect to each parameters is considered.

The log-partial likelihood yields:

$L_p(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left\{ \boldsymbol{Xi'\beta} - ln \left[\sum_{j \varepsilon R_{t(i)}} e^{Xj'\beta} \right] \right\}$ Differentiate the log-partial likelihood with respect to $\boldsymbol{\beta}_k$'s, $k = 1, \ldots, p$

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \beta_k} = \frac{\partial \sum_{i=1}^{m} \left\{ \boldsymbol{Xi'\beta} - ln \left[\sum_{j \varepsilon R_{t(i)}} e^{Xj'\beta} \right] \right\}}{\partial \beta_k}$$

$$= \sum_{i=1}^{m} \left\{ X_{ik} - \frac{\sum_{j \varepsilon R_{t(i)}} X_{jk} e^{Xj'\beta}}{\sum_{j \varepsilon R_{t(i)}} e^{Xj'\beta}} \right\}$$

where X_{ik} is the value of the covariate X_k for the subject with observed ordered survival time $t_{(i)}$.

The above partial derivatives with respect to each β which is coefficient of each covariates equated to zero and yield p equations.

$$\begin{bmatrix} \frac{\partial l_p(\boldsymbol{\beta})}{\partial \beta_1} \\ \frac{\partial l_p(\boldsymbol{\beta})}{\partial \beta_2} \\ \vdots \\ \frac{\partial l_p(\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We obtain the partial likelihood estimates of **β** by using Newton-Raphson algorithm or any other numerical optimization techniques.

To test the significance of each parameter estimate we need to have a measure of variation by defining observed $p \times p$ information matrix as:

$$I_0(\boldsymbol{\beta}) = -\frac{\partial^2 l_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

The estimated variance-covariance matrix is the inverse of the above observed information matrix

$$\hat{\sum} = \left[I_0(\hat{\boldsymbol{\beta}})\right]^{-1}$$

The estimator $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sum)$, and $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased.

However, this estimation procedure is valid only for data in which no more than one event occur at the same time even if it is quite common for survival data to contain tied event times. The maximum partial likelihood estimates for **β** in the presence of ties is obtained in the same manner as in non tied data cases what we have seen except that the derivatives are taken with respect to the unknown parameters in the logarithm of either the Breslow or the Efron approximation to the partial likelihood given below.

The Breslow approximation to the partial likelihood in case of tied data is handled by $l_{pB}(\beta) = \prod_{i=1}^{m} \frac{e^{x_{(i)}+\beta}}{[\sum_{j \in R(t(i))} e^{x_j \beta}]^{di}}$

where *di* denotes the number of subjects with survival time $t_{(i)}$ and $x_{(i)+}$ is the sum of covariate over the *di* subjects.

The Efron approximation to the partial likelihood is more complicated and yields slightly better approximation to the exact partial likelihood than the Breslow approximation. It uses as the partial likelihood

$$l_{pE}(\beta) = \prod_{i=1}^{m} \frac{e^{x_{(i)+}\beta}}{\prod_{k=1}^{di}[\sum_{j\in R(t(i))} e^{x_j\beta} - \frac{k-1}{di} \sum_{j\in D(t(i))} e^{x_j\beta}]}$$ (Hosmer & Lemeshow 1998).

3.6. Model Selection

Proportional hazards regression analysis of survival data requires a number of critical decisions. Usually, we may suspect more covariates/factors that influence the response variable than we can reasonably select to include in the model. So we must decide on a method to select a subset of total number of covariates to be included in the model. When selecting a subset of covariates we must consider clinical importance and statistical significance of the set of covariates.

We made a thorough bivariate analysis of the association between survival time and all important covariates in the set. This includes Kaplan-Meier estimates of group specific survivorship functions, point and interval estimates of the mean survival time and Log-rank test of significance to compare survival experience across the groups defined by the categorical variables. We grouped continuous covariates into clinically meaningful groups and the above methods of categorical covariates will be applied.

All variables significant in the above bivariate analysis at p-value less than 0.25 will be included in the multivariable model. Using 0.25 as level of significance resulted in the inclusion of covariates that have the potential to be either an important confounder or statistically significant in the preliminary multivariable model.

We will fit multivariable Cox proportional hazards regression model using the statistical package SPSS "Method = Backward Stepwise (Likelihood ratio). For the initial multivariable model, we use the p-values from the Wald test of individual coefficients to identify covariates that might be deleted from the model. At this step, since some covariates may not look significant, we should take care not to delete important covariates. This is especially important when a nominal scale covariate with more than one design variable has been selected for deletion. For instance, the categorical predictor clinical stage has four levels and therefore we will include this predictor using three dummy variables with the group CS=IV as the reference group. We do the same for categorical variables such as functional status, reason for ART and types of ART.

When a covariate is selected for deletion, we should assess whether or not removal of the covariate has produced a significant change in the coefficient of the variables retained in the model. If the change in one of the coefficients of the remaining covariates as a result of deletion of a covariate from the preliminary model, is about 20% or more then we considered this a significant change and the variable that was removed should be added back in to the model as it might be an important confounder.

3.7. Assessment of model adequacy

Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid the fitted model must provide an adequate summary of the data upon which it is based (Hosmer & Lemeshow 1998). Methods for the assessment of the fitted

proportional hazards model are based on residuals. Residuals are defined as the difference between the observed and the predicted values of the outcome variable. In proportional hazards model, however, the definition of residuals is not straight forward and causes the development of different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazards model. For the assessment of different aspects of the Cox proportional hazards model, we considered the following residuals.

Martingale residuals are used to check the linearity of continuous covariates included in the model. The martingale residual is defined as: $\hat{M}_i = c_i - \hat{H}(t_i, x_i, \hat{\beta})$, where c_i is censoring indicator and $\hat{H}(t_i, x_i, \hat{\beta})$ cumulative hazard function estimator. These residuals have a mean zero and skewed distribution, even when the fitted model is good.

Schoenfeld residuals with a fitted proportional hazards model are used to check the proportionality assumption based on the individual contributions to the derivatives of the log-partial likelihood with respect to β. The estimator of the Schoenfeld residuals for the i^{th} subject on the k^{th} covariate is obtained by

$$\hat{r}_{ik} = c_i \left[x_{ik} - \frac{\sum_{j \in R_{(ti)}} x_{jk} e^{x_j' \hat{\beta}}}{\sum_{j \in R_{(ti)}} e^{x_j' \hat{\beta}}} \right]$$

From this one can easily understood that the Schoenfeld residuals are defined only at uncensored survival time (when $c_i \neq 0$). Thus, the vector of p Schoenfeld residuals for the i^{th} subject be is:

$$\hat{r}_i = (\hat{r}_{i1}, \hat{r}_{i2}, \hat{r}_{i3}, \ldots\ldots\hat{r}_{ip})$$

The modified forms of the Schoenfeld residuals, scaled Schoenfeld residuals, are recommended to check the proportionality of the covariate over the study time because scaled Schoenfeld residual has a greater diagnostic power than the un-scaled Schoenfeld residuals

(Grambsch & Therneau, 1994). An easily computed approximation for the scaled Schoenfeld residuals $\hat{r}_i^* = \left[\widehat{Var(\hat{r}_i)}\right]^{-1} \hat{r}_i$ is $\hat{r}_i^* = \text{m}\, \widehat{Var(\beta)}\hat{r}_i$.

Now, we are going to assess the adequacy of a model by considering three issues, namely checking model assumption, looking for outliers and testing the goodness of fit of the model.

3.7.1. Checking the assumption of proportional hazards

The next step of any model building procedure is the assessment of the specified model's adherence to the model assumptions. The critical assumption of the Cox model is the extent to which the effect of a covariate on the outcome variable has been the same over the study time. That is, we examine the extent to which the coefficients of a covariate vary over the study time. Although there are various ways to check the proportionality of hazard ratio over time, Grambsch and Therneau (1994) and simulation comparisons by Ngandu (1997) have shown that one numerical test and an association graph yield a powerful and effective method to check this assumption.

As an alternative to the proportional hazards regression model, $h(t, X, \beta) = h_0(t)\, e^{\beta X}$, consider a model with time varying coefficient β as $h(t, X, \beta) = h_0(t)\, e^{\beta(t)X}$. Let the j^{th} covariate coefficient β_j vary over time as $\beta_j(t) = \beta_j + \gamma_j g_j(t)$ where j = 1, 2, 3, …., p, β_j is constant, $g_j(t)$ is some specified function of time (usually $g_j(t)$ is specified as *ln(t)*) and γ_j is coefficient of $g_j(t)$. And the hazard function becomes: $h(t, \mathbf{X}, \boldsymbol{\beta}) = h_0(t)\, e^{\boldsymbol{\beta}(t)\mathbf{X}} = h_0(t)\, e^{(\boldsymbol{\beta} + \gamma \ln t)\,\mathbf{X}} = h_0(t)\, e^{\boldsymbol{\beta}\mathbf{X} + \gamma \mathbf{X} \ln t}$.

The rationale behind this model is that the effect of a covariate may change over the period of follow up. Here, we test whether the coefficient γ_j is zero or different from zero: if it is different from zero it indicates that the proportional hazards assumption fails to hold but if it is zero then the model reduced to the proportional hazard model with satisfied assumption. In line with this we use the plots of the scaled Schoenfeld residuals against the log of time and,

if this plot shows random distribution around the reference line through zero then the assumption are satisfied, however, if it looks to have some systematic pattern the assumption of proportional hazard is violated.

The test procedure to check whether the coefficient is really time varying or not is

$H_0: \gamma_j = 0$ against H_1: Not H_0 $j = 1, 2, 3, \ldots\ldots, p$

The test statistic we use is the Wald test statistic given as

$$Z = \frac{\gamma_j}{S.E(\gamma_j)}$$ which follows a standard normal distribution.

The test rule used is

The coefficient of the j^{th} covariate is not time dependent if $Z < Z_{(\alpha/2)}$ i.e. do not reject H_0.

That is, model assumption is satisfied.

The coefficient of the j^{th} covariate is time dependent if $Z > Z_{(\alpha/2)}$ i.e. reject H_0.

That is, model assumption is not met.

3.7.2. Checking for outliers

Another important aspect of the model evaluation is a thorough examination of regression diagnostic statistic to identify which, if any, observation have an undue influence on the fit of the model. It is important to determine whether the hazard ratio will be affected to a large extent by any one individual observation's data in the model. We do this by fitting the model for all *n* observations in the data set and then refit the same model to the sets of *n-1* observations obtained by omitting each of the *n* observations one at a time. Whenever an observation can exert unusual influence on the hazard ratio, this will be noticeable in significant change in the parameter estimate(s) while removing the observation's data from the data set and refit the same model. That is we can use $\Delta_i\hat{\beta}_j = (\hat{\beta}_j - \hat{\beta}_{j(-i)})$, which is known as delta-beta (Collett 2003), as the statistic to detect an outlier observation of the data.

The change in the parameter estimate is large for an outlier observation while the delta-beta is not large for others.

And we measure the influence of each observation on the estimated regression coefficients. To examine the influence of a j^{th} covariate value of the i^{th} individual on the j^{th} regression coefficient estimate we use delta-beta statistics

$$\Delta_i\hat{\beta}_j = (\hat{\beta}_j - \hat{\beta}_{j(-i)})$$

where $\hat{\beta}_j$ is the j^{th} coefficient based on n observations and $\hat{\beta}_{j(-i)}$ is the j^{th} coefficient based on $n-1$ observations (i^{th} observation is deleted).

The overall summary statistic of the influence of a subject on the estimator of all the coefficients may be approximated using the likelihood displacement statistic ld_i. The statistics ld_i is an approximation to the amount of change in log partial likelihood when the i^{th} subject is deleted. In this context the statistic is called the likelihood displacement statistic. It can be shown that

$$ld_i = 2\,[\, l_p(\hat{\beta}) - l_p(\hat{\beta}_{(-i)})\,]$$

The next step in the modeling process is to identify explicitly the subjects with the extreme values, refit the model deleting these subjects, and calculate the percentage change in the individual coefficients as $\Delta\hat{\beta} = 100(\frac{\hat{\beta}_{reduced} - \hat{\beta}_{all}}{\hat{\beta}_{all}})$. The final decision on the continued use of a subject's data to fit the model will depend on the observed percentage change in the coefficients that result from deleting the subject's data and, more importantly, the clinical plausibility of that subject's data (Hosmer and Lemeshow 1998).

3.7.3. Assessing the goodness of fit of the model

In practice, models may not exactly represent the true relationship between the response variable (survival time of HIV-infected children after ART) with different covariates. Hence,

we should apply a mechanism by which we could assess the goodness of fit of the model what we have chosen. Since the Cox-proportional hazards regression model is one of the models to be used for description and inference about the effects of predictors on the survival time of HIV-infected children after ART initiation, no one could feel satisfied that a particular model of a given form is the only model that best explains reality. Thus we have to assess the goodness of fit of the selected model.

The first step in assessing the fitted model is checking the overall goodness of fit of the model to the observed data. To test the overall goodness of fit of the model we used partial likelihood ratio test. We test H_0: $\boldsymbol{\beta} = 0$ against H_1: Not H_0 at a specified α-level of significance.

The test statistic for this purpose is

$$G = 2[\, l_p(\hat{\boldsymbol{\beta}}) - l_p(\mathbf{0})]$$

Under the null hypothesis this follows a chi-square distribution with *k-1* degrees of freedom.

We decide that the model does not have 'good fit' if the computed value of G based on the available data is $< \chi^2_{(k-1)}$. And we decide to reject the null hypothesis and decide that the model has 'good fit' if the computed value of G based on the available data is $\geq \chi^2_{(k-1)}$.

To test the significance of individual parameter coefficients we used Wald test statistic

We test H_0: $\hat{\beta}_l = 0$ against H_1: Not H_0 at α-level of significance. The test statistics used in this case are $Z = \frac{\hat{\beta}_l}{S.E(\hat{\beta}_l)}$ under the null hypothesis this follows a standard normal distribution.

We decide the l^{th} covariate is not significant if the computed value of Z based on the available data is $< Z_{(\alpha/2)}$ and we do not reject H_0. Whereas we decide to reject the null hypothesis H_0 and consider the l^{th} covariate if the computed value of Z based on the available data is $> Z_{(\alpha/2)}$.

3.8. Interpretation of the regression coefficients

For categorical covariates with k-levels, we apply categorical variable coding with *k-1* dummy variables and one reference level against which all other levels are compared. The comparison is made by taking the ratio of the hazard function of children in each dummy variable to the hazard function of children in the reference group. The resulting hazard ratio e^{β} compares the rate of deaths of children in each group to deaths of children in the reference group. The regression coefficient β is therefore, interpreted as: at any time during the study period the logarithm of the hazard ratio of HIV- infected children with covariate value in a defined dummy variable to HIV- infected children with covariate value equals to zero (reference category) is β, given that all other variables in the model are the same.

For continuous covariate the regression coefficient β is interpreted as the change in the log hazard of death of children for a unit change of a continuous covariate provided that the log hazard ratio is linear with the covariate of interest and all other factors are kept constant. Hence, we should test for linearity of log hazard ratio against the continuous covariates in the model. To test the linearity of log hazard ratio against a continuous covariate, we use the plots of martingale residual. If the plots of the martingale residuals are randomly distributed with no systematic pattern, about a horizontal straight line through zero after excluding the continuous covariate for which we are checking the linearity assumption then the log hazard ratio is linear. If the covariate is not linear then it requires transformation.

Chapter Four

4. Data Analysis

4.1. Univariate Analysis of Data

4.1.1. Univariate Analysis of the Response Variable

A total of 255 HIV infected children (age less than 15) who started ART at Felege-Hiwot referral hospital, in Bahir Dar from 2007 up to early 2009 were involved in this study. Among a total of 255 participants in the study 71 (27.84%) children died due to the disease, 19 (7.45%) were lost to follow up while 165 (64.71%) remained alive during the time of data collection. The response variable was the length of time from start of treatment to death. The mean survival time of the entire observations was found to be 22.381 months with standard deviation 0.693 months. However, these estimates were underestimated because the number of observations that were censored was large.

4.1.2. Univariate Analysis of Explanatory Variables

The distribution of the deaths of children under study for the selected background characteristics, of categorical covariates included in this study is given in Table 4.1.1. Log-rank test for the comparison of survival time of children in different groups are displayed in the same table. The mean, standard deviation and 95% confidence interval of the mean survival time of children in each category defined by a variable are given in Table 4.1.2.

Table 4.1.1. Distribution of deaths of children under study in Bahir Dar Felege-Hiwot hospital according to variables possibly associated with survival time after ART initiation.

Covariate		Total N	N of Events	Censored				
				N	Percent	Chi-Square	Df	Sig.
sex	Male	141	38	103	73.0%			
	Female	114	33	81	71.1%	.134	1	.714
	Overall	255	71	184	72.2%			
PT	No	103	26	77	74.8%			
	Yes	152	45	107	70.4%	.341	1	.559
	Overall	255	71	184	72.2%			
TBtreat	No	204	58	146	71.6%			
	Yes	51	13	38	74.5%	.355	1	.551
	Overall	255	71	184	72.2%			
ReART	CD4 and Clinical reason	155	47	108	69.7%			
	CD4 Only	69	16	53	76.8%	1.367	2	.505
	Clinical Only	31	8	23	74.2%			
	Overall	255	71	184	72.2%			
FS	Working	165	42	123	74.5%			
	Ambulatory	66	20	46	69.7%	3.271	2	.195
	Bedridden	24	9	15	62.5%			
	Overall	255	71	184	72.2%			
CS	Clinical Stage-I	23	5	18	78.3%			
	Clinical Stage-II	49	8	41	83.7%	23.358	3	.000
	Clinical Stage-III	122	28	94	77.0%			
	Clinical Stage-IV	61	30	31	50.8%			
	Overall	255	71	184	72.2%			
TyART	d4t-3TC-NVP	64	17	47	73.4%			
	d4t-3TC-EFV	32	9	23	71.9%			
	AZT-3TC-NVP	100	30	70	70.0%	1.342	3	.719
	AZT-3TC-EFV	59	15	44	74.6%			
	Overall	255	71	184	72.2%			

Table 4.1.2: Results of Kaplan-Meier survival times for children on antiretroviral treatment based on important clinical characteristics (categorical covariate) of HIV-infected children treated with ART in Bahir Dar Felege-Hiwot Hospital.

Covariates	Category	Mean[a]			
		Estimate	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Sex	Male	22.430	.919	20.628	24.231
	Female	22.493	.968	20.595	24.391
	No	22.987	1.023	20.981	24.993
PT	Yes	21.282	.761	19.789	22.774
TB treatment	No	22.159	.773	20.644	23.674
	Yes	21.529	1.145	19.286	23.773
Reason for ART	CD4 and Clinical reason	21.559	.930	19.736	23.381
	CD4 Only	23.849	1.143	21.608	26.090
	Clinical Only	20.139	1.516	17.167	23.111
	Working	23.140	.793	21.586	24.694
Functional Status	Ambulatory	19.367	1.112	17.188	21.547
	Bedridden	16.760	1.943	12.952	20.567
	Clinical Stage-I	23.261	1.240	20.830	25.692
Clinical Staging	Clinical Stage-II	25.317	1.149	23.065	27.568
	Clinical Stage-III	23.740	.874	22.027	25.453
	Clinical Stage-IV	15.252	1.295	12.713	17.790
	d4t-3TC-NVP	23.300	1.202	20.944	25.656
Types of ART	d4t-3TC-EFV	19.717	1.504	16.770	22.665
	AZT-3TC-NVP	19.413	.891	17.667	21.159
	AZT-3TC-EFV	19.828	1.062	17.746	21.911
	Overall	22.381	.693	21.023	23.738

The continuous covariates age, hemoglobin, weight and CD4 count of children were categorized into clinically meaningful groups so that it would be simple to describe and interpret the results. The categories used for analysis have been given in Table 3.1. We used Kaplan-Meier survival function estimator to compare the survival time of HIV-infected children after ART initiation defined by a variable. The distribution of deaths of children and

the estimated mean survival time in each category defined by the variable are shown in Tables 4.1.3 and 4.1.4.

Table 4.1.3. Distribution of deaths of children under study over groups of continuous covariates thought to influence survival time of children treated at Felege-Hiwot hospital in Bahir Dar.

Covariates		Total N	N of Events	Censored		Chi-Square	Df	Sig.
				N	Percent			
Age group	<= 1.5 years of age	29	14	15	51.7%	13.238	2	.001
	between 1.5-5 years of age	89	27	62	69.7%			
	>=5 and <=14 years	137	30	107	78.1%			
	Overall	255	71	184	72.2%			
Hemoglobin Group	<=7g/dl	93	44	49	52.7%	48.689	3	.000
	7-8.5 g/dl	36	13	23	63.9%			
	8.5-10gl	37	7	30	81.1%			
	>=10g/dl	89	7	82	92.1%			
	Overall	255	71	184	72.2%			
CD4 group	Above Threshold	98	27	71	72.4%	.324	1	.569
	Below Threshold	157	44	113	72.0%			
	Overall	255	71	184	72.2%			
Weight group	Normal in body weight	149	49	100	67.1%	4.996	2	.082
	Low in body weight	52	12	40	76.9%			
	Very low in body weight	54	10	44	81.5%			
	Overall	255	71	184	72.2%			

Table 4.1.4: Results of Kaplan-Meier survival times for children on antiretroviral treatment based on important clinical characteristics (continuous covariate) of HIV-infected children treated in Bahir Dar Felege-Hiwot Hospital.

Covariates		Mean[a]			
				95% Confidence Interval	
		Estimate	Std. Error	Lower Bound	Upper Bound
Age group	<= 1.5 years of age	15.002	1.924	11.232	18.773
	between 1.5-5 years of age	21.260	1.006	19.288	23.232
	>=5 and <=14 years	24.040	.805	22.462	25.619
	Overall	22.381	.693	21.023	23.738
Hemoglobin Group	<=7g/dl	16.341	1.064	14.256	18.426
	7-8.5 g/dl	20.151	1.688	16.842	23.460
	8.5-10gl	20.625	1.181	18.311	22.938
	>=10g/dl	27.538	.551	26.458	28.618
	Overall	22.381	.693	21.023	23.738
CD4 group	Above Threshold	23.339	.952	21.474	25.205
	Below Threshold	21.938	.911	20.154	23.723
	Overall	22.381	.693	21.023	23.738
Weight group	Normal in body weight	21.498	.885	19.764	23.232
	Low in body weight	22.681	1.243	20.246	25.117
	Very low in body weight	24.591	1.260	22.121	27.062
	Overall	22.381	.693	21.023	23.738

4.2. Bivariate Analysis of survival time with different explanatory variables

In this section, a bivariate analysis of the association between survival time and all covariates will be done. For categorical covariates, the analysis includes Kaplan-Meier estimates of the group specific survivorship functions, point and interval estimates of the mean survival time and Log-Rank significance tests to compare survival time across the groups defined by the variable used. Continuous covariates are treated as categorical covariates based on the groups we have previously explained.

In this analysis we plotted a Kaplan-Meier survival function graph separately to compare the survival experience for each group of the study subjects as given in the Appendix A. The graphs provide a visual representation of the survival experience of different groups of the study subjects as reported in Table 4.1.1 and Table 4.1.3. The Log rank test was used to test the difference in survival time between groups defined by the variable as shown in Sig. column of Table 4.1.1 and Table 4.1.3.

The bivariate Cox proportional hazards model is used to assess the impact of each covariate on the survival time of children after ART initiation. Covariates are selected based on their contribution to reduce the log partial likelihood function (*-2LL*) for the purpose of further investigation in the multivariable Cox proportional hazards model.

Table 4.2.1 Summary statistics of single covariate analysis of Cox proportional hazards model of survival time of children under study treated at Felege-Hiwot Hospital, Bahir Dar.

	df	B	SE	Wald	Pr > Chi-Sq Sig.	Hazard ratio Exp(B)	95.0% CI for Exp(B)		-2 log likelihood *-2LL*	Reduction in *-2LL due to the inclusion*
							Lower	Upper		
sex	1	-.043	.119	.131	.717	.958	.758	1.210	730.992	.131
PT	1	-.071	.124	.334	.564	.931	.731	1.186	730.785	.338
TBtreat	1	.091	.154	.347	.556	1.095	.810	1.480	730.762	.360
FS	2			3.134	.209					
FS(1)	1	.289	.273	1.121	.290	1.335	.782	2.279	728.206	2.916
FS(2)	1	.602	.368	2.667	.102	1.825	.887	3.758		
ReART	2			1.327	.515					
ReART(1)	1	.141	.174	.653	.419	1.151	.818	1.619	729.725	1.397
ReART(2)	1	-.193	.210	.843	.358	.824	.546	1.245		
TyART	3			1.304	.728					
TyART(1)	1	-.172	.220	.614	.433	.842	.547	1.296	729.802	1.321
TyART(2)	1	.078	.270	.082	.774	1.081	.636	1.835		
TyART(3)	1	.159	.184	.749	.387	1.172	.818	1.681		
CS	3			20.594	.000					
CS(1)	1	-1.203	.484	6.172	.013	.300	.116	.776		

	df	B	SE	Wald	Pr > Chi-Sq Sig.	Hazard ratio Exp(B)	95.0% CI for Exp(B)		-2 log likelihood *-2LL*	Reduction in *-2LL due to the inclusion*
							Lower	Upper		
CS(2)	1	-1.419	.407	12.153	.000	.242	.109	.537	711.450	19.672
CS(3)	1	-.935	.264	12.586	.000	.393	.234	.658		
agegrp	2			11.971	.003				720.767	10.355
agegrp(1)	1	-.773	.331	5.467	.019	.461	.241	.882		
agegrp(2)	1	-1.126	.325	11.970	.001	.324	.171	.614		
hemgrp	3			35.464	.000				682.216	48.907
hemgrp(1)	1	-.550	.324	2.887	.089	.577	.306	1.088		
hemgrp(2)	1	-1.216	.408	8.884	.003	.296	.133	.659		
hemgrp(3)	1	-2.263	.410	30.461	.000	.104	.047	.232		
CD4grp	1	-1.126	.325	11.970	.573	1.148	.710	1.855	730.802	.320
Wgrp	2			4.763	.092				726.034	5.089
Wgrp(1)	1	-.497	.323	2.375	.123	.608	.323	1.145		
Wgrp(2)	1	-.629	.349	3.251	.071	.533	.269	1.056		

Note: Beginning Block, initial Log Likelihood function: -2 Log likelihood: 731.122

The above Table and the plot of Kaplan-Meier survival function graph (given in the Appendix A) showed that the survival time of children after initiation of antiretroviral therapy might be influenced by hemoglobin group, WHO clinical staging of HIV infected children, age group, weight group and functional status at the beginning of treatment. Hence, these variables were included in the preliminary multivariable model, as each variable could be either a potential confounder or a statistically significant covariate at p-value = 0.25. However, if a predictor has a p-value greater than 0.25 in the bivariate analysis, it is unlikely that it will contribute much to a model which includes other predictors. Therefore, we did not reject the null hypothesis of no significant difference in survival time for the possible outcomes of gender, prophylaxis taken, reason for ART initiation, CD4 count category, types of ART treatment and TB treatment at start of ART because these variables had a p-value greater than 0.25 (modest level of significance).

4.3. Multiple regression analysis of survival time with different explanatory variables

As in any regression analysis, when using Cox proportional hazards regression, it is important to include multiple covariates in the model and adjust for all possible imbalances in the observed data. Adjusting the effect can be done through the analysis of interaction effects. We will consider the preliminary model which included all the predictors that had p-values less than 0.25 in the bivariate analysis. This means we include hemoglobin group, WHO clinical stage, age group, weight group and functional status in our model. Multivariable Cox proportional hazards regression model using the response variable, namely survival time of children, with censoring indicator variable along with the covariates was analyzed using the statistical package SPSS "Method = Backward Stepwise (Likelihood Ratio)". The results of the multiple covariates Cox proportional hazards model are displayed in Table 4.3.2.

As depicted in the Table 4.3.1 the categorical predictor WHO clinical stage has four levels and we included this predictor using 3 dummy variables with the group CS = IV as the reference group. The categorical predictor functional status has three levels and we included this predictor using 2 dummy variables with the group functional status = working as the reference group. The continuous predictor variable age is grouped into three categories and we include this predictor using two dummy variables with age group less than 1.5 years of age as the reference group. The continuous predictor hemoglobin is grouped into four categories and we include this predictor using three dummy variables with hemoglobin group less than or equal to 7g/dL as the reference group. The continuous predictor weight is grouped into three categories and we include this predictor using two dummy variables with weight group normal in body weight for age as the reference group

Table 4.3.1. Categorical Variable Codings[b,c] for variables in the model.

		Frequency	(1)	(2)	(3)
CS[a]	0=Clinical Stage-I	23	1	0	0
	1=Clinical Stage-II	49	0	1	0
	2=Clinical Stage-III	122	0	0	1
	3=Clinical Stage-IV	61	0	0	0
FS[a]	0=Working	165	0	0	
	1=Ambulatory	66	1	0	
	2=Bedridden	24	0	1	
agegrp[a]	0=<= 1.5 years of age	29	0	0	
	1=between 1.5-5 years of age	89	1	0	
	2=>=5 and <=14 years	137	0	1	
hemgrp[a]	0=<=7g/dl	93	0	0	0
	1=7-8.5 g/dl	36	1	0	0
	2=8.5-10gl	37	0	1	0
	3=>=10g/dl	89	0	0	1
Wgrp[a]	0=Normal in body weight	149	0	0	
	1=Low in body weight	52	1	0	
	2=Very low in body weight	54	0	1	

a. Indicator Parameter Coding

b. Category variable: CS

c. Category variable: FS

d. Category variable: agegrp

e. Category variable: hemgrp

f. Category variable: Wgrp

Table 4.3.2. Results of the multiple Cox proportional hazards model for survival of children under study treated at Felege-Hiwot hospital, in Bahir-dar that involve covariates significant at P-value = 0.25 in the bivariate analysis.

		B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
Step 1	hemgrp			47.401	3	.000			
	hemgrp(1)	-.884	.353	6.267	1	.012	.413	.207	.825
	hemgrp(2)	-1.608	.418	14.835	1	.000	.200	.088	.454
	hemgrp(3)	-2.616	.423	38.178	1	.000	.073	.032	.168
	CS			26.253	3	.000			
	CS(1)	-2.190	.538	16.588	1	.000	.112	.039	.321
	CS(2)	-1.801	.452	15.865	1	.000	.165	.068	.401
	CS(3)	-.732	.282	6.746	1	.009	.481	.277	.836
	agegrp			23.616	2	.000			
	agegrp(1)	-1.331	.366	13.230	1	.000	.264	.129	.541
	agegrp(2)	-1.717	.354	23.478	1	.000	.180	.090	.360
	Wgrp			1.028	2	.598			
	Wgrp(1)	.000	.348	.000	1	1.000	1.000	.506	1.976
	Wgrp(2)	-.367	.370	.985	1	.321	.692	.335	1.431
	FS			1.052	2	.591			
	FS(1)	-.092	.292	.099	1	.753	.912	.515	1.615
	FS(2)	.335	.386	.754	1	.385	1.398	.656	2.976
Step 2	hemgrp			46.957	3	.000			
	hemgrp(1)	-.851	.351	5.875	1	.015	.427	.215	.850
	hemgrp(2)	-1.576	.416	14.372	1	.000	.207	.092	.467
	hemgrp(3)	-2.624	.424	38.333	1	.000	.073	.032	.166
	CS			27.175	3	.000			
	CS(1)	-2.065	.514	16.145	1	.000	.127	.046	.347
	CS(2)	-1.785	.438	16.605	1	.000	.168	.071	.396
	CS(3)	-.714	.276	6.683	1	.010	.489	.285	.841
	agegrp			23.852	2	.000			
	agegrp(1)	-1.347	.362	13.854	1	.000	.260	.128	.529
	agegrp(2)	-1.722	.354	23.638	1	.000	.179	.089	.358
	Wgrp			1.202	2	.548			

	Wgrp(1)	.013	.346	.001	1	.970	1.013	.514	1.995
	Wgrp(2)	-.388	.366	1.124	1	.289	.678	.331	1.390
Step 3	hemgrp			47.375	3	.000			
	hemgrp(1)	-.789	.340	5.374	1	.020	.454	.233	.885
	hemgrp(2)	-1.570	.415	14.296	1	.000	.208	.092	.469
	hemgrp(3)	-2.627	.420	39.188	1	.000	.072	.032	.165
	CS			28.237	3	.000			
	CS(1)	-2.022	.506	15.977	1	.000	.132	.049	.357
	CS(2)	-1.738	.423	16.875	1	.000	.176	.077	.403
	CS(3)	-.706	.272	6.715	1	.010	.494	.290	.842
	agegrp			23.658	2	.000			
	agegrp(1)	-1.257	.349	12.959	1	.000	.285	.144	.564
	agegrp(2)	-1.693	.349	23.562	1	.000	.184	.093	.365

[1] estimated coefficient $\hat{\beta}$, [2] standard error of the coefficient, [3] Wald statistics, [4]degrees of freedom, [5]significance of Wald statistics, [6] exponent of the coefficient $\hat{\beta}$, [7] 95% CI for Exp($\hat{\beta}$)

The SPSS output revealed that the variable, functional status, is the first variable to be removed from the model, as it caused an increase in *-2LL* of only 1.052 with P-value=0.591. The second variable excluded from the model is weight group, having the smallest contribution to increase *-2LL*, of 1.202 with P-value 0.548. We also assessed whether the removal of the covariate has produced a significant change in the coefficient of the variables remaining in the model. We computed the percentage change in the coefficients remained in the model as a result of deletion of a covariate from the model from Table 4.3.2 above. And no significant change (20% or more) in any one of the remaining coefficients in the model was seen. The results are recorded in Appendix D.

The model containing the variables that were significant in the previous step and a model which eliminated one variable at a time, all the changes are found to be significant.

Therefore, the covariates, hemoglobin group, WHO clinical stage and age group remain in the model at this stage.

Having this preliminary final model (at step 3 from the Table 4.3.2), we need to check the assumption of linearity for the continuous covariates age and hemoglobin contained in the model. Here the plot of martingale residuals versus the excluded covariate for which we are checking the assumption of linearity is used. As the plot of martingale residuals for the covariates of interest displayed the residuals are randomly distributed, with no systematic pattern and the smoothed curves are found to be horizontal straight line through zero as in Fig.4.3.1a and Fig.4.3.1b (see Appendix B). These indicates that the linearity assumption have been fulfilled for all continuous covariates in the model.

In this study we considered all possible interactions of order two, the interactions of hemoglobin group with levels of WHO clinical stages, hemoglobin group with age group and age group with levels of WHO clinical stages. We then fitted a Cox proportional hazards model containing the three covariates and one interaction term at a time. To test the significance of the regression coefficient of the interaction term, Log-rank test statistic was used. As observed in Table 4.3.3 below no coefficient of interaction terms of order two were found significant at 0.05 significant levels. Therefore, we consider the final model which contains the main effects only.

Table 4.3.3. Results of the multiple Cox proportional hazards model which contain one interaction term at a time for Bahir dar Felege-Hiwot hospital data of children ART.

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
hem4	1	-2.62425	0.41996	39.0485	<.0001	0.072
hem3	1	-1.57944	0.41539	14.4573	0.0001	0.206
hem2	1	-0.85617	0.35159	5.9299	0.0149	0.425
hem1	0	0	.	.	.	
CS1	1	-2.45119	0.56512	18.8137	<.0001	0.086
CS2	1	-1.72928	0.42321	16.6965	<.0001	0.177
CS3	1	-0.72173	0.27199	7.0413	0.0080	0.486

CS4	0	0	.	.	.	.
age3	1	-1.71980	0.35397	23.6055	<.0001	0.179
age2	1	-1.21566	0.34984	12.0746	0.0005	0.297
hem4CS1	1	0.24547	0.20160	1.4922	0.0817	1.445
hem4CS2	1	-0.33578	0.23450	0.0004	0.9839	0.689
hem4CS3	1	-0.40342	0.89324	2.4686	0.1161	0.246
hem4CS4	1	0.16420	0.83218	1.9571	0.1618	1.203
hem3CS1	1	-0.87477	0.66045	0.0003	0.9863	0.159
hem3CS2	1	0.19435	0.96771	1.8962	0.1010	1.304
hem3CS3	1	-0.50995	0.82705	0.3802	0.5375	0.601
hem3CS4	1	-0.74187	0.89333	0.6897	0.4063	0.476
hem2CS1	1	0.19277	0.99012	3.6544	0.0859	1.638
hem2CS2	1	0.40310	0.86872	2.6087	0.1063	1.068
hem2CS3	1	-0.28839	0.83125	0.1204	0.7286	0.749
hem2CS4	1	-0.24940	0.68398	3.3367	0.0778	0.287
hem1CS1	1	-0.35160	0.96619	5.9238	0.1049	0.995
hem1CS2	1	-0.12476	0.89729	5.6073	0.1179	0.919
hem1CS3	1	0.97232	0.56906	2.9194	0.0875	1.644
hem1CS4	1	0.56030	0.52141	1.1547	0.2826	1.751
hem4age1	1	0.36412	0.91369	0.1588	0.6903	1.439
hem4age2	1	0.09983	0.82564	0.0146	0.9038	1.105
hem4age3	1	-0.40761	0.88261	0.2133	0.6442	0.665
hem3age1	1	0.42205	0.91872	2.3959	0.1217	1.146
hem3age2	1	-0.90635	0.88829	1.0411	0.3076	0.404
hem3age3	1	0.00417	0.81354	0.0000	0.9959	1.004
hem2age1	1	0.11478	0.71861	0.0255	0.8731	1.122
hem2age2	1	-0.04951	0.67818	0.0053	0.9418	0.952
hem2age3	1	-0.05189	0.67808	0.0059	0.9390	0.949
hem1age1	1	-0.81243	0.66721	1.4827	0.2234	0.444
hem1age2	1	0.33548	0.52310	0.4113	0.5213	1.399
hem1age3	1	0.18077	0.52434	0.1189	0.7303	1.198
CS4age1	1	-0.63634	0.68095	0.8733	0.3501	0.529
CS4age2	1	0.15057	0.52349	0.0827	0.7736	1.162
CS4age3	1	0.21845	0.50299	0.1886	0.6641	1.244
CS3age1	1	0.19149	0.63705	3.4981	0.0914	1.292
CS3age2	1	-0.39376	0.52816	0.5558	0.4559	0.675
CS3age3	1	-0.33341	0.51225	0.4236	0.5151	0.716
CS2age1	1	-0.12315	0.15070	0.6436	0.4224	0.397
CS2age2	1	0.13389	0.80868	1.3336	0.2482	1.544
CS2age3	1	-0.22962	0.76759	0.0895	0.7648	0.795
CS1age1	1	-0.39156	0.98362	0.1585	0.6906	0.676
CS1age2	1	-0.36777	0.97737	0.1416	0.7067	0.692
CS1age3	1	0.08056	0.20334	2.7089	0.0998	1.247

We have now completed the model development stage that involves the three basic activities, namely selection of covariates, checking for linearity of continuous covariates and the issue of interaction and we can begin the second task of assessing the model adequacy.

4.4. Assessment of Model Adequacy

In this section we set out to check the fulfillment of the proportional hazard assumptions, checking for outliers and the goodness of fit of the model.

4.4.1. Proportionality Assumption

The assumption of proportional hazard is a strong assumption and it should be appropriately assessed. We discussed how to assess this assumption in chapter three, and we use one graphical method (the plot of the scaled Schoenfeld residuals) and one numerical test (use of extended Cox model) to assess the adequacy of the preliminary final model. We use the plot of the scaled Schoenfeld residuals versus the covariate and the plot show that the residuals are randomly distributed and the smoothed curve without much departure from the horizontal line. Also the proportional hazards assumption was tested based on the interaction of the covariate with the log of survival time. We found that the interaction coefficients are not significant at 0.05 levels, and the assumption of proportionality holds. Figure 4.4.1 (in the Appendix C) and the numeric test results of the model as given below in Table 4.4.1 depicts that the proportional hazards assumption can be accepted.

Table 4.4.1. SAS output of Cox proportional hazards model of survival of children under study treated at Bahir-dar Felege-Hiwot hospital that involves all the interaction of each covariate with the log of time.

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
hem4	1	-3.70905	1.52819	5.8907	0.0152	0.025
hem3	1	-0.46100	0.92010	0.2510	0.0163	0.631
hem2	1	-1.43394	0.87035	2.7144	0.0994	0.238
hem1	0	0	.	.	.	
CS1	1	-4.63222	2.01617	5.2787	0.0216	0.010
CS2	1	-2.74746	1.32616	4.2921	0.0383	0.064
CS3	1	-1.46485	0.67647	4.6891	0.0304	0.231
CS4	0	0	.	.	.	
age3	1	-2.40727	0.74852	10.3431	0.0013	0.090
age2	1	-1.88756	0.74843	6.3605	0.0117	0.151
age1	0	0	.	.	.	
hem4t	1	0.52632	0.62541	0.7082	0.4000	1.693
hem3t	1	-0.54672	0.47950	1.3000	0.2542	0.579
hem2t	1	0.34539	0.40854	0.7147	0.3979	1.413
hem1t	0	0	.	.	.	
CS1t	1	1.16560	0.82094	2.0160	0.1557	3.208
CS2t	1	0.54764	0.58649	0.8719	0.3504	1.729
CS3t	1	0.41403	0.33700	1.5094	0.2192	1.513
CS4t	0	0	.	.	.	
age3t	1	0.47597	0.40048	1.4125	0.2346	1.610
age2t	1	0.44272	0.39817	1.2363	0.2662	1.557
age1t	0	0	.	.		

4.4.2. Checking for Outliers

Another aspect we need to look at in this study is examination of regression diagnostic statistic to identify observations that exert too much influence on the fit of the model, if any.

In proportional hazards regression model we examine how far each individual covariate value puts unusual impact on the estimated coefficient of the covariates included in the model and on the maximized likelihood in the fitted model by deleting one observation's data at a time from the study. Hence, we reconsider the estimated coefficients of the model by refitting the same model in which one subject (i^{th} subject) deleted at a time. The Delta-Beta statistic, $\Delta\hat{\beta}_i = (\hat{\beta} - \hat{\beta}_{(-i)})$, from the SPSS output have revealed the following fact.

When the data on the 245th child is deleted, the estimated hemoglobin coefficient has shown the largest difference. The deletion decreased the hazard ratio, but the percentage change in the estimate is not much (6.6%) to delete the observation from the analysis. Therefore, we do not eliminate the data of the hemoglobin level of this child as an influential outlier observation.

Deleting the data on the 220th child from the dataset causes the largest change in the parameter estimate for the covariate age. The percentage change in the parameter estimates for age, when this observation is omitted becomes 6.5%. Deleting this observation decreases the relative hazard of death, but again this change is not big. Even if, it is not as such usual to examine the effect of categorical covariates on the estimated coefficients of the model we examined differences in the estimated coefficients for the levels of WHO clinical stages. However, none of percentage changes in the estimated coefficients of levels of clinical stage is that large.

We had examined also the impact of an observation on the overall estimated coefficients of the model using the likelihood displacement statistic ld_i. The statistics ld_i is an approximation to the amount of change in log partial likelihood when the i^{th} subject is deleted. The maximum value of the ld_i is 0.158651 from deletion of the 9^{th} subject. But the change is not much considerable. And we conclude that there is no influential outlier observation.

4.4.3. Assessing the goodness of fit of the model

We also need to check the goodness of fit of the proportional hazards model based on the empirical data. Therefore, for the model fitted in this study the Likelihood Ratio, score and Wald tests are used to compare (at 5% significance level) the goodness of fit of the model. The SAS output in Table 4.4.2 reveal that the log partial likelihood function *(-2LL)* without covariate was 731.122 while the function with covariates was 637.769. This means the model chi-square of 93.3533 with 8 degrees of freedom and p-value <0.0001 shows that the model is appropriate.

Table 4.4.2. SAS output Results of the Likelihood ratio, Score and Wald tests for testing the global null hypothesis of BETA=0.

Model Fit Statistics

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	93.3533	8	<.0001
Score	97.6271	8	<.0001
Wald	79.3802	8	<.0001

4.5. Discussion and Interpretation of the results.

The number of censored observations in the data set used for analysis was large (72.16%), that is we couldn't know the exact survival time of these observations. But we know that these children survived at least their censored time. This large number of censored observations would under estimate the mean survival time of the entire target population.

During 30 months of follow up the mean survival time were 22.381 months. In other similar studies at Zewditu hospital the mean survival time was 27.9 months during the 36 months of study time.

While interpreting the results we use the multiplicative form of the equation headed by $Exp(\hat{\beta})$, which is called the hazard ratio. Values of $Exp(\hat{\beta})$ less than one indicate that the variable under study is a significant factor in decreasing the risk of death of children under ART. The hazard ratio of children at CS(1), CS(2), CS(3) are less than one indicating that the risk of death is lower when compared with children at CS(4) which is the reference category. The risk of death was found to be lower for older children compared with younger age group; similarly the risk of death was higher for children in the lower hemoglobin groups relative to higher groups.

In our dataset, we compared children at different baseline clinical stage with the reference category clinical stage-IV and found that at any time during the study period children with baseline clinical stage-I are only 86.8% less likely to die as compared with children at clinical stage IV given that all other factors kept the same. The 95% confidence interval also suggests that the risk of death of children at clinical stage-I is as much as 35.7% and as small as 4.9% of the risk of death of children at clinical stage-IV. Children with baseline clinical stage-II are 82.4% less likely to die than children at clinical stage-IV. The confidence interval indicates that the risk of death for children at clinical stage-II is as small as 7.7% and as large as 40% of the risks of death of children at clinical stage-IV. Children with baseline clinical stage-III are 50.6% less likely to die than children at clinical stage-IV. The corresponding CI is interpreted as the hazard of death of children at clinical stage-III is as low as 29% and as large as 84.2% of the risk of death of children at clinical stage-IV. Children at clinical stage-III are 3.74 times more likely to die compared with children at clinical stage-I and 2.81 times

more likely to die as compared with children at clinical stage-II. Children with baseline clinical stage-II are 1.33 times more likely to die than children in baseline clinical stage-I.

According to the clinical classification of children based on their hemoglobin level, we considered hemoglobin value less than 7g/dl as the reference category. At any time during the study period the hazard rate of HIV- infected children with hemoglobin value between 7g/dl to 8.5g/dl is 54.6% less than the hazard of death of children with hemoglobin value less than 7g/dl given that other factors constant. The 95% CI also imply that the hazard rate of children with hemoglobin value fromn 7g/dl to 8g/dl is as low as 23.3% and as high as 88.5% of the risk of death of children with hemoglobin value less than 7g/dl. The rate of death of HIV-infected children with hemoglobin value between 8.5g/dl to 10g/dl is 0.208 with 95% CI (0.092, 0.469). This is interpreted as at any time during the study period children in this category are dying 79.2% less likely than children in the reference category, keeping all factors the same. The 95% CI suggests that the rate of death could be as low as 9.2% and as high as 46.9% of the reference category. The hazard ratio for hemoglobin value above 10g/dl is 0.072. This is interpreted as HIV-infected children in this category are dying at a rate which is about 92.8% lesser than the death rate of children dying in the reference category, if all other factors are kept similar. The 95% CI suggests that the rate could be as low as 3.2% and as high as 16.5% of the reference category. Children with hemoglobin value between 8.5g/dl to 10g/dl are 0.458 times less likely to die than children with hemoglobin value between 7g/dl to 8.5g/dl and 2.89 times more likely to die than children with hemoglobin value greater than 10g/dl provided that all other factors are held constant.

HIV-infected children aged between 1.5 to 5 years are 71.5% less likely to die compared with HIV-infected children aged below 1.5 years (the reference category). The 95% CI confirms that this hazard of death for this category could be as low as 14.4% and as high as 56.4%

compared with the reference category. HIV-infected children aged between 5 and 14 years are 81.6% less likely to die than children with age below 1.5 years. The 95% CI verifies that the rate of death could be as small as 9.3% and as high as 36.5%. HIV-infected children aged between 1.5 to 5 years are 1.55 times more likely to die than HIV-infected children aged between 5 and 14 years given that all other factors are constant.

CHAPTER 5

5. CONCLUSION

In this study, we applied the Cox proportional hazards regression model to the survival time of children after ART initiation to identify influential variables/factors which might influence the outcome variable, survival time of children. First we used a nonparametric model, namely the Kaplan-Meier model to estimate survival functions and analyze the data for single factors over different levels. Then we used bivariate Cox proportional hazards method to find the most significant covariates. In a third stage the multivariable Cox Proportional Hazard model was used to fit the data. We checked linearity for continuous covariates as well as possible interactions of order two for covariates in the preliminary final model. We also checked if the assumption of proportional hazard is fulfilled and also examined for the presence of influential observations for those covariates we have just found significant. Finally, we constructed a model which could serve to predict the survival functions of HIV infected children after ART initiation.

The study identified factors/variables which influence the survival of HIV-infected children after beginning ART at Bahir Dar Felege-Hiwot hospital. The result showed that baseline hemoglobin level, WHO clinical stage and age have significant influence on the survival time of children after ART. The mean survival time was found to be 22.381 months with standard deviation 0.693 months. However this estimate underestimates the true population value as much of the observations are not survival time but rather they are censored times.

WHO clinical stage of children was found to be an important predictor for survival of HIV-infected children after ARV treatment. The risk of death could be minimized by initiating ART for children at clinical stage I or II instead of providing ART at clinical stage III and/or IV.

HIV-infected children with lower hemoglobin level and younger age have higher risk of death as compared with HIV-infected children who have higher hemoglobin level and older age.

REFERENCE

AIDS Resource Center, Johns Hopkins University, Bloomberg School of Public Health, Center fo Communication Programs 2005, *Ethiopian National ART Strategic Communication Framework,* website: http://www.etharc.org

Binyam Taye 2008, *Impact of malnutrition in survival of HIV infected children after initiation of ART*, Unpublished M.Sc. Thesis Addis Ababa University.

Bollinger, L., J. Stover, 1999, *The Economic Impact of AIDS*, The Futures Group International, Glastonbury, CT.

Bolton-Moore, M. 2007, 'Clinical outcomes and CD4 cell response in children receiving antiretroviral therapy at primary health care facilities in Zambia', *Journal of the American Medical Association*, vol. 298, no. 16, pp. 1888–1899.

Brady MT, Oleske JM, Williams PL, Elgie C, Mofenson LM, Dankner WM, Van Dyke RB; for the Pediatric AIDS Clinical Trials Group219/219C Team 2010, 'Declines in Mortality Rates and Changes in Causes of Death in HIV-1-Infected Children During the HAART Era', *J Acquir Immune Defic Syndr. 2010 Jan;53(1):86-94*

Collett, D. 2003, *Modeling survival data in medical research*, 2nd edition, Chapman and Hall/CRC, London

Federal HIV/AIDS Prevention and Control Office, Federal Ministry of Health 2007, *Guidelines for paediatric HIV/AIDS care and treatment in Ethiopia*.

Fontana M, Zuin G, Plebani A, 1999, 'Body composition in HIV infected children: relations with disease progression and survival', *Am J clin Nutr*, 69:1282-86

George E, Noel F, Bois G, 2007, 'Antiretroviral therapy for HIV-1- infected children in Haiti', *J Infect Dis*, 195:1411-1418

Gibb DM, Duong T, Dunn D, Chintu C, Mulenga V, Cotton M, Eley B, Zar H, Ellis J, Graham S, Giaquinto C, Nanyonga M, Msellati P, Meyers T, Moultrie H, Hani C, Pinto J, Roux P, Weigel R, Araujo C, Carvalho A, Carvalho I, Diniz A, Ferreira F, Lobato V, Sanchez T, Morelli E, Atai B, Duff C 2008, 'Markers for predicting mortality in untreated HIV-infected children in resource-limited settings**',** AIDS 2008;22(1):97-105.

Global AIDS Program (GAP) in Ethiopia. About the global AIDS program in Ethiopia available at http://www.cdc.gov/nchstp/gap/countries/ethiopia.htm.

Grambsch P.M. and Therneau T.M. 1994, *Proportional hazards tests in diagnostics based on weighted residuals*, Biometrica, 81: 515-526.

Hosmer D.W & Lemeshow S. 1998, *Applied Survival Analysis Regression Modelling of Time to Event Data*, John Wiley and Sons, Inc. New York.

Hussey GD, Reijnhart RM, Sebens AM, Burgess J, Schaaf S, Potgieter S 1998, 'Survival of children in Cape Town known to be vertically infected with HIV-1', S Afr Med J. 1998 May;88(5):554-8.

Isaakidis Petros, Marie-Eve Raguenaud, Vantha Te, Chhraing S Tray, Kazumi Akao, Varun Kumar, Sopheak Ngin, Eric Nerrienet, and Rony Zachariah 2010, 'High survival and

treatment success sustained after two and three years of first-line ART for children in Cambodia', *J Int AIDS Soc*., 2010; 13: 11

Joint United Nations Program on HIV/AIDS (UNAIDS) and World Health Organization (WHO) 2009, *AIDS epidemic update 2009*, Geneva.

Kabue Mark M , Adeodata Kekitiinwa, Albert Maganda, Jan M Risser, Wenyaw Chan and Mark W Kline 2008, 'Growth in HIV-infected children receiving antiretroviral therapy at a pediatric infectious diseases clinic in Uganda', *AIDS Patient Care STDS*, 22(3):245-51.

Kuhn L, Mathews C, Fransman D, Dikweni L, Hussey G. 1998, 'Child feeding practices and attitudes of HIV-seropositive mothers in Cape Town, South Africa', *AIDS*, 13(1): 144-5.

National AIDS resource center 2009, Website
http://www.etharc.org/resources/healthstat/nationalfactsheet/12-nationalfactsheet2009

National Factsheet 2010, Website
http://www.etharc.org/resources/healthstat/nationalfactsheet/13-nationalfactsheet2010

Ng'andu N.H. 1997, *An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model*, statistics in medicine, 16: 611-626

NIAID: National Institute of Allergy and Infectious Diseases.
http://clinicaltrials.gov/ct2/show/NCT00234091

Selamawit Ejigu 2009, First line Antiretroviral Treatment Failure and Factors Associated with it in Addis Ababa, Master's thesis published in Addis Ababa University

Seyoum E, Mekonen Y , Kassa A, Eltom A , Damitew T, Lera M, Felema B, Assefa Y 2009, *ART Acal-Up in Ethiopia*: Success and Challenges, HAPCO, Plan, Monitoring & Evaluation Directorate

Song R. J. 2007 'Efficacy of highly active antiretroviral therapy in HIV-1-infected children in Kenya', *Pediatrics*, vol. 120, no. 4, pp. 856–861.

UNAIDS 2000, *HIV/AIDS and Economic Development in Sub-Saharan African Countries:* African Development forum 2000, USAID.

UNAIDS 2004, *Report on the Global HIV/AIDS epidemic*, Joint United Nations Program on HIV/AIDS, Geneva, Switzerland.

UNAIDS 2008, *Report on Global AIDS Epidemic-Google Book*:
Website:http://www.unaids.org/en/media/unaids/contentassets/dataimport/pub/globalreport/2008/jc1510
Virco 2008, *HIV resistance learning system*, module: Overview of HIV/AIDS, Virci Lab, New Jersey

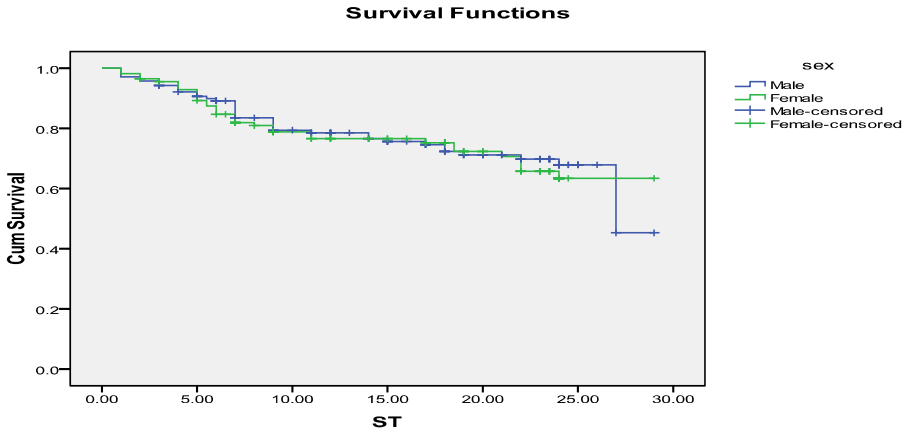
Walker SA, Mulenga V, Ford D 2007, 'The impact of daily cotrimoxazole Prophylaxis and Antiretroviral Therapy Mortality and Hospital Admission in HIV-infected Zambian Children', *Am J Clin infect Dis*, 44:1361-1367

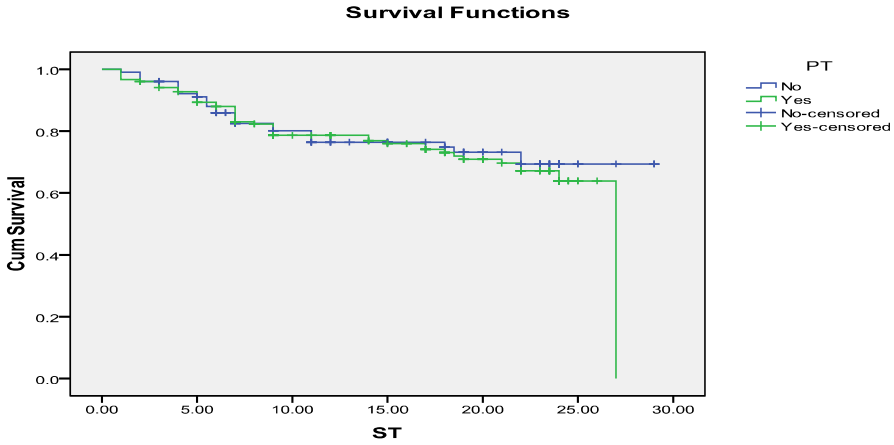
World Health Organization 2009, *Hospital Care for Children*: Guideline for the Management of Common Illnesses with Limited Resources, 20 Avenue Appia, 1211 Geneva 27, Switzerland.

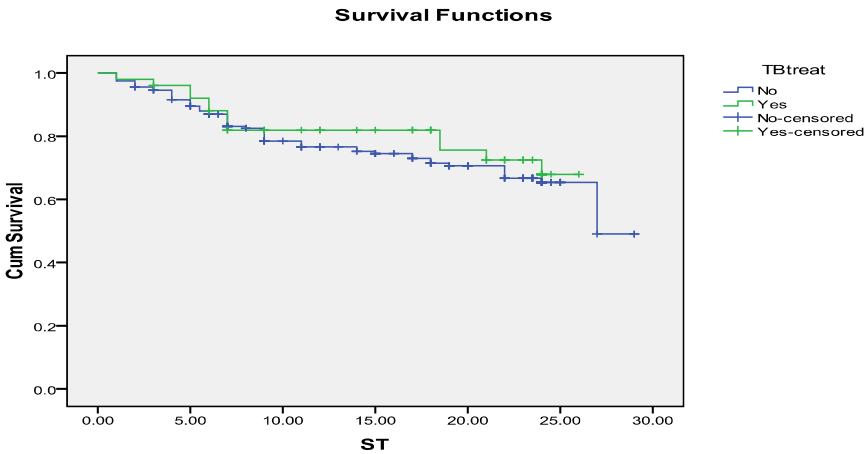
Zhang Xinjian 2007, *HIV/AIDS Relative Survival Analysis*, Master's thesis published in Georgia University.

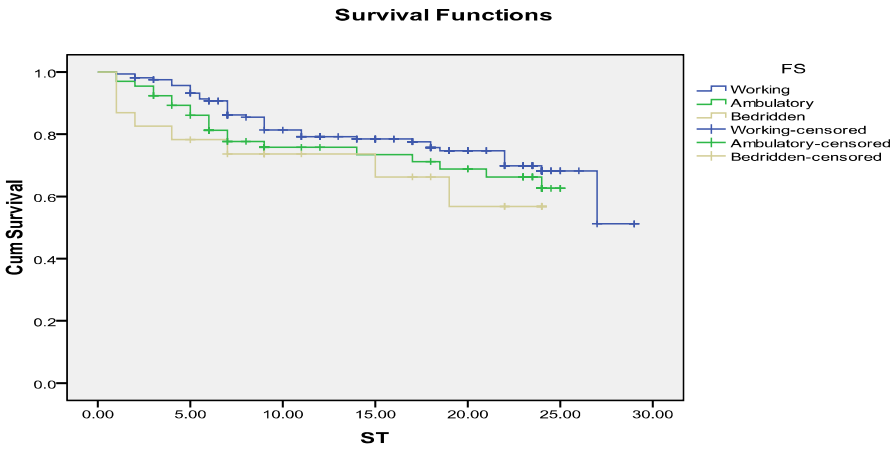
Appendix A

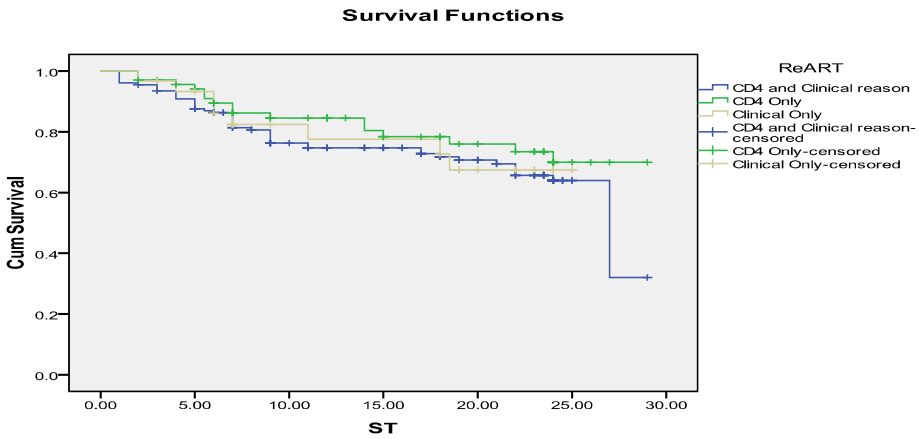
Figure 4.1.1: Plots of Kaplan-Meier survivor function estimate of HIV-infected children treated at Bahir Dar Felege-Hiwot Hospital ART Unit based on different factors, 2010

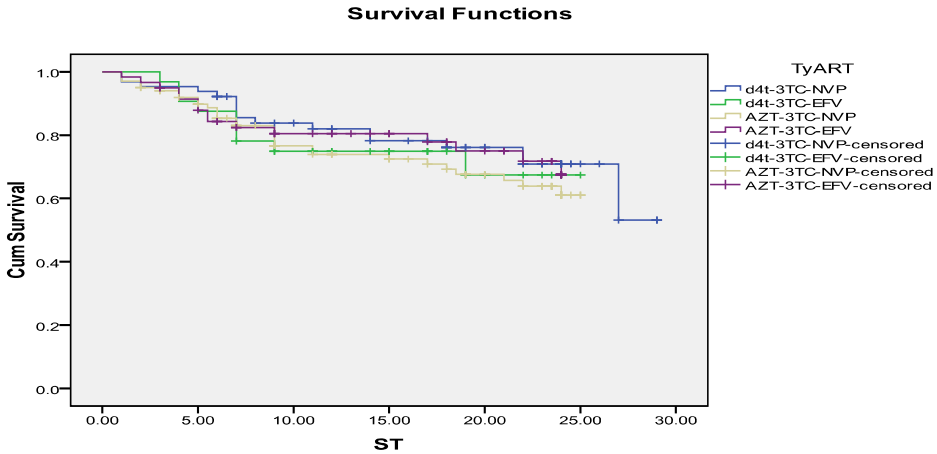

Survival Functions
sex
Male
Female
Male-censored
Female-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST

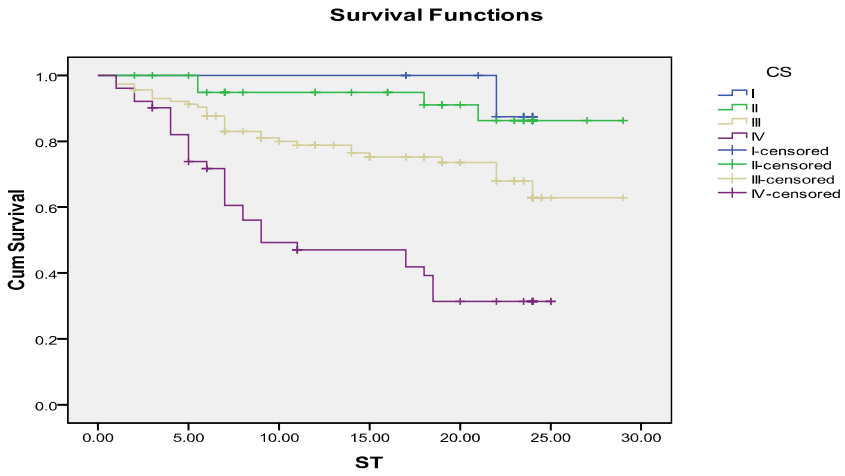

Survival Functions
PT
No
Yes
No-censored
Yes-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST

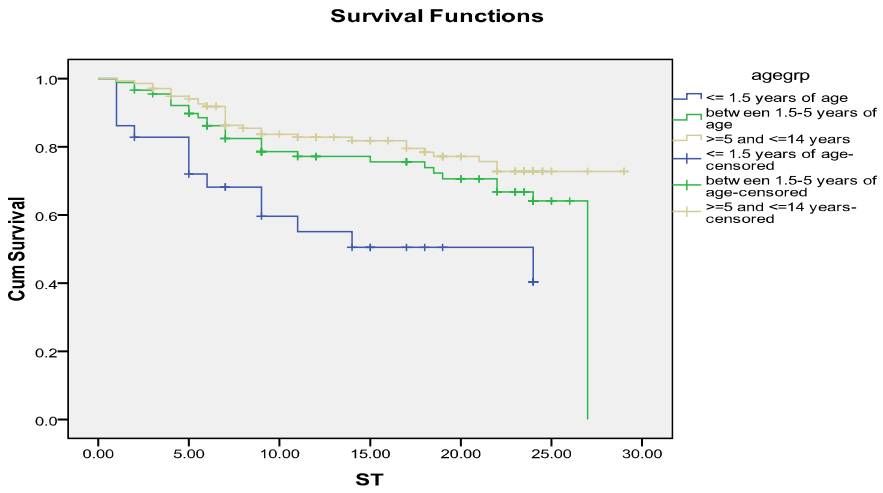

Survival Functions
TBtreat
No
Yes
No-censored
Yes-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST

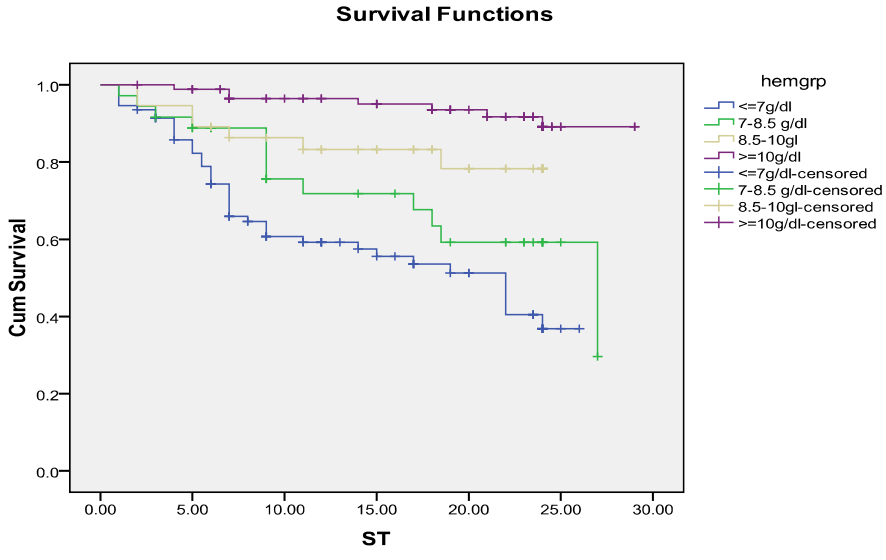

Survival Functions
FS
Working
Ambulatory
Bedridden
Working-censored
Ambulatory-censored
Bedridden-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST

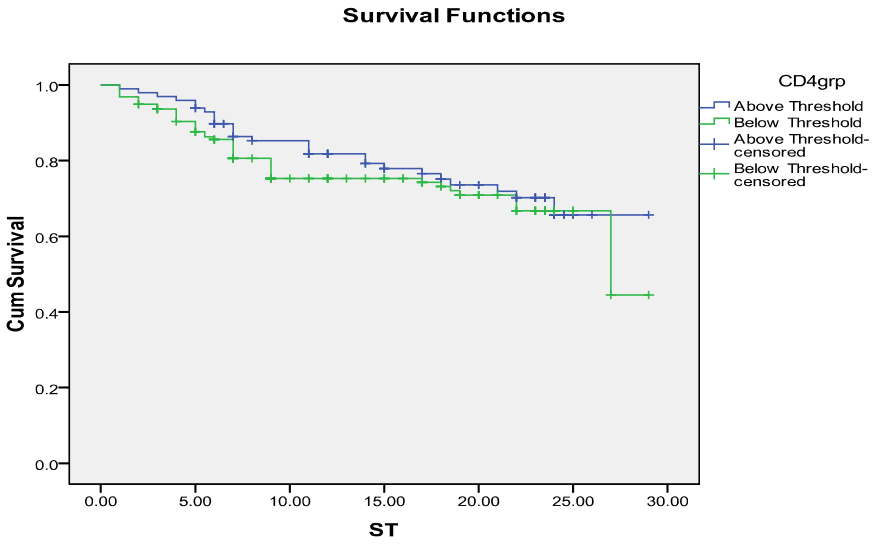

Survival Functions
ReART
CD4 and Clinical reason
CD4 Only
Clinical Only
CD4 and Clinical reason-censored
CD4 Only-censored
Clinical Only-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST

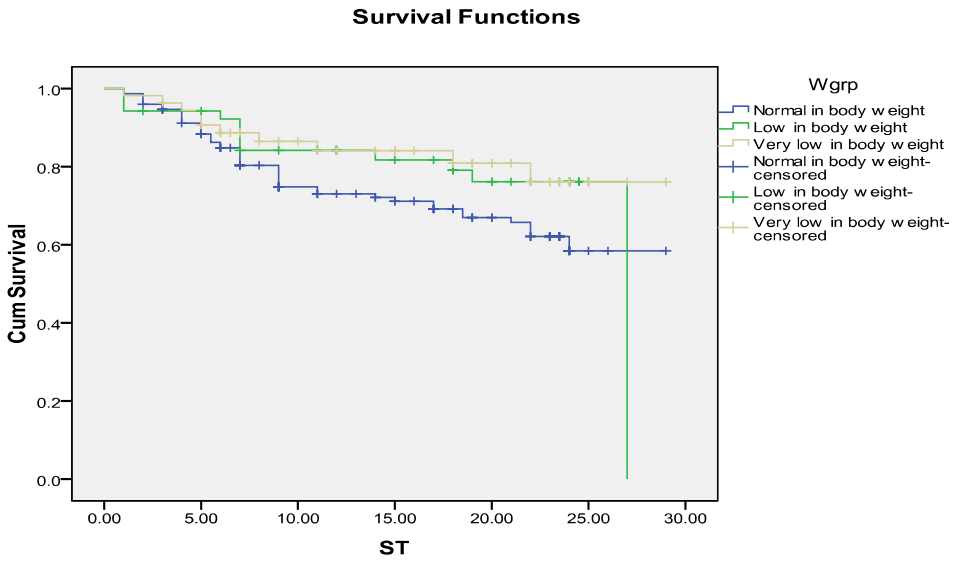

Survival Functions
TyART
d4t-3TC-NVP
d4t-3TC-EFV
AZT-3TC-NVP
AZT-3TC-EFV
d4t-3TC-NVP-censored
d4t-3TC-EFV-censored
AZT-3TC-NVP-censored
AZT-3TC-EFV-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST


Survival Functions
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST
CS
I
II
III
IV
I-censored
II-censored
III-censored
IV-censored


Survival Functions
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST
agegrp
<= 1.5 years of age
betw een 1.5-5 years of age
>=5 and <=14 years
<= 1.5 years of age-censored
betw een 1.5-5 years of age-censored
>=5 and <=14 years-censored


Survival Functions
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST
hemgrp
<=7g/dl
7-8.5 g/dl
8.5-10gl
>=10g/dl
<=7g/dl-censored
7-8.5 g/dl-censored
8.5-10gl-censored
>=10g/dl-censored


Survival Functions
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST
CD4grp
Above Threshold
Below Threshold
Above Threshold-censored
Below Threshold-censored


Survival Functions
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
ST
Wgrp
Normal in body weight
Low in body weight
Very low in body weight
Normal in body weight-censored
Low in body weight-censored
Very low in body weight-censored

Appendix B

Figure 4.3.1: Plots of the martingale residuals for the model excluding a continuous covariate against the values of the excluded covariate

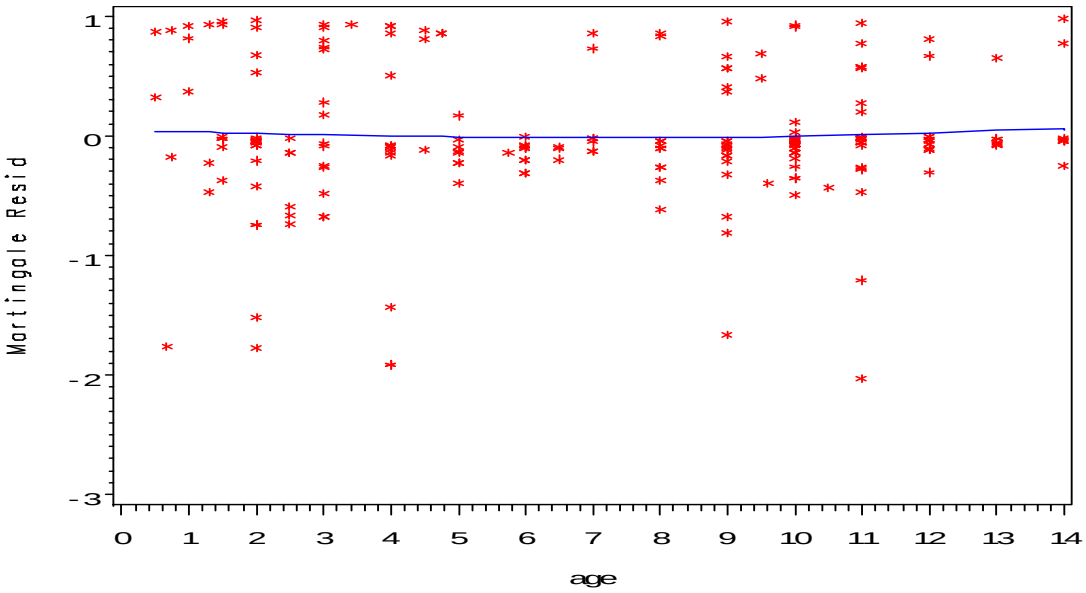


Table 4.3.1a: Martingale residual for the model against excluded age

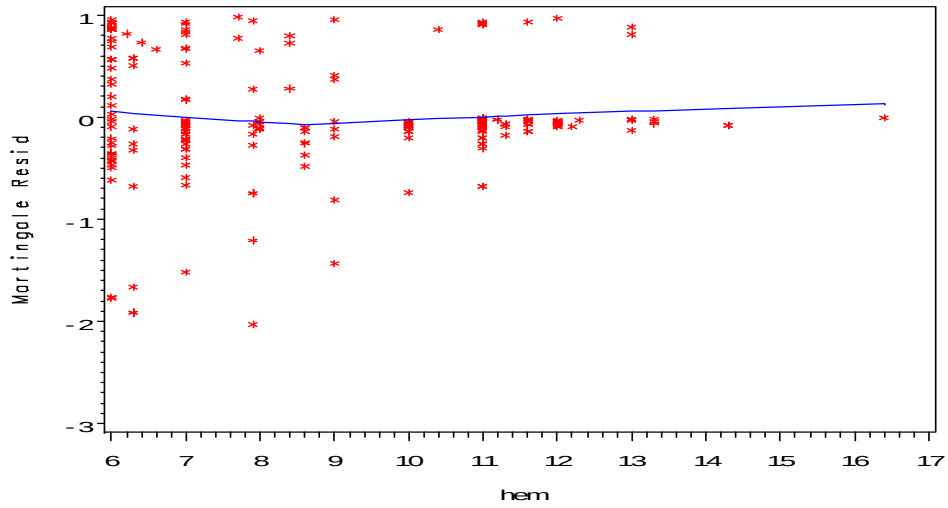
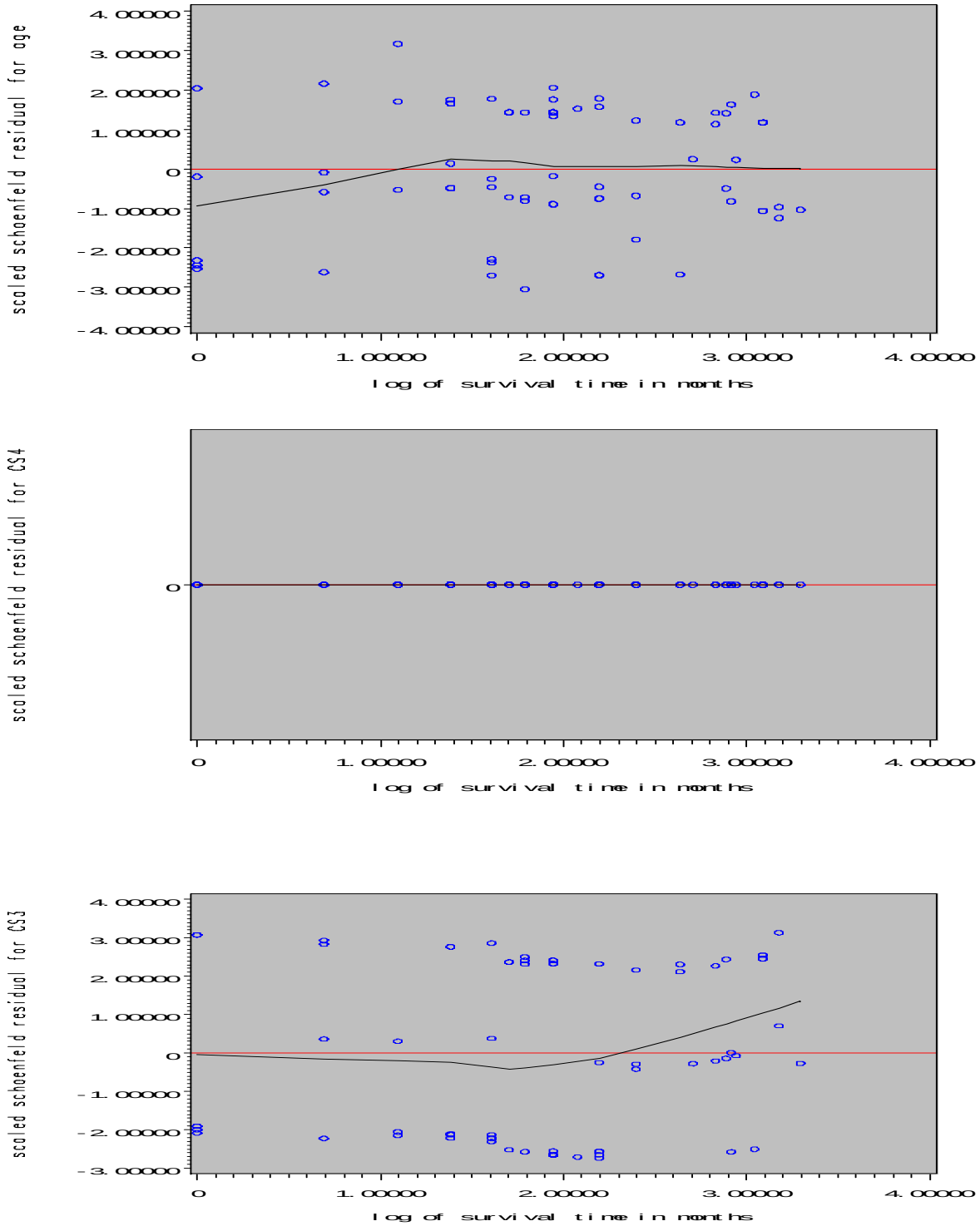
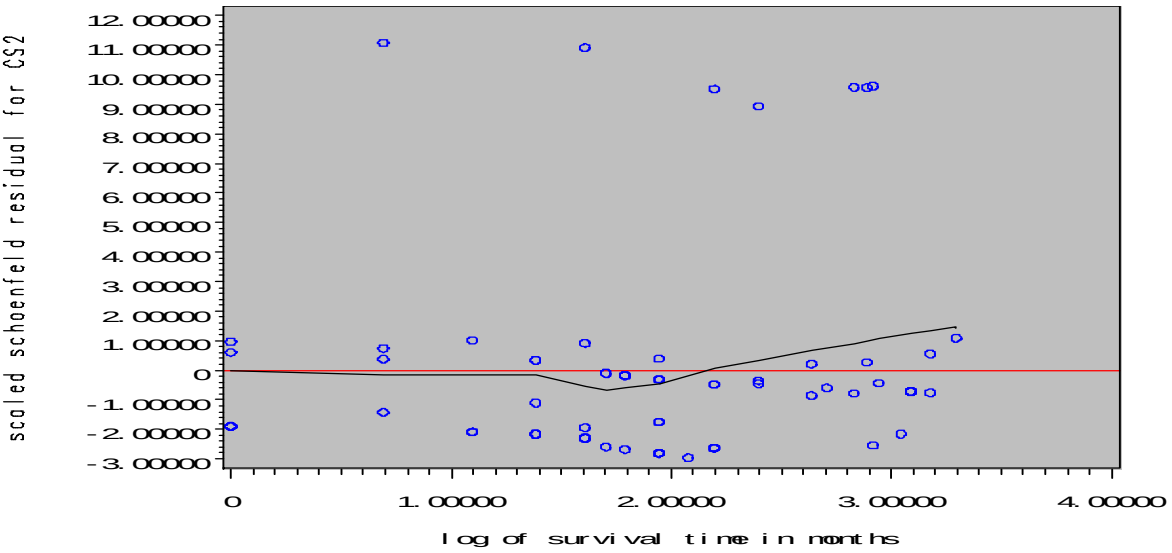


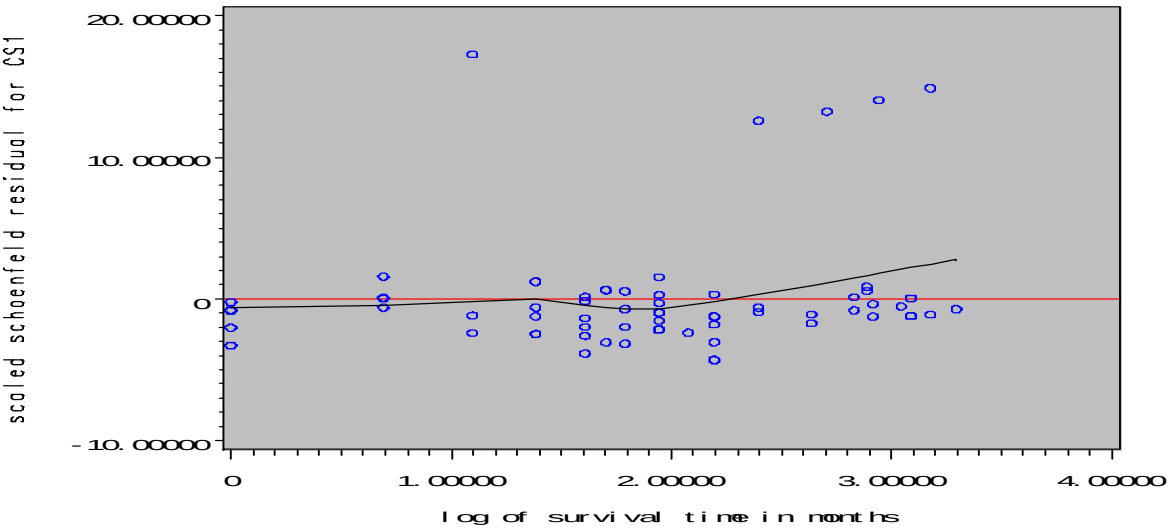
Table 4.3.1b: Martingale residual for the model against excluded hemoglobin

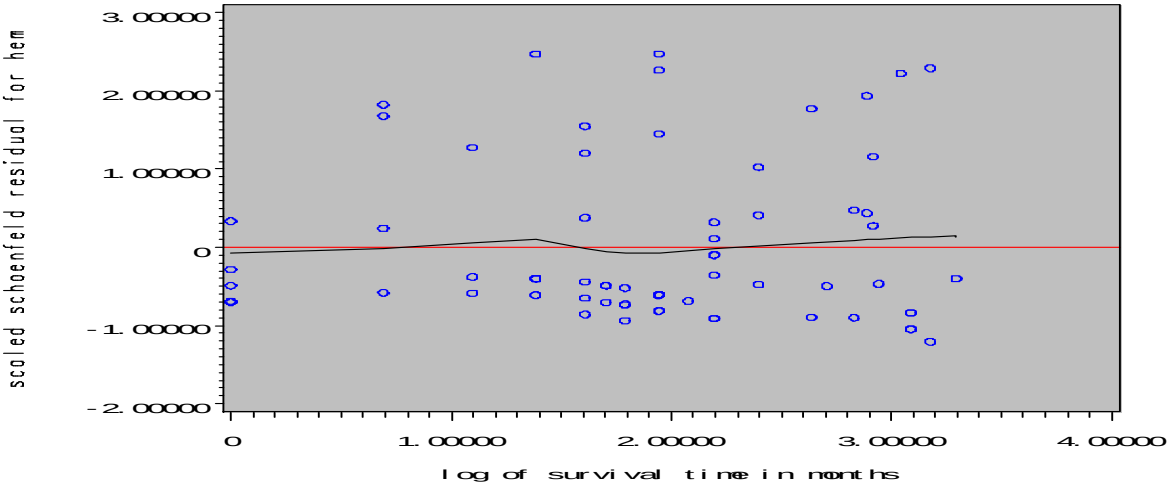
Appendix C

Figure 4.4.1.1: Plots of scaled Schoenfeld residuals for baseline covariate in the preliminary final model against log of survival time




scaled schoenfeld residual for CS2
12.00000
11.00000
10.00000
9.00000
8.00000
7.00000
6.00000
5.00000
4.00000
3.00000
2.00000
1.00000
0
-1.00000
-2.00000
-3.00000
0
1.00000
2.00000
3.00000
4.00000
log of survival time in months


scaled schoenfeld residual for CS1
20.00000
10.00000
0
-10.00000
0
1.00000
2.00000
3.00000
4.00000
log of survival time in months


scaled schoenfeld residual for hem
3.00000
2.00000
1.00000
0
-1.00000
-2.00000
0
1.00000
2.00000
3.00000
4.00000
log of survival time in months