Chapter One

I. Introduction

1.1 Background of the Study

AIDS is one of the most notorious infectious disease epidemics in human history. The report of Centers for Disease Control and Prevention indicates a type of chimpanzee in West Africa was the origin of HIV infection in humans. The virus most likely jumped to humans when humans hunted these chimpanzees for meat and came into contact with their infected blood. Over several years, the virus slowly spread across Africa and later into other parts of the world (CDC 2008).

The first AIDS case was recognized in 1981 among homosexual men in Los Angeles and New York cities of the United States of America. The men had an unusual type of lung infection (pneumonia) called Pneumocystis Carinii[1] Pneumonia (PCP) and rare skin tumors called Kaposi's sarcoma. The patients were noted to have a severe reduction of a type of cell in the blood that is an important part of the immune system, called CD4 cells. These cells, often referred to as T cells, help the body fight infections. In 1983, researchers in the United States and France described the virus that causes AIDS, now known as the human immunodeficiency virus (HIV) and belonging to the group of viruses called retroviruses. In 1985, a blood test became available that measures antibodies to HIV that are the body's immune response to the HIV. This blood test remains the best method for diagnosing HIV infection (Medicine net.com 2009).

HIV stands for the Human Immunodeficiency Virus. H - Human because this virus can only infect human beings, I - Immune deficiency because the effect of the virus is to create a

[1] now known as Pneumocystis jiroveci

deficiency, a failure to work properly within the body's immune system and V - Virus because this organism is a virus, which means one of its characteristics is that it is incapable of reproducing by itself. It reproduces by taking over the machinery of the human cell. Like other viruses, HIV attacks cells in the body. But what makes HIV different is that the immune system can never fully get rid of HIV because the virus attacks the immune system itself - the very mechanism that would normally get rid of a virus.

AIDS stands for acquired immunodeficiency syndrome. It is a set of symptoms and infections resulting from the damage to the human immune system caused by HIV. Each letter stands for A - Acquired because it is a condition one must acquire or get infected with, not something transmitted through the genes. I - Immune because it affects the body's immune system, the part of the body which usually works to fight off microorganisms such as bacteria, viruses, parasites and fungi D - Deficiency because it makes the immune system deficient (that is, the immune system may not function properly). S - Syndrome because someone with AIDS may experience a wide range of different diseases and opportunistic infections.

Two different types of HIV, HIV-1 and HIV-2, cause infection and disease in humans, HIV-1 is thought to have arisen from cross-species transmission of a chimpanzee virus to humans and HIV-2 from cross-species transmission of a Sooty mangabey virus. Three groups of HIV-1 have been described, labeled M, N and O based on genome differences. Most HIV-1 infections are caused by group M viruses and these are divided into 9 subtypes known as clades (A-D, F-H, J, and K).The DNA sequences of viruses in distinct clades can differ by 15%-20%. In comparison with HIV-1, HIV-2 is much less prevalent; infection with HIV-2 is associated with a slower progression to immune deficiency and seems to be less efficiently transmitted even from infected women to their offspring (Myron et.al 2008).

T cells are also called CD4 cells, CD4 T cells or CD4 cell lymphocytes are white blood cells. There are two types, helper T cells and killer T cells. Helper T cells recognize the antigen and activate the killer T cells. Killer T cells then destroy the antigen. When HIV is introduced into the body, this virus is too strong for the helper T cells and killer T cells. The virus then invades these cells and starts to reproduce itself, thereby not only killing the CD4 T cells, but also spreading to infect other healthy cells.

Immediately following infection with HIV, most individuals develop a brief, nonspecific "viral illness". When the immune system is gradually being destroyed by the virus, this destruction will lead the symptoms of AIDS to appear. These symptoms include extreme fatigue, rapid weight loss, appearance of swollen or tender glands in the neck, armpits or groin, shortness of breath, dry cough, persistent diarrhea, high fever or soaking nights sweats appearance of one or more purple spots on the surface of the skin, inside the mouth, anus or nasal passages, whitish coating on the tongue, throat or vagina and also forgetfulness, confusion and other signs of mental deterioration.

Some people infected with HIV may develop a disease that is less serious than AIDS, referred to as AIDS Related Complex (ARC). ARC is a condition caused by the AIDS virus in which the patient tests positive for AIDS infection and has a specific set of clinical symptoms. However, ARC patients' symptoms are often less severe than those with classic AIDS because the degree of destruction of the immune system has not progressed as far as it has in patients with classic AIDS.

The AIDS epidemic has now spanned nearly three decades. The cumulative total number of individuals infected with HIV-1 and death due to AIDS since the pandemic began exceeds 60 million and 25 million people, respectively (Myron et.al 2008). At the end of 2007, the Joint United Nations Program on HIV/AIDS (UNAIDS) and the WHO estimated that there were 33

million [30.3-36.1 million] people living with HIV around the world, that 2.7 million [2.2 to 3.2 million] individuals became newly infected with HIV in 2007 which is slightly below 3 million in 2001, and that 2.0 million [1.8-2.3 million] people died of AIDS in that year.

Sub-Saharan Africa is home to around 26 million people living with HIV. In 2005 it was estimated that 3.2 million became newly infected with HIV, while 2.4 million died of AIDS (Jerene 2007). Countries in Sub-Saharan Africa including South Africa, Botswana and Swaziland, remain the center of the AIDS epidemic. The region has about 67% of all people infected with HIV and 72% of all AIDS deaths UNAIDS (2008). According to UNAIDS report heterosexual intercourse is still driving the epidemic, in sub-Saharan Africa, which shouldered two-thirds of the global AIDS burden and three-quarters of all AIDS-related deaths in 2007. In other parts of the world HIV is mainly affecting people who inject drugs, men who have sex with men (MSM) and sex workers UNAIDS (2008). The Ethiopian HIV/AIDS epidemic shares many features with that elsewhere in Northeast and East Africa. It is currently a general epidemic, defined as more than 1% of the national sexually active adults sero-positive, including more highly infected groups (Kloos et.al 2007).

AIDS is transmitted via three main routes: the most common mode of transmission is the transfer of body secretions through sexual contact. This is accomplished through exposure of mucous membranes of the rectum, vagina or mouth to blood, semen or vaginal secretions containing the HIV virus. Blood or blood products can transmit the virus, most often through the sharing of contaminated syringes and needles. HIV can be spread during pregnancy from mother to fetus.

One cannot get AIDS/HIV from touching someone or sharing items, such as cups or pencils, or through coughing and sneezing. Additionally, HIV is not spread through routine contact in

restaurants, the workplace or school. However, sharing a razor does pose a small risk in that blood from a minor nick can be transmitted from one person to another.

There are two approaches to reducing the burden of sickness and death associated with HIV that lead to acquired immunodeficiency syndrome (AIDS) - Prevention and Treatment.

Gains in saving lives by preventing new infections is the best choice like it is said in Amharic version "Tamo kememakek askedemo metenkeke" [2]

The only way to protect from contracting AIDS sexually is to abstain from sex outside of a mutually faithful relationship with a partner whom the person knows is not infected with the AIDS virus. Otherwise, risks can be minimized if they: don't have sexual contact with anyone who has symptoms of AIDS or who is a member of a high risk group for AIDS, avoid sexual contact with anyone who has had sex with people at risk of getting AIDS unless they know with absolute certainty that their partner is not infected, a latex condom should be used during each sexual act, don't have sex with prostitutes, avoid having sex with anyone who has multiple and/or anonymous sexual partner, avoid oral, genital and anal contact with partner's blood, semen, vaginal secretions, feces or urine; avoid anal intercourse altogether, don't share toothbrushes, razors or other implements that could become contaminated with the blood of anyone who is or might be infected with the AIDS virus, exercise caution regarding procedures, such as acupuncture, tattooing, ear piercing, etc., in which needles or other no sterile instruments may be used repeatedly to pierce the skin and/or mucous membranes.

Long term sustainable provision of treatment is a second choice for people living with HIV, who have already got the HIV infection. There is no cure for HIV infection or AIDS nor is there a vaccine to prevent HIV infection. However, new medications not only can slow the

[2] It is better to protect oneself prior to getting infected and suffering from the disease.

progression of the infection, but can also markedly suppress the virus, thereby restoring the body's immune function and permitting many HIV- infected individuals to lead a normal disease-free life.

Opportunistic infection is an infection that occurs because of a weakened immune system. It is of particular danger for people living with AIDS. The HIV virus itself does not cause death, but the opportunistic infections that occur because of its effect on the immune system can (Medicine net.com (2008)).

For the purpose of prevention and early treatment of infections in Pre-antiretroviral treatment and under antiretroviral treatment people with HIV take a number of medications like cotrimoxazole, fluconazole, TB RX, INH and other medications depending on the type of the opportunistic illness they have. Cotrimoxazole is cheap, effective and widely used for its activity against a wide variety of microorganisms. Patients coinfected with TB and HIV, have a high risk of death not only because of TB but because of increased susceptibility to pneumonia and other infections. Offering cotrimoxazole to all adults with TB at the onset of anti-TB treatment could be an effective, simple, and safe way to reduce mortality, especially in settings where HIV coinfection is common. According to a study supported by the National Institute of Allergy and Infectious Diseases (NIAID) the drug fluconazole significantly delays the onset of fungal infections in people with AIDS (NIAID 1993).

Survival patterns after HIV infection in African populations in the era before antiretroviral therapy (ART) form an important baseline for measuring future successes of treatment programmes (Isingo et.al 2007). To maintain good adherence to ART a Pre-ART assessment of patient's knowledge of ART is essential (Watanabe et.al 2004). Pre-Antiretroviral therapy includes guidelines that attempt to address the factors which are important in the holistic approach to patient management which could also influence the progression and outcome of

disease including natural history of HIV infection, primary prophylaxis and immunization, nutrition, support and counseling (Martin, 2004).

A decade ago, having AIDS was almost equivalent to a death sentence. Since 1996, with the introduction of ART, AIDS has become a chronic, manageable disease. HIV antiretroviral drug treatment is the main type of treatment for HIV or AIDS. Its aim is to keep the amount of HIV in the body at a low level. This stops any weakening of the immune system and allows it to recover from any damage that HIV might have caused already. The drugs are often referred to as antiretroviral, anti-HIV or anti-AIDS drugs, HIV antiviral drugs or ARVs. Standard antiretroviral therapy (ART) consists of the use of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of HIV disease (UNAIDS 2007).Taking two or more antiretroviral drugs at a time is called combination therapy. Taking a combination of three or more anti-HIV drugs is sometimes referred to as Highly Active Antiretroviral Therapy (HAART).

UNAIDS report of 2008 indicated that nearly 3 million people in low-and middle-income countries were receiving ARV treatment and when compared with some 300,000 taking AIDS drugs in the year 2003, the number of people on AIDS medication jumped by 10 times in the last four years. Eventhough AIDS drugs have become much cheaper and more available because of a variety of government and private programs, millions of others still do not have access to the drugs.

According to a new report released on 2008, by the World Health Organization (WHO), the United Nations Joint Program on HIV/AIDS (UNAIDS), and the United Nations Children's Fund (UNICEF) by the end of 2007, nearly half a million HIV positive pregnant women or about 33% worldwide received antiretroviral drugs to prevent mother-to-child HIV transmission (PMTCT) and approximately 200,000 children with HIV worldwide were

receiving antiretroviral treatment. About 1 million people (2.2% of the adult population) were living with HIV in Ethiopia in 2008. In the same year, approximately 290,000 people needed ART (Yibeltal et.al 2009).

Currently there are about 21 Food and Drug Administration (FDA) approved antiretroviral drugs (ARVs) in the world and there are five groups of antiretroviral drugs that attack HIV in different ways. Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs) interfere with the action of an HIV protein called reverse transcriptase, which the virus needs to make new copies of itself, Non-Nuclesoside Reverse Transcriptase Inhibitors (NNRTIs) stop HIV from replicating within cells by inhibiting the reverse transcriptase protein. Protease Inhibitors (PIs) inhibit protease which is another protein involved in the replication process. Fusion or Entry Inhibitors prevent HIV from binding to or entering human immune cells and Integrase Inhibitors interfere with the integrase enzyme, which HIV needs to insert its genetic material in human cells.

WHO recommends that in resource-limited settings HIV infected adolescents and adults should start antiretroviral therapy when the following conditions are met: WHO stage 4 disease (clinical AIDS), regardless of CD4 count, stage 3 diseases with consideration of CD4 count < 350/mm^3 in assisting decision making, stages 1 or 2 diseases with CD4 cell count < 200/mm^3. In a setting where CD4 count is not available, the total lymphocyte count (TLC) can be used, and treatment is recommended for: WHO stage 3 or 4 disease (clinical AIDS) irrespective of the TLC and stage 2 disease with TLC $\leq$ 1200/mm^3.

1.2 Statement of the Problem

The HIV epidemic has resulted in history's single sharpest reversal in human development (UNDP, 2005). Although no cure has yet been found, many advances have been made in the

way HIV /AIDS is identified and treated. In developed countries most of those suffering from acute AIDS have access to effective Anti-retroviral Drugs (ARVs) and AIDS related deaths have decreased significantly since the peak years of the early 1990's (Laura 2004). However, the problem of HIV infection persists even for resource-rich countries, such as the U.S, as socially marginalized populations face the greatest risk of contracting HIV and are among those least likely to be able to afford treatment. For example, homosexuals and needle-users continue to be the groups with highest seroprevalence in the U.S

High HIV transmission rates at the individual and community levels in Africa and other parts of the developing world have been attributed to a wide range of behavioral, biological, socioeconomic, cultural and political factors (Kloos et.al 2007). Sub-Saharan Africa is the hardest-hit region, as HIV/AIDS thrives on conditions such as poverty, migration, economic instability and low levels of literacy.

AIDS is fatal; there is no medicine to cure it. It affects all people, irrespective of color, religion, age, gender etc (Alula Pankhurst 2005). Individuals taking anti-HIV medications can still transmit HIV to others. The most sexually active, those aged 15-49, adults which are able to play key roles in economic production in societies are obviously at a greater risk of infection than others. As more adults become ill or pass away, more children are forced to quit going to school, become involved in child labor (including high-risk activities such as sex work), and are orphaned.

HIV/AIDS epidemic has a political impact as politics can interfere with the potential for real and potent action in providing effective funding, and helping people living with HIV/AIDS. Given the long-term requirements of treating AIDS patients worldwide as well as maintaining prevention and awareness programs, the likelihood of changing political climates is strong.

The AIDS epidemic has often been associated with severe negative public reaction to a person who is assumed to be infected with HIV. For those living with HIV/AIDS, stigma and discrimination can make access to services and treatment more difficult. Stigma is one of the socio-demographic factors that people who live with HIV/AIDS face. It can be better understood when the perspectives are realized both from the outsider's and the insider's viewpoints. Many fear even being tested because of the risk of others finding out about their HIV status has negative impact on the quality of life for those seeking testing and treatment, this kind of discrimination also furthers the spread of the epidemic.

The status of women, particularly in Sub Saharan Africa, also affects the epidemic's nature and the way it is spread. In some countries women are not allowed the freedom to choose their sexual partners and often fall victim to sexual assault and rape. Additionally, women receive less education and have higher illiteracy rates, making information about AIDS prevention and treatment less available, and because of their lack of education, women who have to find work are more likely to turn to high-risk activities such as prostitution. Women are often unable or unwilling to ask their husbands or partners to wear a condom and it has become common for women to be infected by their husbands. Women in the household disproportionately care for children and those that are ill. As a result they will be vulnerable to HIV as more and more mothers and care-givers become ill and succumb to AIDS related deaths. The risk of HIV transmission is high on pregnant women if prevention from mothers to child treatment is not taken.

HIV/AIDS is a significant and almost wholly unrecognized problem among disabled populations worldwide. While individuals with disability are at risk for HIV infection, subgroups within the disabled population – most notably women with disability, disabled members of ethnic and minority communities, disabled adolescents and disabled individuals

who live in institutions, are at especially increased risk. HIV/AIDS educational, testing and clinical programs are largely inaccessible to individuals with disability.

The HIV/AIDS epidemic is a major threat to Africa's economic development. The impact is widespread, affecting individuals, households, firms and governments. The epidemic will harm prospects for African economic growth in both the near and long term. Most importantly, the impact on growth and the economy in turn "feed back" and foster the spread of HIV infection.

It is estimated that HIV/AIDS is already reducing sub-Saharan annual GDP growth by between 0.8 and 1.4 percentage points. Countries with prevalence of over 20 percent may suffer a reduction of up to 2.6 points annually. For a typical sub-Saharan economy this means that after two decades GDP could be less than two-thirds of what it would have been without HIV/AIDS (Avnish et.al 2008).

Eventhough the mortality rate due to AIDS have shown some progress after the introduction of the treatment due to ARVs as one of the effective tools to fight against AIDS and prevention measures, the decline of new HIV infections is not satisfactory. Although large body of evidence that addresses issues about AIDS exists, the understanding of the factors that enhance the mortality rate in Ethiopia is very less and there still lacks sufficient research regarding the medications, the HAART and the factors that influence survival. Previous researches made in Ethiopia by Jerene et.al (2007) and Endale et.al (2006) have identified some factors which have significant influence to survival, however there still lack some evidence on the other factors. As a result the objective of this thesis is to assess and differentiate these variables which have significant influence to survival on HIV-positive people using a method of analysis called survival analysis.

1.3 Objectives of the Study

The general objectives of this thesis are:

- To contribute towards the general overview of the AIDS disease and the treatment.
- To demonstrate the application of survival analysis in HIV/AIDS patients' lifetime data and thereby to identify the factors that influence survival.

Specific objectives of this research include:

- To demonstrate the applications of survival regression models.
- To assess the effect of multiple covariates on survival experience.
- To check whether the research data fulfills statistical assumptions of the Survival Analysis.
- To obtain and compare the mean and median survival experience.

1.4 Significance of the Study

- The results/findings of the study may serve as a basis for further assessment on the antiretroviral treatment for people living with HIV/AIDS.
- The study help both the government and non-governmental organizations who work on the prevention and control of AIDS to take continuous considerations of data collection on the treatment of HIV and factors associated with survival, for further monitoring and evaluation assessment.
- The result of this study provides information to government and other concerned bodies to adjust their intervention programs, in setting additional policies, strategies, and further investigations.
- The study helps to create general awareness on the disease HIV/AIDS, antiretroviral treatment and the factors that influence the survival status of HIV positive people.

1.5 Limitations of the study

The study focuses on the main limitations, some of which are assumed to have trivial effect on the outcome of the research work. One of the major limitations of this study is related with the problem of the availability of complete clinical data. Data available in Black lion Hospital, ART clinic, are recorded in different formats and registers however, it is difficult to obtain full information on the clinical variables especially from the data collection formats available in the early years of the introduction of the antiretroviral treatment. This may be due to the presence of less awareness on the treatment, medications and proper data recording system. Nowadays, even though the data collection system is better there are still problems since data on some of the clinical variables are poorly documented. As a result, insufficient information is available on the covariates which have significant influence to the survival of AIDS patients such as the living conditions, risk behavior, addiction, stigma and discrimination including the clinical variables such as hemoglobin (Hg), and total lymphocytes count (TLC). As observed in the different literatures, there are also different factors that are assumed to have impacts on the mortality of AIDS patients. Variables related to nutrition, body exercise, levels of income and social factors associated with environmental sanitation are some factors that could have impact to survival. As a result had the study included some of these variables the resulting analysis would bring further information and statistical analysis. The second limitation is that the study does not cover HIV positive individuals who take ART out side Black lion hospital. The absence of sufficient literature with regard to HIV/AIDS disease and the treatment in our country is a third limitation of the study.

Chapter Two

Literature Review

2.1 General literature about the model used in the study

Survival analysis involves the modeling of time-to-event data. It has wider applications in different research areas. Hongzhe (2006) employed survival analysis to solve the problem of censoring in genetic analysis, Kuldeep and Adrian (2006) for business failure prediction, Richard et.al (2004) in social work research, Paul and Gary (1985) in Psychosocial study, George (1994) to consider trend testing in reliability engineering, Frances et.al (2006) to study translocation[3], Wanzhen (2001) in socio-economic status study, Fukunari and Takamune (2003) in economic research, Chong et.al (2003) in preterm birth study, Judy (1984) for assessing factors influencing the duration of breast feeding and Carles et.al (2004) in a study that develop and empirically test the impact of political institutions[4].

Medical prognosis has played an increasing role in health care. Reliable prognostic models that are based on survival analysis techniques have been recently applied to a variety of domains, with varying degrees of success. Wong et.al (2006) used survival analysis to study treatment of localized prostate cancer, Yanto et.al (2008) in a study of heart transplantation, Vasantha et.al (2008) in tuberculosis patients treated under dots, Peter and Mark (1995) to study cost relapse in Schizophrenia, Harris (1990) to study improved-short term survival of AIDS patients. John et.al (2008) in breast cancer analysis, Peter et.al (2007) for the calculation of results on surveyed data of sows in a large-sale pig farm, Stephen and Franzcp (2004) to compare the risks of admission between study groups of patients in mental hospital.

[3] the intentional release of captive-propagated and/or wild-caught animals into the wild

[4] presidentialism, federalism and proportional representation on the stability of democratic regimes

Pushkar et.al (2007) used hazard model to estimate the effect of birth spacing on child survival. Moshe et.al (1995) applied the hazard rate functions to characterize aging properties of discrete lifetime distributions.

For time-to-event data with finitely many competing risks, the proportional hazards model has been a popular tool for relating the cause-specific outcomes to covariates. Prentice et.al (1978). It has been applied to different research areas. Frank et.al (2001) used semi-parametric proportional hazard rate (PH) model to include a dynamic specification that approximates a standard ARMA (Autoregressive Moving Average) model in the log integrated baseline hazard, Hindman et.al (2005) to jointly test competing theories of legislative behavior, both those that focus on party influence or committee structure, and those that emphasize the primacy of legislator's individual preferences.

Goko et.al (2006) used semi-parametric cox proportional hazard analysis to asses the effects of fund specific characteristics and the dynamic performance properties on survival probabilities of hedge funds, Natasa et.al (2008) for predicting the time to default in credit behavioral scoring, Mustafa et.al (1999) to determine the association between exercise and progression to AIDS and death with AIDS, adjusting for baseline CD4 count and Kathleen (1997) focused on the addition of cash flow variables to a cox proportional hazards model for bank failure prediction.

Dartiques (2006) used Cox proportional hazards model with time-dependent covariates, to investigate simultaneously the short-term effect of air pollution on health and the effect of individual risk factor on a cohort, Platt et.al (2004) for analysis of fetal and infant death.

Stepanova et.al (2002) used three extensions of Cox proportional hazards model to personal loan data, Toru (2001) applied an extension of (PH) model in a study of marriage market

analysis that should capture the simultaneous determination male and female nuptiality by socio-economic characteristics of both sexes.

The literature mentioned above pointed out the wide applications of survival analysis for various research areas. Particularity the semi-parametric Cox Proportional Hazards model and the non-parmetric Kaplan-Meier methods have played a vital role in the analysis of time-to-event data.

2.2 Specific literature related to the variables used in the study

Health care planning depends upon good knowledge of prevalence, which requires an accurate understanding of survival patterns. Monitoring length of survival after diagnosis is therefore an important component of the surveillance of the acquired immunodeficiency syndrome (AIDS). It provides a basis for evaluating individual prognostic factors. In addition, differences in survival may also reflect differences in access to health care (e.g. access to testing, counseling, preventive treatment (Robert et.al 1995).

Knowledge of the survival times of patients with acquired immunodeficiency syndrome (AIDS) and the factors that influence survival is important both for increasing understanding of the pathophysiology of the disease, clinical decision making and planning health service interventions (Jerene 2007).The survival of patients with AIDS may depend on a variety of factors including host factors, the patterns of diseases present, access to health care, diagnostic routines and therapeutic interventions (Robert et.al 1995). A number of researches have been undertaken to study the impact of different factors on the survival of people who live with HIV/AIDS.

2.2.1 HAART, age, sex, CD4 count, WHO clinical stage

In a resource-limited setting South Ethiopia Jerene et.al (2006) made a four years follow-up study on 180 patients under HAART. The investigators applied statistical methods such as Kaplan-Meier and Cox Regression methods to assess the predictors of disease progression. The result showed a mortality rate of 15.4 per 100 person-year observation (PYO) in the HAART. The treatment also showed a 65% decline in mortality implying that it was the strongest predictor of improved survival. A study made in Brazil, Fransisco et.al (1998) used multivariate analysis and showed that patients who received ART survived longer than those who did not after examining 291 patients cared in public AIDS referral services during the years 1989 to 1992.

To determine the survival of AIDS patient's and related factors, survival analysis was conducted in a Brazilian Research center on a sample of 392 HIV+ individuals registered between1986-99. The study considered the survival time to be the time of the date of AIDS diagnosis until death, and censorship defined as loss to follow up or stay alive. Analysis using the Kaplan-Meier method and Cox model were performed. The result showed that the general median survival was 41 months; antiretroviral therapy was significantly associated with a greater survival prediction (Campos et.al 2002).

Butthum et.al (2008) conducted a retrospective cohort study on HIV- infected patients in a medical school hospital during 1996-2005 and followed-up until June 2007 to determine the impact of antiretroviral therapy (ART) on the long-term survival of AIDS patients. The Kaplan-Meier analysis result revealed that the median survival time was significantly longer in patients receiving ART and the Cox proportional hazards analysis proved that ART significantly improved long-term survival of AIDS patients.

Robert et.al (1995) investigated the association of clinical and demographic factors on survival in Austria by taking 755 AIDS patients diagnosed between the years1988 and 1992. Multivariate survival analysis revealed a shorter survival with higher age at AIDS diagnosis and also 188 of the 755 AIDS patients died within the 3 months after diagnosis of AIDS.

In a study that focused on the survival patterns of HIV infection in African populations in the era before antiretroviral therapy (Pre-ART). Isingo et.al (2007) examined a 12-years dataset from an open cohort study. They used the Weibull model to estimate median survival time and parametric regression methods to investigate the influence of sex and age at infection. The Kaplan-Meier analysis showed 67% surviving 9 years post-infection, and the overall predicted median survival was 11.5 years. The result of regression analysis implied that survival was strongly related to age at infection for each additional year of age, and weakly related to sex. Fried et.al (2001) also showed that age of patients was relevant variable to explain the survival time of people with HIV using the bootstrap technique which is a flexible methodology that does not need the assumption of proportional hazards.

Chaisson et.al (1995) measured disease progression and survival in a cohort of 1372 patients seropositive HIV who were treated at a single urban center. They used the Cox proportional-hazards analysis to examine factors associated with progression to AIDS and death .The result showed that lower CD4 cell count and older age were associated with an increased risk of death but there was no relation between disease progression and sex.

Mota et.al (2002) assessed AIDS patients' survival on 12 years study of 486 adult patients. Of these 362 patients received antiretroviral therapy and 124 did not. In their study Mota et.al used Kaplan-Meier survival analysis method to investigate the impact of variables on patient survival, the log rank test to evaluate possible statistical differences between sub-groups and multivariate analysis using Cox proportional hazards model for assessing the performance of

predictive factors. The result showed that there was no statistically significant difference between median survival of male and female patients, among different age groups but CD4 count and antiretroviral drugs had a significant impact on increased survival of AIDS patients. Horsburgh (1991) also showed that age was not a significant predictor of survival in HIV infection after assessing 1251 patients in the follow-up period of during the years 1985-1989.

In a study conducted to evaluate the effect of transmission category, demographic, clinical and immunological characteristics on the progression to AIDS and survival of zidovudine-treated patients Chiesi et.al (1995) studied a total of 1468 patients enrolled between 1987 and 1990 using the Cox proportional hazards regression analysis to identify independent predictors of progression to AIDS and survival. The result confirmed that age and baseline CD4 count are independent predictors of progression, but do not provide evidence for differences in clinical outcome between the sexes.

A study made on 339,863 persons diagnosed with AIDS at age ≥13 years in the 50 states of the USA and the district of Colombia from 1996 to 2003 Zhang (2007) also showed that excess risk for death was higher for persons with a diagnosis of AIDS at an older age and lower CD4 cell count at diagnosis, younger age favored CD4 or T-cell restoration upon HAART which is consistent with younger persons having good thymic function for CD4 cell generation.

To describe immunological HIV progression, mortality and its predictors Maria et.al (2007) also investigated 974 Zambian adults and showed that CD4 count predicted mortality, the median survival in patients with baseline CD4 count ≥500 cells/mm^3 was 5.62 ,with CD4 count between 200 and 499 cells/mm^3 5.46 years and CD4 count <200 cells/mm^3 3.89 years. The mortality rate also increased significantly with older age and was higher in women.

A 5-year research project has compared psychologically and immunologically long survivors with AIDS as well as HIV+ persons who have remained healthy in the face of very low CD4 cell counts with a group of HIV+ persons without symptomatic AIDS and CD4 counts between 150 and 500/ mm^3 (George et.al 2003).

Morgan et.al (1992) used multivariate survival analysis and showed that CD4 count was laboratory predictor of survival of AIDS patients using the data on 49 symptomatic patients with HIV-1 or HIV-2 registered between January 1987 and June 1990 in the Medical Research Hospital of Gambia.

According to the study conducted by Rabeneck (2000) one hundred seventy-six HIV-infected non-AIDS patients were seen at Special Medicine Clinic between January 1986 and December 1990 and followed for a mean of 22 months. Predictive baseline variables (i.e. those associated with progression to AIDS) were first identified and then examined in Cox proportional hazards model .The final model result showed that CD4 category was significant predictor of progression to AIDS in HIV-infected non-AIDS patients. The proportions of patients who progressed to AIDS clinical stages I (CD4>500 cells/mm^3), stage II (500≥CD4≥200), and stage III (CD4≤200cells/mm^3) were 6/39 (15%), 31/106 (29%) and 17/31 (55%), respectively.

Although the survival of HIV patients has improved following the introduction of HAART, when compared with patients with high-income countries, patients in resource-poor countries have higher mortality rates especially during the first weeks of the treatment. Endale et.al (2006) examined factors predicting mortality in Ethiopian 162 patients treated with HAART. The Kaplan-Meier and Cox regression survival analysis were applied to identify prognostic markers where time to death was the main outcome variable. The result showed that the highest death rate occurred in the first month of the treatment. WHO clinical stage was a

predictor of which stage four was identified as important predictor of death. Weight loss was also seen in about a third of patients who survived up to the fourth week and it was associated with increased death.

Based on a study of Tassie.et.al (2003) which assessed WHO clinical staging by a medical interview and examination on a total of 206 HIV-positive adults in a district hospital of Malawi, 43 patients (21%) were asymptomatic, 27 (13%) at clinical stage 2, 81 (39%) at stage 3 and 55(27%) at stage 4. The clinical stages 3 or 4 or Total Lymphocyte Count (TLC) of less than 1200 cells/µl had a sensitivity of 93% and a specificity of 49% .The clinical stage 4 or stages 2-3 with a TLC of less than 1200 cells/µl presented with a sensitivity of 61%. The WHO clinical stages and TLC addeded little in detecting severe immunodeficiency and, have insignificant effect to survival.

The literature discussed above pointed out the use of survival analysis in studies of HIV/AIDS disease, in particular the use of Kaplan-Meier analysis and proportional hazards models for the identification of clinical and socio-demographic variables which are assumed to have influence on the survival of AIDS patients. However, the predictive effect in some of the variables such as WHO clinical stages was not forwarded briefly.

2.2.2 Marital status, religion, functional status, weight, medication

The prevalence of HIV infection varies with age and marital status with infection rates being influenced by the age-and marial status related frequency of intercourse and most obviously the infectivity of an individual's sexual partner or partners. Thus HIV/AIDS prevalence is generally observed to be higher among the divorced or separated than among either the single or the currently married and higher yet among widows and widowers. For example, in urban Kenya and Zambia, among both women and men aged 15-19 and 20-24 HIV was more

prevalent among the ever-married than the never-married (Glynn et.al 2001). Gregson et.al (2001) more specifically showed that, in rural eastern Zimbabwe, HIV sero-prevalence (adjusting for such factors as age and gender) was 2.7 times higher among married than single people, 5.5 times higher among the divorced and separated, and 7.9 times higher among the widowed. The findings of Jim et.al (2006) showed that there was no significant difference in survival by marital status.

In a retrospective cohort study which involved 4909 patients registered in the clinic Marcos et.al (2008) used the Kaplan-Meier test to evaluate duration of viral suppression prior to virologic failure by marital status and showed that patients who were married or had a steady partner experienced virologic failure later than did those who were separated or widowed.

Marzieh et.al (2008) conducted a cross-sectional study using a convenience sampling method on 139 patients living with HIV in a six month period to determine the health-related quality of life in patients living with HIV and showed that gender, marital status, level of education CD4+ count and clinical stage of the disease had a significant effect on the quality of life of the patients. And also in multivariate analysis the most important predictor of the quality of life was clinical stage of the disease. Being female separated or divorced, having less CD4+ count and being at severe stage of the disease were found to be the most important factors which have association with decreased quality of life of the patients.

The potential of faith-based institutions in combating the HIV/AIDS pandemic is undeniable Agadjanian (2005). Pontes et.al (1996) found out that religious support explain the difference in survival of people living with HIV/AIDS. Long survival is characterized by a collaborative relationship with physician and religious/spiritual coping (George et.al 2003). The study result of Jim et.al (2006) also indicated that there was no significant difference in survival by religion.

John et.al (1998) used a sample of 1784 adults with HIV infection in 10 cities across the United States and measured functional status three times during a 12-month period. Respondents in the sample indicated whether they were limited in their ability to perform six activities, ranging from vigorous activities to bathing and dressing. During a 1-year period, 43% of respondents did not change in functional status, whereas 42% became worse and 15% reported improved functioning. Controlling for prior functional status, multivariate analysis showed that declines in functioning were related to developing acquired immunodeficiency syndrome (AIDS), to prior reports of fatigue and to poor self-rated health ensuing functional status was a predictor of mortality.

Palombi et.al (1997) also showed that functional status had an impact on the survival of AIDS patients after conducting a study on 168 patients with AIDS in Rome and the result implied that lower ability to perform self-sufficiency in Activities of Daily Living (ADLs) were related to shorter survival with a mortality risk ratio by Cox regression. Seage et.al (1997) also made a study on 305 persons with AIDS in Boston by reviewing their medical records and ascertaining the vital status of patients. The result demonstrated that measures of activities of daily living, functional status, had an impact to predict the survival of people with AIDS. Casalino et.al (1998) also showed that short and long term survival are strongly associated with the preadmission health status, functional status and weight loss of people with HIV after examining 421 patients in two years study period.

Moderate weight loss before a diagnosis of AIDS is by itself, associated with a reduced survival time after the onset of AIDS (Marck 1995). Tang et.al (2003) assessed the weight of 678 HIV-positive people living in Boston who were enrolled in the Nutrition for Healthy Living study between April 1995 and August 1997. The investigators used Cox model and found that weight loss was significantly associated with an increased risk of death. Dickerson

et.al (1994) followed up 112 patients with AIDS over a 6-year period. The Kaplan-Meier survival estimate showed that patients with weight loss <10% of usual body weight survived 12 months, that is 2.5 times longer than patients with >20% weight loss. The result confirmed that length of survival is reasonably predictable based on percent of weight loss.

Patients with advanced HIV infection have overlapping superimposed risks for developing infections due to protozoa, fungi, viruses, bacteria and mycobacteria. Although it is unlikely that a one-or two-drug regimen will provide complete prophylactic coverage, agents that protect against several classes of pathogens might simplify prophylaxis protocols and improve clinical outcomes. Careful selection of antimicrobial agents that results in synergistic or complementary activity may reduce toxicity and thus increase compliance and perhaps prolong the disease-free period of survival for patients with advance HIV disease (Richard, 1996).

Theo (2005) conducted a study based on the data from 1321 adults treated with TB between June 2001 and June 2002 in South Africa to evaluate whether cotrimoxazole (CTX) benefits TB patients irrespective of HIV status, in a setting where the malaria risk is low and where bacterial resistance to cotrimoxazole is common. It was found that Cotrimoxazole prophylaxis reduced the risk of death in TB patients, that mortality was 29% lower in the patients treated with cotrimoxazole than in the historic controls. Using the data collected on concomitant medications including opportunistic infection prophylaxis and adverse events Rabaud et.al (2000) showed that stopping CTX for any reasons have an association with subsequent shorter survival in patients with AIDS before HAART and under HAART.

In a study that aimed to assess the effects of routinely administered cotrimoxazole on death and illness episodes in HIV infected adults three trials (1416 people) were studied on heterosexual men and women in West Africa. Meta-analysis of the three African trials showed

that CTX prophylaxis had a beneficial effect in preventing death and illness episodes in adults with both early and advanced HIV disease **(**Cochrane Database, 2006).

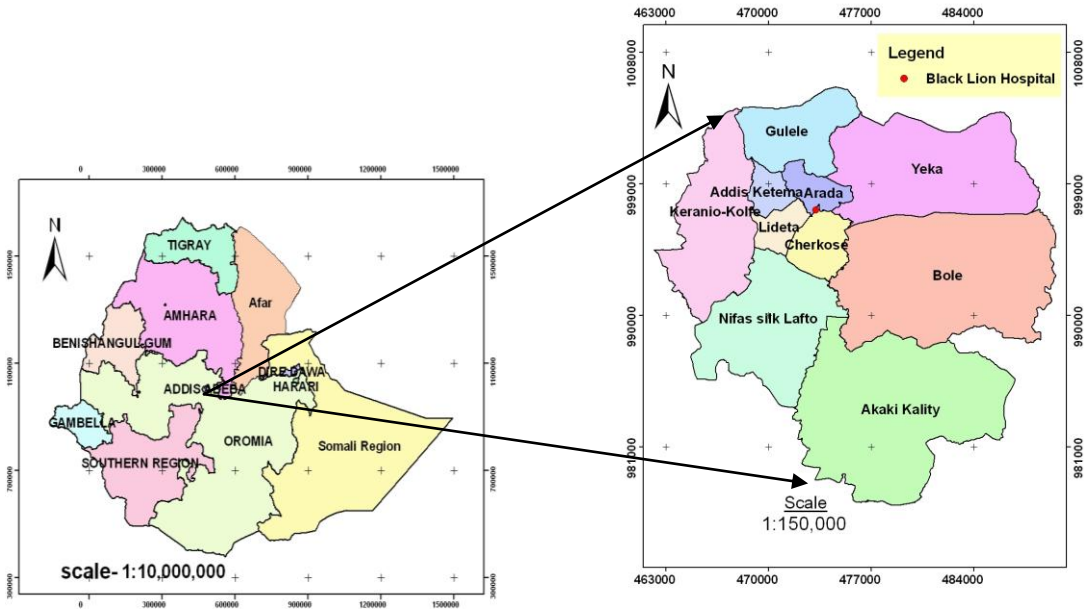
The literature forwarded above shows the effect of marital status, religion, functional status, weight loss and medication on the survival of people with HIV. Nevertheless, the addressing of these factors influence on survival is not sufficient. Besides, there is little information on the predictive effect of number of medications taken by HIV positive people.

Chapter Three

Data and Methodology

3.1 Study Area Background

Ethiopia is located in Eastern Africa. The country has an estimated population of 77 million which has about 80 nationalities speaking 80 languages. Addis Ababa is the capital city of Ethiopia, the head quarter of the African Union and its predecessor the OAU; Addis Ababa is located at the geographic center of the nation (See the Map of Ethiopia). It is also the largest city in Ethiopia with a population of 2.7 million, consisting of 1.3 million men and 1.4 million women (Ethiopian Population and Housing Census (2008)). The city is populated by people from different regions of Ethiopia. Almost all Ethiopian ethnic groups are represented in Addis Ababa due to its position as capital of the country. This ethnic blend gives the city a diverse of culture, making the capital even more attractive. The city possesses a complex mix of highland climate zones, with temperature differences of up to 10 °C, depending on elevation and prevailing wind patterns.



Map of Ethiopia showing the study area, Addis Ababa and catchment area of Black Lion Specialized Hospital

3.2 Data

There are private and public hospitals, clinics and health centers in Addis Ababa. The data for this study were collected from Black Lion Specialized Hospital which is one of the government Hospitals that provide general outpatient and regular services, medical, pediatric, surgical and obstetric emergency and also general in-patient services. It is a place where the Faculty of Medicine of Addis Ababa University is located. The Hospital has different health centers at which patients can get specialized examination and treatment. ART clinic is one of the health facilities in the Hospital to provide general voluntary counseling and testing (VCT) services, follow-up, pre-antiretroviral and antiretroviral treatment for people living with HIV/AIDS (PLWHA).

The ART clinic in Black Lion Specialized Hospital started to properly record patient's information with the support of John Hopkins University[5], in 2005. Data about patients are recorded using the standardized data collection formats and registers prepared by the Ministry of Health of Ethiopia (MoHE). Nurses and health officers who are working in the ART clinic collect primary data starting from the date patients have started HIV- chronic care in the clinic till it is confirmed that patients have experienced one of the events such as death, lost to follow-up, dropped from the clinic and transferred to other health centers. Information on patients who transferred to the clinic is also recorded after encompassing their past history from the referring Hospitals or Heath Centers. The examining medical doctors also record patient's follow-up information. Generally a patient's record is documented in a folder.

According to the MoHE-HIV Prevention & Control Office (HAPCO) report last updated at the end of February 2006 patients ever enrolled for chronic HIV/Care, ever started ART and on ART in Addis Ababa public and private Hospitals were 21,459, 13,739 and 10,888,

[5] It is one of the universities working on HIV in Ethiopia.

respectively. Out of these, patients in Black Lion were 1,594, 1,258 and 1,163, respectively. The data for the intended research were collected from this Hospital which comprises a sample of 1,000 patients enrolled in the clinic between the year 2005 and 2008. It is collected by the nurses in the clinic using a designed format for the purpose of the study, and then entered into a computer and further processed by using software packages like SPSS, SAS, MS Excel and Stata.

3.3 Variables in the study

3.3.1 The response Variable:

Time: Length of time (in days) from date enrolled in chronic-HIV care till the event occurs is used as a dependent variable. The event includes death and censored times (lost and dropped from follow-up, and transferred to other health centers or hospitals).

3.3.2 Predictor Variables:

The study includes predictor covariates from the social, demographic, medical and clinical background of patients. These are listed below:-

1. Age
2. Sex (Male, Female)
3. Marital status (Never Married, Married, Separated, Divorced, widowed)
4. Religion (Muslim, Orthodox, Protestant, Catholic, others)
5. Weight
6. Functional status (Working, Ambulatory, Bedridden)
7. WHO clinical stage (Stage 1, Stage 2, Stage 3, Stage 4)
8. Number of medications taken (0,1,2,3 and 4)
9. CD4 cell count (in mm^3)
10. HAART (yes, no)

3.4 Methodology

The origin of survival analysis goes back to mortality tables centuries ago. The new era of survival analysis emerged not until the end of World War II. This new era was stimulated by interest in reliability or failure time of military equipment. At the end of the war these newly developed statistical methods emerging from strict mortality data research to failure time research quickly spread through private industries. Survival analysis was developed by Biostatisticians in the Medical and Biological sciences. Its application extends to other fields of natural and social sciences.

Survival or Failure Time Data Analysis involves the modeling and analysis of data that have a principal end point, the time until an event occurs (time-to-event data). Such types of data frequently arise from medicine, public health, demography, etc where the analysis is usually referred to as Survival data analysis and industrial studies in engineering fields as Reliability analysis.

Survival time is the duration of time till an event occurs. The inherent aging process that is present when subjects are followed over time is what distinguishes survival time from other dependent variables. Start and end of survival time depends on the context of the problem under investigation. For example, the lifetimes of machine components in industrial reliability, the durations of strikes or periods of unemployment in economics, the time taken by subjects to complete specified tasks in psychological experimentation, the lengths of tracks on a photographic plate in particle physics, the performance of a certain task in a learning experiment in psychology or a change of residence in a demographic study and the survival times of patients in a clinical trial.

Survival data analysis involves a dependent variable, time-to-an event, which is always non-negative and has a positively skewed distribution. It considers conditional information on the remaining time of subject's survival given current survival time. Moreover, the data are censored or have incomplete information since subjects in a study may not experience the event of interest due to different reasons. The existence of covariates that change over time is also a distinguishing feature in survival analysis.

Survival data could be incomplete due to censoring and truncation. Censoring is a form of missing data problem which is common in survival analysis. It has three basic forms: right, left and interval censoring. Truncation is the selection process inherent in the study design. Right censoring is further classified in to Type I, Type II and Random censoring according to whether the time is fixed or random.

The presence of censoring in the data makes the study of survival time more interesting from a statistical research perspective. The sample data collected for the intended research involve one of the most common types of censoring called right censoring for the reason that individuals might survive beyond the study period, lost to follow-up, dropped out from the study, transfer to other places, transfer in from other places and also might die due to the disease other than HIV. It particularly follows a right random censoring mechanism since the patients entry into the study period is random. The failure time is the same as the exact failure time when an event occurs, and censored time is the difference between the time when an event occurs and patients' enrollment in a study. As a result survival analysis is important for the research under study, for data that demonstrate the features of survival analysis in general.

As the merits of Survival Analysis grew, parametric models gave way to nonparametric and semi parametric approaches for their appeal in dealing with the ever-growing field of clinical trials in medical research. Survival analysis is well suited for such work because medical

intervention follow-up studies could start without all experimental units enrolled at start of observation time and could end before all experimental units had experienced an event.

Survival data can be analyzed using: non-parametric (distribution-free) methods, parametric and regression models. The non-parametric method includes descriptive methods; parametric models are concerned with fitting the survival data set with distributions like exponential, Weibull, gamma, log-normal, log-logistic. Regression model analysis consist of semi-parametric (Cox proportional hazard) and parametric regression models.

3.4.1 Descriptive methods

Descriptive analysis for survival data is one type of the non-parametric methods. It includes univariate statistical analysis, and comparison of the survivorship function between two or more groups.

Univariate description of the data is the beginning of statistical analysis in any applied setting. The fundamental building block of the analysis is estimation of the cumulative distribution function. The distribution of T or the cumulative distribution function of T is the probability that an individual will die before time t.

$$F(t) = P[T \leq t], \quad t \geq 0 \quad \text{or} \qquad [3.1]$$

$$F(t) = \int_0^t f(u)du$$

To obtain estimates of the survival function, standard estimators such as the sample mean, variance, median, etc will not yield estimates of the desired parameters because of the presence of censored observation. As a result the life table, Nelson-Aalen, Fleming-Harrington and Kaplan-Meier estimators are applicable for survival analysis.

The Kaplan-Meier estimator of the survivorship function [Kaplan and Meier (1958)] also called product limit estimator, is the estimator used by most software packages. This 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.

Let T be a non-negative random variable representing survival/failure time of subjects from a homogenous population and let *t* is equal to the realization of T. The probability distribution of T can be specified in many ways, three of which are particularly useful in survival applications: the survival function, the probability density function, and hazard function.

The object of primary interest is the survival function also called survivorship function, conventionally denoted S, which is defined as the probability that an individual or a subject will survive at time t or beyond.

$$S(t) = P[T > t] = 1 - P[T \leq t] = 1 - F(t) \qquad [3.2]$$

Related quantities are defined in terms of the survival function. The lifetime distribution function, conventionally denoted F, is defined as the complement of the survival function,

$$F(t) = 1 - S(t) \qquad [3.3]$$

- The relationship between the probability density function and survival function at time *t* is given as $f(t) = \frac{dF(t)}{dt} = \frac{d[1 - S(t)]}{dt} = \frac{-dS(t)}{dt}$ [3.4]

The probability density function mostly denoted by *f* is sometimes called the event density; it is the rate of death or failure events per unit time.

The hazard function is the instantaneous rate of failure. Suppose *h* (t) is the rate of failure at time *t*. The hazard function is the conditional probability density function, which is:-

$$h(t) = \frac{f(t)}{S(t)} \qquad [3.5]$$

Let H $(\cdot)$ be the cumulative hazard function.

$$H(t) = \int_0^t h(u)du \qquad [3.6]$$

- The survival function in terms of the hazard function is given by

$$S(t) = e^{-\int_0^t h(u)du} = e^{-H(t)} \qquad [3.7]$$

- The probability density function in terms of the hazard function can also be derived to be

$$f(t) = h(t)\exp(-\int_0^t h(u)du) \qquad [3.8]$$

The functions described above in [3.1] to [3.8] can be derived for different distribution functions which represent survival time. The function [3.4], alone is important for further estimation such as the mean, median, quartile of survival times and also mean residual life times. The mean survival time is given by

$$\text{E (T)} = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt \text{ and} \qquad [3.9]$$

The median survival time is given by

$$\text{S } (t_{med}) = 0.5,\ t_{med} = S^{-1}(0.5) \qquad [3.10]$$

There are different estimators of the functions. The most common method namely the Kaplan-Meier analysis is described below.

Assume we have a sample of *n* independent observations, their survival times denoted by $t_1, t_2, \cdots, t_n$ and indicator of censoring by $\delta_1, \delta_2, \ldots \delta_n$ where $\delta_i = 1$, event/death and $\delta_i = 0$ for censored observations. The survival data are denoted by (t_i, δ_i), $i = 1,2,\ldots n$.

The first step to obtain the Kaplan-Meier estimator of the survival function is to order the survival times as $t_{(1)} < t_{(2)} < \ldots < t_{(n)}$. Assume that among the *n* observations there are $m \leq n$ failures occurred at distinct *m* times. Then the rank- ordered failure times $t_{(1)} < t_{(2)} <, \ldots, < t_{(m)}$.

Let n_i= the number at risk of dying or failure at $t_{(i)}$

d_i= the number of failures at $t_{(i)}$

Then the Kaplan-Meier estimator of the survival function at time *t* is obtained from the equation.

$$\hat{S}_{KM}(t) = \prod_{t(i) \leq t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{t(i) \leq t} \hat{p}_i \qquad [3.11]$$

with the convention that $\hat{S}_{KM}(t) = 1$ if $t < t_{(1)}$.

In most, if not all, applied settings we will need a confidence interval estimate for the survivorship function as well as point and confidence interval estimates of various quantiles of the survival time distribution. The variance of the Kaplan-Meier estimators which is referred as Greenwood's formula is given as:

$$\hat{Var}(\hat{S}_{KM}(t)) = [\hat{S}_{KM}(t)]^2 \sum_{t(i) \leq t} \frac{d_i}{n_i (n_i - d_i)} \qquad [3.12]$$

The (1-α)$\times$ 100% confidence interval for the survival function S $(\cdot)$ at time t_0 is

$$\hat{S}_{KM}(t_0) \pm Z_{\alpha/2} S.e(\hat{S}_{KM}(t_0)) \qquad [3.13]$$

When the confidence interval of the survival function gives a value less than 0 and greater than 1, the survival function should be adjusted using logarithmic transformation.

The survival function changes only at failure times. That means decreasing step function takes a jump at failure times and also for censored times the conditional probability of survival is always 1.00. As a result of this, the survival probability is unchanged from the previous survival probability.

When we have a large study in which mortality experience is presented in calendar time units (such as quarterly, semi-annually, etc), the life table estimator of the survivorship function may be used as an alternative to the Kaplan-Meier estimator. The life table estimator has been used for more than 100 years to describe human mortality experience and is among the earliest examples of the application of statistical methods. It is dependent on the grouped data while the Kaplan-Meier estimator of the survival function is based on individual observations.

The Nelson-Aalen and Fleming- Harrington are estimators of the survival function which are based on the hazard function. As compared to the Kaplan-Meier estimator, the two estimators perform better for small samples.

After a description of the overall survival experience in a study, comparing the survivorship experience between two or more groups is important in descriptive methods of survival data. When comparing groups of subjects a graphical display of an estimator of the survivorship function such as the Kaplan-Meier estimator for each group is important. In general, the

pattern of one survivorship function lying above another means, the group defined by the upper curve lived longer or had a more favorable survival experience, than the group defined by the lower curve.

To check whether the difference in survival experience seen from the graphical presentation is significant or not standard tests like t-test, one-way analysis of variance (ANOVA) and non-parametric rank sum test are essential for the comparison between two or more groups but, since the survival data includes censored observation the standard tests cannot be applied directly. There are various tests for comparing survival functions. Some of this are log-rank test, generalized Wilcoxon test, Taron-Ware test, Peto-Prentice test, and Harrington-Fleming test and these are dependent on ratios of weighted survival experience. The calculation of each test is based on a contingency table of group by status at each observed survival time. The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times. For comparison of survival functions between two groups the test statistic Q may be defined as follows:

$$Q = \frac{[\sum_{i=1}^{m} w_i \left(d_{1i} - \hat{e}_{1i} \right)]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \sim \chi^2_{(1)} \qquad [3.14]$$

where w_i = weights, d_{1i} is number of deaths, $\hat{e}_{1i}$ is the expected number of deaths estimate and $\hat{v}_{1i}$ is the variance of number of deaths estimate for group one. The formula can further be modified when comparison of the survival function for multiple groups is made.

In log-Rank test the weights equal to one (w_i=1) and generalized Wilcoxon test the number of subjects at risk at time $t_{(i)}$ ($w_i = n_i$). The choice of weight influences the type of differences in

the survivorship function. The generalized Wilcoxon test, since it uses weights equal to the number at risk, will put relatively more weight on differences between the survivorship functions at smaller values of time. The log-rank test, since it uses weights equal to one, will place more emphasis than does the Wilcoxon test on differences between the functions at larger values of time. The Tarone-Ware test uses a weight function intermediate between these, $w_i = \sqrt{n_i}$. Other tests such as Peto-Prentice and Harrington-Fleming tests also use weights for the comparison of survival experience between groups.

If the estimated survivorship functions cross one another, which mean if interaction exists between functions, it is not possible to undertake the comparison with the test mentioned above. As a result regression modeling is suggested as a remedy for addressing interaction.

3.4.2 Regression model for survival data

There are three main types of regression models for survival data which differ in the interpretation of the coefficients and the assumptions made about the relationship between the covariates and survival experience/distribution. These are Accelerated Life Model, Proportional Hazards Model and Proportional Odds Model. In all of the three models the log of survival time, log of the hazard rate and log of odds of survival, respectively, have an additive relationship with the linear combination of covariates.

3.4.2.1. Proportional Hazards / Cox Model

Cox proportional hazards regression model is a multiple regression method to examine the relationship between survival time and one or more predictors, covariate(s), which usually termed in survival analysis. The covariates might be time-dependent or independent.

Suppose the relationship between time T>0, a vector of covariates associated with time $\underline{x} = (x_1, x_2, \ldots x_p)'$ and $\underline{\beta} = (\beta_1, \beta_2 \ldots \beta_p)'$ is given by:

$$\underline{x}'\underline{\beta} = x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p \qquad [3.15]$$

where $\underline{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)'$ is a vector of regression coefficients.

A regression model for the hazard function is the product of two functions is $h(t, \underline{x}, \underline{\beta}) = h_0(t) r(\underline{x}', \underline{\beta})$ where $h_0(t)$ characterizes how the hazard function changes as a function of survival time and $r(\underline{x}', \underline{\beta})$ as a function of subject covariates. $h_0(t)$ is the base line hazard function, equal to the hazard function when $r(\underline{x}', \underline{\beta})$ =1.

Cox (1972) suggested, $r(\underline{x}', \underline{\beta}) = \exp(\underline{x}'\underline{\beta})$, and hence the hazard ratio

$$h(t, \underline{x}, \underline{\beta}) = h_0(t) \exp(\underline{x}'\underline{\beta}) \qquad [3.16]$$

The above model is referred to as Cox model or the Cox Proportional hazards model or simply proportional hazards model. It refers that the hazard functions are multiplicatively related, that is their ratio is constant over survival time, which means that the risk of failure is the same no matter how long the subject has been followed. This is the main assumption of the Cox model and β characterizes the effect of the covariates.

The hazard function in Cox model is called semi-parametric function since it does not explicitly describe the baseline hazard function, $h_0(t)$. The survival function is given by:

$$S(t, x, \beta) = e^{-H(t, x, \beta)} \qquad [3.17]$$

where $H(t, x, \beta)$ is the cumulative hazard function at time *t* for a subject with covariate x.

Since we have assumed that survival time is absolutely continuous, the value of the

cumulative hazard function is expressed as:

$$H(t,x,\beta)=\int_0^t h(u,x,\beta)du=r(x,\beta)\int_0^t h_0(u)du=r(x,\beta)H_0(t) \qquad [3.18]$$

The survivorship function for the general semi-parametric hazard function is:

$$S(t,x,\beta)=e^{-r(x,\beta)H_0(t)}=[e^{-H_0(t)}]^{r(x,\beta)}=[S_0(t)]^{r(x,\beta)} \qquad [3.19]$$

where $S_0(t)=e^{-H_0(t)}$ is the baseline hazard survivorship function. Under the Cox model, the survivorship function is

$$S(t,\underline{xi},\underline{\beta})=[S_0(t)^{\exp(\underline{x}'\underline{\beta}}] \qquad [3.20]$$

Proportional hazards regression model is the most popular and commonly used regression model. The estimation of the baseline hazard function is not required, time-dependent and time independent covariates can be handled, and the statistical analysis can be easily computed using standard computer software such as SPSS, SAS, STATA, R and S-Plus etc.

3.4.2.2. Fitting the Proportional Hazards Regression Model

Suppose the survival data is represented by $(t_i,\delta_{i,}\underline{x}_i)$ for i=1,2,…,n where t_i the length of time a subject is observed (survival time), δ_i an indicator of censoring for the i^{th} individual and $\underline{x}_i$ a vector of covariates for the i^{th} individual.

The likelihood for right censored data includes both the survival and hazard functions is given as

$$L(\beta)=\prod_{i=1}^{n}[h(t_i,\underline{x}_i;\underline{\beta})]^{\delta_i}\ S(t_i,\underline{x}_i;\underline{\beta}) \qquad [3.21]$$

with $h(t_i, \underline{x}_i; \underline{\beta}) = h_0(t_i)e^{\underline{xi}'\underline{\beta}}$ and $S(t_i, \underline{x}_i; \underline{\beta}) = [S_0(t_i)]^{\exp(\underline{x}_i'\underline{\beta})}$

The proposed partial likelihood function suggested by Cox (1972) avoids specification of the baseline hazard function, treating it as a nuisance parameter and removing it from the estimating equation. It assumes that there were no tied values among the observed survival times. Suppose we have *m* distinct failure times and let $\underline{x}_{(i)}$ is the vector of covariates at ordered failure time $t_{(i)}$.We define the Partial Likelihood as

$$L_p(\underline{\beta}) = \prod_{i=1}^{m}\left[\frac{e^{\underline{x}'(i)\underline{\beta}}}{\sum_{j\in Rt(i)} e^{\underline{x}j\underline{\beta}}}\right]^{di} \qquad [3.22]$$

where *d*i is the number of deaths, *d*i=1 and Ri is the set of subjects at risk at time just prior to t_i (t_{i-0}).

3.4.2.3 Estimation of the regression parameters β 's

Estimates of the regression parameters β's will be obtained by maximizing the partial likelihood function and solving the score equations by the Newton-Raphson or any other numerical optimization techniques.

When there are more than one death/failure times the partial likelihood should be modified. There are three approaches that incorporate for tied survival times. These are Exact, Breslow and Efron approximations. Efron approximation gives better approximation of the partial likelihood than Exact and Breslow approximations. The three approximations will reduce to the partial likelihood when there are no ties. The study will employ one of the approximation techniques since it involves tied observations.

The estimator of covariate adjusted survival function is an estimator of the survival function of the proportional hazards regression model which describes the survival experience of subjects adjusted for the covariates and is given by:-

$$\hat{S}(t_i, \underline{x}_i; \hat{\beta}) = [S_0(t_i)]^{\exp(\underline{x}'\hat{\beta})} \quad [3.23]$$

Interpretation of a fitted proportional hazards regression model is based on the hazard ratio; $e^{\hat{\beta}}$ after all statistical assumptions are fulfilled. The (1-α)$\times$100% confidence interval for the estimated parameter is given as

$$\hat{\beta} \pm Z_{\alpha/2}\, s.\hat{e}(\hat{\beta}) \quad [3.24]$$

and for the hazard ratio is obtained as

$$Exp(\hat{\beta} \pm Z_{\alpha/2}\, s.\hat{e}(\hat{\beta})) \quad [3.25]$$

We use partial likelihood ratio test for comparing two models and to test overall goodness of fit of the model. The test statistic is given by

$$G = [\ell_p(\hat{\beta}) - \ell_p(\underline{0})] \quad [3.26]$$

where $\ell_p(\hat{\beta})$ and $\ell_p(\underline{0})$ is log-likelihood of the model with and without covariates, respectively.

G has a Chi-square distribution with *p* degrees of freedom. If the value of G is greater than the Chi-square value with *p* degrees of freedom, then the null hypothesis will be rejected. The formulation of the null and alternative hypothesis depends upon the problem of the study. There are also other tests such as Wald and Score tests which we usually use to test the significance of individual parameters.

3.4.2.4 Model Development

Performing a proportional hazards regression analysis of survival data requires a number of critical decisions. Model development in proportional hazards regression analysis require decisions in selecting subsets of covariates as it is likely that more covariates are present in real life problems, selection of interaction terms to be included in the model and checking the linearity of continuous covariates and choosing the appropriate transformation for non-linear covariates.

The methods available to select a subset of covariates to be included in a proportional hazards regression model are essentially the same as those used in any other regression models. The three most common methods include purposeful selection, step-wise and best sub-set selections. Purposeful selection method is completely controlled by the data analyst, while stepwise and best subset selections of covariates are statistical methods.

Survival analysis using Cox regression method begins with a thorough bivariate analysis of the association between survival time and all important covariates. For categorical covariates this should include Kaplan-Meier estimates of the group-specific survivorship function, point and interval estimates of the median and/or other quantiles, survival time and use of one or more of the significance tests to compare survivorship experience across the groups defined by the variable. For descriptive purposes, continuous covariates could be broken in to quartiles, or other clinically meaningful groups and the methods for categorical covariates could then be applied.

I. Purposeful selection technique.

It involves the following steps:-

1. Include all variables significant in the bivariate analysis at the 20-25 percent level and also any other variables which are judged to be of clinical importance to fit the initial multivariable model.

2. Check the *p*-values <0.05 from the Wald tests of the individual coefficients to identify covariates that might be deleted from the model.

3. Check the *p*-values of the partial likelihood ratio test in order to confirm whether the deleted covariate is significant or not. This is especially important when a nominal covariate with more than one design variables has been selected for deletion, since the overall significance is based on the significance levels of the individual coefficients of the design variables.

4. Assess whether the deletion of the covariates in the reduced model have important change in the coefficient of the variables remaining in the model. A value of about 20% is used as indicator of an important change in a coefficient. If the variable excluded is an important confounder, it should be added back in to the model.

5. Confirm the variables excluded from the initial multivariable model are neither statistically significant nor an important confounder by adding them back in to the model. This will lead to" Preliminary Main Effects Model."

6. Finally determine whether interactions are needed in the model. The selection process begins by forming a set of biologically plausible interaction terms from the main effects in the model. The significance of each separate interaction is assessed by adding it to the main

effects model and using the partial likelihood ratio test. The reason for selecting interaction terms is to improve inference and obtain a more realistic model. As a result all interactions terms significant at the 5-10 percent level will be added jointly to the main effects model.

II. Checking the Linearity of Continuous Covariates

The assumption of linearity can be checked by using the plot of the log-hazard versus the midpoints of the groups. In this method the first step is to group the continuous covariates into few classes. Then based on the classes fit proportional hazards model to obtain the coefficients for each of the classes and plot the midpoints of the classes against the corresponding log hazard ratios (regression coefficients). The grouping of continuous covariates differs from one researcher to another. If the resulting plot is a straight line then the covariate is said to be linear in the model otherwise it is not. In such cases appropriate transformation to linearity is required. Linearity assumption can also be checked using the plot of martingale residuals. Martingale residuals are defined as

$M_i = \delta_i - \hat{H}(ti, \underline{x}i^{(-)}, \hat{\beta}^{(-)})$ where δ_i is an indicator of censoring, $\underline{xi}^{(-)}$ and $\hat{\beta}^{(-)}$ refer the set of covariates and their corresponding coefficients after excluding the covariate for which we are checking the assumption of linearity. And $\hat{H}(ti, \underline{xi}^{(-)}, \hat{\beta}^{(-)})$ is the cumulative hazard after excluding the covariate of interest. If the resulting plot is random showing no systematic pattern and the smoothed plot is straight line. This indicates that the covariate is linear in the model. The use of fractional polynomials is also other method for checking the linearity assumptions.

3.4.3.5. Model Assessment

The preliminary final model which fulfills the model development stages will not be identified as the final model until it has been critically examined for the adherence to key assumptions (e.g. proportional hazards) and for the presence of undue influence or outliers on the fitted model.

I. Checking the assumption of Proportional Hazards

A residual is the difference between the observed value of the outcome variable and that value predicted by the model. The two key assumptions in this definition of a residual are the value of the outcome is known and the fitted model provides an estimate of the mean of the dependent variable or systematic component of the model. The two assumptions are not true when using partial likelihood to fit the proportional hazards model to censored survival data. The absence of an obvious residual has lead to the development of several different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazard model. These include the Schoenfeld, Cox-Snell (martingale) and Score residuals. The Schoenfeld residual is based on individual contribution to the log partial likelihood. For greater diagnostic power the Scaled Schoenfeld residual is preferable. If the plot of scaled Schoenfeld residuals versus the logarithm of time is a random, smooth, straight line around zero the proportional hazards assumption will be satisfied. Furthermore, the assumption of proportional hazards would be fulfilled if the interaction of logarithm of time with the covariate is found to be insignificant (the regression coefficients are not time varying). This means that the study covariates have values that remained fixed over the follow-up period.

The Schoenfeld residuals are based on the individual contributions to the derivative of the log partial likelihood. The log-likelihood from the model in [3.22] is

$$L_p(\underline{\beta}) = \sum_{i=1}^{m} di\left[\underline{x}_{(i)}\underline{\beta} - \ln\left[\sum_{j\in Rt(i)} e^{\underline{x}' j\beta} \right]\right] \qquad [3.27]$$

The first derivative with respect to $\beta_i's$

$$\frac{\partial L_p(\underline{\beta})}{\partial \beta_k} = \sum_{i=1}^{m} di\left[x_{(ik)} - \frac{\sum_{j\in Rt(i)} x_{jk} e^{\underline{x}' j\underline{\beta}}}{\sum_{j\in Rt(i)} e^{\underline{x}' j\underline{\beta}}} \right]$$

$$\frac{\partial L_p(\underline{\beta})}{\partial \beta_k} = \sum_{i=1}^{m} d_i(x_{(ik)} - \bar{x}_{wik}) \qquad \text{k= 1, 2,..., } p \qquad [3.28]$$

where $\bar{x}_{wik} = \sum_{j\in Rt(i)} w_{ij}(\underline{\beta}) x_{jk}$ and $w_{ij}(\underline{\beta}) = \frac{e^{\underline{x}' j\underline{\beta}}}{\sum_{j\in Rt(i)} e^{\underline{x}' j\beta}}$

The estimator of the Schoenfeld residual for the i^{th} subject on the k^{th} covariate is obtained from [3.28] by substituting the partial likelihood estimator of the coefficient $\hat{\beta}$ and is

$$\hat{r}_{ik} = di(x_{ik} - \hat{\bar{x}}_{wik}) \qquad [3.29]$$

where $\hat{\bar{x}}_{wik} = \frac{\sum_{j\in R(ti)} x_{jk} e^{x' j\hat{\beta}}}{\sum_{j\in R(ti)} e^{x' j\hat{\beta}}}$ which is the estimator of the risk set conditional mean of the covariate.

The scaling of Schoenfeld residuals by an estimator of its variance which yields a residual with greater diagnostic power than the unscaled residuals is suggested by Grambsch and Therneau (1994). Suppose the vector of p Schoenfeld residuals for the i^{th} subject is denoted as:

$$\hat{r}_i^{'} = (\hat{r}_{i1}, \hat{r}_{i2}, ...\hat{r}_{ip}) \qquad [3.30]$$

Let the variance of the vector *p* Shoenfeld residuals for the i^{th} subject be $\hat{Var}(\hat{r}_i)$. The vector of scaled Schoenfeld residuals is the product of the inverse of the covariance matrix times the vector of residuals, which is

$$\hat{r}_i^{*} = [\text{var}(\hat{r}_i)]^{-1}\hat{r}_i \qquad [3.31]$$

The components of the *p* by *p* diagonal matrix is

$$\frac{\partial^2 L_p(\beta)}{\partial \beta^2{}_k} = -\sum_{i=1}^{m}\sum_{j\in Rt(i)} w_{ij}(x_{jk} - \bar{x}_{wik})^2 \quad k=1, 2 ..., p \qquad [3.32]$$

The components of the off diagonal elements

$$\frac{\partial^2 L_p(\beta)}{\partial \beta_k^2 \beta_\ell} = -\sum_{i=1}^{m}\sum_{j\in Rt(i)} wij(x_{jk} - \bar{x}_{wik})(x_{j\ell} - \bar{x}_{wi\ell}) \qquad k \neq \ell = 1, 2..., p \qquad [3.33]$$

where $\hat{w}_{ij} = \dfrac{e^{\underline{x}' j\hat{\underline{\beta}}}}{\sum_{\ell\in Rt(i)} e^{\underline{x}'\ell\hat{\beta}}}$

An easily computed approximation for the scaled Schoenfeld residuals is based on the experience that the matrix $\hat{Var}(\hat{r}_i)$ tends to be fairly constant over time and if this matrix is constant its inverse may be approximated by multiplying the estimator of the covariance matrix of the estimated coefficients by the number of events (that is the observed number of uncensored survival times *m*).

$$\left[\hat{\text{var}}(\hat{r}i)\right]^{-1} = m\,\hat{\text{var}}(\hat{\beta}) \qquad [3.34]$$

The approximate scaled Schoenfeld residuals which are computed by software packages, namely

$$\hat{r}_i^* = m\hat{\text{var}}(\hat{\beta})\hat{r}_i \qquad [3.35]$$

II. Subjectwise Diagnostic Checks

Another important aspect of model evaluation is a thorough examination of regression diagnostic statistics to identify which subjects have an unusual configuration of covariates exert an undue influence on the estimates of the parameters and on the fit of the model. An outlier is an extreme observation. Examining scaled score residuals and normal probability plots are helpful in identifying outliers. The effect of outliers on the regression model may be easily checked by dropping these points and refitting the regression equation.

III. Measures of goodness of fit

Suppose we have G groups, the score test in proportional hazards model requires an introduction of G-1 dummy/design variables, and then fitting proportional hazards model including the G-1 dummy variables. The log of partial likelihood for the fitted model with *p* covariates is identified as ℓ_p and the log partial likelihood for model zero, the model with no covariates as ℓ_0 and *n* number of subjects. The measure of goodness of fit R_p^2 based on partial likelihood is given by:-

$$R_p^2 = 1 - \left\{\exp\left(\frac{2}{n}[\ell_0 - \ell_p]\right)\right\} \qquad [3.36]$$

The next chapter (Chapter four) will deal with all of the methods mentioned above the descriptive analysis and regression models of the proportional hazards if the entire model has passed all the model development and assessment steps.

Chapter Four

Data Analysis

4.1 Descriptive analysis

Descriptive analysis is the beginning of any statistical analysis before proceeding to more complicated models. As a result we start with the findings that include the non parametric estimation of the survival function with the Kaplan-Meier method, test of equality across the different levels of covariates using log-rank test, comparison of mean survival experience among each level for the covariates, description of the graphs of survival and hazard functions.

This research involves observations on 1000 HIV positive people that were followed during the time from 2005 to 2008. Of these 90% are censored and 10% are uncensored. Since patients might survive beyond the study period, lost to follow-up and might die due to other causes, the observations follow right censoring mechanism, random type in particular. Summary of censored and uncensored individuals for each level of the covariates is described in Table 4.1.1.

People who live with HIV/AIDS have different survival experience depending on various reasons related with each individual. The result presented in Table 4.1.2 shows that they can live for an average of as low as 4 years and as high as 8 years and also have median survival time that ranges from 3 to 5 years. The general mean and median survival time is 6 and 4 years respectively. In order to investigate each individual's estimate of the survival time we use estimation techniques such as Kaplan-Meier. Since estimated survival function changes only at failure/death times most software put the estimated result for uncensored observations. The graphs of survival and hazard functions is depicted in Figure 4.1.1 (a) and (b) indicating

that a decrease in survival experience and an increase in the hazard rate have direct relation with the increase of time.

The *p*-values of the log-rank test shows difference in survival experience between two or more levels of the covariates. The result obtained from Table 4.1.3 points out that all covariates except marital status, religion and age categories have differences in the levels of their survivorship function. This in turn indicates the significant effect of each covariate on the dependent variable, namely survival time.

From the results in Table 4.1.4 the comparison of survival experience of HIV positive people between sex categories indicates that males live for an average of 4 years longer than females. The graph in Figure 4.1.2 (a) also show the same result in that the upper curve of the survival and hazard functions is for males indicating greater survival experience and less hazard rate, respectively, as compared to females.

The functional status of HIV positive people has significant influence to survival experience; those who are actively working live for an average of 1 year longer than those who are able to work and stay in bed sometimes, on ambulatory status, and 2 years longer than those staying in bed. The average survival time of patients with ambulatory functional status is also 1 year greater than those who are staying in bed. The plots in Figure 4.1.2 (b) show that the uppermost and lowest curves in the survival function are for the HIV positive people having functional status of working and bedridden, respectively. The hazard function curve also shows that the hazard rate for patients in bedridden condition is higher than that of the working and ambulatory functional status.

People living with HIV take a number of medications depending upon the different kinds of opportunistic infection they have. This implies that the number of medications taken has

direct or indirect relations with the existence of the number of infections. The result of the study indicates that the taking no or at least one medication has some effect on survival experience. Patients who take one medication have longer survival time than the others; they live 4 years longer on average than people who do not take any medications and those taking two medications; 5 years longer than those who receive three and four medications. Those patients who take no and two medications have better survival time on average as compared to those who take more than three medications. Taking 4 medications is associated with the lowest survival experience indicating the presence of a number of opportunistic infections regardless of the severity of the disease. In Figure 4.1.2 (c) the upper curve of the survival function indicates greater survival for patients taking one medication and the hazard rate is also less for these patients.

The clinical stages of people with HIV ensure the time of starting treatment with and without taking into account the CD4 cell count. The result verifies that patients under WHO clinical stage 3 have an average of 2 years longer survival experience than those under stages of 1, 2 and 4. Furthermore, those patients of the clinical stages 1, 2 and 4 have the same survival experience having an average of 3 years survival time. Figure 4.1.2 (d) indicates that the upper curve of the survival and hazard function is for patients under WHO clinical stage 3 having greater survival and less hazard rate as compared to stages 1, 2 and 4.

The major goal of the pre-antiretroviral treatment is to follow the patients' health condition, so as to provide the necessary treatment and medical examination follow-up prior to starting the antiretroviral drug. As the results of the study confirm, those patients who are eligible to ART but have not taken the treatment have less survival experience as they live for an average of 5 years less than those receiving the ART. The graphical presentation in Figure 4.1.2 (e) shows that the upper curve of the survival and hazard functions is for patients taking

ART indicating greater survival experience and less hazard rate as compared to patients who do not receive the treatment.

The comparison among the different weight groups shows that patients having higher weight (in kilograms) is not an indication for greater survival since patients having weight between 45 and 55 live for an average of four years longer than those having between 55 and 80. Patients who have weight less than 35 and between 35 and 45 live shorter with an average of 5 and 6 year, respectively, when compared with those people who have weight in the range 45 and 55 kilogram. Besides, HIV positive people having weight between 55 and 80 have better survival experience; they live longer for an average of 1 and 2 years as compared to those having weight in the ranges of 20 and 35; 35 and 45, respectively. Those patients who are in the smallest weight group can live for an average of one year longer than those who have weight that range between 35 and 45.This indicates that the latter weight group is associated with lowest survival experience as compared to all other weight groups. The plot of the different groups of weight in Figure 4.1.2 (f) indicate that the upper curve of the survival function is for patients having weight between 45 and 55; the lowest curve is for the weight group between 35 and 45. The upper curve of the hazard function is for the patients having weight that ranges from 45 and 55 indicating the least hazard rate as compared to others.

CD4 cell count has a significant impact on the survival of people with HIV; those having greater than $200/mm^3$ live 3 years longer on average than those people with less than $200/mm^3$. The plots in Figure 4.1.2 (g) show that the upper curve of the survival and hazard functions indicates CD4 cell count greater than $200/mm^3$ which suggests greater survival and less hazard rate as compared to patients having less than $200/mm^3$.

In order to test the homogeneity of each level of the variables on each treatment (ART) group, we use log-rank test. The values of log-rank test in Table 4.1.5 indicate that there is a strong

significant difference in survival experience among the levels of number of medications taken in people with HIV who take ART or not and functional status in those who take the treatment. In addition there are differences in the levels of weight and age for those who do not take the treatment. The graphs of the Kaplan-Meier estimator of the survival functions which indicate differences in the levels of the variables under the groups of ART are given in Figure 4.1.3 (a) and (b). It shows that as time increases survival decreases irrespective of whether patients take the treatment or not. However, the decrease of survival time with increase of time is lower for HIV positive people who take ART. This suggests that the treatment contributes to the increase of survival time.

There is a difference in the survival status of HIV positive people with regard to the number of medications taken by patients under ART and those who do not take the treatment. The plots shown in Figure 4.1.3 (c) and (d) suggest that those taking one medication have greater survival as compared to the others. Besides the plots in Figure 4.1.3 (e) show highly significant differences in the survival rate with regard to the levels of functional status of patients who take ART; those who are working survive longer. The significant difference in survival experience among weight groups in patients who do not take the ART is shown in Figure 4.1.3 (f), the upper curve which is weight group that ranges between 55 and 80 has greater survival as compared to weight group less than 55. The lower curve corresponds to the weight group less than 35. The difference in the age groups is also shown in Figure 4.1.3 (g). The upper most and lowest curve are for patients under the age group 30 to 45 and greater than 60, respectively, indicating that patients in the age group of 30 to 45 have better survival chances than elderly patients among those patients who do not take the treatment.

4.2 Results of Cox Proportional Hazards Model

In order to study the relationship between survival experience and covariates, a regression modeling approach to survival analysis using proportional hazards regression model can be used with the purposes of estimating the regression coefficients, making interpretation based on the hazard function, conducting statistical tests and constructing confidence intervals.. Model development and assessment of model adequacy must be done before making interpretation from the results of the fitted model. The major goal of model development is to obtain a model which describes the data optimally.

The first step in the model development process is to select covariates which are important to the study. When a study involves multiple covariates, appropriate statistical techniques must be used in order to select covariates which have significant effect on survival and which are judged to be clinically important for inclusion in the initial multivariable model. Purposeful selection is one of the methods useful for selecting covariates. It begins with a model that contains all variables that are significant in the bivariate analysis in relation to time of enrollment in HIV-chronic care to the occurrence of some event/death due to HIV at the 20-25 percent level. The results related to this are presented in Table 4.2.1 for discrete covariates and in Table 4.2.2 for continuous covariates.

All variables except religion categorized into four groups and age are significant at the 20 percent level in Table 4.2.1 and Table 4.2.2, respectively, and they are candidates for inclusion in the initial multivariable model.

The variables weight and CD4 count are significant in both of the coding schemes, and need to be added to the list for inclusion in the multivariable model and used in its continuous form.

The log-rank test and partial likelihood ratio test *p*-values presented in Table 4.2.1 show similar results between the significance levels. This is as expected since among the tests in partial likelihood, the score test is algebraically related to the log-rank test for a discrete covariate and the performance of the score test is quite similar to the partial likelihood ratio test. This implies that log-rank test is an acceptable choice for selection of covariates to be included in the initial multivariable model.

The initial multivariable model resulted from a purposeful selection of covariates includes all variables significant in the bivariate analysis at the 20% level as well as variables which are judged to be important in the analysis. Following the fit of the initial multivariable model, the *p*-values of the Wald statistic identifies the covariates that might be deleted from the model. The *p*-values of the partial likelihood ratio test should also confirm the non significance of the deleted covariate(s). After fitting the reduced model we assess whether or not the removal of the covariate has produced an important change in the coefficients of the variables remaining in the model. A value of 20% change is generally considered as an important change in a coefficient.

Table 4.2.3 presents the results of fitting the initial multivariable proportional hazards model containing all variables significant at the $p < 0.20$ level in the bivariate analysis and the variable age in the categorical form, which is important for the analysis from previous studies. Examining the *p*-values of the Wald statistic with the goal of trying to simplify the model, we note that all variables are significant except marital0, marital2 and marital4; functional status of ambulatory (function1), patients taking 2 medications (nummedic2), and the WHO clinical stage one (WHO1). We next fit a model excluding the design variables of marital status except marital3.

From the *p*-values of the Wald statistic presented in Table 4.2.4 we find that all the variables are significant except divorced patients (marital3), HIV positive people having a functional status of ambulatory (function1), patients who take 2 medications (nummedic2), CD4 count and all the design variables of the WHO clinical stages. The coefficient for age3 is also not significant but, due to its importance in the study, we keep it in the model. The partial likelihood ratio test for comparing the models in Tables 4.2.3 and 4.2.4 is G=3.13 which, with 3 degrees of freedom, is less than chi-square value 7.82 supporting our decision to remove the design variables of marital status except marital3. The maximum change in the coefficient of the variable remaining in the model is 18 percent for sex which is judged not to be an important change to warrant inclusion of the design variables of marital status in the model. We then proceed fitting the proportional hazards model excluding marital3.

Examining the *p*-values of the Wald statistic in Table 4.2.5 we find that all the variables except people with HIV having ambulatory functional status (function1), who take 2 medications (nummedic2), and all the design variables of WHO clinical stages are significant. The coefficient for age is also insignificant but we keep it in the model. The partial likelihood ratio test comparing the models in Tables 4.2.4 and 4.3.5 is G=3.70 which is less than 3.84, the chi-square value with one degrees of freedom, supports our decision to remove the variable marital3. Examining the change in the coefficients of the models with and without marital3, the maximum change in the coefficient is 14 percent for sex. This is judged to be not enough change to warrant including marital3 in the model. We next fit the model excluding function1.

The *p*-values of the Wald statistic in Table 4.2.6 show that all the variables except patients taking two medications (nummedic2) and the design variables of WHO clinical stages are

significant. One of the design variables for age is also not significant but due to its importance in the model it will be kept in the model. The partial likelihood ratio test comparing the Tables 4.2.5 and 4.2.6 is G=1.26 which is less than a chi-square value with one degrees of freedom 3.84. The maximum change in the coefficient for any variable remaining in the model is 8 percent for CD4 cell count, which is not judged to be important enough change to warrant including function1 in the model. As a result we proceed fitting the model excluding nummedic2.

When the *p*-values of the Wald statistic in Table 4.2.7 is examined we find that all the variables except the design variables of WHO clinical stages and age3 are significant. Comparing the models in Tables 4.2.6 and 4.2.7, the partial likelihood ratio test is G=1.96 which is less than when compared to the chi-square value with one degree of freedom, supporting our decision of removing nummedic2. When the change in the coefficient of the reduced model is assessed, the maximum change is 13 percent for nummedic4 which is judged as not an important change to include nummedic2. We then proceed fitting the proportional hazards model excluding the design variables of WHO clinical stages.

When the *p*-value of the Wald statistic in Table 4.2.8 is examined all variables except age3 are significant. The value of the partial likelihood ratio test comparing the models in Tables 4.2.7 and 4.2.8 is G=5.55 which is less than the chi-square value with three degrees of freedom. As a result removing the design variables of WHO clinical stages is important. The maximum change in the coefficients of the reduced model is 8 percent for CD4 count. This change is judged as not significant for including the design variables of WHO clinical stages. Finally, we proceed with reducing one of the design variables of age, age3 (age between 45 and 60), since it has become not significant in the final model of selecting the variables

important for the study. The *p*-values of the Wald statistic in Table 4.2.9 indicate that all of the variables are significant. Comparing the models in Tables 4.2.8 and 4.2.9 the value of the partial likelihood is G=0.10 which is less than the chi-square value with one degree of freedom. The inclusion of age3 is not important since the maximum change is 2 percent for sex which is judged as not a significant change.

The second step in the modeling process is to check the assumption of linearity for the two continuous variables in the model: Weight and CD4 count using the smoothed plots of martingale residual. The plots of the martingale residuals demonstrate the linearity of continuous covariates after excluding the covariate for which we are checking the assumption of linearity. The plots of martingale residuals indicated in Figure 4.2.1 are random showing no systematic pattern, and the smoothed plots approximate a straight line. As a result the continuous covariates CD4 count and weight are linear in the model.

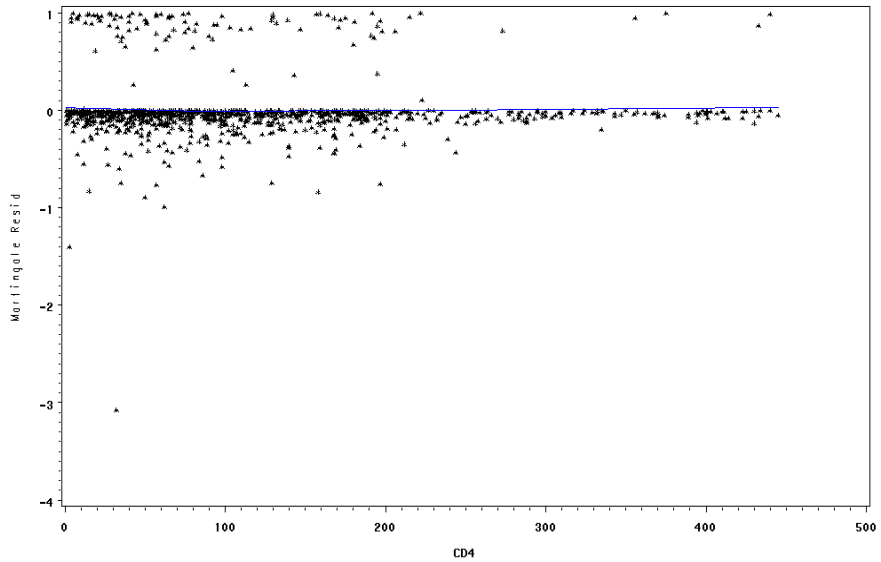


Figure 4.2.1 Martingale Residuals and Lowess Smoothed Residuals for CD4 count.

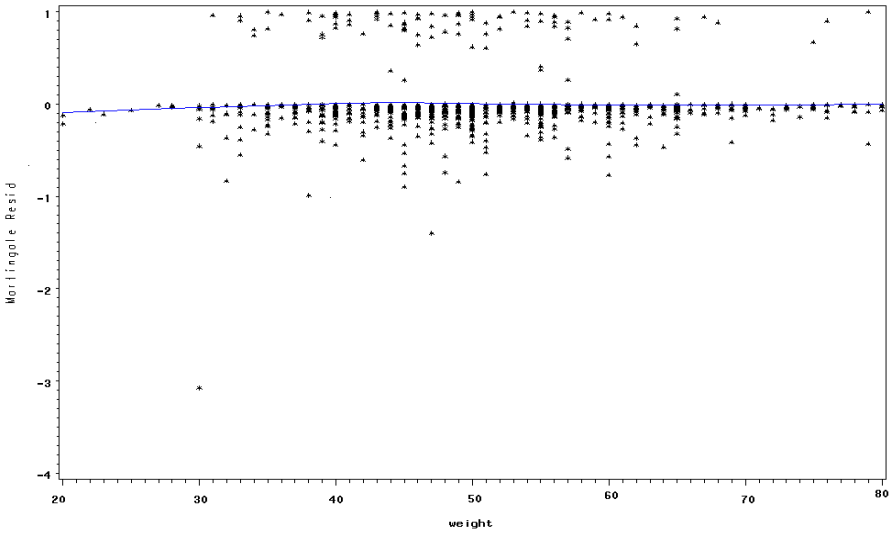


Figure 4.2.1 Martingale Residuals and Lowess Smoothed Residuals for weight.

The third step in the model building process is to add the design variables of - marital status and WHO, nummedic2, function1 and age3 back into the model following the step of exclusion to be sure that they are neither significant in their own right nor confounders of other main effects. The partial likelihood ratio test for the inclusion of the first excluded design variables of marital0, marital2 and marital4 in the model is G=3.00 which with three degrees of freedom has a Chi-square value 7.82. The maximum percent change in a coefficient is less than 20 percent for all main effects in Table 4.2.9. The comparison of percentage change in coefficients for the two models is described in Table 4.2.10. As a result we conclude that the three design variables of marital status are not required in the model.

We next fit the model including the second excluded variable - marital3 to the model presented in Table 4.2.9. The partial likelihood ratio test for the inclusion of marital3 is G=3.18 which is less than chi-square value with one degrees of freedom 3.84. The change in the coefficients is less than 20 percent. Thus, we conclude that marital3 is not important to be included in the model. Comparison of the percentage change in coefficients is presented in Table 4.2.11.

The next step is to fit the model including the third excluded variable, function1, to the model presented in Table 4.2.9. The partial likelihood ratio test for the inclusion of function1 in the model is G=0.95 which is less than 3.84, the chi-square value with one degree of freedom. The change in the coefficients is also less than 20 percent. As a result we conclude that the inclusion of function1 in the model is not necessary. The comparison of change in coefficients is depicted in Table 4.2.12.

The final step is to fit the proportional hazards model by including nummedic2, all the design variables of WHO and age3 into the model consecutively. Nummedic2 is not important in the model since the partial likelihood ratio test for including it is 1.62 which is less than the chi-square value with one degree of freedom. The design variables for WHO should be excluded from the model since the partial likelihood ratio test for including it is G=5.49 which is less than 7.81, the chi-square value with three degrees of freedom. The same test for the inclusion of age3 into the model provides that G=0.10, that is less than the chi-square value with one degree of freedom suggesting excluding it from the model. The changes in the coefficients are also less than 20 percent. Detail description of the changes in the coefficients of the variables in Table 4.2.9 of nummedic1, the design variables of WHO and age3 is presented in Tables 4.2.13, 4.2.14 and 4.2.15, respectively. The results presented in Table 4.2.16 also showed that the design variables of religion are not important for the inclusion in the model since the value of the partial likelihood ratio test G= 3.00 is less than the chi-square value with three degrees of freedom. The maximum change in the coefficient of the variables in Table 4.2.9 is also less than 20% confirming the exclusion of the design variables of religion.

All covariates excluded from the models mentioned above have not brought any changes of the covariates in the model presented in Table 4.2.9. The model with variables presented in

Table 4.2.9 which is attached at the back of this thesis is referred to as “Preliminary Main Effects Model”.

The final step in the model development process is the consideration of interaction terms to be included in the model for the purpose of improving inferences and obtain a more realistic model. This step begins with the creation of a list of plausible interactions formed from the main effects in Table 4.2.9. The significance of each separate interaction is assessed by adding interaction terms to the main effects model and using the partial likelihood ratio test.

Table 4.2.17 presents two interaction terms formed from the main effects in the model and the *p*-value for the partial likelihood ratio test comparing the models with and without interaction. It is located at the back of this thesis together with all the other tables prior to Table 4.2.17. The interaction between antiretroviral treatment and taking of one medication is found to be significant in the model. The interaction between sex and patients in age group2 (age2) is also another significant interaction. These interactions were added in the “Preliminary Main Effects Model” in Table 4.2.9 and the resulting fitted model is shown in Table 4.2.18.

Examining the *p*-values of the Wald statistic in Table 4.2.18 the interactions between nummedic1 and ART; sex and age2 are not significant. The comparison of the partial likelihood ratio test between a model with and without interaction is G= 2.37, this is, less than 5.99 which is the chi-square value with two degrees of freedom. This implies that the larger model which includes the interaction does not fit the data better than the smaller model. Hence, the interaction is not a significant predictor. The maximum change in the coefficients of the reduced model is also less than 20 percent. As a result the model in Table 4.2.9 is remains as “Preliminary Final Model” but interpretation based on this model should not be made until its fit and adherence to model assumptions has been checked.

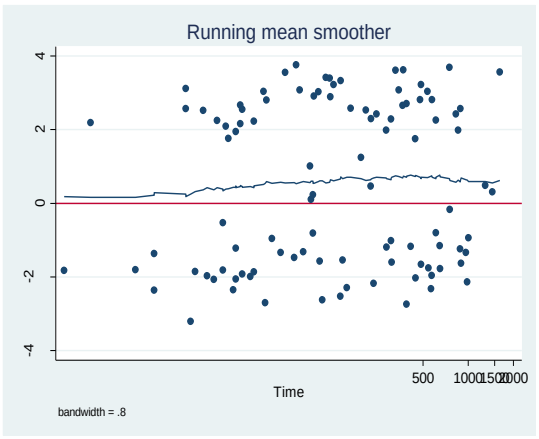
Next assessing the adequacy of the model should be done in order to evaluate how well the fitted regression describes the data set. Requirements for model assessment include testing the assumption of proportional hazards, checking for the presence of leverages and measuring the overall goodness of fit of the model. The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model. The basic assumption of the Proportional Hazards Model is that the hazard ratios are constant overtime. That means the risk of failure is the same no matter how long subjects have been followed.

One method of checking this assumption is to form an interaction between each variable in the model with logarithm of time and assess their significance using the partial likelihood ratio test, score test or Wald test and plot the scaled and smoothed scaled Schoenfeld residuals obtained from the model without the interaction terms. Thus, if the interaction effect is found to be not significant it shows the absence of time varying covariates. This means there are no covariates which show some pattern with the time. The hazard ratios in turn will be constant thereby fulfilling the proportional hazards model assumption. The value of the partial likelihood ratio test comparing the model in Table 4.2.9 to the model containing 10 interactions with log-time is G=13.46 which is less than 18.31 the chi-square value with 10 degrees of freedom. This result suggests that the model may have proportional hazards in each of the ten covariates.

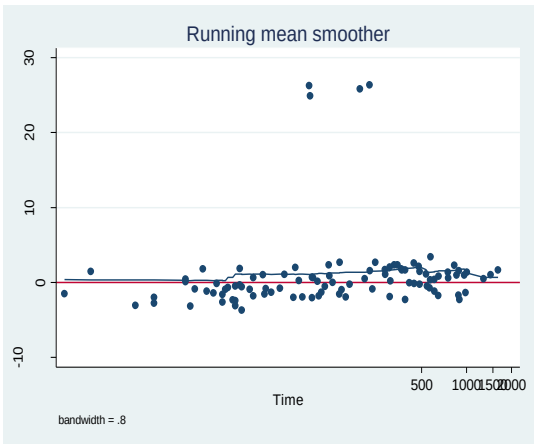
Table 4.2.19 shows the results of fitting a model obtained by including the interaction of each covariate with log-time. This shows that none of the interactions are significant at the 0.05 level indicating that none of the variables have some pattern with log-time. This table is attached at the back of the thesis.

The assumption of proportional hazards should be checked for each covariate in the model. The second step is to examine the plots of scaled Schoenfeld residual and the lowess smooths. Figures 4.2.2(a)-(h) show the plot of the scaled Schoenfeld residuals versus the covariates for each of the ten terms in the model. The plots shown suggest whether the assumption of proportional hazards is fulfilled or not. That is, each subplot in the figure has slope essentially equal to zero.

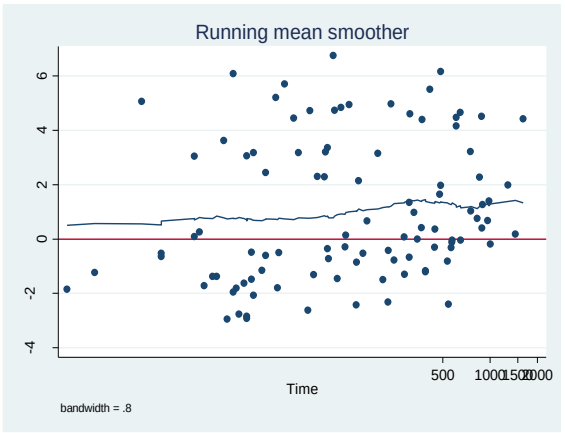
a



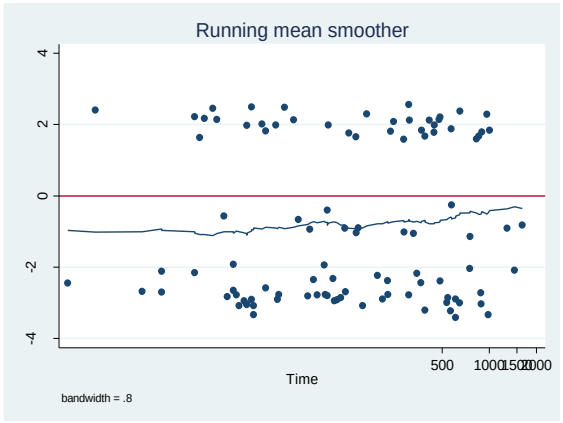
b



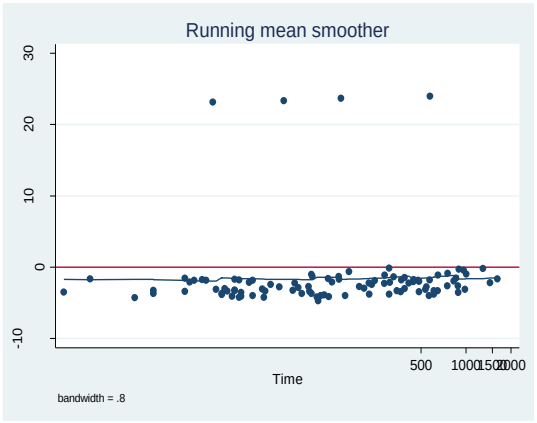
c



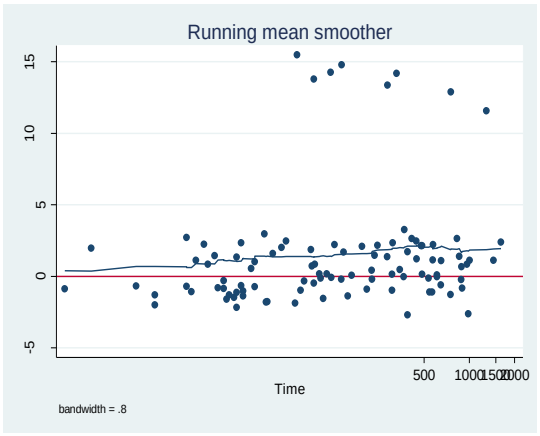
d



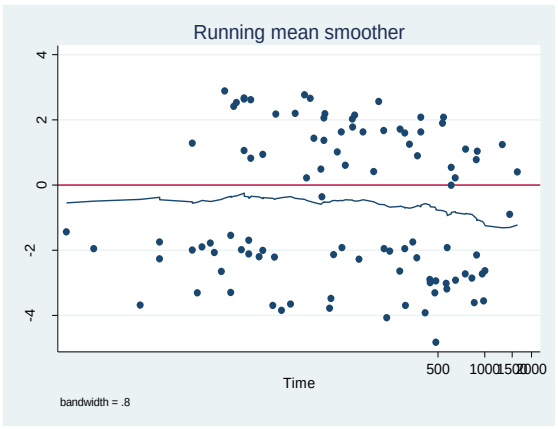
e



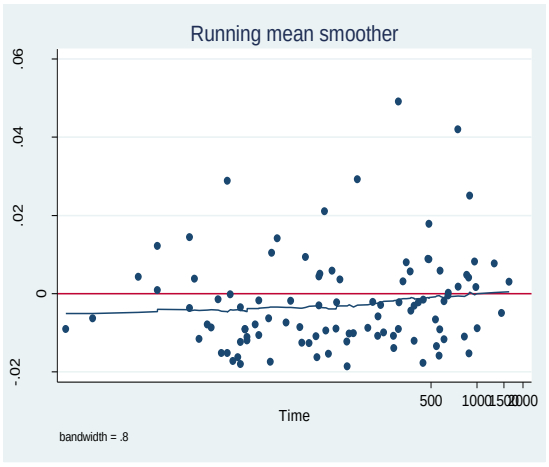
f



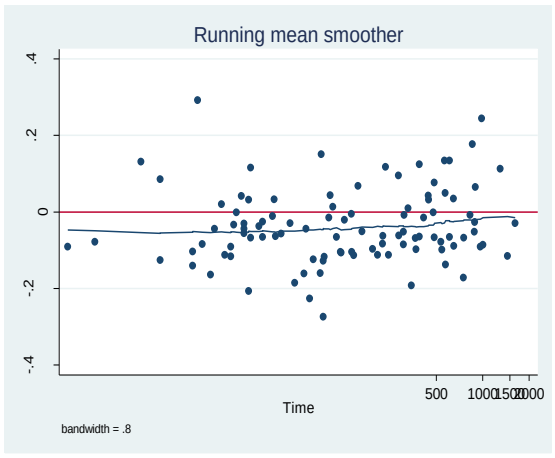
g



h



i



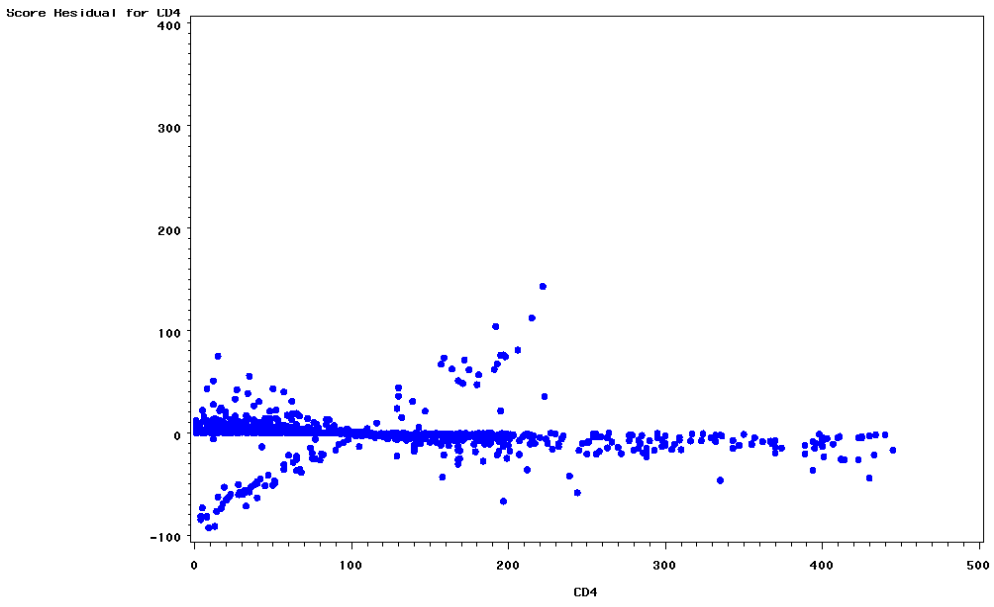
j

Figure 4.2.2 (a)-(j). Graphs of the Scaled Schoenfeld residuals and their lowess smooth obtained from the model in Table 4.2.9 for the covariates sex, age2, age4, function2, nummedic1, nummedic3, nummedic4, ART, CD4 count and weight .The line that passes through zero is a reference line.

The above Figure 4.2.2 shows that each of the ten plots are random, smooth and have zero-slope. The plots have parallel shape pattern showing no trend with time. Consequently there is no covariate which has interaction with log of time supporting the proportional hazards assumption.

The next step is to evaluate the model using a thorough examination of regression diagnostic statistic to identify leverage. This is a diagnostic statistic that measures how "unusual" the values of the covariates are for an individual. In some sense it is a residual in the covariates. This can be shown through the plot of Score residuals (partial leverage residuals) which have the linear regression leverage property for continuous covariates.

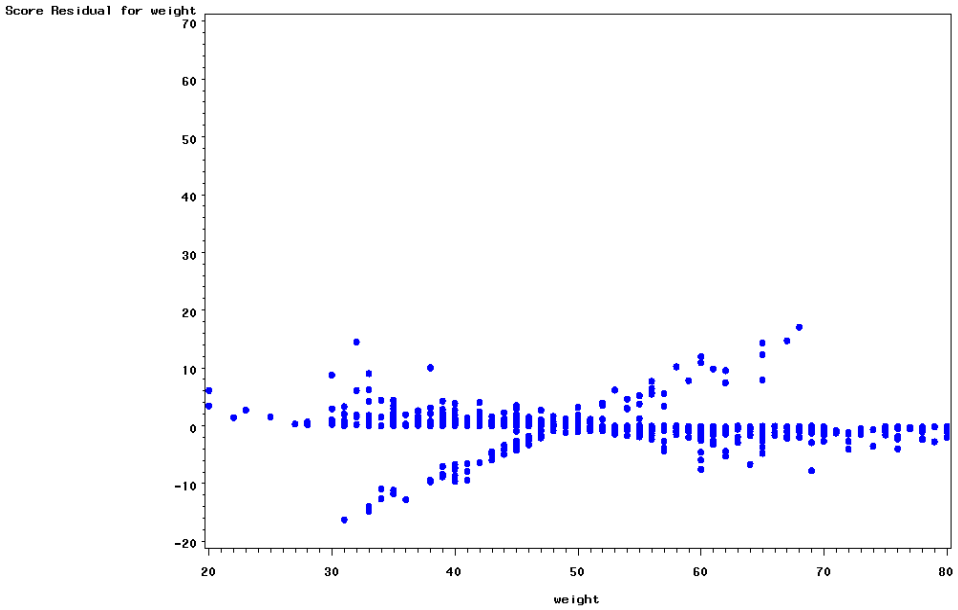
The graphs of the score residuals for the covariates CD4 and weight obtained from the fitted model in Table 4.2.9 are shown in Figure 4. 3.3 (a) and (b).



(a) Score residuals for CD4

Figure 4.2.3 (a). Graphs of the score residuals computed from the model in Table 4.2.9 for CD4 count.

The score residuals for the CD4 count in figure 4.2.3(a) display the fan shape, being smallest near the mean value 131 and increasing in absolute value for patients having CD4 count greater than or less than 131. In the plot we see that there is one point in the top right and two points at the bottom right that fall a bit away from the rest of the points. However, the distance between these points and the others is not striking. Patients having higher CD4 count $\geq 400/mm^3$ have score residuals that are well within the observed range of values. As a result we conclude that there are no high leverage values for CD4



(b) Score residuals for weight

Figure 4.2.3 (b) .Graphs of the score residuals computed from the model in Table 4.2.9 for weight.

The score residuals for weight in figure 4.2.3 (b) also has the fan shape being smallest near the mean weight 51 kilogram and monotonically increasing for weight less or greater than the average weight. The purpose of the plot is to see whether there are individuals whose weight yield unexpectedly large values. The plot indicates that there are two points at the top left and one point at the bottom right which are a little bit away from the other points. However, the

distance between these points and the others is not striking. As a result we conclude that there is no high leverage value for weight.

The graphs for the categorical covariates are less important in that all the values fall on two vertical bands at zero and one, the two covariate values for dichotomous covariates.

The final step in the model assessment is to measure the overall goodness of fit. All measures depend on the proportion of values that are censored. A perfectly adequate model may have low R^2 due to high percent of censored data. The model in Table 4.2.9 presented the value of the -2Log-Likelihood with covariates which is equal to 1083.482 and without covariates equal to 1175.033. The resulting goodness of fit is given as:

$$R_p^{\,2} = 1 - \left\{ Exp\left[\frac{2}{n}(-(-L_0) + (-L_1)) \right] \right\} = 1 - \left\{ Exp\left[\frac{-1175.033 + 1083.482}{1000} \right] \right\} = 0.087$$

As a result the model displayed in Table 4.2.9 has passed all the tests for a good fitting model and is regarded as a final model to make interpretation from the results.

In order to test whether there is significant difference between the coefficients of people with HIV who take one and three medications, let β_1 and β_3 be the coefficients of HIV positive people who take one and three medications, respectively. The variance-covariance matrix for the proportional hazards model is presented in Appendix 1. The hypothesis to be tested is:

H_0: $\beta_1 = \beta_3$ versus H_1: $\beta_1 \neq \beta_3$

The result of the test is $z = \dfrac{\hat{\beta}_3 - \hat{\beta}_1}{s.e(\hat{\beta}_3 - \hat{\beta}_1)} = \dfrac{-1.569 + (0.768)}{0.531} = -1.51$.

The absolute value of z is less than $z_{0.025}$=1.96. As a result the null hypothesis will not be rejected and we conclude that there is no statistically significant difference between the hazard ratios of patients taking one and three medications.

To test whether there is a significant difference in the coefficients of HIV positive people who take one and four medications, the test of hypothesis would be:

H_0: $\beta_1 = \beta_4$ versus H_1: $\beta_1 \neq \beta_4$

The result of the test is $z = \frac{\hat{\beta}_4 - \hat{\beta}_1}{s.e(\hat{\beta}_4 - \hat{\beta}_1)} = \frac{1.401 - (-0.768)}{0.405} = 5.36$.

The null hypothesis will be rejected since $z > z_{0.025}$. This indicates that there is a statistically significant difference between the hazard ratios of patients taking one and four medications.

The test of hypothesis H_0: $\beta_3 = \beta_4$ versus $H_{1=} \beta_3 \neq \beta_4$ also checks whether there is significant difference between the coefficients of the number of taking three and four medications. For this test we get $z = \frac{\hat{\beta}_4 - \hat{\beta}_3}{s.e(\hat{\beta}_4 - \hat{\beta}_3)} = \frac{1.401 - (-1.569)}{0.628} = 4.73.$

The null hypothesis will be rejected since $z > z_{0.025}$ implying that there is a statistically significant difference between the hazard ratio of patients taking three and four medications.

In order to test the existence of significant difference between the age group two (age2) and four (age4), suppose the coefficients for age2 and age4 be δ_2 and δ_4. The hypothesis to be tested will be H_0: $\delta_2 = \delta_4$ versus $H_{1:} \delta_2 \neq \delta_4$.

The computed value of the test statistic would be $z = \frac{\hat{\delta}_4 - \hat{\delta}_2}{s.e(\hat{\delta}_4 - \hat{\delta}_2)} = \frac{1.051 - 0.500}{0.531} = 1.04$

The null hypothesis will not be rejected since the value of the test statistic is less than $z_{0.025}$=1.96. As a result we conclude that there is no statistically significant difference between HIV positive people in the age group between 30 and 45 and between 60 and 75.

4.2.1 Interpretation of the Proportional Hazards Regression Model

The interpretation from the results of the final model is based on the hazard ratios. Comparison is made with the reference category and in between groups for the categorical covariates.

The hazard ratio for sex is 0.55. This means that HIV positive males are dying at a rate 45% lower than females. The 95% confidence interval suggests that the rate could be as much as 64% lower to only 16% lower.

The reference category for the age group is age between 15 and 30. The estimated hazard ratio for age2 is 1.65. This implies that people living with HIV who are in the age group of 30 to 45 are dying at a rate 65% greater than those who are in the younger age group (age between 15 and 30). The confidence interval suggests that the hazard ratios are as low as 1.08 and as high as 2.5.

The hazard ratio for age4 is 2.86. The interpretation of this is that HIV positive people who are under the older age group (ages between 60 and 75) are dying at a rate which is about 2.86 times greater than those who are in the younger age group. The 95% confidence interval implies that the rate could be as low as 1.01 and as high as 8.14.

The reference category for functional status is the working group (function0). The estimated hazard ratio for bedridden group (function2) which is 2.67 means HIV positive people who have bedridden functional status die as the result of AIDS at a rate which is about 2.67 times greater than patients who are working. The 95% confidence interval shows that the hazard ratio is as low as 1.66 and as high as 4.30.

The reference category for the design variables of number of medications is patients who do not take any medications (nummedic0). The estimated hazard ratio for taking one medication is 0.46. The interpretation of this is that people living with HIV who take one medication are

dying at a rate that is 54% lower than those who do not take any medications. The 95% confidence interval indicates that the rate could be as much as 70% lower to only 28 % lower.

The hazard ratio for taking three medications is 0.21 which means that people living with HIV who take three medications die at a rate which is 79% lower than people who do not take any medications. The confidence interval suggests that the rate could be as much as 93% to only 42% lower.

The estimated hazard ratio for taking four medications is 4.06. The interpretation of this is that people who live with HIV taking four medications are dying at a rate which is about 4.06 times greater than patients who take no medications. The 95% confidence interval suggests that an increase in the rate of death due to HIV could actually be as low as 1.87 times and as high as 8.82 times.

To compare HIV positive people who take one and four medications the estimated hazard ratio is $\hat{HR}(4,1) = e^{\hat{\beta}4-\hat{\beta}1} = e^{1.401+0.768}$=8.75. This means patients taking four medications are dying at a rate 8.75 times greater than those who take one medication. The confidence interval (1.38,2.96) also suggests that an increase in death due to AIDS is as high as 1.38 times or even increased rate of 2.96 times is consistent with the data.

Comparing HIV positive people taking three and four medications the estimated hazard ratio is $= \hat{HR}(4,3) = e^{\hat{\beta}4=\hat{\beta}3} = e^{1.401+1.569}$=19.5. This implies that patients taking four medications are dying at a rate which is about 19.5 times greater than those who take three medications. The 95% confidence interval also suggests that the hazard rate is as low as 1.74 and as high as 4.20.

The estimated hazard ratio for ART is 0.50 implying that patients who are under ART die at a rate 50% lower than patients who do not take the treatment. The confidence interval (0.33, 0.76) suggests that the rate could actually be as much as 64% lower to 24% lower.

For a 10/mm^3 increase in CD4 count the hazard ratio is exp (10*-0.002) =0.98. The interpretation is that people living with HIV whose CD4 count increases by 10/mm^3 are dying at a rate which is 2% less than the people having a decrease in CD4 cell count. The 95% confidence interval estimate of the hazard ratio is (0.955, 0.999), which implies that for every 10/mm^3 increase in CD4 count of HIV positive people the hazard rate is as much as 5% lower to only 0.1% percent lower.

The estimated hazard ratio for 5kg increase in weight is 0.83= exp (5*-0.037). The interpretation is that people with HIV having a 5kg increase in weight are dying at a rate which is 17% less than those people having a decrease in weight. The 95% confidence interval estimate of the hazard ratio is (0.75, 0.92) suggesting that for every 5kg increase in weight of people living with HIV the hazard rate could be as much as 25% lower to only 8 percent lower.

4.3 Discussion of the Main Results

Studying the history of disease progression due to HIV and the treatment is useful for clinicians, public health experts and policy makers for the purpose of developing treatment guidelines, modeling the epidemic and prioritizing and allocating resources. A lot of effort has been made to improve the health quality of HIV positive patients and extend the time interval from HIV infection/AIDS diagnosis to death in Ethiopia. For instance, activities like prevention of the disease through effective use of the treatment by implementing the use of many prophylaxis, intervention strategies and methodologies, such as mass media campaigns, peer education about HIV transmission, treatment of other sexually transmitted infections, safe blood provision, and prevention from mother to child treatment have been undertaken. The appropriate analytical tool should be chosen in order to get results useful for decision making and policy formulation. Survival analysis is not only a conventional technology but

also a frequently implemented technique to evaluate the efficiency of new treatment and prevention methods.

The study employed the descriptive methods and proportional hazards regression model. Our result of the Kaplan-Meier method and log-rank test showed that female HIV positive people have on average shorter survival as compared to males. There is nothing that avoids HIV positive people to do any work and live a normal life like those who are HIV negative. As a result patients who have regular jobs or spent their time in a working environment have better survival experience. Patients who are in bedridden functional status have the least survival experience than the working and ambulatory patients.

HIV positive patients may or may not take medications depending on the number and types of opportunistic infections they have. The common primary prophylaxes include Cotrimoxazole, Fluconazole, INH and TB RX. There are also other medications for treating infections and side effects. Our result showed that patients who take one medications live longer on average than those who take no, two, three and four medications. Patients who take four medications have the least survival experience as compared to others. This is associated with the existence of greater number of infections which would have an impact to lower the rate of survival.

The weight of patients have strong influence on survival but higher weight does not imply longer survival since our study result justifies that those patients who have weight that ranges between 45Kg and 55Kgs live longer on average than having more than 55Kg and less than 45Kgs. And those who have weight in between 20Kg and 35Kgs have an average of better survival experience than that of patients weight group 35Kg and 45Kgs.

Improved quality of life in HIV positive people due to the introduction of highly active antiretroviral treatment is also justified by our results. PLWHA who take ART have better survival time on average than that of patients who are eligible to ART and do not take the treatment. CD4 or T cells count have strong influence on the survival status of HIV patients, those having more than 200/mm^3 have greater survival experience than having less than 200/mm^3.The restoration of increase in CD4 count is achieved by the ART which further improves the quality of life.

The comparison of survival estimates by the treatment group suggests that in both groups, those who took the treatment or those who did not, there are significant differences in the survival status of HIV positive people with regard to the number of medications taken and differences exist in the functional status in those who take ART. In addition there are differences in the levels of weight and age for those who do not take the treatment.

The proportional hazards regression model results showed that the variables number of medications taken, functional status of bedridden, CD4 count, ART age, categories, sex and weight have strong influence for the survival of HIV positive people.

Age is one of the demographic variables and it is mostly a significant predictor of survival in most research studies. The results of regression analysis of Isingo et.al (2007) and Chiesi et.al (1995) showed that survival was strongly related to age. The study result of Chaisson et.al (1995) and Zhang (2007) also suggested that older age was associated with an increased risk of death. Similar results are also forwarded by our study, that is, age is a significant predictor of survival in the regression analysis and the hazard rate is higher in older HIV positive people than in the younger age group.

The regression analysis made by Isingo et.al (2007) showed that the significant effect of sex on survival is not significant. The results obtained by Chaisson et.al (1995) and Chiesi et.al (1995) also suggest the insignificant association of sex to disease progression and survival. However, our result shows that there is a significant difference in survival between HIV positive males and females in that females have higher hazard rate than males. Difference in sex and significant relationship with survival can be seen from different angles. A study reported by Marzieh et.al (2008) also shows the significant influence of gender on the survival experience of people living with HIV, is similar to the results obtained in our study.

Findings regarding the effectiveness of ART or HAART are mostly from industrialized countries. The comparisons made for testing the treatment were population-based or on selected group of patients. Some studies have measured the direct effect of ART. Fransisco et.al (1998) reported that patients taking the treatment have longer survival than those who do not take the treatment. Jerene et.al (2006) also showed that HAART created a decline in mortality by 65%. We have also similar results which represent the actual situation that HAART is associated with improved quality of life and the interaction of the treatment with clinical variables is not significant indicating that HAART prevents death irrespective of the patient's weight, CD4 count and WHO clinical stages. The decline in mortality due to the treatment is 50%. With successful taking of the ART, CD4 counts rise. Sometimes they rise quickly. Other times they can go up slowly. If CD4 cell counts fall while taking ART, this indicates the requirement of changing the medications being taken. The Kaplan-Meier result of Campos's et.al (2002) indicated that antiretroviral treatment was strongly associated with greater prediction and the general median survival was 41months. Similar analysis result of Butthum et.al (2008) also revealed that the median survival time was significantly longer in

patients receiving ART. The Kaplan-Meier analysis result of our study suggests that the general median survival time is 48 months.

Body weight is one of the clinical and laboratory markers that are associated with AIDS progression and survival. Weight loss or gain in HIV positive people have significant influence on survival. Our study shows that patients having a weight gain of 5 kg have hazard rate which is 17% less than patients having a weight loss. Similar results made by Tang et.al (2003) suggest that weight loss is associated with reduced survival time. Moreover, the study reported by Dickerson et.al (1994) showed the significant relation of weight loss with an increased risk of death.

CD4 counts are used together with the viral load to estimate how long someone will stay healthy, to indicate when to start antiretroviral therapy (ART), when to start drugs to prevent opportunistic infections and monitoring treatment success. The studies made using 4 years adult patients by Morgan et.al (1992) and Chiesi et.al (1995); 12 years data study report of Kitchen et.al (2007) showed that CD4 count is a laboratory predictor of mortality. Furthermore, Zhang (2007) suggested that excess risk for death was higher on HIV diagnosed people with lower CD4 cell count. Our study is also based on a 4 years data on adult patients. The Kaplan-Meier result showed that patients having a CD4 cell count of greater than 200/ mm^3 have a better survival experience than those having less than 200/mm^3.The proportional hazards regression analysis also showed that it has significant relation with survival and also patients whose CD4 count increases by 10/mm^3 have less hazard rate as compared to those having a reduction in their CD4 cell count.

Researches made for instance by Casalino et.al (1998) had showed that functional status of HIV positive patients is strongly associated with short and long term survival, Palombi et.al (1997) also showed that lower ability to perform self-sufficiency in Activities of Daily Living (ADLs) were related to shorter survival and Seage et.al (1997) demonstrated that measures of activities of daily living, functional status, had an impact to predict the survival of people with AIDS. Our result is also the same but the highest significant effect accounts to the specific category of bedridden patients. As a result HIV positive people involved under this category have highest mortality rate than those having the functional status of working and ambulatory.

People living with HIV take a number of medications depending upon the number of opportunistic infections exist in their body. There are no such relevant studies that considered the number of medications taken except Rabaud et.al (2000) that showed shorter survival due to stopping one of the medications CTX. Our study result indicates that HIV positive people taking one and three medications have less hazard rate as compared to patients taking no medications. Patients taking four medications have greater hazard rate as compared with those taking no, one and three medications. This suggests that the greater the number of medications taken corresponds to the presence of a large number of opportunistic infections which resulted for shorter survival. Besides, patients who do not take medications have greater hazard rate as compared to patients taking one and three medications implying that having an opportunistic infection and not taking the necessary medications is strongly associated with shorter survival time since it is obvious for HIV positive people to have at least one infectious diseases and also it is rare to find patients with no opportunistic infections at all regardless of their taking the treatment.

In this study marital status is found to be a factor that does not have influence on survival. The survival experience of HIV positive people under in any of the categories of the marial status depends on each individual's thinking, behavior, following of the treatment in a proper manner and so on. The effect of being married, unmarried, divorced, separated or widowed has social influence but not related to the survival time. Eventhough, there are not many studies relevant to ours Glynn et.al (2001) reported that the prevalence of HIV is higher among divorced or separated than among singles, the currently married and among the widows.

.

Faith-based institution in religion may support for the survival status of HIV positive people. Agadjanian (2005) and Pontes et.al (1996) found out that religious support explains the difference in survival of people living with HIV/AIDS. George et. al (2003) reported that the advantage of collaborative relationship with religious/spiritual coping for longer survival. Our study suggests that religion is not a significant predictor of survival time. The log-rank test also justifies that there is no significant difference among the types of religion. The faith of HIV positive individuals depends on the type of religious institutions they follow and the intensity of faith on their religion which would not have a significant impact to survival experience. The report of researchers mentioned above explained the general effect of religion but not specific to each types of religious institutions. Jim et.al (2007) reported a similar result as in our study that there were no significant differences between the religious institutions and the social variable religion do not have impact on survival.

The WHO staging sytem, which has clinical and laboratory components, has rarely been used in routine clinical practice in Ethiopia. It has been used in selected research projects in the country including a validation study which described the clinical usefulness of the WHO

staging system. The log-rank test result showed that there are significant differences between the levels of the clinical stages. The mean estimate of survival time from the Kaplan-Meier analysis shows that patients under clinical stage 3 have longer survival time on average as compared to stages 1, 2 and 4. The treatment is mostly recommended for patients who are under the clinical stages 3 and 4. As a result the better survival of patients under the clinical stages 3 is associated with the treatment taken. However, the regression analysis justified that there is no statistically significant influence on survival experience due to the WHO clinical stages. The study made by other researchers involves the study of the stages in relation with other clinical variables but not in the comparison of the different clinical stages.

4.6 Policy Implications

Antiretroviral treatment has been changing the health quality of the life of HIV positive people by prolonging the survival time of patients since the cure of the disease is not possible now. Thus, strong efforts should be made for sustainable distribution of the treatment. Many patients need ART in Ethiopia. This study evaluates the treatment program that started ART at Black Lion Specialized Hospital in 2005. The treatment reduced deaths among AIDS patients by 50%. About 43% of the patients followed the treatment. This thesis provides information that is important in identifying the significant factors influencing the survival status of HIV positive people including carrying out of ART in Ethiopia.

The main challenge is the sustainability of providing of ART and number of medications at district hospitals in Ethiopia. The Ethiopian Health Policy focuses on thorough and integrated primary health care with emphasis on community based services. The Health Service Extension Program which was introduced in 2004, in Ethiopia was intended to bring the health service close to people. According to the package, both health centers and district hospitals should provide HIV care, medications and treatment including ART. It is therefore encouraging to note the Integrated Management of Adolescent and Adult Illness (IMAI)

developed by the WHO that provides practical guidance on decentralized and integrated delivery of HIV care.

Policy documents are necessary to convince the world to take the issue of HIV/AIDS more seriously. The policies made by the policy makers persuade governments and funders to prioritize the issue, take action, and allocate funds to HIV/AIDS interventions.

This research is conducted in the capital city of Ethiopia. Such research should be viewed as an opportunity for strengthening research capacity at different Hospitals, Health Centers in and outside Addis Ababa. The sustainability of such efforts needs both enabling policy environment and resources.

Chapter Five

Conclusions and Recommendations

5.1 Conclusions

- The study involves significant predictors of survival experience which are age categories, sex, number of medications taken, ART, CD4 count, weight and functional status. The variables religion, marital status and WHO clinical stages were found to have no significant influence to the survival status of HIV positive people.

- Age is a significant predictor of survival; older age is associated with an excess risk of death since the hazard rate is higher than the younger age group.

- Female HIV positive people have on average less survival time and greater hazard rate as compared to males.

- HIV positive people having working functional status have greater opportunity of living longer time as compared to ambulatory and bedridden. On the contrary patients having the functional status of bedridden have excess hazard rate.

- The number of medications has significant influence for longer survival. However, taking more medication is strongly associated with shorter survival. Taking no medication is not an indication of being healthy. The risk of dying is higher as compared to patients taking no less than 3 medications.

- Having greater weight is not an indication of longer survival and having less weight may not be associated with shorter survival. The hazard rate of patients in the highest weight group is very high.

- HAART or ART plays a greater role in improving the life quality of HIV+ people. Those taking it live longer and the risk of death is reduced as compared to those who do not take the treatment. This reinforces the importance of universal access to antiretroviral therapy to extend survival.

- Higher CD4 cell count (>200/mm^3) is associated with an increase rate of survival time. On the contrary less CD4 cell is associated with an increased rate of death.

- HIV positive people can live for an average of as low as 4 years and as high as 8 years and also have median survival time that ranges from 3 to 5 years.

- The proportional hazards regression model was an appropriate approach for estimating the survival of people living with HIV who started the pre-treatment and proceeded with the ART.

- There are statistically significant differences in survival experience with regard to the number of medications taken and functional status for HIV positive people taking the treatment. Besides, in the group of patients who do not take the treatment significant differences occur with regard to number of medications, age and weight.

- The plot of martingale residual confirmed the linearity of continuous covariates in the model. Proportional hazards regression model assumption was also checked using the plots of Scale-Schoenfeld residual and the interaction of each covariate with log-time. The result justified that the assumption was fulfilled. Moreover, the plots of Score residuals indicated the absence of leverages for the continuous covariates.

- The significance of interaction effects was also assessed but none of which are significant in the final model.

5.2 Recommendations

For clinical practice:-

- Strict follow-up should be made on patients who are eligible to ART but not taking the treatment.
- Documenting the number of medications taken by patients, CD4count, weight and the functional status should be viewed as important components of the routine clinical care for patients on antiretroviral therapy.
- For the successful improvement of survival with ART, pre-treatment care should be given appropriately.
- Hospice / home based care/ are necessary for HIV positive people having bedridden as well as ambulatory functional status.
- Since HIV positive people might die due to opportunistic infections that exist in their body in spite of the HIV virus. Careful attention should be provided for patients who do not take any medications.

For research:-

- Factors that cause higher hazard rate of females HIV positive people than males should be investigated.
- Factors that affect the shorter survival of patients having high weights should be examined.
- Further research regarding updated policy issues and guidelines of ART is needed.

For policy:-

- The revision of antiretroviral treatment guideline should incorporate simple laboratory and clinical markers of survival experience.
- New research results should be considered in formulating new policies and upgrading the existing ones.
- The Ethiopian HIV/AIDS antiretroviral treatment policy should be adapted to the changing environment of HIV treatment and care.

REFERENCES

Agadjanian V. (2005). Gender, religious involvement and HIV/AIDS prevention in Mozambique. Social Science and Medicine, Volume 61, No.7, PP1529-1539.

Alula Pankhurst (2005). Conceptions of and responses to HIV/AIDS: Views from twenty rural Villages. A thesis work from the Department of Sociology and Anthropology. Addis Ababa University, PP 45-48.

Avnish K.C., Tang B.A, Soumya T.C (2008). Economic Impact of HIV/AIDS in Sub-Saharan Africa. IMF Working Papers. http://www.imf.org/external/pubs/ft/wp/2002/wp0238.pdf

Butthum B., Sungkanuparph S., Chakriyanuyok T. (2008). Antiretroviral therapy in AIDS patients: Impact on the survival and long-term treatment outcome. Journal of Infectious diseases. Volume 56, No.1, PP 40-43.

Campos D.P, Gadelha A.M, Veloso V.G, Grinsztejn B.G, Ribeiro S.R, Bandeira T.S, Valente J.G (2002). Survival analysis of AIDS patients cared for in research center Evandro Chagas Hospital Rio de Janeiro/Brazil. 1986 – 2000. International Conference on AIDS. 2002 Jul 7-12; 14: abstract no. C10727.

Carles E.M, Boix S.T and Alicia T.S, Adsera R.B (2004). Democratic stability and political institutions survival analysis. Paper presented at the annual meeting of the Midwest Political Science Association, Palmer House Hilton, Chicago, Illinois, April 15, 2004. PP 23-28.

Casalino E., Medoza G., Wolff M., Bedos J., Caudebout C., Regnier B. and Vachon F. (1998). Predictors of short-and long-term survival in HIV- infected patients. Revue. vol. 113, No 2, pp. 421-429.

Centers for Disease Control and Prevention (2008). Basic information topics CDC HIV-AIDS. http://www.Basic information on AIDS.

Chaisson R.E, Keruly J.C, Moore .RD (1995). Race, sex, drug use, and progression of human immunodeficiency virus disease. Medline. Volume 33 No.12 PP 751-756.

Chiesi A., Vella S., Giuliano M., Floridia M., Tomino C., Seeber A., Barcherini S. Bucciardini R., Mariotti S. (1995). Effect of sex, age, and transmission category on the progression to AIDS and survival of zidovudine-treated symptomatic patients of AIDS. Revue. Vol. 9, No1, PP. 51-56.

Chong Yau Fu and Shih Hua Liu (2003). Using Survival Analysis in Preterm Birth Study. Source: Recent Advances in Statistical Methods. Vol 5, No. 3 PP 107-113.

Cochrane Database of Systematic Reviews (2006). Cotrimoxazole prophylaxis for opportunistic infections in adults with HIV. (www. Cochrane Database.com)

Cox, D.R and D. Oakes. (1984). Analysis of survival Data London: Chapman and Hall.

Cox, D.R. (1972). "Regression models and life Tables (with Discussion)." Journal of the Royal statistical society.

Dartigues, Jean-Francois (2006). Survival analysis to estimate association between short-term mortality and air pollution. Environmental Health Prespectives.Vol 4, No5, PP 23-25.

Dickerson C., Salvato P. and Thompson C. (1994). Malnutrition and survival in AIDS. International Conference on AIDS. 1994 Aug 7-12; 10: 195 (abstract no. PB0795).

Endale A., Jerene D., Hailu Y, Lindtjorn B. (2006). Predictors of early death in a cohort of Ethiopian patients treated with HAART. BMC infectious diseases. Vol 6, No.1 PP 136-137.

Ethiopian Population and Housing Census (2008). Summary and Statistical Report of the 2007 Population and Housing Census. PP 1-5.

Frances J.S, William F.F, Peter J.U (2006). Using survival analysis to study translocation success. Journal of Ecological Applications, Vol. 16, No. 5, PP 1771-1784.

Frank G. and Nikolaus H. (2001). Semi parametric Autoregressive Conditional Proportional Hazard Models. Working paper available on http://ssrn.com/abstract No.= 289645

Fransciso A., Cibele C., Mark D. (1998). Health Care Utilization and Survival among patients with AIDS in Brazil. Cadernos de Saúde Pública vol.14 no.4 PP 112-115. Rio de Janeiro Oct./Dec. 1998.

Fried E., Ostraford J. and Nomino V. (2001). A New Model for AIDS Survival Analysis. AIDS working paper. Vol 5, No 8, PP 10-14.

Fukunari K., Takamune F. (2003). Globalizing Activities and the Rate of Survival: Panel Data Analysis on Japanese Firms. Journal of the Japanese and International Economies, 2003, Vo17 (4, Dec), 538-560.

George A. (1994). Trend testing in reliability engineering: International Journal of Quality & Reliability Management, Vol. 13, No. 2, PP 45-54.

Glynn J.R, Carael M.T, Auvert B.W (2001). Why do young women have a much higher prevalence of HIV than young men? A study paper in Kisumu, Kenya and Ndola, Zambia.

Goko H. and Baba. N (2006). Survival Analysis of Hedge Funds. International Finance Research and studies Statistics. Vol 4, No.2, PP 12-15.

Grambsch P.M and Therneau T.M (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika, Volume 81, No. 3 PP. 515-526.

Gregson S., Mason P., Garnett G., Zhuwau T., Nyamukapa C., Anderson R. and Chandiwana S. (2001). A rural HIV epidemic in Zimbabwe. Findings from a population based survey. International Journal of STD and AIDS, Vol 12, No. 2, PP 189-196.

Harris, J.E. (1990). Improved short-term survival of AIDS patients initially diagnosed with Pneumocystis carinii pneumonia, 1984 through 1987. The Journal of the American Medical Association. Vol 5, ISSN: 0098-7484.PP 67-70.

Hindman, M. and Adamic, L. (2005). Lawmaking and Law Unmaking: A Survival Model of Federal Law in the United States. Paper presented at the annual meeting of the Midwest Political Science Association 67th Annual National Conference, The Palmer House Hilton, Chicago.

Hongzhe Li (2006). Survival Analysis Methods in Genetic Epidemiology. Biostatistics Working Paper 7.http://biostats.bepress.com/upennbiostat/papers/art7.

Horsburgh C.R, Hanson D., Fann S.A, Havlik J.A, Thompson S.E (1991). Predictors of survival in HIV infection include CD4+ cell count. European Journal of Epidemiology Vol. 13 No.2 PP 125-132.

Hosmer D.W, Lemeshow S., (1989). Applied Survival Analysis. John Wiley and Sons, Inc., New York.

Isingo R., Zaba B., Marston M., Ndege M., Mngara J., Mwita W., Wringe A., Beckles D., Changalucha J., Urassa M. (2007). Survival after HIV infection in the pre-antiretroviral therapy era in a rural Tanzanian cohort. AIDS, Vol 6, No. 5 PP 13-17.

Jerene D. (2007). HIV Antiretroviral Therapy. Center for International Health University of Bergen. Thesis submitted in partial fulfillment of the requirements for the degree of philosophiae Doctor (PhD) at the University of Bergen, Norway.

Jerene D., Noess A. and Lindtjorn B. (2006). Antiretroviral therapy at a district hospital in Ethiopia prevents death and tuberculosis in a cohort of HIV patients. AIDS research April7, Volume3, No1, Page 10.

Jim T., Ivan K., Dermot M., Heiner G. (2006). Time from HIV Sero-conversion to death: The factors that affect survival. AIDS 2006, Vol 21, No 3, PP 55-63.

John A. and Stephen C. (1998). Functional status, Transitions and Survival in HIV Disease. Journal of evidence from the AIDS cost. Medicine, Vol 36, No. 4, PP 533-543.

John K., Martin B., Denis R. Patrick C., Mina J., Bissell, Marcia V. (2008). Prognostic breast cancer signature identified from 3D Culture Model accurately predicts clinical outcome across independent datasets. American Society of Cancer treatment. Vol 3 No.4, PP. 111-114.

Judy R. (1984). An Introduction to Survival Analysis: Factors Influencing the Duration of Breast Feeding. Journal of Sociology, Vol. 20, No. 2, 183-196.

Kalbfleish J.D and R. L.Prentice (1980). The Statistical Analysis of Failure time data. John Wiely and Sons, Inc., New York.

Kathleen L. (1997). A test of the temporal stability of proportional hazards models for predicting bank failure: Journal of Financial and Strategic Decisions, Vol. 10, No.3 PP 245-249.

Kitchen M., Quigley M., Mwinga A., Fuchs D. Lisse I., Porter J., McAdam K., Godfrey P. (2007). HIV Progression and Predictors of Mortality in a Community-Based Cohort of Zambian Adults. Journal of the International Association of Physicians in AIDS Care Volume 7, No.1 PP 17-20.

Kloos H., Lindtjorn B. and Damen Haile Mariam (2007). The AIDS Epidemic in a Low-Income Country: Ethiopia. Human Ecology Review Vol 14, No.1, PP 39-42.

Kuldeep K. and Adrian G. (2006). Business Failure Prediction using Survival Analysis. Conference paper presented at the International Conference of Business, Economics and Management Disciplines, 23-24 June 2006, Beijing.

Marc K. (1995). Weight loss before and onset predicts reduced survival. National Institute of Allergy and Infectious Diseases. Vol4, No.2, PP 211-215.

Marcos M., Alcino A., Arnaldo E. and Ricardo S. (2008). Characterization of Virologic Failure After an Initially Successful 48-Week Course of Antiretroviral Therapy in HIV/AIDS Outpatients Treated in Santos, Brazil. The Brazilian Journal of Infectious Diseases, Vol. 12 No.3, PP 162-166.

Martin D., Spencer D., Lynne W., McDonald L. (2004). Pre-ART guideline. South African Journal of HIV medicine, Vol 34, No.8, PP 20-23.

Marzieh N., Khatereh A. and Mitra R. (2008). Health-Related Quality of Life in Patients with HIV/AIDS. Journal of Iranian Medicine, Vol. 11, No. 6, PP 608 – 612.

MedicineNet.com.Webster's New World Medical Dictionary, 3rd Edition (May, 2008). Wiley Publishing.Inc

Morgan G., Whittle H., Egboga A., Todd J., Corrah T., Wilkins A., and Demba E., Rolfe M., Berry N. and Tedder R. (1992). Clinical and laboratory predictors of survival in Gambian patients with symptomatic HIV-1 or HIV-2 infection. AIDS, Volume 6 No.7 PP 685-689.

Moshe S., Shanthikumar J.G, Valdez-Torres J.B (1995): Discrete hazard rate functions. Journal of computers & operations research, Vol. 22, No.38, PP 391-402.

Mota R.S, Guerreiro M.F, Ligia R.S, Franca M.C, Caminha F.C (2002). Survival of adult AIDS patients in a reference hospital of a metropolitan area in Brazil. AIDS. Vol.36 No.3 PP 110-115.

Mustafa F., Saide A., Tanor B. (1999). Association of exercise and progression of AIDS. Annals of Epidemiology, Vol.9 No.2, PP 127-131.

Myron S., Nick H., Jay A.L, Kevin B. and Joep L. (2008). The spread, treatment, and prevention of HIV-1: evolution of a global pandemic. Journal of Clinical Investigation, Vol. 118 No. 4, PP 123-134.

Natasa S., Mirta B., Marijana Z. (2008). Comparison procedure of predicting the time to default in behavioral scoring. Working paper on behavioral scoring in Brazil PP (12-13)

National Institute of Allergy and Infectious Diseases (NIAID 1993). Infectious diseases. Vol5, No.1, PP 24-25.

Palomb L., Mancinelli S., Liotta G., Narciso P., Marazzi M. (1997). The impact of socio-economic factors, mental health and functional status on survival in a sample of AIDS patients. Journal of AIDS Care, Vol. 9, No. 6, PP 671 - 680.

Paul J. and Gary R. (1985). Psychiatric recidivism in a psychosocial rehabilitation setting survival analysis. American Psychiatric Association. Hosp Community Psychiatry Vol 36, No1, PP 521-524.

Peter B., Imre E., Sandor K. (2007). Survival analysis of culling reasons and economic examination of production period in sow culling. Annals of Statistics. Volume 37, Number 1, PP 394-426.

Peter J. and Mark O. (1995). Cost of Relapse in Schizophrenia National Institute of Mental Health. Schizophrenia Bulletin, Vol. 21, No. 3, PP 123-129.

Platt R.W, Joseph K.S, Ananth C.V, Grondines J., Abrahamowicz M., Kramer M.S (2004). A proportional hazards model with time-dependent covariates and time-varying effects for analysis of fetal and infant death. Epidemiology. Vol 160, No.3 PP 199-206.

Pontes k., Thomas S., Tandor F. (1996). Social characteristics of HIV/AIDS patients. A working paper study in Tanzania.PP 3-5.

Prentice R. and Kalbfleish J. (1978). Biometrics, Vol, 34, No. 10 PP 541-554.

Pushkar M. and Sarmistha P. (2007). Birth Spacing, Fertility Selection and Child Survival: Analysis Using a Correlated Hazard Model. Paper provided by Institute for the Study of Labor (IZA) in its series IZA Discussion Papers with number 2878.

Rabaud C.,Charreau I., Izard S.,Raffi F.,Meiffredy V., Leport C., Aboulker J.P (2000). Adverse Reactions to Cotrimoxazole in HIV Infected Patients: Predictive Factors and

Subsequent Disease Progression. Antimicrobial Agents and Chemotherapy. Vol 40. No.6 PP 33-35.

Rabeneck L., Martin M.C, Jan M.H., Chris E.L and Nelda P.W (2000). A simple clinical staging system that predicts progression to AIDS using CD4 count, oral thrush, and night sweats. Journal of General Internal Medicine, Vol 8, No. 1, PP 5-9.

Richard E. (1996). Working paper on potential Role of Prophylaxis for TB and infections diseases PP 3-6.

Richard P. and Peter J. (2004). Survival Analysis in Social Work Research. British Journal of Social Work: Vol. 34, No. 6, PP. 907-914.

Robert Z., Gilbert R. and Jean P. (1995). Survival Differences in Austrian Patients with the Acquired Immunodeficiency Syndrome European Journal of Epidemiology, Vol. 11, No. 5, PP. 519-526.

Seage G. R, Gatsonis C., Weissman J. S , Haas J .S, Cleary P. D , Fowler F. J, Massagli M. P Stone V. E, Craven D. E, Makadon H., Goldberg J., Coltin K., Levin K .S and Epstein A.M (1997). The Boston AIDS Survival Score (BASS): a multidimensional AIDS severity instrument. Volume 5. No.3 pages 546-548.

Stepanova M., Thomas L. (2002). Survival Analysis Methods for Personal Loan Data. Journal of Operations Research, Vol 50, No.2, PP 277-289.

Stephen R., Franzcp F., Jianguo X., Neil J.P (2004). Impact of compulsory community treatment on admission rates. Epidemiology Vol 11, No.1, PP 496-501.

Tang A.M, Forrester J., Spiegelman D., Knox T.A, Tchetgen E., Gorbach S.L (2003). Weight loss and survival in HIV-positive patients in the era of Highly Active Antiretroviral Therapy: Journal of Acquired Immune Deficiency Syndromes, Vol. 31, No.10, PP 230-236.

Tassie A., Ikeda K., Ogane M., Fukuyama Y., Takeda K. (2003). Factors that influence the predictive behavior of HIV positive people. AIDS, Vol 5, No2, PP 12-14.

Theo (2005). News about AIDS. (www. AIDS Map News).

Toru S. (2001). Two sex proportional hazard model and marriage market analysis. Demography, Vol. 34, No.3 PP 331-341.

UNAIDS (2007). Report Calls for realignment of HIV prevention. (www. The 2007 World AIDS Campaign)

UNAIDS (2008). Faith & AIDS: Connections releases new stats on HIV/AIDS worldwide. (www. faith & AIDS connections).

UNDP (2005). HIV epidemic history. (www.AIDS epidemic history).

Vasantha M., Gopi P.G. and Subramani R. (2008). Survival of Tuberculosis Patients Treated Under Dots in a Rural Tuberculosis Unit (TU), South India. Indian Tuberic.Vol 55, No2, PP 64-69.

Wanzhen G.(2001). The application of survival analysis to LIPID study: the effect of socio-economic status on cardiovascular death. Australasian Biometrics Vol 10, No 1 PP 10-13.

Watanabe M., Ikeda K., Ogane M., Fukuyama Y., Takeda K., Ito S., Yamada Y., Oka S., Kimura S. (2004). Preparation of pre - antiretroviral therapy (ART) assessment manual to lead good adherence to ART in limited HIV- treatment experience situations. International Conference on AIDS. 2004 Jul 11-16; 15: abstract no. B11612. AIDS Clinical Center, Tokyo, Japan.

WHO/UNAIDS/UNICEF (2008). Towards Universal Access: Scaling Up Priority HIV/AIDS Interventions in the Health Sector Progress Report. Media release June 2, 2008.

Wong Y., Wan F., Montagnet C., Localio R., Hudes G., Armstrong K. (2006). Treatment of localized prostate cancer. Prostate Cancer Symposium. Journal of Clinical Oncology Vol 4, No 9, PP 271-275.

WWW. Medicine net.com (February 2009). History of HIV/AIDS.

Yanto S., Geert J., Gero T., Diederic E., Reiner K. (2008). Survival analysis in heart transplantation. European Journal of Cardio-Thoracic Surgery. Vol 33 No.5 PP 856-861.

Yibeltal Assefa, Degu Jerene, Sileshi Lulseged, Gorik O., Wim V.D (2009). Rapid Scale-Up of Antiretroviral Treatment in Ethiopia: Successes and System-Wide Effects. Master's Thesis published by Addis Ababa University.

Zhang Xinjian (2007). HIV/AIDS Relative Survival Analysis. Masters thesis published in Georgia University.

Table 4.1.1 Summary of the number of censored and uncensored values for each category of the predictor variables/factors

Sr. no	Variable	Category	Total	Failed	Censored	Percent censored
1	Age	15-30	273	20	253	93
		30-45	399	46	353	88
		45-60	298	30	268	90
		60-75	30	4	26	87
2	Sex	Female	472	56	416	88
		Male	528	44	484	92
3	Marital status	Never married	310	37	273	88
		Married	371	39	332	89
		Separated	26	3	23	88
		Divorced	135	9	126	93
		Widow/widower	158	12	146	92
4	Religion	Muslim	86	11	75	87
		Orthodox	808	78	730	90
		Protestant	95	11	84	88
		Catholic	2	0	2	100
		Other	9	0	9	100
5	Number of Medications	0	265	34	231	87
		1	485	37	448	92
		2	110	17	93	85
		3	120	4	116	97
		4	20	8	12	60
6	Functional status	Working	423	33	390	92
		Ambulatory	343	35	308	90
		Bedridden	234	32	202	86
7	WHO clinical stages	Stage 1	29	2	27	93
		Stage 2	102	3	99	97
		Stage 3	474	53	421	89
		Stage 4	395	42	353	89
8	ART	Yes	428	53	375	88
		No	572	47	525	92
9	Weight	20-35	45	6	39	87
		35-45	220	26	194	88
		45-55	372	41	331	89
		55-80	363	27	336	93
10	CD4 count/mm^3	CD4 $\leq 200/mm^3$	170	13	157	92
		CD4>$200/mm^3$	830	87	743	90

Table 4.1.2 Means and Medians for Survival Time (years)

Mean				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
2062(6)	365	1346 (4)	2777(8)	1451(4)	185	1089(3)	1813(5)

Table 4.1.3 Log-Rank Test *p*-values for Categorical Covariates in the Black Lion HIV data (n=1000)

Sr. no	Variable	DF	Chi-square	Log-Rank Test *p*-Value
1	Sex	1	3.80	0.04
2	Marital Status	4	8.26	0.08
3	Religion	4	2.57	0.63
4	Functional Status	2	16.69	<0.0001
5	Number of medications	4	32.87	<0.0001
6	WHO Clinical stages	3	9.69	0.02
7	ART	1	11.14	0.001
8	Age	3	3.48	0.32
9	Weight	3	15.58	0.001
10	CD4 count/mm^3	1	7.85	0.04

Table 4.1.4 Means for Survival Time (years)

Variable	Category	Estimate		Std. Error	95% Confidence Interval	
		Days	Years		Lower Bound	Upper Bound
Sex	Female	1126	3	67	995	1257
	Male	2589	7	519	1571	3607
Marital status	Never married	1110	3	81	951	1268
	Married	1796	5	577	665	2926
	Separated	878	2	98	686	1070
	Divorced	1487	4	50	1389	1586
	Widow	1139	3	43	1054	1224
Functional status	Working	1997	5	477	1061	2932
	Ambulatory	1363	4	59	1248	1479
	Bedridden	1031	3	108	819	1243
Number of medications	Nummedic0	919	3	52	818	1021
	Nummedic1	2674	7	420	1852	3497
	Nummedic2	1071	3	88	900	1243
	Nummedic3	860	2	29	804	916
	Nummedic4	709	2	166	384	1033
Weight	20-35	1134	3	140	859	1408
	35-45	802	2	46	712	893
	45-55	2917	8	287	2355	3479
	55-80	1324	4	69	1189	1459
WHO clinical stage	WHO stage1	1132	3	95	945	1318
	WHO stage2	967	3	22	923	1011
	WHO stage3	1989	5	439	1130	2849
	WHO stage4	1164	3	77	1013	1315
CD4 category	CD4>=200 /mm^3	2126	6	370	1401	2851
	CD4<200/ mm^3	1167	3	41	1088	1247
ART	Yes	3377	9	380	964	2455
	No	1317	4	81	1158	1476

Table 4.1.5 Log-Rank Test for each Treatment group

Category	ART-Yes			ART-No		
	Chi-Square	Df	sig.	Chi-Square	df	Sig
Sex	1.88	1	0.17	1.02	1	0.31
Number of medications	10.28	4	0.04	54.05	4	0.00
Weight	5.58	3	0.13	13.10	3	0.00
Functional status	12.71	2	0.00	18.01	2	0.10
WHO clinical stage	5.94	3	0.12	3.64	3	0.30
CD4	2.34	1	0.13	3.13	1	0.08
Age	2.788	3	0.4	11.101	3	0.01
Marital status	5.86	4	0.21	6.53	4	0.16
Religion	0.87	3	0.83	4.03	4	0.40

Table 4.2.1 Log-Rank Test and Partial Likelihood Ratio Test *p*-values for Categorical Covariates in the Black Lion HIV data (n=1000)

Sr. no	Variable	DF	Log-Rank Test *p*-Value	Partial Likelihood Ratio Test *p*-Value
1	Sex	1	0.041	0.052
2	Marital Status	4	0.082	0.066
3	Religion	4	0.633	0.373
4	Functional Status	2	<0.0001	<0.0001
5	Number of medications	4	<0.0001	<0.0001
6	WHO Clinical stages	3	0.021	0.006
7	ART	1	0.001	0.001
8	Age	3	0.323	0.347
9	Weight	3	0.001	0.000
10	CD4 count/mm^3	1	0.043	0.000

Table 4.2.2 Estimated Hazard Ratio for time to death/event on HIV disease with 95% Confidence Intervals, Wald Test and Partial Likelihood Ratio Test *p*-values for Continuous Covariates in the Black Lion HIV data (n=1000)

Sr. no	Variable	Change	Hazard Ratio for Change (95% CIE)	DF	Wald Test *p*-value	Partial Likelihood Ratio Test *p*-value
1	Age	5 year	0.995(0.899,1.101)	1	0.922	0.921
2	Weight	5kg	0.686(0.563,0.836)	1	0.000	0.000
3	CD4	10/mm^3	0.981(0.970,0.993)	1	0.001	0.000

Table 4.2.3 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model Containing Variables Significant at the 20% Level in the Bivariate Analysis for the Black Lion HIV data (n=1000)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.435	0.230	3.567	0.059	0.647
age2	1	0.799	0.289	7.609	0.006	2.222
age3	1	0.459	0.336	1.864	0.172	1.582
age4	1	1.473	0.595	6.137	0.013	4.362
marital0	1	0.148	0.263	0.316	0.574	1.159
marital2	1	0.665	0.616	1.166	0.280	1.945
marital3	1	-0.661	0.382	2.994	0.084	0.516
marital4	1	-0.360	0.345	1.088	0.297	0.698
function1	1	0.274	0.258	1.132	0.287	1.316
function2	1	1.153	0.294	15.340	<.0001	3.167
nummedic1	1	-0.987	0.250	15.532	<.0001	0.373
nummedic2	1	-0.389	0.317	1.504	0.220	0.678
nummedic3	1	-1.788	0.533	11.242	0.001	0.167
nummedic4	1	1.078	0.415	6.758	0.009	2.939
WHO1	1	0.388	0.754	0.265	0.607	1.474
WHO2	1	-0.904	0.614	2.170	0.141	0.405
WHO3	1	0.288	0.224	1.643	0.200	1.333
ART	1	-0.662	0.219	9.098	0.003	0.516
CD4	1	-0.002	0.001	3.519	0.061	0.998
Weight	1	-0.042	0.011	13.791	0.000	0.959

-2Log-Likelihood=1066.884

Table 4.2.4 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model Reducing marital0, marital2 and marital4

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.512	0.222	5.521	0.019	0.599
age2	1	0.739	0.284	6.744	0.009	2.093
age3	1	0.349	0.305	1.310	0.252	1.418
age4	1	1.270	0.569	4.977	0.026	3.559
marital3	1	-0.642	0.361	3.168	0.075	0.526
function1	1	0.288	0.256	1.267	0.260	1.333
function2	1	1.132	0.291	15.171	<.0001	3.101
nummedic1	1	-0.986	0.251	15.473	<.0001	0.373
nummedic2	1	-0.412	0.316	1.701	0.192	0.662
nummedic3	1	-1.789	0.533	11.271	0.001	0.167
nummedic4	1	1.104	0.412	7.191	0.007	3.015
WHO1	1	0.423	0.752	0.317	0.574	1.527
WHO2	1	-0.953	0.612	2.425	0.119	0.385
WHO3	1	0.279	0.221	1.595	0.207	1.321
ART	1	-0.650	0.219	8.843	0.003	0.522
CD4	1	-0.002	0.001	3.704	0.054	0.998
Weight	1	-0.043	0.011	14.330	0.000	0.958

-2Log-Likelihood=1070.013

Table 4.2.5 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model Reducing Marital3

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.592	0.219	7.285	0.007	0.553
age2	1	0.687	0.283	5.884	0.015	1.987
age3	1	0.305	0.304	1.008	0.315	1.356
age4	1	1.241	0.569	4.752	0.029	3.459
function1	1	0.287	0.256	1.262	0.261	1.333
function2	1	1.101	0.289	14.480	0.000	3.007
nummedic1	1	-0.973	0.251	15.085	0.000	0.378
nummedic2	1	-0.435	0.315	1.905	0.168	0.647
nummedic3	1	-1.764	0.533	10.965	0.001	0.171
nummedic4	1	1.152	0.410	7.886	0.005	3.163
WHO1	1	0.434	0.751	0.333	0.564	1.543
WHO2	1	-0.944	0.612	2.382	0.123	0.389
WHO3	1	0.252	0.220	1.305	0.253	1.286
ART	1	-0.660	0.219	9.079	0.003	0.517
CD4	1	-0.002	0.001	4.018	0.045	0.998
Weight	1	-0.043	0.011	14.154	0.000	0.958

-2Log-Likelihood=1073.712

Table 4.2.6 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model reducing function1

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.585	0.219	7.115	0.008	0.557
age2	1	0.686	0.283	5.898	0.015	1.987
age3	1	0.297	0.302	0.966	0.326	1.346
age4	1	1.211	0.568	4.547	0.033	3.358
function2	1	0.943	0.249	14.287	0.000	2.568
nummedic1	1	-0.969	0.250	15.005	0.000	0.380
nummedic2	1	-0.432	0.315	1.883	0.170	0.649
nummedic3	1	-1.770	0.533	11.046	0.001	0.170
nummedic4	1	1.188	0.409	8.431	0.004	3.280
WHO1	1	0.319	0.743	0.185	0.667	1.376
WHO2	1	-0.977	0.612	2.550	0.110	0.377
WHO3	1	0.219	0.218	1.014	0.314	1.245
ART	1	-0.653	0.219	8.868	0.003	0.521
CD4	1	-0.003	0.001	4.712	0.030	0.997
Weight	1	-0.044	0.011	14.919	0.000	0.957

-2Log-Likelihood=1074.972

Table 4.2.7 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model reducing nummedic2

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.577	0.219	6.959	0.008	0.562
age2	1	0.698	0.284	6.041	0.014	2.011
age3	1	0.310	0.303	1.045	0.307	1.363
age4	1	1.246	0.568	4.808	0.028	3.477
function2	1	0.954	0.249	14.637	0.000	2.596
nummedic1	1	-0.799	0.224	12.781	0.000	0.450
nummedic3	1	-1.612	0.523	9.511	0.002	0.199
nummedic4	1	1.339	0.398	11.319	0.001	3.814
WHO1	1	0.337	0.744	0.205	0.651	1.400
WHO2	1	-0.940	0.612	2.363	0.124	0.390
WHO3	1	0.211	0.218	0.936	0.333	1.234
ART	1	-0.691	0.217	10.098	0.002	0.501
CD4	1	-0.002	0.001	3.981	0.046	0.998
Weight	1	-0.041	0.011	13.822	0.000	0.960

-2Log-Likelihood=1076.929

Table 4.2.8 Estimated Coefficients, Standard Errors, Chi-Square, Hazard Ratio and *p*-values for the Proportional Hazards Model Reducing the Design Variables of WHO Clinical stages

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.576	0.214	7.200	0.007	0.562
age2	1	0.673	0.284	5.635	0.018	1.961
age3	1	0.300	0.302	0.985	0.321	1.350
age4	1	1.117	0.566	3.891	0.049	3.056
function2	1	0.967	0.243	15.860	<.0001	2.631
nummedic1	1	-0.783	0.221	12.532	0.000	0.457
nummedic3	1	-1.572	0.523	9.040	0.003	0.208
nummedic4	1	1.395	0.396	12.388	0.000	4.034
ART	1	-0.698	0.213	10.682	0.001	0.498
CD4	1	-0.002	0.001	4.703	0.030	0.998
Weight	1	-0.039	0.011	13.308	0.000	0.962

-2Log-Likelihood=1082.483

Table 4.2.9 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model without age3 /Preliminary Main Effects Model

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio	Confidence Interval for the Hazard Ratio	
Sex	1	-0.589	0.214	7.616	0.006	0.55	0.365	0.843
age2	1	0.500	0.215	5.378	0.020	1.65	1.080	2.514
age4	1	1.051	0.533	3.888	0.039	2.86	1.006	8.140
function2	1	0.984	0.243	16.385	<.0001	2.67	1.661	4.305
nummedic1	1	-0.768	0.221	12.100	0.001	0.46	0.301	0.715
nummedic3	1	-1.569	0.523	9.005	0.003	0.21	0.075	0.580
nummedic4	1	1.401	0.396	12.529	0.000	4.06	1.869	8.823
ART	1	-0.697	0.214	10.627	0.001	0.50	0.328	0.757
CD4	1	-0.002	0.001	4.278	0.039	1.00	0.996	0.999
Weight	1	-0.037	0.011	12.368	0.000	0.96	0.944	0.984

-2Log-Likelihood (with covariates)=1083.482 (without covariates) =1175.033

Table 4.2.10 Comparison of the Model for the inclusion of the Design variables marital0, marital2 and marital4 with the Model in Table 4.2.9

Proportional Hazards Model without marital0,marital2 and marital4			Proportional Hazards Model including marital0,marital2 and marital4			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.539	-9.3
age2	1	0.500	age2	1	0.511	2.2
age4	1	1.051	age4	1	1.051	-0.1
function2	1	0.984	function2	1	1.019	3.4
nummedic1	1	-0.768	nummedic1	1	-0.774	0.7
nummedic3	1	-1.569	nummedic3	1	-1.575	0.4
nummedic4	1	1.401	nummedic4	1	1.399	-0.2
ART	1	-0.697	ART	1	-0.711	1.9
CD4	1	-0.002	CD4	1	-0.002	-3.9
Weight	1	-0.037	Weight	1	-0.036	-3.7
			marital0	1	0.124	
			marital2	1	0.944	
			marital4	1	-0.232	
-2Log-Likelihood=1083.482			-2log-Likelihood=1080.486			

Table 4.2.11 Comparison of the Model for the inclusion of the marital3 with the Model in Table 4.2.9

Proportional Hazards Model without marital3			Proportional Hazards Model including marital3			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.532	-10.8
age2	1	0.500	age2	1	0.532	6.1
age4	1	1.051	age4	1	0.928	-13.2
function2	1	0.984	function2	1	1.014	3.0
nummedic1	1	-0.768	nummedic1	1	-0.779	1.4
nummedic3	1	-1.569	nummedic3	1	-1.594	1.6
nummedic4	1	1.401	nummedic4	1	1.359	-3.1
ART	1	-0.697	ART	1	-0.686	-1.6
CD4	1	-0.002	CD4	1	-0.002	-4.4
Weight	1	-0.037	Weight	1	-0.037	0.0
			marital3	1	-0.595	
-2Log-Likelihood=1083.482			-2log-Likelihood=1080.304			

Table 4.2.12 Comparison of the Model for the inclusion of the function1 with the Model in Table 4.2.9

Proportional Hazards Model without function1			Proportional Hazards Model including function1			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.594	0.8
age2	1	0.500	age2	1	0.499	0.0
age4	1	1.051	age4	1	0.956	-10.0
function2	1	0.984	function2	1	1.112	11.6
nummedic1	1	-0.768	nummedic1	1	-0.761	-1.0
nummedic3	1	-1.569	nummedic3	1	-1.559	-0.7
nummedic4	1	1.401	nummedic4	1	1.368	-2.4
ART	1	-0.697	ART	1	-0.693	-0.6
CD4	1	-0.002	CD4	1	-0.002	-8.7
Weight	1	-0.037	Weight	1	-0.036	-3.5
			function1	1	0.247	
-2Log-Likelihood=1083.482			-2log-Likelihood=1082.529			

Table 4.2.13 Comparison of the Model for the inclusion of the nummedic2 with the Model in Table 4.2.9

Proportional Hazards Model without nummedic2			Proportional Hazards Model including nummedic2			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.592	0.4
age2	1	0.500	age2	1	0.491	-1.8
age4	1	1.051	age4	1	0.899	-17.0
function2	1	0.984	function2	1	0.970	-1.4
nummedic1	1	-0.768	nummedic1	1	-0.919	16.4
nummedic3	1	-1.569	nummedic3	1	-1.710	8.2
nummedic4	1	1.401	nummedic4	1	1.266	-10.7
ART	1	-0.697	ART	1	-0.664	-4.9
CD4	1	-0.002	CD4	1	-0.003	9.2
Weight	1	-0.037	Weight	1	-0.040	6.2
			nummedic2	1	-0.395	
-2Log-Likelihood=1083.482			-2log-Likelihood=1081.867			

Table 4.2.14 Comparison of the Model for the inclusion of the Design Variables of WHO with the Model in Table 4.2.9

Proportional Hazards Model without the Design Variables of WHO			Proportional Hazards Model including the Design Variables of WHO			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.591	0.3
age2	1	0.500	age2	1	0.520	3.9
age4	1	1.051	age4	1	1.061	0.9
function2	1	0.984	function2	1	0.971	-1.3
nummedic1	1	-0.768	nummedic1	1	-0.783	1.9
nummedic3	1	-1.569	nummedic3	1	-1.607	2.3
nummedic4	1	1.401	nummedic4	1	1.344	-4.2
ART	1	-0.697	ART	1	-0.692	-0.8
CD4	1	-0.002	CD4	1	-0.002	-10.7
Weight	1	-0.037	Weight	1	-0.039	4.9
			WHO2	1	-1.226	
			WHO3	1	-0.080	
			WHO4	1	-0.294	
-2Log-Likelihood=1083.482			-2log-Likelihood=1077.989			

Table 4.2.15 Comparison of the Model for the inclusion of the age3 with the Model in Table 4.2.9

Proportional Hazards Model without age3			Proportional Hazards Model including age3			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.576	-2.4
age2	1	0.500	age2	1	0.573	12.8
age4	1	1.051	age4	1	1.117	5.9
function2	1	0.984	function2	1	0.967	-1.7
nummedic1	1	-0.768	nummedic1	1	-0.783	1.9
nummedic3	1	-1.569	nummedic3	1	-1.572	0.2
nummedic4	1	1.401	nummedic4	1	1.395	-0.5
ART	1	-0.697	ART	1	-0.698	0.1
CD4	1	-0.002	CD4	1	-0.002	4.8
Weight	1	-0.037	weight	1	-0.039	5.2
			age3	1	0.300	
-2Log-Likelihood=1083.482			-2log-Likelihood=1082.483			

Table 4.2.16 Comparison of the Model for the inclusion of the Design Variables of Religion with the Model in Table 4.2.9

Proportional Hazards Model without the Design Variables of Religion			Proportional Hazards Model including the Design Variables of Religion			Percentage Change in Coefficient
Variable	DF	Coefficient	Variable	DF	Coefficient	
Sex	1	-0.589	Sex	1	-0.591	0.2
age2	1	0.500	age2	1	0.451	-10.8
age4	1	1.051	age4	1	0.955	-10.1
function2	1	0.984	function2	1	0.963	-2.1
nummedic1	1	-0.768	nummedic1	1	-0.788	2.5
nummedic3	1	-1.569	nummedic3	1	-1.580	0.7
nummedic4	1	1.401	nummedic4	1	1.324	-5.9
ART	1	-0.697	ART	1	-0.704	1.0
CD4	1	-0.002	CD4	1	-0.002	-3.0
Weight	1	-0.037	Weight	1	-0.037	-0.3
			religion1	1	-0.155	
			religion2	1	0.218	
			religion3	1	-13.086	
			religion4	1	-12.873	
-2Log-Likelihood=1083.482			-2log-Likelihood=1080.496			

Table 4.2.17 *p*-values of interaction terms among two variables

Interaction	Variables	Df	*p*-Value	Interaction	Variables	Df	*p*-Value
Nummedic1	Function2	1	0.4285	Function2	Sex	1	0.7324
	Sex	1	0.3541		ART	1	0.4852
	ART	1	0.0006*		CD4	1	0.1415
	CD4	1	0.4573		Weight	1	0.3404
	Weight	1	0.2605		age2	1	0.0780
	age2	1	0.9617		age4	1	0.0943
	age4	1	0.9831	Sex	ART	1	0.7674
Nummedic3	Function2	1	0.2043		CD4	1	0.7802
	Sex	1	0.1607		Weight	1	0.1506
	ART	1	0.3616		age2	1	0.0055*
	CD4	1	0.4697		age4	1	0.0769
	Weight	1	0.0874	ART	CD4	1	0.4054
	age2	1	0.9747		Weight	1	0.1133
	age4	1	0.9821		age2	1	0.3246
Nummedic4	Function2	1	0.9245		age4	1	0.9764
	Sex	1	0.0905	CD4	Weight	1	0.6817
	ART	1	0.9799		age2	1	0.1259
	CD4	1	0.4512		age4	1	0.0736
	Weight	1	0.6052	Weight	age2	1	0.1935
	age2	1	0.101		age4	1	0.4870
	age4	1	0.320				

Table 4.2.18 Estimated Coefficients, Standard Errors, Chi-Square, *p*-values and Hazard Ratio for the Proportional Hazards Model Including Interaction Terms

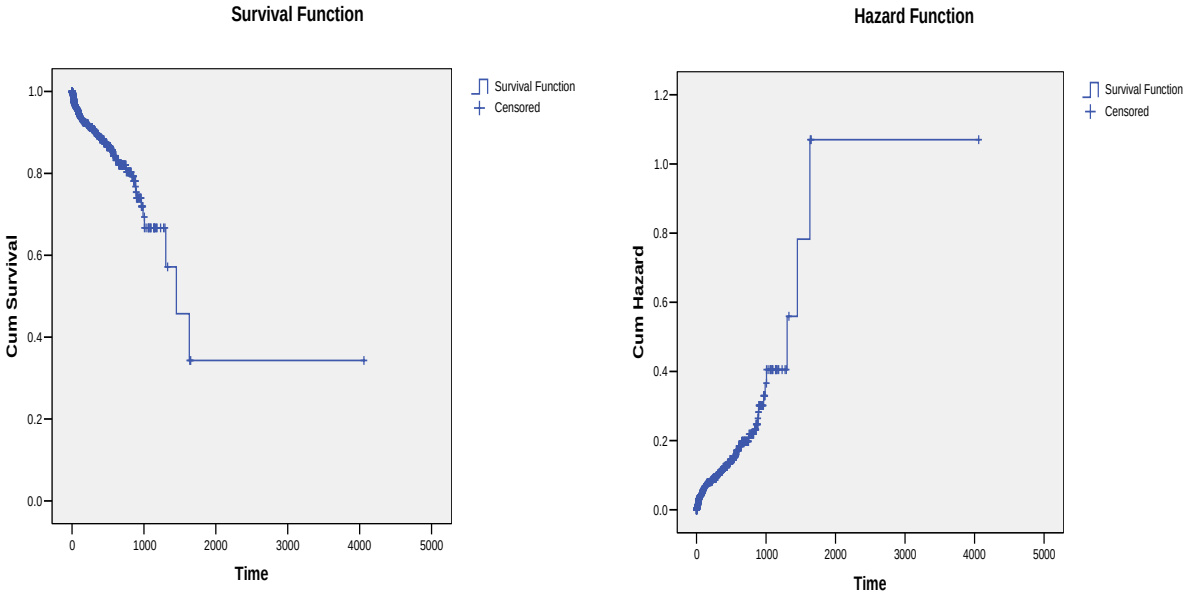
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sex	1	-0.582	0.303	10.479	0.001	0.559
age2	1	0.164	0.301	0.295	0.587	1.178
age4	1	0.563	0.217	6.697	0.010	1.755
function2	1	0.865	0.248	12.219	0.001	2.376
nummedic1	1	-0.160	0.288	0.309	0.578	0.852
nummedic3	1	-1.491	0.523	8.111	0.004	0.225
nummedic4	1	1.241	0.400	9.602	0.002	3.459
ART	1	-0.210	0.275	0.583	0.445	0.811
CD4	1	-0.002	0.001	3.586	0.058	0.998
Weight	1	-0.037	0.010	12.383	0.000	0.964
nummedic1ART	1	0.822	0.534	2.369	0.124	2.275
sexage2	1	0.854	0.430	3.939	0.047	2.348

-2Log-Likelihood=1081.116

Table 4.2.19 Estimated Coefficients, Standard Errors, Chi-Square Value for the ten interactions with Log-time Added to the Model in Table 4.2.9 for the Black lion HIV Data (n=1000)

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
Sext	1	0.035	0.146	0.058	0.810	1.036
age2t	1	0.152	0.145	1.096	0.295	1.164
age4t	1	0.612	0.419	2.126	0.145	1.843
function2t	1	0.331	0.165	4.019	0.085	1.392
nummedic1t	1	0.069	0.149	0.213	0.644	1.071
nummedic3t	1	0.134	0.394	0.116	0.733	1.144
nummedic4t	1	0.538	0.322	2.801	0.094	1.713
ARTt	1	-0.090	0.147	0.377	0.540	0.914
CD4t	1	0.001	0.001	2.554	0.110	1.001
Weightt	1	0.006	0.007	0.639	0.424	1.006

-2Log-Likelihood=1070.023



(a) (b)

Figure 4.1.1 (a) & (b).Graphs of the Kaplan-Meier estimator of the survival & hazard functions. (Time in days)

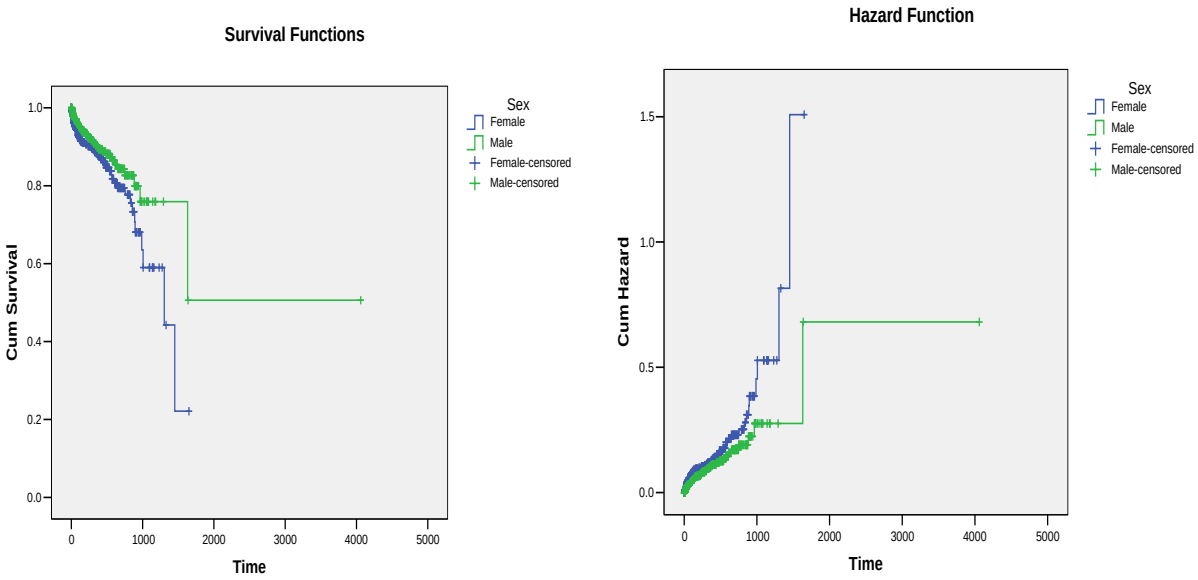


Figure 4.1.2a. Graph of the Kaplan-Meier Estimator of the Survival and Hazard Functions for sex. (Time in days)

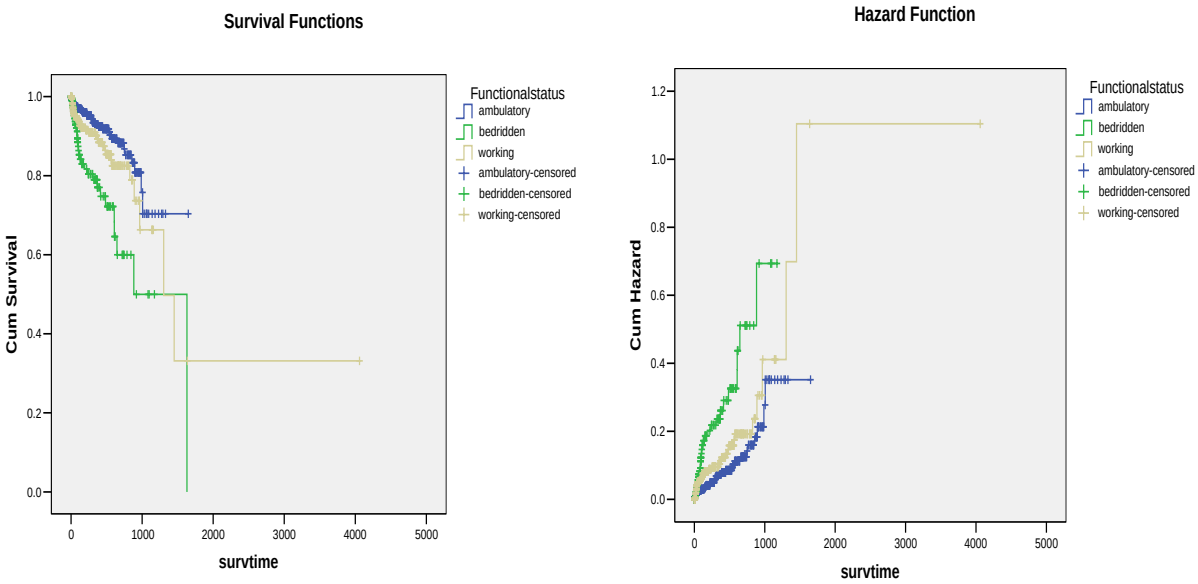


Figure 4.1.2b. Graph of the Kaplan-Meier Estimator of the Survival and Hazard Functions for functional status. (Survival time in days)

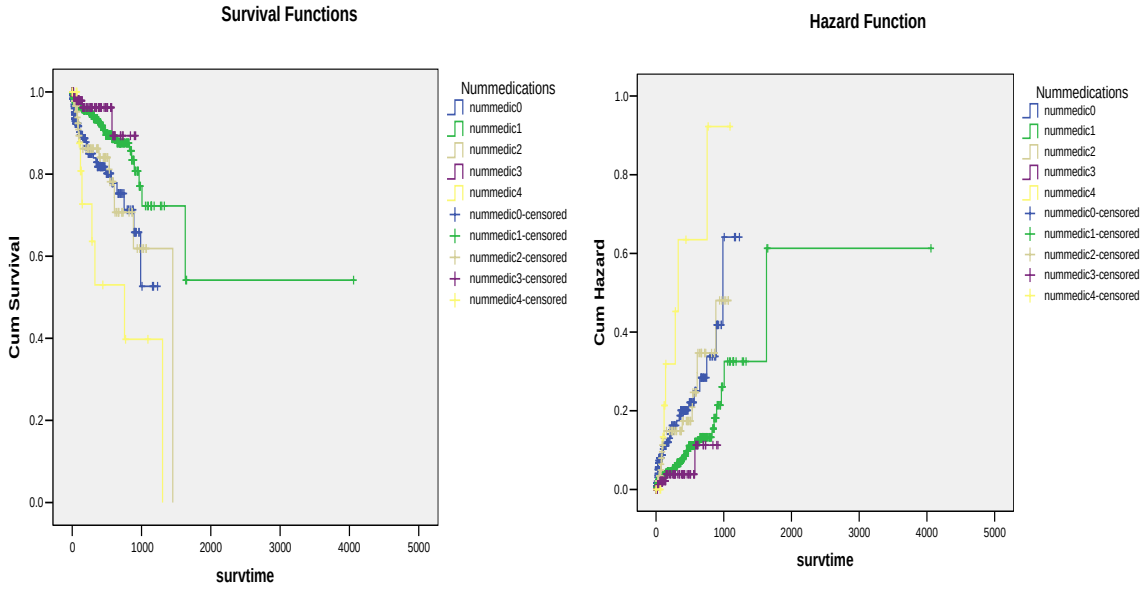


Figure 4.1.2c. Graphs of the Kaplan-Meier Estimator of the Survival and Hazard Functions for number of medications. (Survival time in days)

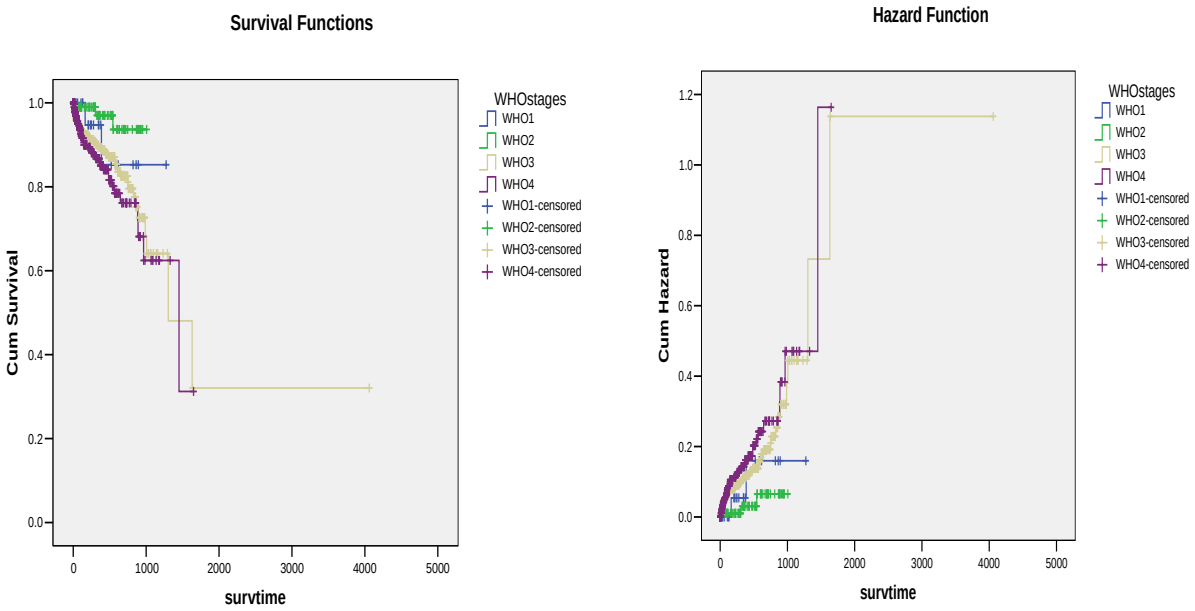


Figure 4.1.2d. Graphs of the Kaplan-Meier Estimator of the Survival and Hazard Functions for WHO clinical stages. (Survival time in days)

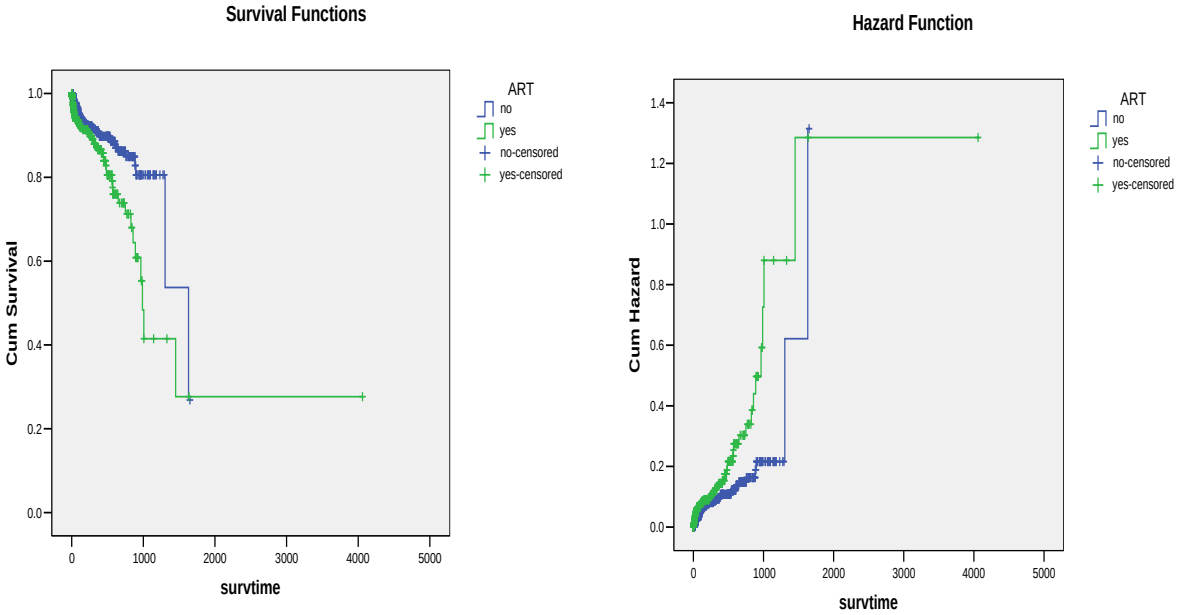


Figure 4.1.2 e. Graphs of the Kaplan-Meier Estimator of the Survival and Hazard Functions for ART. (Survival time in days)

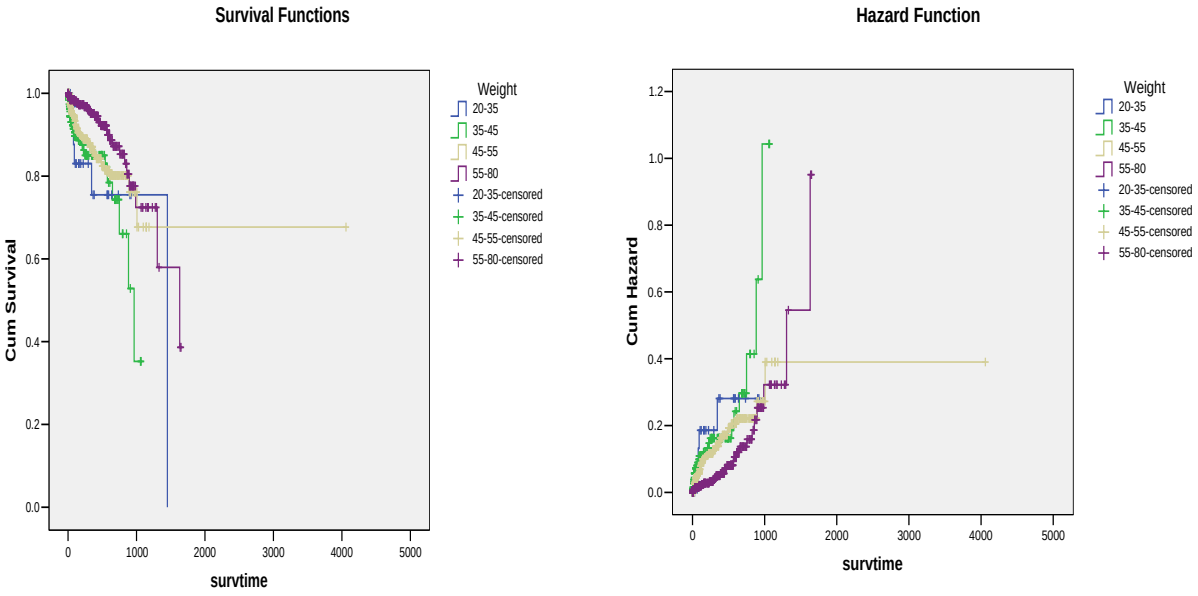
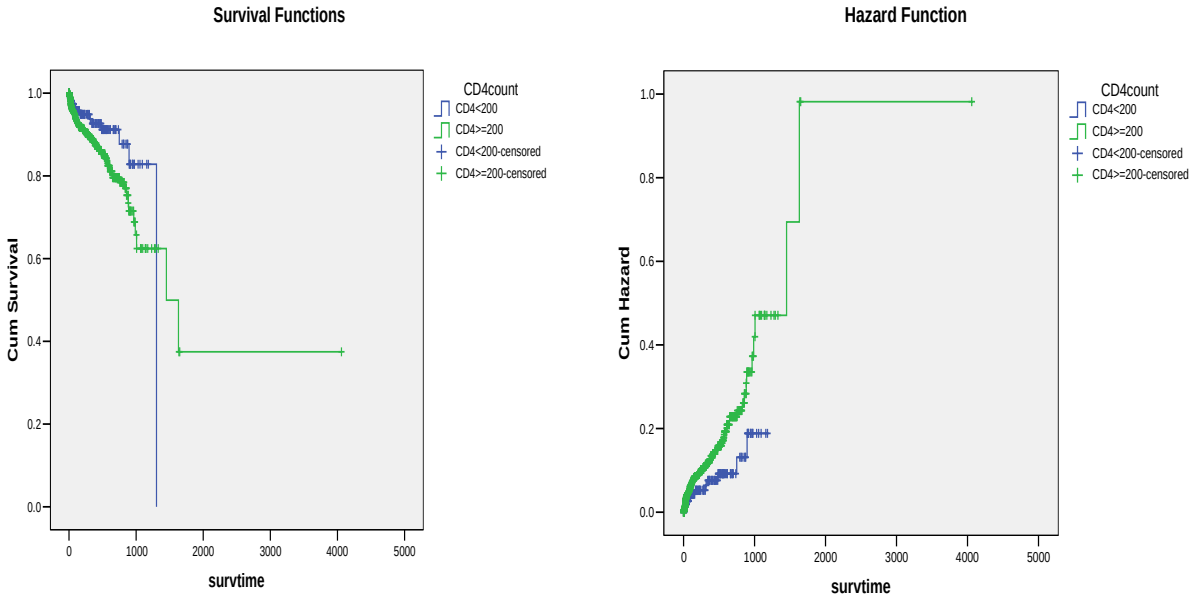
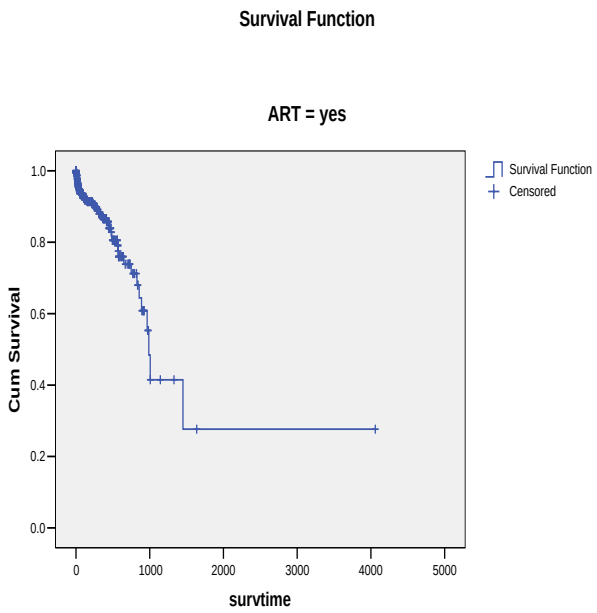


Figure 4.1.2f. Graphs of the Kaplan-Meier Estimator of the Survival and Hazard Functions for weight. (Survival time in days)

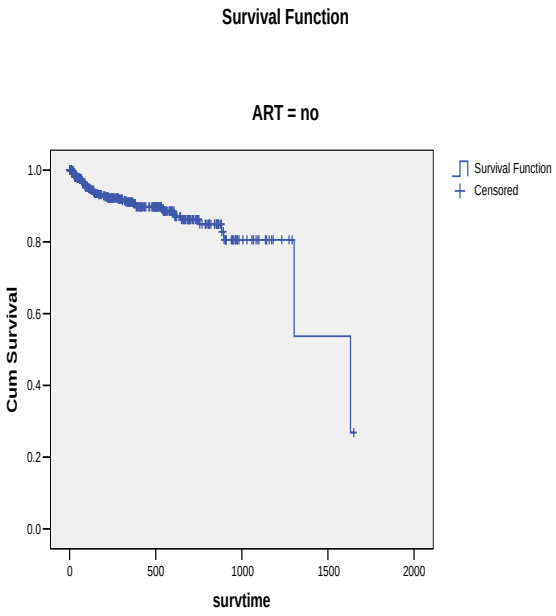


(g)

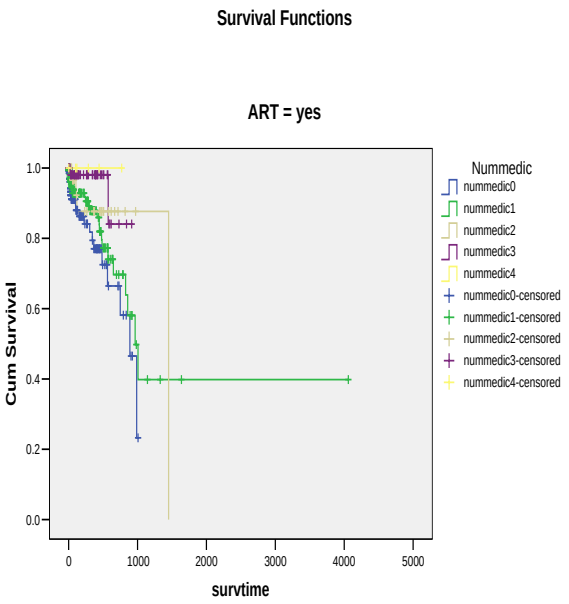
Figure 4.1.2g. Graphs of the Kaplan-Meier Estimator of the Survival and Hazard Functions for CD4 count categories


Survival Function
ART = yes
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
survtime
Survival Function
Censored

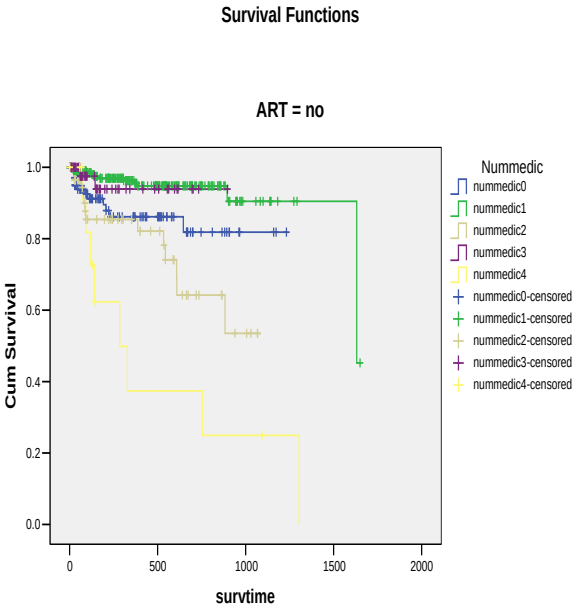
(a)


Survival Function
ART = no
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
survtime
Survival Function
Censored

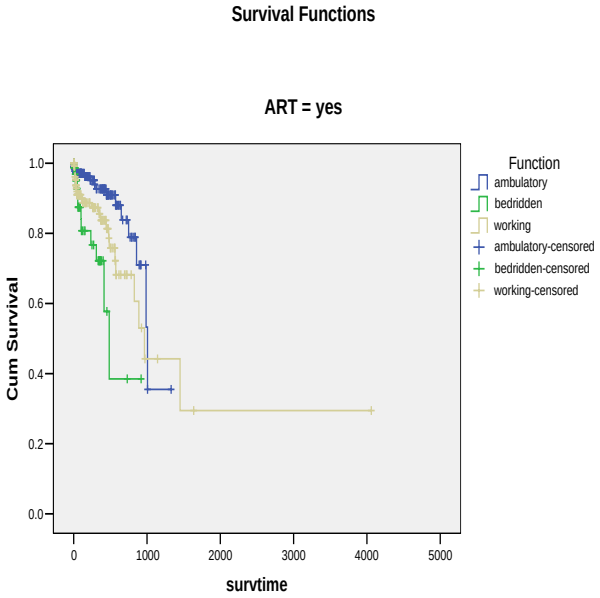
(b)


Survival Functions
ART = yes
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
survtime
Nummedic
nummedic0
nummedic1
nummedic2
nummedic3
nummedic4
nummedic0-censored
nummedic1-censored
nummedic2-censored
nummedic3-censored
nummedic4-censored

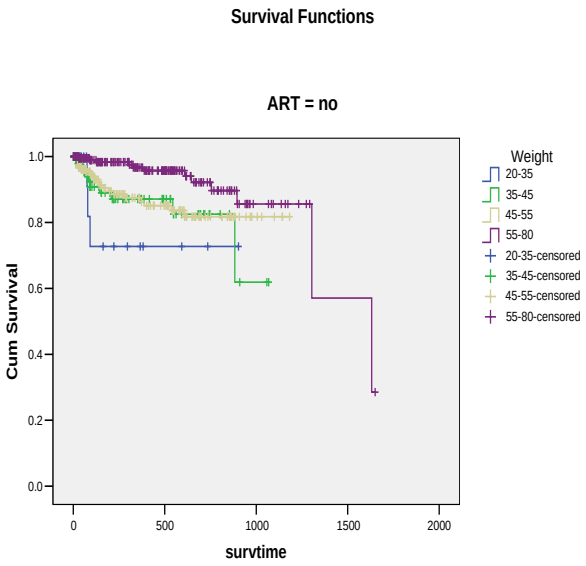
(c)


Survival Functions
ART = no
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
survtime
Nummedic
nummedic0
nummedic1
nummedic2
nummedic3
nummedic4
nummedic0-censored
nummedic1-censored
nummedic2-censored
nummedic3-censored
nummedic4-censored

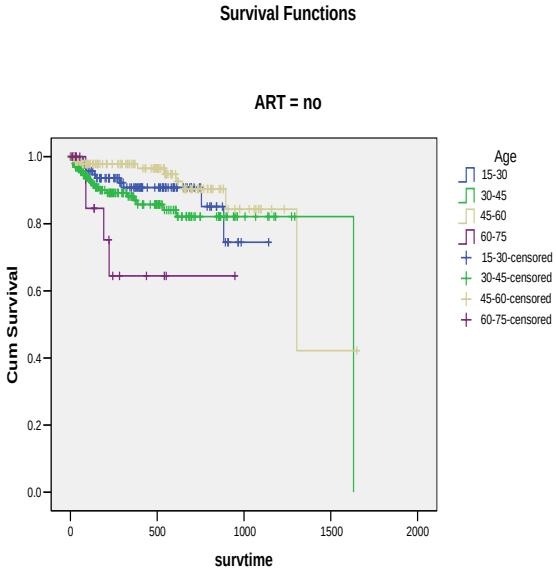
(d)



(e)



(f)



(g)

Figure 4.1.3(a)-(g). Graphs of the Kaplan-Meier Survival Estimates for the covariates under each of the treatment groups. (Survival time in days)

APPENDIX I

Variance -Covariance Matrix for the preliminary final model in Table 4.2.9

Obs.	Name	Sex	age2	age4	function2	nummedic1
1	Sex	0.046	-0.007	0.000	-0.004	0.000
2	age2	-0.007	0.046	0.024	0.007	0.001
3	age4	0.000	0.024	0.284	-0.003	0.008
4	function2	-0.004	0.007	0.003	0.059	0.001
5	nummedic1	0.000	0.001	0.008	0.001	0.049
6	nummedic3	0.001	-0.002	0.009	-0.005	0.020
7	nummedic4	-0.004	0.012	0.024	0.013	0.021
8	ART	-0.002	0.001	0.008	-0.014	-0.002
9	CD4	0.000	0.000	0.000	0.000	0.000
10	Weight	0.001	0.000	0.001	0.000	0.000

Obs.	Name	nummedic3	nummedic4	ART	CD4	weight
1	Sex	0.001	-0.004	0.002	0.000	0.001
2	age2	-0.002	0.012	0.001	0.000	0.000
3	age4	0.009	0.024	0.008	0.000	-0.001
4	function2	-0.005	0.013	0.014	0.000	0.000
5	nummedic1	0.020	0.021	0.002	0.000	0.000
6	nummedic3	0.273	0.018	0.007	0.000	0.000
7	nummedic4	0.018	0.157	0.007	0.000	0.000
8	ART	0.007	-0.007	0.046	0.000	0.000
9	CD4	0.000	0.000	0.000	0.000	0.000
10	Weight	0.000	0.000	0.000	0.000	0.000

APPENDIX II

Description of Variables in the study of Black lion HIV Data

Variable	Description	Codes/Values
Patient ID	Identification Code	1-1000
Age	Age at enrollment	1=(15-30), 2=(30-45), 3=(45-60), 4=(60-75)
Sex	Patient's sex	1=Male 0=Female
Marital	Marital status	0=Unmarried,1 =Married, 2= Separated 3=Divorced , 4= Widow
Religion	Patient's religion	0=Muslim, 1=Orthodox, 2=Protestant 3= Catholic, 4= Other
Nummedic	Number of medication taken	0=Nummedic0,1=Nummedic1,2=Nummedic2 3=Nummedic3, 4=Nummedic4
Weight	Patient's weight	In Kilograms
Function	Patient's functional status	0=Working 1=Ambulatory 2=Bedridden
WHO	WHO clinical stages	1=Stage1, 2=Stage2, 3=Stage3, 4=Stage4
CD4	CD4 count	Per mm^3
ART	Antiretroviral treatment	1=Yes, 0=No
Survtime	Time to death (measured from enrollment)	In days
Censor	death/lost to follow-up/dropped transferred out	1=Death, 0= Otherwise

APPENDIX III

SAS AND STATA CODES

Checking the Assumption of Linearity

SAS CODES

```
Generating the plots of the martingale residuals and the lowess smoothed
residuals: for CD4 count.
Proc phreg data=Blackhiv noprint;
model time*censor(0)=sex age2 age4 function2 nummedic1  nummedic3 nummedic4
ART weight;
Output out=residuals resmart=resmart;
run;
proc sort data=residuals;
  by time;
run;
data CD4;
  set Blackhiv;
  keep time CD4;
run;
proc sort data=CD4;
  by time;
run;
data residCD4;
  merge residuals CD4;
  by time;
run;
ods listing close;
proc loess data=residCD4;
  model  resmart = CD4 /smooth=0.8;
  ods output OutputStatistics=myout;
  run;
quit;
ods listing;
proc sort data=myout;
  by CD4;
run;
goptions reset=all;
symbol1 c = black i=none v=star h=0.7;
symbol2 c=blue i=join v=none;
axis label=(a=90 'Martingale Resid');
proc gplot data=myout;
  format DepVar f4.0 CD4 f4.0;
  plot DepVar*CD4=1 Pred*CD4=2 /overlay vaxis=axis1 ;
run;
quit;
```

STATA CODES

```
stcox , sex age2 age4 function2 nummedic1 nummedic3 nummedic4 ART weight,
nohr mgale(mgale)
graph twoway (lowess mgale CD4) (scatter mgale CD4), ylabel(-2.642 .997)
xlabel(20 56)
drop mgale
```

```
stcox , sex age2 age4 function2 nummedic1  nummedic3 nummedic4 ART CD4,
nohr mgale(mgale)
graph twoway (lowess mgale weight) (scatter mgale weight), ylabel(-2.642
.997) xlabel(20 56)
```

Checking the Assumption of Proportional Hazards Regression Model

```
proc phreg data=Blackhiv;
model time*censor(0) = sex age2 age4 function2 nummedic1 nummedic3
nummedic4 ART CD4 weight sext age2t age4t function2t nummedic1t nummedic3t
nummedic4t ARTt CD4t weightt;
```

SAS CODES

```
   sext=sex*log(time);
   age2t=age2*log(time);
   age4t=age4*log(time);
   function2t=function2*log(time);
   nummedic1t=nummedic1*log(time);
   nummedic3t=nummedic3*log(time);
   nummedic4t=nummedic4*log(time);
   ARTt=ART*log(time);
   CD4t=CD4*log(time);
   weightt=weight*log(time);
    run;
```

STATA CODES

```
stcox sex age2 age4 function2 nummedic1  nummedic3 nummedic4 ART weight
CD4, nohr nolog noshow tvc(sex age2 age4 function2 nummedic1  nummedic3
nummedic4 ART weight CD4) texp( ln(_t) )
stcox sex age2 age4 function2 nummedic1  nummedic3 nummedic4 ART weight
CD4, nohr sca(sca*)
* Stata 9 code and graph.
estat phtest, log plot(sex) yline(0)
```

Checking outliers using score residuals

SAS CODES

```
proc phreg data=Blackhiv noprint;
  model time*censor(0) = sex age2 age4 function2 nummedic1 nummedic3
nummedic4 ART CD4 weight;
  output out= res ressco=rsex rage2 rage4 rfunction2 rnummedic1 rnummedic3
rnummedic4  rART rCD4 rweight;
run;
goptions reset=all;
symbol v=dot h=.3 c=blue;
proc gplot data=res;
  plot rCD4*CD4;
  plot rweight*weight;
run;
quit;
```

STATA CODES

```
stcox sex age2 age4 function2 nummedic1  nummedic3 nummedic4 ART weight
CD4 nohr nolog esr(scr*)
graph twoway scatter scr1 weight, ylabel(-16.26 30.76) xlabel(20 80)
```

Declaration

I declare that this thesis is my original work and has not been presented for a degree in any University. All the sources of material used for the thesis are duly acknowledged.

Name: _____

Signature: _____

Date: _____

Place: _____

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

Name: _____

Signature: _____

Date: _____

Place: _____