ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

DEPARTMENT OF STATISTICS

DETERMINING FACTORS THAT AFFECT THE SURVIVAL RATE OF HIV-INFECTED PATIENTS ON ART: THE CASE OF ARMED FORCES GENERAL TEACHING HOSPITAL, ADDIS ABABA, ETHIOPIA

BY

KETEMA KEBEBEW

A THESIS SUBMITTED TO THE OFFICE OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS (BIOSTATISTICS)

JUNE, 2011

ADDIS ABABA

ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

DEPARTMENT OF STATISTICS

DETERMINING FACTORS THAT AFFECT THE SURVIVAL RATE OF HIV-INFECTED PATIENTS ON ART: THE CASE OF ARMED FORCES GENERAL TEACHING HOSPITAL, ADDIS ABABA, ETHIOPIA

BY

KETEMA KEBEBEW

Approved by the board of examiners:

Ato Yilma Tefera	----------
Chairman, Department graduate committee	Signature
Ato Mekonnen Tadesse	----------
Examiner	Signature
Butte Gotu (Ph.D)	----------
Examiner	Signature

TABLE OF CONTENTS

Tables

Figures

ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
ARV	Antiretroviral
CI	Confidence Interval
D4T	Stavudine
EFV	Efavirenz
HAART	Highly Active Antiretroviral Therapy
HGB	Haemoglobin
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
AFGTH	Armed forces general teaching hospital
CSWs	Commercial sex workers
EDHS	Ethiopian Demographic and Health Survey
STDs	Sexually transmitted diseases
MoH	Ministry of Health
LR	Likelihood Ratio
MLE	Maximum Likelihood Estimate/Estimator
OIs	Opportunistic Infections
PLWHA	People Living with HIV/AIDS
PMTCT	Prevention of Mother to Child Transmission
TB	Tuberculosis
AA	Addis Ababa
TLC	Total Lymphocyte Count
WHO	World Health Organization
ZDV	Zidovudine
3TC	Lamuvidine
IQR	Interquartile range

ACKNOWLEDGMENT

My heartfelt gratitude goes to Professor Eshetu Wencheko, my advisor and instructor, for his excellent scientific guidance and tireless efforts to make this work a reality.

My sincere appreciation also goes to my friends and colleagues whoever explicitly or implicitly backed my work in whatsoever means.

ABSTRACT

Since the first detection of AIDS in Ethiopia, it claims about 67,000 lives of people in 2007. Of course, the introduction of ART dramatically improved survival and health quality of HIV-infected patients in the industrialized world; and the survival benefit of ART has been well studied too. However, in resource-poor settings, where such treatment was started only recently, limited data exist on treatment results. Moreover, mortality has been high particularly in the first month of initiating ART and factors contributing to this high mortality are poorly understood. The study is designed to identify determinant factors (demographic and health related) that affect the survival of HIV-infected patients in the Ethiopian Armed Forces based on the data from Armed Forces General Teaching Hospital (AFGTH).

A retrospective cohort study was conducted in AFGTH located in Addis Ababa, Ethiopia. Records of patients enrolled between September 2003 and August 2007 were reviewed continuously using patients' ART unique identification numbers as reference. Kaplan-Meier survival curves and Log-Rank test were used to compare the survival experience of different category of patients, and proportional hazards Cox model was employed to identify independent predictors of mortality. 734 patients on ART were followed for a median of 38.5 months (IQR 10.75, 53). Of these 86 died during the follow up time of whom 28 (32.6%), 43 (50%) and 61 (70.9%) deaths occurred within three months, six months and twelve months of ART initiation, respectively. The independent predictors of mortality were low CD4 cell count at baseline, (HR = 0.995, 95% CI: 0.991 - 0.999), ambulatory and bedridden functional status, (HR=2.011, 95%CI: 1.018 - 3.973) and (HR=3.358, 95%CI: 1.734 - 6.500), respectively, WHO clinical stages III and IV (HR=7.052, 95%CI: 1.677- 29.658) and (HR=12.64, 95%CI: 3.003 - 53.199), respectively, TB co-infection, (HR=1.734, 95% CI: 1.039 - 2.893) and OIs (HR=8.985, 95% CI: 1.240 - 65.085).

In nut shell, there has been a high mortality of the cohort in the earlier months of treatment. Thus, a careful monitoring of patients with low CD4 cell count, advanced WHO staging, history of OIs, co-infection with TB and being employed is necessary in order to improve the survival of AIDS patients.

CHAPTER ONE

INTRODUCTION

1.1 Background and statement of the problem

Two dozen years after the first clinical cases of acquired immunodeficiency syndrome (AIDS) were reported, AIDS has become the most devastating disease humankind has ever faced. AIDS is one of about 30 new infectious diseases, including Legionnaires' disease, hepatitis C, bovine spongiform encephalopathy/variant Creutzfeld-Jakob disease, several viral hemorrhagic fevers, severe acute respiratory syndrome (SARS) and, most recently, avian influenza, that have emerged as a result of profound world-wide changes in human ecology (Kloos *et al.*, 2007).

AIDS is a disease of the human immune system caused by the human immunodeficiency virus (HIV). Each letter stands for A = Acquired (not inherited), I = Weakens the immune system, D = Creates a deficiency of CD4+ cells[1] in the immune system, and S = Syndrome, or a group of illnesses taking place at the same time. HIV infects primarily vital cells in the human immune system such as helper T cells (to be specific, CD4+ T cells), macrophages, and dendritic cells that are necessary to activate B-lymphocytes and induce the production of antibodies. The letter of HIV stands for Humanbeings, Immunodeficiency Virus weakens the immune system (and increases the risk of infection), and Virus that attacks the body respectively.

HIV infection leads to low levels of CD4+ T cells through three main mechanisms: direct viral killing of infected cells, increased rates of apoptosis in infected cells and killing of infected CD4+ T cells by CD8 cytotoxic lymphocytes that recognize infected cells. When CD4+ T cell numbers decline below a critical level, cell-mediated immunity is lost, and as a result the body becomes progressively more susceptible to opportunistic infections. The two known types of HIV are HIV-1 and HIV-2. HIV-1 is more virulent, more infective, and is the cause of the majority of HIV infections globally. The lower infectivity of HIV-2 compared to HIV-1 implies

[1] The CD4 T-cells are white blood cells that help coordinate the immune system's response to infection and disease.

that fewer of those exposed to HIV-2 will be infected per exposure. Because of its relatively poor capacity for transmission, HIV-2 is largely confined to West Africa.

Most untreated people infected with HIV-1 eventually develop AIDS. These individuals mostly die from opportunistic infections or malignancies associated with the progressive failure of the immune system. Based on the nucleotide acid sequence of complete viral genomes, HIV-1 is subsequently defined and classified into three groups: M (major), O (outlier), and N (non-M or O). The M group of HIV-1, which includes over 95% of the global virus isolates, consists of at least eight discrete clades A, B, C, D, F, G, H, and J. HIV progresses to AIDS at a variable rate affected by viral, host, and environmental factors; most will progress to AIDS within 10 years of HIV infection (or some will have progressed much sooner while some will take much longer).

The AIDS epidemic has now spanned nearly three decades. AIDS was first recognized by the U.S. Centers for Disease Control and Prevention in 1981 though its cause, HIV, was identified in the early 1980s. According to the UNAIDS (2007) report, it was estimated that 33.2 million people lived with the disease worldwide, and that AIDS killed an estimated 2.1 million people, including 330,000 children. More than three-quarters of these deaths occurred in sub-Saharan Africa. In 2008, an estimated 33.4 million people were living with HIV/AIDS worldwide; nearly 70% of these were found in sub-Saharan Africa (UNAIDS, 2009). Similarly, the first evidence of an HIV/AIDS epidemic in Ethiopia was detected in 1984. Since then, AIDS has claimed the lives of millions and has left behind hundreds of thousands of orphans. Ethiopia has an estimated 2 million people living with HIV and the third highest number of infections in Africa, according to UNAIDS (2007). Moreover, according to the Ethiopian Demographic and Health Survey (EDHS), the estimated adult HIV prevalence in 2005 was 1.4%. In response to the epidemic, the government of Ethiopia issued an HIV/AIDS policy in 1998. The following tables summarize the above mentioned facts.

Table A: Global Estimates of HIV/AIDS

Global Estimates (in millions)							
	2001	2002	2003	2004	2005	2006	2007
People living with HIV	29.0	30.0	30.9	31.6	32.1	32.7	33.2
New infections	3.2	3.1	3.0	2.9	2.8	2.7	2.5
Deaths	1.7	1.9	2.0	2.1	2.2	2.1	2.1

Source: UNAIDS (2007)

Table B: HIV/AIDS: A Comparison of Global and African Situations (2007)

	Global	Africa
People living with HIV/AIDS	33.2 millions	22.5 millions (68%)
New infections	2.5 millions (6,800/day)	1.7 millions (4,700/day)
Death from HIV/AIDS	2.1 millions (5,700/day)	1.6 millions (4,400/day)

Source: UNAIDS (2007)

Table C: Information on Ethiopia

Total population, 2007	73,918,505
People living with HIV/AIDS, 2007	980,000
Women (aged 15+) with HIV/AIDS, 2007	530,000
Children with HIV/AIDS, 2007	92,000
Adult HIV prevalence (%), 2007	2.1
AIDS deaths, 2007	67,000

Source: Population Reference Bureau and UNAIDS (July 2009)

HIV is transmitted through direct contact of the mucous membrane or the bloodstream with a bodily fluid containing HIV, such as blood, semen, vaginal fluid, preseminal fluid, and breast milk. This transmission can involve anal, vaginal or oral sex, blood transfusion, contaminated hypodermic needles, exchange between mother and baby during pregnancy, childbirth, breastfeeding or other exposure to one of the above bodily fluids. The most common mode of transmission of HIV in Ethiopia is through unprotected sex with an infected person. Thus, to prevent HIV/AIDS transmission, it is recommended that people practice safe sex through the much-advocated ABC method (abstinence, being faithful to one uninfected partner, and condom use). An HIV infected person may experience fever, weight loss, diarrhea, fatigue, skin rashes, shingles thrush, or memory problems.

Although treatments for AIDS and HIV can slow the course of the disease, so far there is no known cure or vaccine. The provision of antiretroviral treatment (ART) has decreased morbidity and mortality in people living with HIV/AIDS (PLWHA). Putting it in other words, treatment with anti-retroviral increases the life expectancy of people infected with HIV. Even after HIV has progressed to diagnosable AIDS, the average survival time with antiretroviral therapy is estimated to be more than 5 years. Without antiretroviral therapy, someone who has AIDS typically dies within a year. Due to the difficulty in treating HIV infection, preventing infection is a key aim in controlling the AIDS pandemic, with health organizations promoting safe sex and needle-exchange programmes in attempts to slow the spread of the virus.

Antiretroviral drugs can reduce mother-to-child transmission (MTCT) of HIV in one or more of the following ways:

1) By reducing viral replication and thus lowering plasma viral load in pregnant women;
2) Through pre-exposure prophylaxis of babies by crossing the placenta; and
3) Through post-exposure prophylaxis of babies after delivery.

In developed countries, highly active antiretroviral therapy (HAART) has reduced the vertical transmission rates to around 1- 2% but HAART is not yet widely available in low and middle income countries. In these countries, various simpler and less costly ARV regimens have been offered to pregnant women and/or their newborn babies. The use of antiretroviral medicines dramatically reduced AIDS related illnesses and death in countries where these drugs are widely accessible. HAART that is believed to improve the health of HIV/AIDS patients is a combination of antiretroviral therapy (ART) and three or more drugs.

In the case of Ethiopia, the national HIV adult prevalence for 2007 is estimated at 2.1%, of which 7.8% is urban and 1.0% rural. People living with HIV/AIDS number about 777,500, of which 242,500 are in need of ART. About 61% of the people in need of treatment live in urban areas, yet urban dwellers constitute an estimated 15.5% of the total population. Unsurprisingly, a large proportion of the 67,235 individuals who ever started the treatment are from urban areas. In 2006, about 2,434 children were on ART, falling short of the national target of 4,000. The total

number of children under 15 in need of ART in 2006 was 14,396 (or 5.9%) of all people in need, while children accounted for 3.6% of individuals on treatment. An estimated number of 13,970 HIV-positive births occurred in 2005, of which 6,827 were rural. It had been projected that 26,053 children will be in need of treatment in 2010, or 6.5% of all people in need.

In July 2003, the government adopted the policy of ARV drug supply and use, paving the way for additional initiatives that facilitate access to free and low-cost ARVs. The epidemic has rigorously affected different sectors of development and also challenged the severely constrained health care system of the country. According to Ministry of Health (MoH) report, about 40-60% of hospital beds were occupied by AIDS patients in 2001. In January 2005 the government launched the free ARV treatment initiative.[2]

HIV/AIDS affects society and economies at various levels, from the family and community to national and international levels - particularly by eroding the human capital. It is, for example, noted that particularly in Sub-Saharan Africa, HIV/AIDS continues to slow or even reverse improvements in life expectancy and distort the age-sex structure of the entire population (UNAIDS, 2006). The 2005 Human Development Report identifies AIDS as the factor inflicting the single greatest reversal in human development history (UNDP, 2005). The direct economic impact of HIV can be observed due to a reduction in labor force as a result of AIDS. Apart from such labor force reduction, the medication cost and the related opportunity costs and switching of expenditure to meet a higher medication cost also entail another adverse impact to the economy. This will, in turn, affect the saving and the steady state path of the economy. AIDS stigma exists around the world in a variety of ways, including ostracism, rejection, discrimination and avoidance of HIV infected people. In other words, HIV-related stigma refers to all unfavorable attitudes, beliefs, and policies directed toward people perceived to have HIV/AIDS as well as toward their significant others and loved ones, close associates, social groups, and communities.

Many factors can affect how quickly HIV infection progresses to AIDS. Factors such as age, co-infections (infections other than HIV), ethnicity, poverty, illiteracy, gender inequality, geographic location, genetics, infection route (how the disease was transmitted), nutrition,

[2] HAPCO, 2007. Accelerated access to HIV/AIDS prevention, care and treatment in Ethiopia.

pregnancy, stress, healthcare provider's experience in treating HIV patients and whether or not the patient smokes or uses recreational drugs can affect the rate at which an HIV patient develops AIDS.

Several human rights standards may be relevant to protect the rights of women specifically in the context of HIV - both in terms of prevention and response. For instance, the international convention on economic, social and cultural rights (ICESCR) requires state parties to "recognize the right of everyone to the enjoyment of the highest attainable standards of physical and mental health" (Article 12). In Ethiopia, women account for a larger share of those directly affected by HIV/AIDS. The national HIV prevalence in 2005 is estimated to have been 3% among males and 4% among females while 55% of the estimated 1.32 million PLWHA were females. Females also accounted for 54.5% of AIDS deaths and 53.2% of new infections in 2005. In the age group most affected by the pandemic, i.e., 15-29 years, there were more women living with HIV/AIDS than men as compared to less affected age groups. In fact, women and girls constitute the first group of people at high risk of infection (Fikremarkos Merso).

During the early stage of the HIV epidemic in Ethiopia, female commercial sex workers (CSWs), truck drivers and soldiers appeared to be among those first infected as HIV appeared to spread to towns along major roads. In 1988 and 1989, mean sero-positivity rates among more than 6,000 CSWs in 24 communities throughout Ethiopia were 18 and 29%, respectively, with rates in individual communities ranging from 1% to 38% and the highest rates in the war zone (Mehret *et al*., 1990). Similarly, the analysis by Fraser *et al.* (2008) identified the eight vulnerable[3] and most-at-risk sub-populations: military and other uniformed forces; long-distance truck drivers and other transport workers; fishermen and fisherwomen; female sex workers; refugees and internally displaced persons; prisoners; and females affected by sexual and gender based violence.

There is a claim that military personnel have a high risk of exposure to sexually transmitted diseases (STDs), including HIV/AIDS. In 1997, UNAIDS reported that sexually transmitted

[3] "Vulnerability factors" were defined as social and contextual factors describing the individual's condition in society (e.g. living in gender exclusive environment, low level of empowerment).

disease rates among armed forces are generally 2 to 5 times higher than in civilian populations; the difference can be even greater in times of conflict. By the same token, HIV/AIDS has been referred to as a pre-eminent security threat. The United Nations Security Council has met twice regarding regional security instability in Africa due to the AIDS crisis. Studies in the USA, the UK, and France showed that soldiers from these countries have a much higher risk of HIV infection than equivalent age/sex groups in the civilian population. Recent figures from Zimbabwe and Cameroon show military HIV infection rates 3 to 4 times higher than in the civilian population.

Based on the study in 21 countries of Sub-Saharan Africa (SSA), it was concluded that HIV/AIDS prevalence in military[4] (males) aged 15-24 is much higher than the general population of the same age-sex group (Oumar *et al.*, 2008). The investigation also outlined factors that contribute for HIV infection and its transmission as demographics and the circumstances of the deployment. The factors are a direct result of the nature of military populations (i.e., composed of young, male and sexually active individuals), elongated periods of departing from home and deployment to conflict zones. The study did not skip stating the challenge that persisted throughout militaries and suggesting more researches and campaigns to improve HIV/AIDS in SSA.

The military preparedness (effectiveness)[5] has suffered as a result of HIV/AIDS, which has caused illness and death among army personnel. There is a psychological effect, due to the anxiety and discrimination associated with HIV, both from inside the army and from the families of army personnel in their own communities. There are additional burdens for the medical and social services in the army, including the cost of campaigns to prevent new infections, counseling services for groups at risk and for those infected, care for people living with HIV/AIDS, and social support measures for them and their families. Military planners have long recognized the link between soldiers' well-being and military effectiveness, particularly the "health and fitness of troops before and during deployment" (Smith, 1992). "AIDS in the military, as well as in the national environment, is no longer an academic issue; it is a reality that

[4] Military populations consist of members of national armed forces, including regular army, navy, and air force contingents, militia and reserve units, and paramilitary/guerrilla groups.
[5] Military preparedness (effectiveness) is the process by which armed forces convert resources into fighting power.

has to be tackled with all the vigour and effort that is commensurate with its ramifications" (Major General Matshwenyego Fisher, the then Chief of Staff, Botswana Defence Force).

In 2000 the U.S. National Intelligence Council estimated the following HIV/AIDS military prevalence rates: Angola 40 to 60 percent, Eritrea 10 percent, DRC (Zaire) 40 to 60 percent, Nigeria 10 to 20 percent, Tanzania 15 to 30 percent, Cote d'Ivoire 10 to 20 percent, and Congo Brazzaville 10 to 25 percent. Other recent estimates include the following: Lesotho 40 percent, Malawi 50 percent, Zimbabwe 55 percent, Botswana 40 percent, Namibia 33 percent, Zambia 35 percent, and South Africa 21 percent (Sagala, 2008). Though HIV/AIDS is undoubtedly a threat to individual security, the argument/notion that it risks national defense, and further, international security, is unwarranted (Sato, 2008).

In the case of Ethiopia, in 1996, it has been discovered that the army had an infection rate of 6% (higher among senior officers); the then Chief of Staff, General Tsadkan Gebretinsae designated fighting AIDS as its number one priority. Six years later, while Ethiopia's national HIV prevalence has risen to over 7% and the national AIDS campaign has stagnated, the army still has a prevalence of just 6%.

Moreover, a study conducted by Kloos *et al* in 2007 used a cros-sectional national data to show the utilization of ART services in Ethiopia and found out a very high ART dropout rate of about 20.7% which has been considered as a challenge of the national scale up program. This has provided a direct remark that underlying patient retention mechanisms are not mature enough and could give clue for underlying survival outcome to be poor. Likewise, the extent of relevance of demographic, clinical and immunological correlates for survival in Ethiopia setting is not yet well described by previous studies. This paper aims to identify the major factors that affect the survival status of HIV-infected patients so that the results of the study can be utilized by policy makers of the sector.

1.2 Objectives of the study

The general objective of this study is to identify determinant factors (demographic and health related) that affect the survival of HIV-infected patients on ART in the Ethiopian Army Hospital (or AFGTH).

The specific objectives are:

- To identify predictors of survival for patients with AIDS.
- To assess the seven years survival pattern of patients on ART among the HIV/AIDS patients in the Armed Forces.
- To suggest the way that might support to scale up the life span of HIV-infected patients on ART.

1.3 Application of the result

- The result of this study provides information to government and other stake-holders in setting policies, strategies, and further investigation for reducing death of HIV-infected patients.
- This result will help both donors and government to appreciate factors that influence the survival status of HIV/AIDS patients and adjust their intervention programs accordingly.
- The finding of this study may enable clinicians to target services for patients at exceptionally high risk for early mortality and to provide patients with more accurate prognostic information.
- The results also contribute a lot for further study in the field.

1.4 Limitation of the study

- The study is conducted based on secondary data which might have incomplete and prejudiced information.
- The study was merely based on adults due to insignificant number of infants and children so that the results do not cover this portion of population.
- The study presumed that all deaths are caused by HIV/AIDS.
- High percentage of censored observation which might be due to lost to follow up.

- Moreover, the study is based on baseline values of the variables of interest such as CD4 cell count, weight, WHO clinical stage, functional status and regimen.

CHAPTER TWO

LITERATURE REVIEW

The main focus of this literature review is to bring to light relevant information from previous studies about the determinants of AIDS mortality, distribution of HIV/AIDS, socioeconomic and demographic status of different groups of individuals dying of AIDS.

Literature in relation to specific variables of the study (ART, CD4, WHO stage, TB, gender, opportunistic infections and socio-demographic characteristics)

The primary goals to initiate antiretroviral therapy (ART) are mainly to suppress plasma HIV viral load, to reduce HIV-associated morbidity and prolong survival, to improve quality of life, to restore and preserve immunologic function, and to prevent HIV transmission (Office of AIDS Research Advisory Council 2009). The investigation based on a prospective cohort of 1691 HIV seropositive women who enrolled between October 1994 and November 1995 in the United States, indicated that the use of antiretroviral therapies led to improved immunological function, suppressed HIV disease activity, and dramatic declines in morbidity and mortality.

There are major differences in the risk of infection faced by different population groups. That is to say, the distribution of HIV infection rates by age groups and gender can indicate transmission levels in the sexually active population and thus guide prevention programs. The Kenya Demographic and Health Survey (KDHS 2003) indicated that almost 9 percent of women are infected with HIV compared with 4.6 percent of their male counterparts. Moreover, the survey depicted those young women aged 15–24 were 5.5 times more likely to become infected with HIV than young men of the same age category. It also provided evidence that the current estimate of infection levels among urban residents is 8.3% while infection levels among rural residents is 4.0% (UNGASS 2008). By the same token, Fraser *et al.*, 2008 substantiated the

claim that prevalence is higher in urban than rural areas as well as indentifying vulnerable and most-at-risk sub-populations.

Similarly, in 2003, the highest HIV infection rates in Ethiopia reportedly occurred in the 15-34 age group (Kloos *et al.*, 2007). The study also indicated that the highest rates in female ANC attendees were in the 15-24 age group (8.6%). This furnishes the ground for an argument that proclaims vulnerability is greater in young females than males. Moreover, the South Africa's 2004 Reproductive Health Research Unit (RHRU) survey of 15 to 24 years old found that among aged 15–19, HIV prevalence was 2.5% and 7.3% for males and women, respectively (Whiteside, 2006). Moreover, evidence from Eritrea suggested that men aged 18-25 had lower HIV rates than women of the same age range (De Waal, 2005).

A case-control study that was carried out in Addis Ababa by Asefa *et al*. (2005) provided evidence that substance abuse, particularly alcohol, was found to be a significant risk factor for HIV infection. The study suggested the need for health education to bring about behavioral changes and further study to identify the prevalence and role of substance in exposure to HIV infection in the community.

Johnson & Dorrington (2006) used survival analysis to describe the impact of HIV/AIDS and the effects of HIV/AIDS prevention and treatment programmes. They found out that HAART is expected to have a significant impact on HIV prevalence, due to the improved survival prospects of infected individuals. In addition the study advocated the reduction of HIV prevalence in 2005 roughly by 2% due to prevention programmes. Similarly, a study conducted on 1691 HIV seropositive women in USA revealed that HAART improved immunological function, suppressed HIV disease activity, and reduced morbidity and mortality (Gange *et al*., 2002).

In Recsky *et al.* (2004), Pearson's X^2, the Cochran-Armitage and the Wilcoxon rank-sum tests have been used to determine the degree to which antiretroviral resistance may contribute to mortality among HIV-infected individuals enrolled in the centralized HIV/AIDS Drug Treatment Program in British Columbia, Canada, who had died between July 1997 and December 2001. In

the investigation, of a total of 637 deaths[6], 83 (13.0 %) were attributed to accidental causes; and the remaining 554 deaths (87.0%) were attributed to non-accidental causes. The accidental causes were illicit-drug overdose (57.8 %), concussion (18.1%) and the remaining percentage accounted to suicide, traffic accidents, assaults, and other in-juries. The non-accidental causes were identified as 383 (69.1%) directly related to HIV infection (JCD-9 categories 042-044 and ICD- 10 categories B20-B24) and 34 (6.1%) related to liver disease, 25 (4.5%) to various cardiac conditions, 20 (3.6 %) to viral and/ or bacterial infections, 18 (3.2%) to malignant neoplasms, 43 (7.8%) to other causes, and 31 (5.6%) to unknown causes. The study concluded that not only treatment failure due to antiretroviral resistance was a major factor influencing mortality in this cohort but co-morbidities, and other factors had got a lion share as well.

A longitudinal survey of HIV-positive patients treated with ART at Felege-Hiwot Hospital showed that the ART-naïve HIV patients were from low levels of education and with minimum monthly income (Bayeh *et al*., 2010). Further the study recommended the implementation of appropriate interventions in order to promote and enable HIV positive individuals to enter into ART programs as early as possible.

Volmink *et al.* (2007) used survival analysis to estimate the probability of infants infected with HIV (the observed proportion) at various specific time-points based on 14,398 participants[7] conducted in 16 countries. The intervention was any antiretroviral regimen with the specific aim of decreasing the risk of mother-to-child transmission of HIV infection. This review of trials confirms that antiretroviral treatment (ZDV with 3TC) administered in the perinatal period compared with placebo lowers the risk of mother-to-child transmission of HIV.

A study conducted in Cambodia between January 2003 and December 2007 on 670 HIV-positive children, used Kaplan-Meier analysis for survival and Cox proportional hazards model to identify risk factors associated with treatment failure (Isaakidis *et al.,* 2010). A result of the study revealed the survival probability at 24 and 36 months of ART initiation was 0.93 (95% CI: 0.91-0.95) and 0.91 (95% CI: 0.88-0.93), respectively. The study found that better survival,

[6] Causes of death were classified according to coding in the International Classification of Diseases, 9th Revision (1995-1999 [JCD-9]) and 10th Revision (2000-2001 [ICD-10]).
[7] Pregnant women with HIV infection or infants born to mothers with HIV infection.

immunological restoration and viral suppression can be sustained after two to three years of ART among children in resource-constrained settings. In addition, the study recommended the routine access to virological measurements so that a timely diagnosis of treatment failure can be achieved.

A prospective, cross-sectional, study was conducted on 100 patients between November 2001 and September 2002 in Phnom Penh, Cambodia, with the objective to assess the spectrum of HIV-associated complications and disease stage among individuals presenting for first-time care (Kong et al., 2007). The study identified the common AIDS-defining illness such as tuberculosis (43%), cryptosporidiosis (13%), severe bacterial infections (12%), cryptococcosis (12%), and Pneumocystis jiroveci pneumonia (10%). Increased screening for HIV and routine health maintenance (i.e., to facilitate management of opportunistic infections) for those infected were recommended by the study.

Between August 1998 and April 2002, 404 HIV-1-infected adult patients (or age greater than 15 years) were enrolled in the Senegalese antiretroviral drug access initiative in Dakar and included in an observational cohort study (Etard *et al.*, 2006). The study used the Kaplan-Meier method to estimate survival probabilities up to 60 months after starting HAART and Cox proportional hazards model with fixed covariates at baseline to estimate risk factors associated with death. The result of the study indicated the decrement in death rate after initiation of HAART. That is a cumulative probability of dying was 17.4% (95% CI, 13.9 - 21.5%) and 24.6% (95% CI, 20.4 - 29.4%) at 2 and 5 years of treatment, respectively. The study found tuberculosis, mycobacterial and neurological infections as a leading cause of death after initiation of HAART. Thus, an earlier initiation of HAART in the course of HIV infection was suggested as a better approach to diagnosis and management of tuberculosis and opportunistic infections in order to reduce high mortality. Similarly, a study by Laurent *et al.*, 2005, shared part in the claim that a long term survival improvement of HIV patients under HAART in Africa analogous to developed countries.

A study conducted for five years (between September 2002 and April 2007), in Nyanga, Cape Town, South Africa, on 2,196 patients under ART of whom 67% (n=1,479) were females

included demographic and clinical data, income, CD4 cell count and WHO staging as variables of interest (Cornell *et al.*, 2009). The study employed Wilcoxon rank sum test for medians, the chi-square test for proportions and Cox proportional hazards models to estimate the crude and adjusted impact of covariates on the hazard of death on ART and the hazard of loss to follow-up. A result of the study, in multivariable analysis, revealed the existence of a strong association of survival with age (HR = 1.05, 95% CI: 1.02 - 1.09; $p < 0.001$), CD4 cell count >150 cells/µl versus <50cells/µl (HR = 0.35, 95% CI: 0.14 - 0.87; $p = 0.023$) and any monthly income versus none (HR = 0.47, 95% CI: 0.25 - 0.88; $p = 0.018$). It provided evidence that men were older and had more advanced disease, with lower CD4 cell counts and higher viral loads which resulted in higher mortality rate among men than women. Thus, the study recommended to enroll men into care earlier in HIV disease and to reduce socio-economic inequalities in ART programme.

A study based on 4670 patients receiving ART in Malawi between June 2004 and December 2006 used Kaplan-Meier estimates to calculate gender differences in survival and Poisson regression to calculate death rate (Taylor-Smith *et al.,* 2010). A result of the study revealed that the death rate for men is nearly 2 times that of women (i.e., HR = 1.90; 95% CI: 1.57-2.29). Moreover, the result showed Kaposi's sarcoma (HR = 2.29; 95% CI: 1.7 - 2.96) and ART drop out (HR = 2.07; 95% CI: 1.67 - 2.56) as prognostic factors of mortality. Thus, the study recommended early access to ART in order to reduce mortality.

A study was conducted to determine rates, risk factors and causes of death among patients accessing a prospective community-based antiretroviral treatment (ART) cohort programme in Cape Town, South Africa (Lawn *et al*., 2005). The study found baseline CD4 cell count less than 50 cells/ml, WHO clinical stage IV, wasting syndrome, tuberculosis, acute bacterial infections, malignancy and immune reconstitution disease as major predictors of mortality. Furthermore, the study strongly recommended the early initiation of ART. Similarly, a study conducted in Western Cape Province, South Africa based on a cohort analysis of 12, 587 adults and 1,709 children used Kaplan–Meier estimates to describe the proportion of patients remaining in care (Boulle *et al.,* 2008). A result of the study indicated that the high early mortality is mediated by the extreme disease advancement (higher WHO clinical stage) and extremely low CD4 cell count at enrolment.

A study conducted on 119 adult HIV/AIDS patients for 8 years (from 1997 to 2005), at the Hospital das Clínicas, Brazil, was designed to find various AIDS related infections and death (Wolmer de Melo *et al.*, 2008). The study used the Kaplan-Meier method to construct survival curves and to calculate median survival time, and the log-rank test to compare the estimated survival curves. The investigation found that about 90% of deaths were due to AIDS related diseases. The most frequent were respiratory infections of unknown etiology (40%), tuberculosis (either pulmonary or extra-pulmonary) (18.5%), cerebral toxoplasmosis (15.5%), P*neumocystis jiroveci* pneumonia (7.7%), diarrhea (6.2%), Kaposi's sarcoma (4.5%) and non-Hodgkin's lymphoma (4.5%). By the same token, the study identified male gender, hemoglobin of less than 10 mg/dl (or a total lymphocyte count of less than 1,000/mm^3) and using antiretrovirals for less than one or six months as a major factors that associated with lower survival status of the patients.

In Seage et al. (1997), data for the study was collected from Boston hospital through personal interview and medical records. The study employed the Kaplan-Meier method to calculate median survival time and the Cox proportional hazards regression techniques to develop multivariable regression models. The result of the study indicated functional status and recent opportunistic diseases as the major predictors of survival time.

The World Health Organization (WHO) reported in 1999 that of a total 53.9 million deaths, 1.5 million deaths was caused by TB. In addition, it was claimed as TB co-infection is the leading cause of mortality among those infected with HIV worldwide. A finding of a cross-sectional study based on 241 cases reported from nine domestic hospitals throughout mainland China was in agreement with the stated claim. The patients in the study were followed from January 2003 to December 2005. In spite of the fact that treatments for TB and HIV were provided to the patients, mortality attributable to co-infection was reported for 15.8% of the cases. As a result, the study concluded that HIV/TB co-infection was related to high mortality even when HAART and/or drug therapy for TB was provided (Xueyan et al., 2008).

A study in North Carolina utilized surveillance data for all HIV seropositive TB on 543 cases that were reported between January 1, 1993 to December 31, 2003 from TB Control Program. TB cases that did not die prior to December 31, 2004 were right-censored as of that date. The study employed the Kaplan-Meier method to estimate long-term survival, log-rank test to compare categories within group in survival and the Cox Proportional Hazards method to compute hazard ratios. The result of the study showed that, in multivariable survival analysis, age greater than 45 years (HR: 1.44, 95% CI: 0.96 - 2.15), baseline CD4 cell count (HR: 0.76 per 100 cell increase, 95% CI: 0.65–0.88), and having HAART started during TB treatment (HR: 0.43, 95% CI: 0.26 - 0.72) were independently associated with long-term survival after TB diagnosis. The study recommended further research on interventions to enhance utilization of HIV-related health care and integration of TB and HIV services (Gadkowski et al., 2009). A prospective study in Guinea-Bissau, West Africa, between 1994 and 1997 on 280 hospitalized patients with pulmonary TB substantiated the fact that pulmonary TB co-infection escalate mortality of the patients under treatment (Norrgren et al., 2010).

In Delpierre et al. (2008), the study was conducted on the database contained data on 6805 patients diagnosed with HIV infection between 1 January 1996 and 1 July 2006. The study employed Kaplan-Meier curves to compare the survival time of categories and the Cox proportional hazards regression model to analyze factors associated with the use of HAART. The median follow up was 44 months (IQR: 17 - 78). A variable that was statistically significant at 0.10 level in bivariable analysis was included into a multivariable Cox proportional hazards regression model. The result of the study showed, in multivariable analysis, that death rate was higher for patients without employment (HR = 3.75, 95%CI: 2.11 - 6.66), and for patients diagnosed late (HR = 9.18, 95% CI: 4.32 - 19.48).

A retrospective study of 790 HIV-infected patients that employed between 16 May 1985 and 31 December 2001 was conducted in Singapore. The study used Kaplan-Meier method to construct survival curves and the Cox proportional hazards model to determine independent predictors of disease progression. Both univariable and multivariable analyses showed that patients of younger age and higher baseline CD4 cell count associated with a lower risk of progression to AIDS (Chow *et al*., 2005).

A study conducted in Sub-Saharan Africa based on data from 18 published cohort studies containing 39,536 HIV/AIDS patients had employed the Kaplan-Meir method to assess the proportion of survival time and random-effects model to find hazard ratio of prognostic variables (Lawn et al., 2008). Thus, a result of the study suggested advanced WHO clinical stage and low CD4 cell count as indicator of high mortality. Similarly, a study in Malawi based on 1308 patients employed Kaplan-Meier method to assess the probability of survival and the Cox proportional hazards model to assess the potential predictors of death. The study found low body-mass index, WHO clinical stage IV, male gender, and baseline CD4 count lower than 50 cells/ml as independent determinants of death (Ferradini et al., 2006).

There was a study that was conducted in south Ethiopia between August 2003 and August 2005 on two cohorts of patients: the pre-HAART cohort (185 patients) and the HAART (180 patients) cohort. In the study, the Kaplan-Meier method was used to assess the event-free survival, the log-rank test was employed to test for the statistical significance thereof, and the Cox proportional hazards model was used to find out the effect of HAART on mortality and on tuberculosis incidence rates. Thus, the study indicated that the HAART improved survival and decreased tuberculosis incidence. Furthermore, the study recommended the importance of strengthening tuberculosis prevention efforts with the scale-up of treatment programmes (Jerene et al., 2006[a]).

A study on 272 HIV/AIDS patients on ART in Shashemene and Assela Hospitals employed Kaplan Meier method to construct survival curves and the Cox proportional hazards model to determine predictors of mortality (Andinet and Sebastian, 2010). The median survival time of the study was 104.4 weeks. The findings of the study showed WHO clinical stage IV, hemoglobin 510 g/dL, and cotrimoxazole prophylaxis therapy (CPT) initiation as the independent determinants of mortality. By the same token, Jerene *et al.* (2006[b]) based on 162 patients, who were enrolled and treated between August 2003 and January 2005, ascertained that advanced disease stage (WHO clinical stage IV) and having total lymphocyte count (TLC) of up to 750/mcL were the major prognostic factors of mortality. The study also recommended identifying and treating patients early through improved counseling and testing strategies.

CHPTER THREE

DATA AND METHODOLOGY

3.1 Data

This study is a retrospective cohort study[8] based on data from the ART clinic in the Armed Forces General Teaching Hospital (AFGTH), Addis Ababa, Ethiopia. The hospital serves as a tertiary level teaching and referral hospital delivering health services to the defense forces, civilians in the Ministry of National Defense (MoND) and their dependants, as well as public patients referred by other specialized hospitals. AFGTH serves as a teaching center in parallel with Defense University Health Sciences College in various fields of specialization such as Medical Doctors (MD) and Masters level in Surgery. It also provides a training service for practitioners/students (i.e., for apprenticeship) from various health institutions including Defense Health Sciences College itself.

The ART clinic has its own separate facilities such as pharmacy, dermatology, ART laboratory, etc. It gives both adherence and testing counseling, and provides ART to HIV patients as well as delivering screening, follow up and referral services for TB patients.

Patients are eligible for ART on the basis of the 2002 World Health Organization (WHO) guidelines (WHO stage IV disease or CD4 cell count below 200 cells/μL). All patients older than 15 years (i.e., both adolescents and adults) who never received any antiretroviral therapy, and started ART between September 2003 and August 2007 at the ART clinic of AFGTH were eligible for this retrospective study.

This study is based on a review of the patients' intake forms and follow-up cards of HIV patients on ART at AFGTH ART clinic. The patient's forms are designed by Federal Ministry of Health

[8] A retrospective cohort study, also called a historic cohort study, is a medical research study in which the medical records of groups of individuals who are alike in many ways but differ by a certain characteristic (for example, female nurses who smoke and those who do not smoke) are compared for a particular outcome (such as lung cancer).

(FMOH) for uniformity for use in the country so that those forms can be used to document almost all relevant clinical and laboratory variables. A total of 734 patients in the clinic who started ART between September 2003 and August 2007 were included for the study. The patients were followed up until August 2010. The data were collected by a statistician and public health officer and coded, cleaned and analyzed by using SPSS, SAS, and STATA.

3.2 Variables in the study

3.2.1 The response Variable:

The response (dependant) variable is the survival time of HIV patients, the length of time from ART start date until the date of death (or censor) measured in months.

3.2.2. Predictor (independent) Variables:

The following predictor (covariate) variables are considered for the study.

- Age in years
- Gender (Male, Female)
- Marital status (Never Married, Married, Others)
- Level of Education (No education, Primary, Secondary and above)
- Religion (Muslim, Coptic orthodox, Others)
- Employment status (Employed, Not employed)
- Opportunistic infections (no, yes)
- TB co-infection (no, yes)
- Household size (one, two or more)
- Number of rooms (one, two or more)
- Risk behavior (Regular sexual partner, Casual or both regular and casual sexual partner)
- Substance (Tobacco, Alcohol, Soft drugs) use (no, yes)
- Functional status (Working, Ambulatory and Bedridden)
- WHO clinical stage (Stage I/ II, Stage III and Stage IV)
- CD4 cells count
- Body weight(kg)
- Regimen (D4T-based, AZT-based)

Note: All predictor variables were taken at baseline values.

3.3 Methodology

The Cox proportional-hazards regression model (introduced in a seminar paper by Cox, 1972), is a broadly applicable and the most widely used method in survival analysis. It is used to delineate the role of factors called covariates that influence the duration of survival. The Kaplan–Meier estimator (named after Edward L. Kaplan and Paul Meier), also known as the product limit estimator, estimates the survival function from life-time data. The Kaplan-Meier (K-M) method of survival analysis is used when the subjects are continuously observed and the exact duration of reaching to the end-point or at the time of dropout is known.

The term "survival analysis" pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study. In other words, survival analysis is an important statistical technique used to describe and model time–to–event data. The purpose of survival analysis is to model the underlying distribution of the failure time variable and to assess the dependence of the failure time variable on covariates.

The term survival analysis suggests that the event is death, but that is not necessarily so. Events could also denote success, such as recovery from therapy. Survival time then describes the time from a certain origin to the occurrence of an event. Time–to–event data can be found in many disciplines, for example in medicine:
– the time to death for patients having a certain disease (this explains the term survival analysis),
– the time to relapse of or cure from a certain disease, and
– the time to death of HIV patients after retroviral therapy.

Survival data are different from other types of continuous data because over the period of study the endpoint of interest is not necessarily observed in all subjects. This may occur because:
(a) Some patients are lost to follow-up, that is, they are not followed to the end of the study and, when last seen, have not experienced the event of interest, or
(b) The event has not occurred in some patients by the time the study ends for analysis. Such data are referred to as censored survival times and are different from missing data in that they

provide a lower bound for the actual non-observed survival times. Any analysis carried out on survival data should use statistical methods that do not disregard censored data and, indeed, make the fullest possible use of it to avoid loss of information.

There are three common forms of censoring:

a. Right Censoring: The most common form of incomplete data is right censoring. A survival time is said to be right censored if it is recorded from its beginning until a well defined time before its end time. It means a subject's follow-up time terminates before the outcome of interest is observed. For instance, if an HIV-1 patient is followed until he has a viral load high than 1000 copies/μl and is followed without experiencing this scenario until the end of the observation period. In other words, a survival time is said to be right censored if it begins at time $t = 0$ and terminates before the outcome of interest is observed.
b. Left Censoring: A survival time is said to be left censored if an individual developed the event of interest prior to the beginning of the study. This situation is less common in survival studies and is often not a focus.
c. Interval Censoring: A survival time is categorized as interval censored if it is only known that the event of interest occurs within an interval of time without the knowledge of when exactly it occurs. Interval censoring occurs in clinical trials where patients have periodic follow-ups and in industrial experiments where equipment items are inspected periodically, etc.

3.3.1 Descriptive methods for survival data

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

Survivor function

Survival data are not amenable to standard statistical procedures used in data analysis due to censoring.

The survival time in days, weeks or months, whichever is the most appropriate, can then be calculated. The survivor function and hazard function are the two functions of central interest in summarizing survival data. The actual survival time of an individual, *t*, can be regarded as the value of a random variable *T*, which can take any non-negative value. The different values that *T* can take have a probability distribution, and we call *T* the random variable associated with the survival time. When the random variable *T* has a probability distribution with underlying probability density function f (*t*), the cumulative distribution function (cdf) of *T*, denoted F_T (*t*), is then given by

$$F_T(t) = P_T(T \leq t) = \int_0^t f(u)du, \; t > 0 \qquad (1)$$

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

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

$$S(t) = 1 - F(t) \qquad (2)$$

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

Median Survival Time

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

$$t_{med} = S^{-1}(0.5). \qquad (3)$$

Hazard function

The term has different meanings in different field of studies: it is known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic processes, the age-specific failure rate in epidemiology and the inverse of the Mill's ratio (the hazard rate) in economics.

The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization) in an interval after time *t*, conditional on the subject having survived to time *t*.

It is the probability that an individual dies somewhere between t and $t + \Delta t$, divided by the probability that the individual survived beyond time *t*. The hazard function seems to be more intuitive to use in survival analysis than the pdf because it quantifies the instantaneous risk that an event will take place at time t given that the subject survived to time *t*.

The hazard function h (*t*) can be formulated as:

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{p\{t \le T \le t + \Delta t / T \ge t\}}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{p(t \le T \le t + \Delta t, T \ge t) \big/ p(T \ge t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{p(t \le T \le t + \Delta t)}{\Delta t} \times \frac{1}{p(T \ge t)} = f(t) \times \frac{1}{S(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \tag{4}$$

Let H (.) be the cumulative hazard function

$$H(t) = \int_0^t h(u)du \tag{5}$$

Similarly, the survival function can be given in terms of the hazard function as:

$$\begin{aligned} S(t) &= e^{-\int_0^t h(u)du} = e^{-H(t)} \\ \Rightarrow h(t) &= \frac{-d \log s(t)}{dt} = \frac{dH(t)}{dt} \end{aligned} \tag{6}$$

Estimation of the survivor function

Among the other estimators of the survivor function the Kaplan-Meier estimator is the most common one. 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. This method is non-parametric or distribution-free, since it does not

require specific assumptions to be made about the underlying distribution of the survival times (Hosmer and Lemeshow, 1999).

Assume we have a sample of n independent observations, their survival times denoted by $t_1, t_2, ..., t_n$ and indicator of censoring by $\delta_1, \delta_2, ..., \delta_n$ where δ_i = 1, if an event/death occurs and δ_i = 0 otherwise. Thus, the survival data are denoted by (t_i , δ_i), i=1,2,…,n. The first step to obtain the Kaplan-Meier estimator of the survival function is to order the survival times as $t_1, t_2, ..., t_n$. Assume that among the n observations $m \leq n$ failures occurred at distinct m times. Then the rank- ordered failure times $t_1, t_2, ..., t_m$.

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

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

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

$$\hat{S}_{KM}(t) = \prod_{t_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i}\right) = \prod_{t_{(i)} \leq t} \hat{p}_i$$ with the convention that $\hat{S}_{KM}(t) = 1$ if $t < t_{(1)}$. Thus,

$$\hat{H}_{KM} = -\sum_{t_{(i)} \leq t} \hat{p}_i \qquad (7)$$

The variance of the Kaplan-Meier estimators which is referred to as Greenwood's formula is given as:

$$\hat{\text{V}}\text{ar}(\hat{S}_{KM}(t)) = [\hat{S}_{KM}(t)]^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i (n_i - d_i)} \qquad (8)$$

Similarly, Aalen (1975, 1978), Nelson (1969, 1972) and Altshuler (1970) have proposed an alternative estimator of H (t) that refers to as the Nelson-Aalen estimator. It is formulated as:

$$\tilde{H}(t) = \sum_{t_{(i)} \leq t} \frac{d_i}{n_i} \text{ and it implies } \tilde{S}(t) = e^{-\tilde{H}(t)} = \prod_{t_{(i)} \leq t} \exp\left(-\frac{d_i}{n_i}\right) \qquad (9)$$

It is merely in the case of small samples that the Nelson-Aalen estimate of the survivor function prevails over the KM estimate (Hosmer and Lemeshow, 1999).

Comparison of survival curves

In clinical research one is concerned not only with estimating the survival function but, more often, with the comparison of the life experience of two or more groups of subjects differing for a given characteristic or randomly allocated to different treatments. After providing a description of the overall survival experience in the study, we usually turn our attention to a comparison of the survivorship experience in key subjects in the data. The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for these groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due to chance. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests. Since survival data are typically right skewed, we would likely use rank-based non-parametric tests followed by estimates and confidence intervals of the medians or other quantiles within groups. Modifications of these procedures are required when censored observations are present in the data. When we compare groups of subjects, it is always good to begin with a graphical display of the data in each group. Among the various non-parametric tests one can find in the statistical literature, the Mantel-Haenzel (1959) test, currently called the "log-rank" test will be used. Nowadays, the Kaplan-Meier method for estimating survival curves and the log-rank test for comparing two estimated survival curves are the most frequently used statistical tools in medical reports on survival data (Hosmer and Lemeshow, 1999).

Log-rank test

The log rank test, developed by Mantel and Haenszel, is a non-parametric test for comparing two or more independent survival curves. Since it is a non-parametric test, no assumption about the distributional form of the data is required. This test is however most powerful when used for non-overlapping survival curves. This test can be generalized to accommodate other tests that are equally used sometime in practice such as Generalized Wilcoxon test, Tarone-Ware test, and Peto-Peto-Prentice test. Each of these tests uses different weights to adjust for censoring that is often encountered in survival data. For instance, the Wilcoxon test weights the j^{th} failure time by n_j (the number still at risk), the Tarone–Ware test weights the j^{th} failure time by $\sqrt{n_j}$ and the Peto-Peto-Prentice test weights the j^{th} failure time by the survival estimate $\tilde{S}(t_j)$ calculated over all

groups combined (Kleinbaum and Klein, 2005 and Hosmer and Lemeshow, 1999). The log rank test statistic for comparing two groups is given by:

$$Q = \frac{\left[\sum_{i=1}^{m} w_i (d_{1i} - \hat{e}_{1i})\right]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \qquad (10)$$

where:

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

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

$\hat{e}_{1i} = \frac{n_{1i} - d_i}{n_i}$ is the expected no of events (death) corresponding to d_{1i}.

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

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

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

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

Under the null hypothesis that two survival functions are equal, the log rank test statistic *Q* has an approximation of chi-square distribution with one degree of freedom $(\chi_{(1)}^{\ 2})$ for large samples. The null hypothesis of equality of survival functions will be rejected for large values of Q. The most frequently used test is based on weights equal to one, w_i =1.

Note that the log-rank test can be extended for comparing three or more groups of survival experience.

3.3.2 Regression Models for Survival Data

In most medical studies which give rise to survival data, supplementary information is collected on each individual so that the relationship between the survival experience of individuals and

various explanatory variables may be investigated (Hosmer and Lemeshow, 1999). A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches. Semi-parametric models are models that parametrically specify the functional relationship between the lifetime of an individual and his/her characteristics (demographic, socio-economic, etc.) but leave the actual distribution of lifetimes arbitrary. The most popular of the semi-parametric models is the proportional hazards model. It has the property that the ratio of the hazards depends on the values of their explanatory variables but does not depend on time *t*. A hazard model is a regression model in which the "risk" of experiencing an event (death in our case) at a certain time point is predicted with a set of covariates.

Proportional Hazards Model

For each individual *i* we observe a vector of covariates x_i with values fixed across time. The hazard at time *t* is now a function of *t* and x_i , which we denote by $h(t, x_i)$. The hazard at $x_i = 0$ (i.e., $h(t, x_i = 0)$) is denoted by $h_0(t)$. If all covariates are categorical, $h_0(t)$ is the hazard for individuals in the reference (baseline) category of each variable. For this reason $h_0(t)$ is often referred to as the *baseline hazard*. A proportional hazards (PH) model is written as

$$h(t, x_i) = h_0(t)g(x_i), \tag{11}$$

where $g(x_i)$ is some function of the covariates (Hosmer and Lemeshow, 1999).

If the values of the covariates are changed from their reference categories (or, more generally, from zero) to a value **x***, then the hazard is multiplied by g(**x***). Therefore, the covariates are assumed to have a multiplicative effect on the hazard. The PH assumption implies that the effect of a change in **x** on the hazard is the same for all values of *t*. To see this, consider the hazard functions for two different sets of covariate values, x_1 and x_2. From (11) the ratio of the hazards at these two values of **x** is

$\frac{h(t, x_1)}{h(t, x_2)} = \frac{g(x_1)}{g(x_2)}$, which does not depend on t.

The Cox Proportional Hazards Model

We can specify the density function of a parametric distribution or we can specify the hazard function. The advantage of the latter approach is that we directly address the aging process, but as shown previously, it does not easily lend itself to the use of scatter plots to motivate regression models. The latter approach may also be preferred in a setting where the end products of the statistical analysis are estimated parameters that compare the survival experience of the selected subgroups. By specifying a model through the hazard function, we may address specific questions such as how survival is related to the subject's characteristics or the covariates.

Cox (1972) took a different approach to standard parametric survival analysis and extended the methods of the non-parametric Kaplan-Meier estimates to regression type arguments for life-table analyses. Cox advanced to prediction of survival time in individual subjects by only utilizing variables co-varying with survival and ignoring the baseline hazard of individuals. He did this by making no assumptions about the baseline hazard of individuals and only assumed that the hazard functions of different individuals remained proportional and constant over time. When there are several explanatory variables, and in particular when some of these are continuous, it is much more useful to use a regression method such as Cox rather than a KM approach (Hosmer and Lemeshow, 1999).

Cox introduced the semi-parametric proportional hazards model to cater for covariate effects for single event failures. This model is valid under the assumption of proportional hazards. Cox (1972) observed that if proportional hazards assumption holds (or is assumed to be hold), then it is possible to estimate the effect parameter(s) without any consideration of the hazard function. There are several features of the Cox proportional hazards model for which it was chosen to explain the effect of covariates on time until event.

Features of the Cox proportional hazards model

An important feature of the Cox proportional hazards model, which concerns the proportional hazards assumption, is that the baseline hazard is a function of *t*, but does not involve the **x**'s. In contrast, the exponential expression, involves the **x**'s, but does not involve *t*. The **x**'s, here, are assumed to be time-independent. Moreover, the baseline hazard function $h_0(t)$ is left completely

unspecified thus the Cox model is a semi-parametric model. The other assumption of the proportional hazards refers to the fact that the effects of covariates are the same for all values of *t*. Putting it in other words, the Cox proportional hazards model assumes that changes in the hazard of any subject over time will always be proportional to changes in the hazard of any other subject and to changes in the underlying hazard over time (Kleinbaum and Klein, 2005).

A key reason for the popularity of the Cox model is that, even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations. Saying this in another way is that the Cox PH model is a "robust" model, so that the results from using the Cox model will closely approximate the results for the correct parametric model (Kleinbaum and Klein, 2005).

The other point is the simple interpretation given by the Cox model as "relative risk" type ratio. For instance, when a covariate is dichotomous, say gender, with a value of x_1=1 for males and x_0 = 0 for females, the hazard ratio becomes e^{β}. If the value of the coefficient is $\beta = \ln(2)$ (or e^{β} = 2), then it is simply saying that males are *dying* at twice the rate of females. If β = 0 (or $e^{\beta} = 1$), then **x** has no effect on the hazard. $e^{\beta} > 1$ implies that the group with **x** =1 has a higher hazard, or shorter event time, than the group with **x** = 0. By the same token, $e^{\beta} < 1$ implies that the hazard is higher among the group with **x** = 0 (Hosmer and Lemeshow, 1999).

The hazard function

The most commonly applied survival analysis model is the Cox proportional hazards model. In the Cox model, the function $g(x) = \exp(\beta' x_i)$ so that (11) becomes

$$h(t, x_i, \beta) = h_0(t) \exp\left(\beta' x_i\right) \qquad (12)$$

where $h_0(t)$ is the baseline hazard function at time *t*, $x_i' = (x_{1i}, x_{2i}, ..., x_{pi})$ for i= 1,2,…,*n* is a vector of measured covariates for the i[th] individual at time *t*, and β' is a *p* vector of unknown regression parameters that are assumed to be the same for all individuals in the study. They measure the influence of the covariate on the survival experience with β_i representing the

increase in the log hazards as x_i increases one unit relative to the baseline hazard function. This model is referred to in the literature by a variety of terms, such as the Cox model or the Cox proportional hazards model or simply the proportional hazards model. The hazard function in equation (12) depends on both time and the associated covariates, but through two separate factors: the first is a function of time only which is left arbitrary, but is assumed to be the same for all the subjects, the second is a quantity which depends on the individual covariates (Hosmer and Lemeshow, 1999).

From the representation in equation (12) one can notice a couple of features. First, if the vector of covariate is a zero vector, then the hazard function for the ith individual is the baseline hazard function. It is the hazard function in the absence of covariates or when all of the coefficients of the covariates are assumed to be zero. Second, if we divide both sides by $h_0(t)$, we get equation (13) below which indicates where the term proportional comes from. Since for each individual, $e^{x_i' \beta}$ is constant across time, equation (15) below shows that at every value of t, the ith individual's log hazard ratio is constant. Loosely speaking, this implies that each individual's hazard function is "parallel" to the $h_0(t)$.

$$\frac{h_i(t, x_i)}{h_0(t,0)} = \frac{h_0(t)\exp(\beta' x_i)}{h_0(t)} = e^{\beta' x_i} \tag{13}$$

The Cox model is often called proportional hazards model because, if we look at two independent subjects with covariate values $\mathbf{x}_1$ and $\mathbf{x}_2$, the ratio of their hazard functions at time *t* is:

$$\frac{h(t,x_1)}{h(t,x_2)} = \frac{h_0(t)\exp(\beta' x_1)}{h_0(t)\exp(\beta' x_2)} = \exp[\beta'(x_1 - x_2)] \tag{14}$$

which is constant and does not vary over time, that is, the ratio does not depend on *t* and the hazard rates are proportional. The Cox proportional hazards model can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$\log\{\frac{h(t,x)}{h_0(t)}\} = \beta' x \tag{15}$$

Note that the cumulative hazard function is given by:

$$H(t) = H_0(t)\exp(\beta' x) \tag{16}$$

Consequently, from the proportional hazard function, we obtain the survivor function given by:

$S(t,x,\beta) = [S_0(t)]^{\exp(\beta' x)}$ where, $S_0(t)$ is the baseline survival function (Hosmer and Lemeshow, 1999).

Fitting the Cox Proportional Hazards Model

The data in survival analysis based on the sample size *n* are denoted by the triplet (t_i, δ_i, x_i), $i = 1,2,...,n$ where t_i is the time at which the i[th] individual dies from the disease of interest, δ_i is the event indicator $\delta_i = 1$ if the event has occurred and $\delta_i = 0$ if it is censored (the lifetime may be right, left or interval censored), and x_i is the vector of covariates or the risk factors for the i[th] individual.

The Cox model will be fitted by estimating the unknown regression coefficients through the maximum likelihood method. The actual likelihood function is constructed by considering the contribution of the probability that a subject with covariate value **x** dies from the disease of interest at time *t* (i.e., $f(t, \beta, \mathbf{x})$), and the probability that a subject with covariate value **x** survives at least *t* time units (i.e., $S(t, \beta, \mathbf{x})$). That is, under the assumption of independent observations, the full likelihood function is obtained by multiplying the respective contributions of the observed triplets, a value of $f(t, \beta, \mathbf{x})$ for a noncensored observation and a value of $S(t, \beta, \mathbf{x})$ for censored observations.

Thus, the contribution of each triplet to the likelihood is the expression

$$[f(t, \beta, \mathbf{x})]^{\delta_i} \times [S(t, \beta, \mathbf{x})]^{1-\delta_i} \tag{17}$$

Since the observations are assumed to be independent, the likelihood function is the product of the expression in (17) over the entire sample and is formulated as:

$$l(\beta) = \prod_{i=1}^{n} \left\{ [f(t_i, x_i, \beta)]^{\delta_i} \times [S(t_i, \mathrm{x}_i, \beta)]^{1-\delta_i} \right\} \tag{18}$$

It can be further simplified as

$$\begin{aligned} l(\beta) &= \prod_{i=1}^{n} \left\{ [h(t_i, \mathrm{X}_i, \beta) \times (S(t_i, \mathrm{X}_i, \beta)]^{\delta_i} \times [S(t_i, \mathrm{X}_i, \beta)]^{1-\delta_i} \right\} \\ &= \prod_{i=1}^{n} \left\{ [h(t_i, \mathrm{X}_i, \beta)]^{\delta_i} \times [S(t_i, \mathrm{X}_i, \beta)] \right\} \end{aligned} \tag{19}$$

Cox (1972) proposed using an expression he called a partial likelihood function due to the fact that the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who are censored. In other words, the likelihood for the Cox model does not consider probabilities for all subjects. Let us consider a sample of *n* subjects and suppose a total of *m* failures occur, with *m* smaller than *n*, due to the presence of censoring. Let $t_1 < t_2 < \ldots < t_m$ be the *m* distinct ordered failure times observed and let $R(t_i)$ be the set of individuals at i^{th} failure time, which consists of all subjects with survival or censored times greater than or equal to the specified time (Hosmer and Lemeshow, 1999).
Thus the partial likelihood is given by the expression:

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{x_{(i)}\beta}}{\sum_{j \in R(t_{(i)})} e^{x_j \beta}} \right]^{\delta_i} \tag{20}$$

The expression assumes that there are no tied times, and designed in such a way that it excluded terms when $\delta_i = 0$. As a result the equation in (20) becomes,

$$l_p(\beta) = \prod_{i=1}^{m} \frac{e^{x_{(i)}\beta}}{\sum_{j \in R(t_{(i)})} e^{x_j \beta}} \tag{21}$$

To obtain the maximized likelihood with respect to the parameters of interest, β, we maximize the log partial likelihood function as

$$L_p(\beta) = \sum_{i=1}^{m} \left\{ x_{(i)}\beta - \ln\left[\sum_{j \in R(t_{(i)})} e^{x_j \beta} \right] \right\} \tag{22}$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of (22) with respect to β, setting the derivatives equal to zero and solving for the unknown parameters. This is known as the Newton-Raphson iterative method.

That is, for each derivative

$$U(\beta) = \frac{\partial L_p(\beta)}{\partial \beta} = \sum_{i=1}^{m} \left\{ x_{(i)} - \frac{\sum_{j \in R(t_{(i)})} x_j e^{x_j \beta}}{\sum_{j \in R(t_{(i)})} e^{x_j \beta}} \right\} = \sum_{i=1}^{m} \left\{ x_{(i)} - \sum_{j \in R(t_{(i)})} w_{ij}(\beta) x_j \right\} = \sum_{1}^{m} \left\{ x_{(i)} - \bar{x}_{w_i} \right\} = 0 \tag{23}$$

where $w_{ij}(\beta) = \dfrac{e^{x_j \beta}}{\sum_{l \in R(t_{(i)})} e^{x_l \beta}}$ and $\bar{x}_{w_i} = \sum_{j \in R(t_{(i)})} w_{ij}(\beta) x_j$

U (β) is called the score or gradient vector. The solution to the equation (23) is denoted by $\hat{\beta}$.

The estimator of the variance of the estimator of the coefficient is obtained in the same manner as variance estimators are obtained in most maximum likelihood estimation applications. The estimator is the inverse of the negative of the second derivative of the log partial likelihood at the value of the estimator. Derivation of the expression in (23), will result in

$$\frac{\partial^2 L_p(\beta)}{\partial \beta^2} = -\sum_{i=1}^{m} \left\{ \frac{\left[\sum_{j \in R(t_{(i)})} e^{x_j \beta} \right] \left[\sum_{j \in R(t_{(i)})} x_j{}^2 e^{x_j \beta} \right] - \left[\sum_{j \in R(t_{(i)})} x_j e^{x_j \beta} \right]^2}{\left[\sum_{j \in R(t_{(i)})} e^{x_j \beta} \right]^2} \right\} \tag{24}$$

The expression in (24) shall be simplified using $w_{ij}(\beta)$ in equation (23) above.

That is,

$$\frac{\partial^2 L_p(\beta)}{\partial \beta^2} = -\sum_{i=1}^{m} \sum_{j \in R(t_{(i)})} w_{ij} (x_j - \bar{x}_{w_i})^2 . \tag{25}$$

The negative of the 2nd derivative of the log partial likelihood in either (24) or (25) is known as the observed information and denoted by

$$I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta^2} \tag{26}$$

If we consider models that contain more than one covariate, the result in (26) becomes

$I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta \partial \beta'}$ which is known as the observed information matrix (Hessian matrix).

According to the Newton-Raphson procedure an estimate of β at the $(j+1)^{th}$ of the iterative procedure, $\hat{\beta}_{j+1}$, is $\hat{\beta}_{j+1} = \hat{\beta}_j + I^{-1}(\hat{\beta}_j)U(\beta_j)$, j = 0, 1, 2, ...

As a result, the estimator of the variance of the estimated coefficient is the inverse of (26) evaluated at $\hat{\beta}$ and is

$$\hat{Var}(\hat{\beta}) = I(\hat{\beta})^{-1} \tag{27}$$

After fitting the regression model, we go for assessing the significance of the coefficient and the construction of the confidence interval as well. The three different tests used to assess the significance of the coefficient are explained below (Hosmer and Lemeshow, 1999).

i) The partial likelihood ratio test

The partial likelihood ratio test, G, is computed as twice the difference between the log partial likelihood of the model containing covariates and the log partial likelihood of the model not containing the covariates. Mathematically,

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\}, \tag{28}$$

where $$L_p(0) = -\sum_{i=1}^{m} \ln(n_i) \tag{29}$$

and n_i denotes the number of subjects in the risk set at observed survival time $t_{(i)}$.

Under the null hypothesis, that the coefficients are equal to zero, that is $H_0 : \beta = \bar{0} = (0,0,...,0)'$, the statistic, G, follows a chi-square distribution with 1 degree-of-freedom for "sufficiently" large sample size.

ii) The Wald test

The Wald statistic ($z = \frac{\hat{\beta}}{\hat{SE}(\hat{\beta})}$) is defined as the ratio of the estimated coefficient to its estimated standard error. Under the null that is a single parameter $\beta_i = 0$, $H_0 : \beta_i = 0$, the Wald

statistic follows a standard normal distribution (i.e., $z \sim N(0,1)$). Obviously, the square of Wald statistic follows a chi-square distribution with 1 degree-of-freedom.

Similarly, a 100(1-α) % Wald-statistic-based confidence interval for β will be $\hat{\beta} \mp z_{\alpha/2} \hat{SE}(\hat{\beta})$, $Z_{\alpha/2}$ is the upper α/2 percentile point of the standard normal distribution.

iii) The score test

The score test which is obtained by computing the ratio of the derivative of the log partial likelihood in (23) to the square root of the observed information in (26), all evaluated at β=0. Thus, the equation will be:

$$z^* = \left. \frac{\partial L_p / \partial \beta}{\sqrt{I(\beta)}} \right|_{\beta=0} . \qquad (30)$$

Under the null hypothesis that each parameter is equal to zero, $H_0 : \beta_i = 0$, this statistic follows a standard normal distribution. The score test may be reported as the squared by some statistical packages that will follow a chi-square distribution with 1 degree-of-freedom under the null hypothesis.

When there is a disagreement among the three tests of the significance of the coefficient, the partial likelihood ratio test will prevail.

3.4 Model Development

Model development in proportional hazards regression analysis requires critical 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 of selecting a subset of covariates in a proportional hazards regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections. Survival analysis using Cox regression method begins with a thorough univariable analysis of the association between survival time and all important covariates (Hosmer and Lemeshow, 1999).

Recommendable procedure in selecting variables in the study

According to Hosmer and Lemeshow (1999) and Collett (2003) it is recommended to follow the steps given below.

1. Include all variables that are significant in the univariable analysis at the 20 to 25 percent level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.

2. The variables that appear to be important from step 1 are then fitted together in a model. In the presence of certain variables others may cease to be important. As a result, backward elimination is used to omit nonsignificant variables (i.e., those variables that do not significantly increase the value of $-2\log\hat{L}$) from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined. Here, $\hat{L}$ denotes the maximized likelihood under an assumed model and computed from the partial likelihood equation (20) by replacing the β's by their maximum likelihood estimates under the model.

3. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method (i.e., any that reduce $-2\log\hat{L}$ significantly are retained in the model). This process may result in terms in the model determined at step 2 ceasing to be significant.

4. A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of $-2\log\hat{L}$, and that no term not included significantly reduces $-2\log\hat{L}$.

3.5 Assessment of Model Adequacy

Model-based inferences depend completely on the fitted statistical model. For these inferences to be *valid* in any sense of the word, the fitted model must provide an adequate summary of the data upon which it is based. Some of the methods for the assessment of a fitted proportional hazards model can equally used for parametric regression models with the exception that assessing the adequacy of survival models have to cope with the occurrence of censored survival times.

Residual analysis

Many model checking procedures are based on quantities known as residuals. 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 the 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. However, the two assumptions are not valid 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 Cox-Snell, martingale and Schoenfeld residuals (Collett, 2003 and Hosmer & Lemeshow, 1999).

Cox-Snell residuals (rc_i) are residuals most widely used in the analysis of survival data. The Cox-Snell residual for the i^{th} subject is given by

$$rc_i = \hat{H}_i(t) = -\hat{S}_i(t), \tag{31}$$

where $\hat{H}_i(t)$ and $\hat{S}_i(t)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} subject at time *t*, respectively. In nut shell, Cox-Snell residuals are useful in assessing an overall model fit.

Martingale residuals ($r\hat{M}_i$) are also called modified Cox-Snell residuals and, expressed as

$$r\hat{M}_i = \delta_i - \hat{H}_i(t) = \delta_i - rc_i \tag{32}$$

where $\delta_i = 1$ for uncensored observations and zero otherwise, and rc_i are Cox-Snell residuals. These residuals have similar properties to the error components in other models, in addition to

the properties that its mean is equal to zero under the correct model. In large samples, the martingale residuals are uncorrelated with one another and have an expected value of zero. However, the martingale residuals are not symmetrically distributed about zero.

Plot of these residuals versus explanatory variables is used to indicate whether any particular variable needs to be transformed before incorporating it in the model. In other way round, martingale residuals are useful in determining the functional form of covariate to be included in the model. If after plotting the residuals versus explanatory variables, the plot does not show an obvious relationship, then the variable is not important in the model to be included. Therefore, if most of the points fall horizontally about zero, in the plots of the martingale residuals versus the values of the independent variables, then the fitted model is taken as satisfactory.

Schoenfeld residuals (rs_{ik})

Schoenfeld (1982) proposed residuals for use with a fitted proportional hazards model and packages providing them refer to as the "Schoenfeld residuals", which are based on the individual contributions to the derivative of the log partial likelihood. It is obtained by taking the first derivative of the log of the partial likelihood function for the k^{th} covariate as follows:

$$\frac{\partial L_p(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} \delta_i \left\{ x_{ik} - \frac{\sum_{j \in R(t_{(i)})} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_{(i)})} e^{x'_j \beta}} \right\} = \sum_{i=1}^{n} \left\{ x_{ik} - \bar{x}_{w_i k} \right\}, \tag{33}$$

$$\text{where } \bar{x}_{w_i k} = \frac{\sum_{j \in R(t_{(i)})} x_{jk} e^{x'_j \beta}}{\sum_{j \in R(t_{(i)})} e^{x'_j \beta}}. \tag{34}$$

The estimator of the Schoenfeld residual for the i^{th} individual on the k^{th} covariate is obtained from (33) by substituting the partial likelihood estimator of the coefficient, $\hat{\beta}$, and is

$$\hat{rs}_{ik} = \delta_i (x_{ik} - \bar{\hat{x}}_{w_i k}), \tag{35}$$

where $\bar{\hat{x}}_{w_i k} = \dfrac{\sum_{j \in R(t_{(i)})} x_{jk} e^{x'_j \hat{\beta}}}{\sum_{j \in R(t_{(i)})} e^{x'_j \hat{\beta}}}$ is the estimator of the risk set conditional mean of the covariate.

Since the partial likelihood estimator of the coefficient, $\hat{\beta}$, is the solution to the equations obtained by setting (33) equal to zero, the sum of the Schoenfeld residuals is zero.

It is suggested that (Grambsch and Therneau, 1994) scaling the Schoenfeld residuals by an estimator of its variance yields a residual with greater diagnostic power than the unscaled one. Let the vector of *p* Schoenfeld residuals for the i^{th} subject be denoted as

$\hat{rs}_i' = (\hat{rs}_{i1}, \hat{rs}_{i2}, ..., \hat{rs}_{ip})$,

where $\hat{rs}_{ik}$ is the estimator in (36), with the convention that $rs_{ik} = \text{missing}$ if $\delta_i = 0$.

Thus, the vector of scaled Schoenfeld residuals is given as the product of the inverse of the covariance matrix and the vector of residuals;

$$\hat{r}_i^* = \left[\hat{V}ar(\hat{r}_i)\right]^{-1} \hat{r}_i, \tag{36}$$

where $\hat{V}ar(\hat{r}_i)$ is the estimator of the $p \times p$ covariance matrix of the vector of residuals for the i^{th} subject.

However, Grambsch and Therneau (1994) suggest based on their experience that the matrix, $\hat{V}ar(\hat{r}_i)$, tends to be fairly constant, the use of an easily computed approximation for the scaled Schoenfeld residuals. 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 (in our study number of death *m*).

That is, $\left[\hat{V}ar(\hat{r}_i)\right]^{-1} = m\hat{V}ar(\hat{\beta})$.

Consequently, the approximate scaled Schoenfeld residuals are obtained by substitution as

$$\hat{r}_i^* = m\hat{V}ar(\hat{\beta})\hat{r}_i. \tag{37}$$

Each of these residuals provides a useful tool for examining one or more aspects of model adequacy.

A. Testing for the form (linearity) of covariates

After identification of a particular set of explanatory variables on which the hazard function depends, it is important to check that the correct functional form has been adopted for the continuous covariates. When incorrect functional form exists, an improvement for the fit of a model may be achieved by applying some transformation of the values of a variable. The plot of martingale residuals obtained from fitting the model, excluding the covariate whose functional form needs to be determined, against the excluded covariate display the functional form required for the covariate. In such a way that, LOESS smoothed curve can be superimposed on the scatter plots to give interpretation. If the functional form suggested in using the above plots has some pattern, which is non linear, the covariate can be so transformed and the martingale residuals again should be plotted against the transformed covariate. In so doing, a straight line would then provide evidence that the utilization of appropriate transformation to the covariate of interest (Hosmer and Lemeshow, 1999).

B. Identification of influential subjects

Another important aspect of model evaluation is a thorough examination of the regression diagnostic statistics to identify which, if any, subjects have an unusual configuration of covariates, exert an illegitimate influence on the estimates of the parameters or have an undue influence on the fit of the model. Such observations may be termed as influential (aberrant) observations and the data from such individuals will need to be the subject of further scrutiny. Conclusions from survival analyses are often framed in terms of estimates of the relative hazard, which depends on the estimated values of the coefficients in the Cox regression model. For that reason, it has particular importance to examine the influence of each observation on these estimates (Hosmer and Lemeshow, 1999).

In many occasions, the influence that each observation has on the estimated hazard function will be of interest, and it will then be important to identify observations that influence the complete set of parameter estimates in the model. In other words, it may happen that the structure of the fitted model is particularly sensitive to one or more observations in the data set. Such observations can be analyzed through diagnostics that are designed to highlight observations that influence the complete set of parameter estimates in the linear predictor. This could be done by

fitting the model to all *n* observations in the data set, and then fitting the same model to the sets of *n*-1 observations obtained by omitting each of the *n* observations in turn.

To achieve this purpose, to examine influence in the proportional hazards setting, we need to use statistics analogous to Cook's distance in linear regression. This is denoted as

$$\Delta\hat{\beta}_{ki} \approx \hat{\beta}_k - \hat{\beta}_{k(-i)}, \tag{38}$$

where $\hat{\beta}_k$ denotes the partial likelihood estimator of the coefficient computed using the entire sample of size n and $\hat{\beta}_{k(-i)}$ denotes the value of the estimator if the ith subject is removed.

Cain and Lange (1984) show that an approximate estimator of (38) is the kth element of the vector of coefficient changes

$$\Delta\hat{\beta}_i = \left(\hat{\beta} - \hat{\beta}_{(-i)}\right) = \hat{V}ar(\hat{\beta})\hat{L}_i, \tag{39}$$

where $\hat{L}_i$ is the vector of score residuals and $\hat{V}ar(\hat{\beta})$ is the estimator of the covariance matrix of the estimated coefficients.

These are commonly referred to as the scaled score residuals and their values may be obtained from some software packages, for instance, SAS (Collett, 2003 and Hosmer and Lemeshow, 1999).

C. Methods for Assessing the Proportional Hazards Assumption

The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model. However, there are various grounds for which the model may not have proportional hazards (or constant hazard ratio over time). If hazards are not proportional, this means that the linear component of the fitted model varies with time in some manner. As a result, we need to plot the logarithm of the Kaplan-Meier cumulative hazards function based on different factors so that it helps in assessing the proportional hazards assumption before fitting a Cox model (Collett, 2003 and Hosmer and Lemeshow, 1999). Obviously, if the assumption of proportional hazards is met, the plots should be parallel (or the two curves are equidistant over time).

There are a number of ways in which the proportional hazards model can be changed to non-proportional hazards functions or log-hazard functions that are not equidistant. For instance, if the j^{th} time-independent variable is denoted as x_j, then we can define the j th product term as $x_j \times g_j(t)$ where $g_j(t)$ is some function of time for the j^{th} variable. Likewise, Grambsch and Therneau (1994) also considered a specific form of time-varying coefficient:

$$\beta_j(t) = \beta_j + \gamma_j g_j(t), \tag{40}$$

where $g_j(t)$ is a specific function of time and γ_j is a coefficient of the same.

Thus, the extended Cox model that simultaneously considers all time-independent variables of interest can be formulated as:

$$h(t, x, \beta) = h_0(t) \exp\left(\sum_1^p \beta_j x_j + \sum \gamma_j x_j g_j(t) \right) \tag{41}$$

In order to check the proportional hazards assumption, we consider the null hypothesis that all the γ terms are equal to zero so that the model reduces to the proportional hazards model. The hypothesis all γ_i's are zero ($H_0 : \gamma_i = 0$) is tested via the partial likelihood ratio test, score test or Wald test. Furthermore, the plot of scaled Schoenfeld residuals of each covariate versus the logarithm of analysis time (i.e., the time variable in survival analysis) used to confirm whether there is some departure from proportional hazards or not.

D. Overall Goodness of Fit

A number of plots based on residuals can be used in the graphical assessment of the adequacy of a fitted model. For instance, if the fitted model is correct, the Cox-Snell residuals were shown to have an exponential distribution with unit mean. Putting it in other words, plots of these residuals against the survival times, the rank order of the survival times, or explanatory variables may indicate whether there are particular survival times, or values of the explanatory variables, where the model does not fit well (Collett, 2003).

In addition, it is possible to use some measure analogous to standard R^2 as the case of all regression analysis as a measure of model performance. 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. Thus, the measure of goodness of fit based on partial likelihood is given by:-

$$R_p^{\,2} = 1 - \left\{ \exp\left[\frac{2}{n} \left(L_0 - L_p \right) \right] \right\}, \tag{42}$$

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1. Baseline characteristics of the study

A total of 2217 patients were treated with ART in the hospital AFGTH during the study period from Sep 2003 to Aug 2010. However, the study included 734 ART patients for whom data for variables of interest are complete. Of these 88.3% are censored and 11.7% are uncensored.

The socio-demographic characteristics of the cohort are summarized in Table 1 of the Appendix. The majority of patients were males (548). Of the total (734) patients, 268 were never married and 381 had at least secondary school education. The same table depicts that 632 were Coptic orthodox, 530 were employed, 413 had one room of residence, 479 had at least two members in the household, 515 were naïve to substance use, 196 had bedridden functional status and only 28 had regular sexual partner.

The clinical characteristics of the patients are summarized in Table 2 of the Appendix. Among the patients, the regimen d4T was frequently prescribed (for 403 patients) and 210 were staged clinically as IV. Of the total of 734 patients, TB and opportunistic infections (OIs) were prevalent among 403 and 648 patients, respectively. The median age, CD4 count and weight of patients at start of ART are 34 years (interquartile range 29 - 40 years), 83 cells/mm^3 (interquartile range 36 - 139.25 cells/mm^3) and 53 kg (interquartile range 48 - 59 kg), respectively (Table 3 Appendix).

4.2. Descriptive survival analyses

The patients were followed up for a median of 38.5 months. The minimum follow up time was 1 month and the maximum was 82 months. Among the total of 734 study subjects, 86 (11.7%) died during the follow up time of whom 28 (32.6%), 43 (50%) and 61 (70.9%) deaths occurred within three months, six months and twelve months of ART initiation, respectively. The overall mean estimated survival time of patients under the study was 72 (95% CI: 70- 74) months. In order to get a closer look at estimate of the survival time we use the Kaplan-Meier and Nelson-Aalen estimation techniques. Since estimated survival function changes only at failure/death times, most statistical packages give the estimated result for uncensored observations. The graph

of hazard functions is depicted in Figure 1 below showing that an increase in the hazard rate has direct relation with the increase of time.

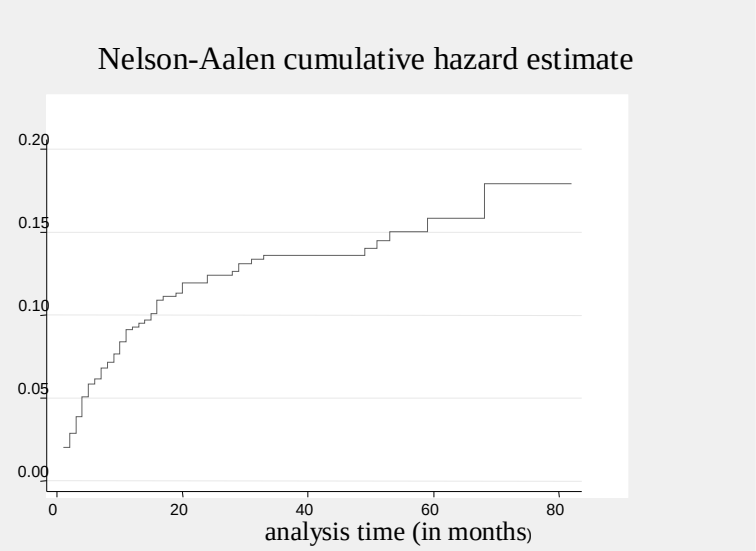


Figure 1: The Nelson-Aalen estimated cumulative hazard function of HIV patients under ART in AFGTH, AA, 2011.

Females and males have almost the same survival times (i.e., 72.0 months for females and 71.8 months for males). Likewise, as can be seen from Figure 1.1 (a - n), the survival curve of females slightly lies above the survival curve of males showing better survival of females than males. The mean survival time of patients based on different socio-demographic and clinical characteristics are summarized in Table 4 of the appendix.

The graph, Figure 2, of the estimate for overall Kaplan-Meier survivor function depicted that, relatively, a large number of the deaths occurred at the earlier months of ART treatment; and the same graph showed the decrement over a follow up period. A separate graph of the estimates of the Kaplan-Meier survivor functions is constructed for different covariates; in so doing it is possible to see the existence of difference in survival experience between the indicated categories of individuals. In general, the pattern of one survivorship function lying above another means the group defined by the upper curve had a better survival than the group defined by the lower curve. Some of the graphs did not show clear differences between the intended categories.

However, among others, graphs of opportunistic infections (OIs), WHO stage, TB and functional status manifest relatively larger gaps and convey similar information as Table 4. For instance, patients in working functional status have longer experience of survival time than those who are on ambulatory status and those who stay in bed. By the same token, the result substantiated that patients under WHO clinical stage I/II have longer survival time than those who are in stage III and stage IV. The graphs of Kaplan-Meier survival estimates based on different categories of covariates are displayed in Figure 1.1 (a - n) of the Appendix.

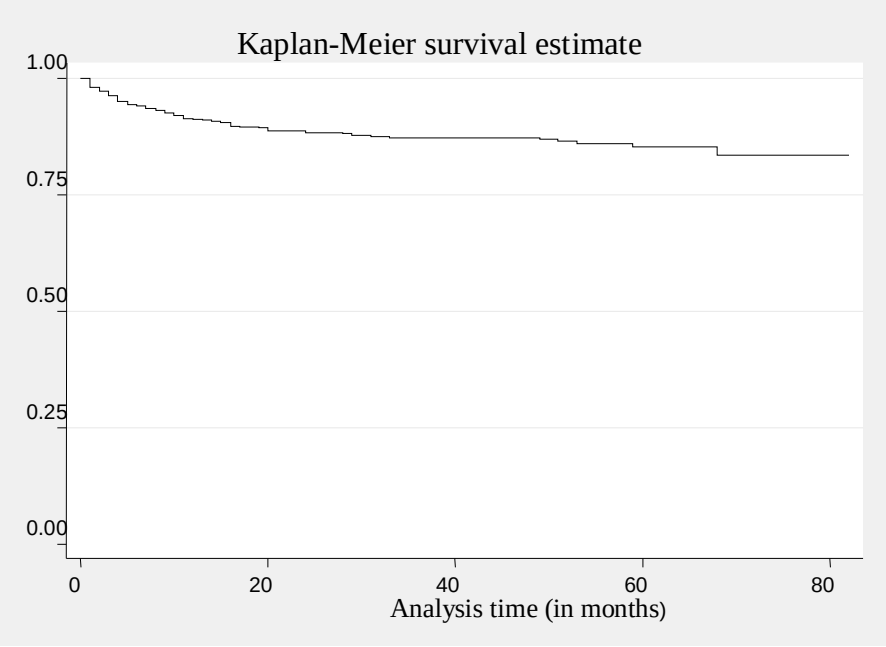


Figure 2: The plot of the overall estimate of Kaplan-Meier survivor function of HIV patients under ART in AFGTH, AA, 2011.

To check for significance differences among categories of factors that are shown using the Kaplan-Meier estimates of the survivor functions, we employ a log-rank statistical test. Based on the log-rank test, there was no significant difference in survival experience between the various categories of gender, marital status, religion, level of education, number of rooms, risk behavior, ART regimen and substance use. However, the log-rank test showed that the survival experience of patients in different categories of employment status, household size, functional status, OIs, TB and WHO clinical stage differ significantly. The results are shown in Table 5 of

the Appendix. A close examination of Table 4, Figure 1.1 (a - n) and Table 5 reveal that patients who were unemployed, who lived in households of size two or more, those who had working functional status, who did not suffer from opportunistic infections, who had no TB co-infection and who were under WHO clinical stage I/II had better survival time.

4.3. Results of the Cox proportional hazards model

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

The aim of model development is to obtain a model that satisfactorily describes the data at hand. For the same purpose, the first step is to select covariates which are important in a study at some *relaxed* level of significance. In this study, a model that contains all variables that are significant in the univariable analysis (Table 6 Appendix) in relation to time of enrollment in ART treatment to the occurrence of event (or death) due to HIV at the 20-25 percent level of significance is used for selection of important covariates. The univariable analysis indicates that not all of the 17 explanatory variables are statistically important to be included in the multivariable analysis stage. Among these, the candidate predictors for further analysis are educational level, employment status, number of rooms, household size, functional status, CD4 cell count, WHO clinical stage, OIs, TB, ART, age and weight of the patients.

Consequently, the most appropriate subset of these predictors to be included in the multivariable model will be selected based on their contribution to the maximized log partial likelihood of the model ($-2LL$). The highest reduction in $-2LL(\hat{\beta})$ is observed for WHO clinical stage that reduced the value for the null/empty model, which is 1087.061 to 1029.514. This difference is 57.547 and it is statistically significant (p-value <0.0001) when compared with percentage points of the χ^2 distribution on 1 degree of freedom. The next highest change is obtained for functional status where the difference equal to 46.721 and significant at p-value < 0.001**.**

All potential variables that are supposed to have significant impact (at *p-value* < 0.25) on the survival time of patients at univariable analysis will be included in the initial multivariable proportional hazards model which leads to a value of $-2LL(\hat{\beta})$ to 959.683 (Table 7 Appendix).

Thus, the covariates will be eliminated in order of the magnitude in which they increase the $-2LL(\hat{\beta})$. As shown in Table 7 of the Appendix, the least important covariate in the model is number of rooms in which a patient resides since the removal of the same brings insignificant increment (p-value< 0. 971) in the value of $-2LL(\hat{\beta})$. Thus, number of rooms becomes the first to be removed from the multivariable model.

The next step is to fit a multivariable model of the remaining 11 covariates (Table 8 Appendix). Of the total eleven covariates, the prescribed ART regimen insignificantly (p-value < 0.633) increase $-2LL(\hat{\beta})$ if removed from the model. Following the same procedure weight, educational level, age and household size were removed at p-values of 0.389, 0.289, 0.217 and 0.069, respectively (Tables 9, 10, 11 and 12, respectively). Thus, we obtain a multivariable model that includes six covariates, namely employment status, functional status, WHO clinical stage, OIs, TB and CD4 cell count. These are significant at 5% level of significance (Table 13 Appendix).

The other important step is considering variables that are non-significant at univariable analysis for possibility of confounders. This can be achieved by adding (one at a time) to the model containing the variables employment status, functional status, WHO clinical stage, OIs, TB and CD4 cell count. The percentage changes in the coefficients of the variables included in the model containing the main effects were found to be small (less than 20%) which reveals that none of them are significantly confounder (Table 14 Appendix). Moreover, Table 15 of Appendix presents the case when the variables that were not significant in the multivariable proportional hazards Cox regression model are added one at a time, and the results show the non-existence of confounders. Thus, variables that were neither significant at univariable analysis nor at multivariable analysis are not confounders of the main factors in the, preliminary, model.

Therefore, both Tables 14 and 15 confirm that the potential covariates to be kept in the multivariable proportional hazards Cox regression model are employment status, functional status, WHO clinical stage, OIs, TB and CD4 cell count.

In the process of model developing, it is relevant to check the correct functional form of a continuous covariate that is/are included in the multivariable model. The plots of the martingale residuals demonstrate the linearity of a continuous covariate after excluding the covariate for which we are checking the correct functional form thereof. Thus, the resulting plot, for the only continuous variable in the model CD4 count, together with a LOESS smoothed curve superimposed to ease interpretation is presented in Figure 3 below. It can be seen that the plots of martingale residuals are random showing no systematic pattern, and the LOESS smoothed curve appears nearly a horizontal line through zero. Consequently, the assumption of linearity for CD4 cell count is fulfilled for the model.

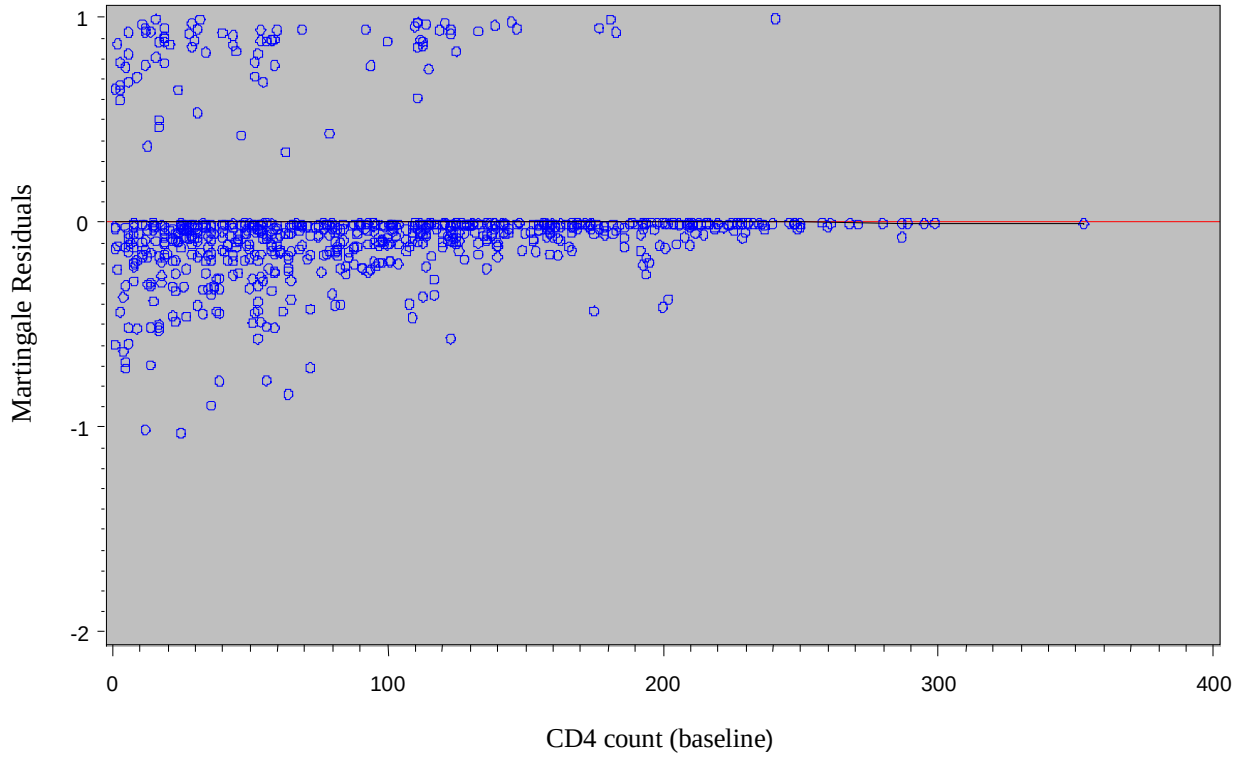


Figure 3: Plot of the martingale residuals for the model excluding CD4 count against the values of CD4, with LOESS smoothed curve superimposed.

Finally, we take into account the possible interactions among covariates that are significant at multivariable level of analysis (i.e., by taking one covariate at a time to the preliminary model). The Wald statistic is employed for the same task and the result verifies that none of the interaction terms were significant at 5% level (Table 16 of Appendix). Once again, the result ensures that the preliminary model of the study will contain only the six covariates in Table 13. The parameter estimates and hazard ratios of the covariates are shown in Table 4.1 below. Nevertheless, before interpreting the estimates, it is appropriate and mandatory to check whether the assumption of the proportional hazards Cox regression model is met or not.

Table 4.1: Estimated values of the coefficients, hazard ratios, 95% CI for the hazard ratio and P-values of the explanatory variables on fitting the proportional hazards model to the data from HIV patients under ART in AFGTH, AA, 2011.

Covariates	Parameter estimate	Standard Error	Wald χ^2	DF	P-Value	Hazard Ratio	95.0% CI for the Hazard Ratio	
							Lower	Upper
Employment	0.837	0.314	7.111	1	0.008	2.310	1.248	4.275
Functional			13.823	2	0.001			
Ambulatory	0.698	0.347	4.040	1	0.044	2.011	1.018	3.973
Bedridden	1.211	0.337	12.917	1	0.000	3.358	1.734	6.500
WHO Stage			16.448	2	0.000			
Stage III	1.953	0.733	7.103	1	0.008	7.052	1.677	29.658
Stage IV	2.537	0.733	11.969	1	0.001	12.640	3.003	53.199
OIs	2.196	1.010	4.722	1	0.030	8.985	1.240	65.085
TB	0.550	0.261	4.437	1	0.035	1.734	1.039	2.893
CD4 count	-0.005	0.002	5.697	1	0.017	0.995	0.991	0.999

Note that the value of -2log L for the model containing the covariates in this table is 967.978.

4.4. Diagnosis of the model

Since fitting a model is not the end of the story, we need to asses some requirement of the model of the study. In other words, the preliminary final model shall be diagnosed for describing our

data optimally or not. In this setting, the requirement is all about the diagnosis for the final proportional hazards model that consists of testing the assumption of proportional hazards, checking for the presence of leverages (influential observations) and measuring the overall goodness of fit of the model.

4.4.1 Assessing the proportional hazards assumption

The proportional hazards assumption, which asserts that the hazard ratios are constant overtime, is vital to the interpretation and use of a fitted proportional hazards model. That means the risk of failure must be the same no matter how long subjects have been followed. In order to test the said assumption above, the extended Cox model is employed and graphical display is used to substantiate the same. For that reason, all interactions of covariates with the logarithm of survival times are modeled together with the main effects; and Wald statistic is used to test the significance of the interaction terms at 5% level of significance.

Table 17 of the Appendix shows p-values of the Wald statistic that indicate none of the coefficients of the interaction terms are significant at 5% level. Since the interaction effect is found to be non- significant (meaning high p-values), this is a confirmation for the absence of time varying covariates. Put differently, there are no covariates which show a trend/pattern with the time and therefore the hazard ratios will be constant over the study time. Therefore, there is no sufficient evidence to reject the null hypothesis that the coefficients of the time varying variables (interaction terms) are zero. This ascertains the validity of the assumption of the proportional hazards holds.

Furthermore, plotting the scaled Schoenfeld residuals of each covariate against log time will be used to check whether the assumption of proportional hazards is violated or not. Figure 1.2 (a - h) of the Appendix shows the plot of scaled Schoenfeld residuals of each covariate for the final model. The graphs depict that each of the eight plots are random, smooth and approximate the horizontal through zero. Thus, there is no covariate which has interaction with log of time revealing the proportional hazards assumption is met.

4.4.2 Assessing for influential observations

At this step of model examination, we look for the existence of leverage observations that have illegitimate impact on inferences made on the basis of model fitted to an observed set of survival data. Leverages, similar to what is obtained in linear and logistic regression, are also adapted into proportional hazards regression through the DFBETA to examine if there are subjects with undue influence on the fit. As a result, the DFBETA is employed to examine if there is disproportionate influence of an observation on the parameter estimate in the fitted Cox regression model (Collett, 2003 and Hosmer and Lemeshow, 1999).

The first five largest changes in parameter estimates are shown in Table 18 of the Appendix. To begin with the largest difference for covariate employment status is observed for patient numbered 396. The result reveals that the change in the parameter estimate if the data for this patient is discarded is 0.08884. Obviously, the omission of this patient increases the hazard rate in relation to the baseline hazard rate. However, the question to be posed is whether the increment in the hazard rate is striking or not. The same can be judged by taking standard error of the parameter estimate of the employment status (0.314) in the full data set into account. That is the percentage change in parameter estimate if the observation is removed is about 28.29% of the standard error (i.e., less than one standard error). Thus, removing this observation cannot bring a significant change on employment status.

Similarly, omitting data for patient 695 and patient 422 from the full data set brought the largest change in parameter estimates of baseline CD4 cell count and TB co-infection, respectively. However, the largest changes in the parameter estimates for CD4 cell count and TB co-infection when the said observation is omitted in turn is 0.00063 (12.6% of the standard error) and 0.05754 (22.05% of the standard error), respectively. Since the percentage changes in the parameter estimates are less than one standard error, there is no high leverage value for both CD4 cell count and TB co-infection. Moreover, the maximum change in the parameter estimates, if observation of the largest difference is removed, for ambulatory, bedridden, WHO stage III, WHO stage IV and OIs are 26.58%, 28.20%, 64.03%, 63.63% and 32.24% of their respective standard error, respectively. Thus, deleting the observations of the highest difference in parameter estimates had no significant impact on the parameters of the covariates and on the fit

of the model. Therefore, it can be concluded that there is no aberrant observation in the data set that illegitimately inflate the estimates of the parameters of the covariates in the final model.

4.4.3 Checking for overall goodness of fit

The final step in the model assessment is to measure the overall goodness of fit. For this objective we use the Cox-Snell residuals and R^2. The plot of the Nelson-Aalen estimate of the cumulative hazard function of the Cox-Snell residual against the Cox-Snell residuals is presented in Figure 4 below.

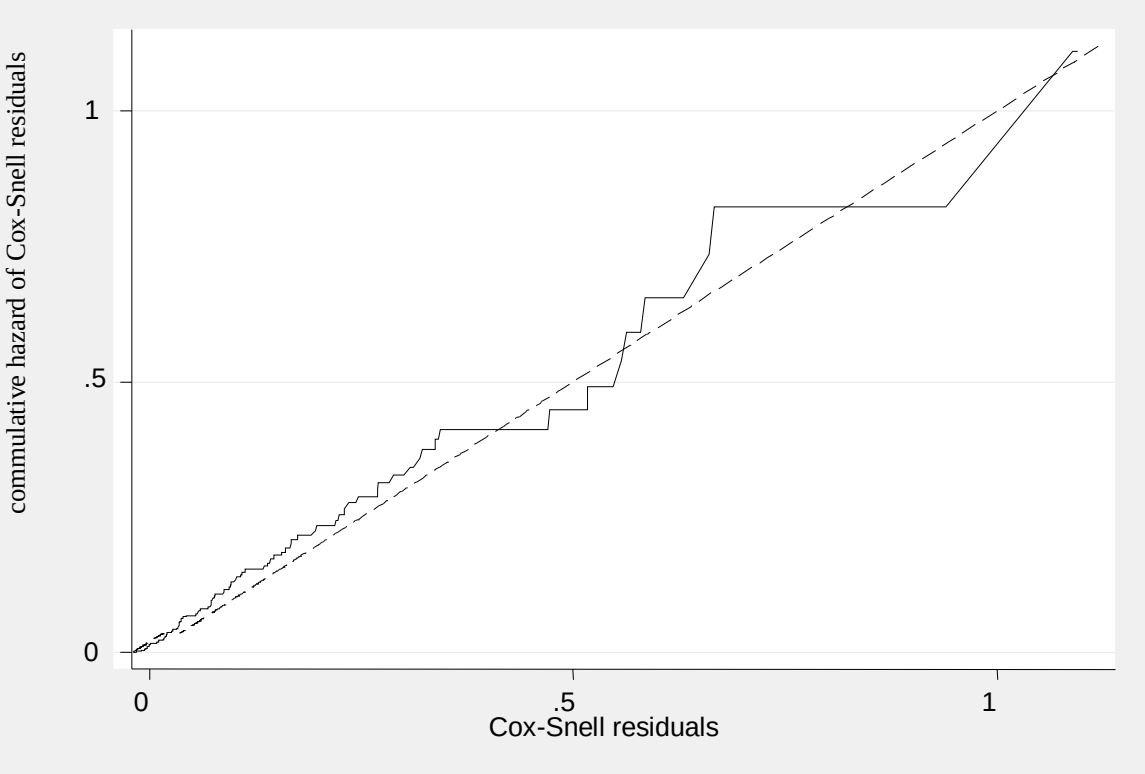


Figure 4: Cumulative hazard plot of the Cox-Snell residuals of the proportional hazards Cox regression model in table 4.1. The 45^0-striaght line through the origin is drawn for reference.

It can be seen that the plot of the residuals in Figure 4 is fairly close to the 45^0 straight line through the origin. Thus, the plot is evidence that the model fitted to the data is satisfactory.

Moreover, an adequate model may have low R^2 due to high percent of censored data. The model in Table 4.1 presented the value of the $-2LL(\hat{\beta})$ with covariates which is equal to 967.978and without covariates equal to1087.061. Thus, the model fitted for the study has the value of $R_p^{\,2} = 1 - \exp\left[\frac{1}{734}(-1087.061 + 967.978)\right] = 0.150$, which is small, indicating that the model fit the data well. In addition, results of the Likelihood ratio, Score and Wald tests for model goodness of fit displayed in Table 19 suggest that the model is good fit (i.e. significant at 5% level of significance). Therefore, the model with estimates as given in Table 4.1 is the final model.

4.5. Interpretation and discussion of results

4.5.1 Interpretation

The interpretation from the results of the final model that consists of the main effects is based on the hazard ratios. That is, the coefficient of a categorical explanatory variable in the model can be interpreted as the logarithm of the ratio of the hazard of death to the baseline (reference group) hazard. In other words, comparison is made with the reference category and between groups for the categorical covariates. By the same token, the coefficient for a continuous explanatory variable is the estimated change in the logarithm of the hazard ratio for a unit increase in the value of the explanatory variable provided that the other variables in the model are kept fixed. Consequently, the interpretation of covariates that are included in the final proportional hazard model of HIV patient under ART is as follows.

Let us begin with baseline CD4 cell count of the patient that is supposed to be significant both clinically and statistically. In this study, a baseline CD4 cell count has been found to have a significant impact on the mortality (or survival) of HIV infected patients (i.e., the estimated Hazards Ratio (HR) = 0.995, 95% CI: 0.991 - 0.999, p=0.017). The estimated hazard ratio for a 50 cells/mm^3 increase in the baseline CD4 cell count is $0.779 = \exp(-0.005 \times 50)$, and its

corresponding 95 % CI is 0.640 to 0.947. The result reveals that patients whose CD4 cell count is larger by 50cells/mm^3 die at lower hazard rate of 22.1%.

Similarly, variables that are found to be significantly associated with the survival of patients in the fitted Cox regression model are TB co-infection, employment status and opportunistic infections (OIs). The hazard ratio for TB co-infected patients in relation to those who are uninfected with TB is 1.734 (95% CI: 1.039 - 2.893). It means the patients infected with TB have about 73.4% higher mortality rate than patients without TB infection. The 95% confidence interval indicates that the hazard rate goes to a maximum of 2.893 and a minimum of 1.039. The hazard ratio for OIs is 8.985 (95% CI: 1.240 - 55.085). The interpretation is that the HIV patients who have one or more history of opportunistic infections manifest about 9 times the hazard faced by patients who have no OIs. The confidence interval suggests that the hazard ratios is as low as 1.240 and as high as 55.085. The employment status of patients is found to be a strong prognostic factor of an overall mortality of a patient (HR = 2.310, 95%CI: 1.248 - 4.275). The result reveals that the rate of dying for an employed patient is 2 times that of unemployed patient.

The estimated hazard ratios for a patient with ambulatory and bedridden functional status in comparison with working group are 2.011(95%CI: 1.018 - 3.973) and 3.358(95%CI: 1.734 - 6.500), respectively. This implies that the risk of death of a patient in ambulatory and bedridden categories of functional status is, approximately, 2 times and 3.4 times higher than that of the working patients, respectively. It makes sense to compare hazard rates between none referent groups. In so doing, we can find the hazard ratio for a patient with bedridden functional status when compared to ambulatory functional status as $\exp(1.211 - 0.698) = 1.67$ (or 95% CI: 1.023 – 2.726). It means a patient with bedridden functional status has a mortality rate that is roughly 2 times a patient with ambulatory functional status. Similarly, the end points of confidence interval shows the hazard ratio is as minimum as 1.023 and as maximum as 2.726.

The reference category for the design variables of WHO clinical stage is patients who are in stage I/II. The estimated hazard ratio for clinical stage III is 7.052. Thus, a patient whose clinical stage is III has approximately 7 times the hazard faced by patient whose clinical stage is I/II. The 95% confidence interval indicates that the rate of dying for patients in clinical stage IV is as high

as 29.658 and as low as 1.677 that of patients in clinical stage I/II. Similarly, the hazard ratio for clinical stage IV is 12.64, which implies the rate of dying of patients in advanced clinical stage (IV) is 12.6 times the rate of dying of patients in primitive clinical stage I/II. The end points of the 95% CI of the same suggest that risk of death for patients in the advanced clinical stage is high as 53.199 times and as low as 3.003 times. Now, it is useful and legitimate if one finds the estimated hazard ratio of clinical stage IV compared to stage III which equals $\exp(2.537 - 1.953) = 1.79$ (or 95% CI: 1.165 – 2.76). Since the confidence interval does not contain 1, an individual in clinical stage IV has a significantly higher hazard rate than patients in clinical stage III. Thus, patients in stage IV are 79% more likely to die than patients in stage III.

4.5.2 Discussion of the results

HIV infection prevention is one of the most challenging tasks for clinicians and public health workers at any level. In order to improve the quality of life of HIV-infected patients and lengthen the survival interval from HIV infection /AIDS diagnosis to death, Ethiopia has implemented the use of numerous prophylaxis, strategies and methodologies. Some of these are mass media campaigns, peer education about HIV transmission, treatment of sexually transmitted infections, condom social marketing, safe blood provision, and prevention from mother to child transmission. Survival analysis is not only a conventional methodology but also a frequently implemented technique to evaluate the efficiency of new treatment, prevention methods, and assess prognostic factors.

This 7 years retrospective cohort study of HIV/AIDS patients on ART gives an insight into survival and its determinants in an army hospital setting in Ethiopia. The study found low CD4 cell count, advanced WHO clinical stages (III and IV), TB co-infection, OIs, being bedridden and ambulatory and being employed as statistically significant and strong predictors of mortality of HIV patients under ART. In this study, 11.7% of the patients died, about 50% of the deaths occurred during the earlier months (less than 6 months) of ART. A study conducted by Andinet and Sebastian (2010) suggested that higher death of an HIV infected patients on ART are realized within the first 4 months. Similarly, the study result of Sieleunou *et al*. (2009), Jerene *et al*. (2006), Etard *et al.* (2006), Lawn *et al*. (2006), Laurent *et al*. (2005) and Johannessen *et al*. (2008) substantiated the argument that higher death rate was experienced in the earlier time of

treatment of the patients. The possible reason for the findings might be that the patients started ART at the advanced stage of the disease.

The CD4 cell count, similar to the plasma viral load, is the most important indicator of HIV disease progression and a strong predictor of survival. That may be due to the fact that HIV attacks CD4 cells, and as time goes by people with HIV often see their CD4 cell counts drop. That is the lower the CD4 cell count the greater the chances of getting a number of very serious diseases. The significant impact of CD4 cell count on survival rate has been assessed by many studies. A study by Lawn *et al*. (2008) reported that low baseline CD4 cell count was a strong risk factor for early mortality; and it also discovered that the hazard ratio for the CD4 cell count of less than 50 cells/mm^3 (versus CD4 >50 cells/ mm^3) was 2.5 (95% CI, 1.9 – 3.2). That means, mortality of patients having CD4 cell count of less than 50 cells/ mm^3 is more than 2.5 times when compared to those patients having a CD4 cell count greater than 50 cells/ mm^3. Similarly, a cohort study that was conducted in Cameroon revealed that the mortality was about two times for patients who began ART with a severe immune-depression (CD4 count <50 cell /mm^3) (Sieleunou *et al.,* 2009). A study in Malawi reported CD4 cell count as an independent determinant of death of patients (Ferradini *et al*., 2006). The result of our study is in agreement with the above cited findings.

The WHO Clinical Staging system, like CD4 cell count, has been shown to be a practical and accurate way to manage HIV-infected patients. In this study, we found that the advanced WHO clinical stages III and IV were independent markers of mortality for patients on ART. The possible justification for the finding is that the advanced clinical stage of the disease is the cause for HIV-associated complications. Similar to our finding, studies by Andinet and Sebastian (2010), Ferradini *et al*. (2006), Sieleunou *et al*. (2009) and Lawn *et al*. (2008) provided evidence that the advanced clinical stages (III and IV) had a strong association with high mortality of an HIV infected patients on ART. By the same token, a result of cohort study from Western Cape Province, South Africa has also revealed that the high mortality were due to the higher WHO clinical stage of the disease (Boulle *et al.,* 2008). Moreover, the study in southern Ethiopia found the WHO clinical stage as an independent indicator of mortality in patients (Degu *et al.*, 2006).

It is an undeniable fact that employment status of a patient is a major factor in maintaining his/her income levels and living conditions. The Cox regression analysis results of Delpierre *et al.* (2008) showed that the probability of death was higher for patients without employment (HR= 3.75, 95% CI 2.11 - 6.66) when compared with those employed. Moreover, a study by Cornell *et al*. (2010) revealed better survival time for employed patients. However, a result from our study disagrees with the above findings and associates being unemployed with better survival time. The possible justification is linked to working environment of the patients; it means type of work that demands high effort of the patients which causes stress. For instance, military and police forces carry out their duties in stressful environment. A study by Leserman *et.al* (1999) that demonstrated more stress and less social support as the cause of HIV disease progression is evident for the said possible reason.

The past history of opportunistic infections of an HIV infected patient is one of statistically significant predictors of death in our study. This study indicates that patients who suffered from OIs have higher risk of dying than those free of the infections. A result of a study in Senegal agrees with our finding since it found OIs as main causes of death for HIV patients (Etard *et al.*, 2006). Similarly, a study in USA suggested OIs as a paramount predictor of mortality and ground for severity of illness. Studies by Lawn *et al*. (2008), Johannessen *et al*. (2008) and Wolmer de Melo *et al.* (2008) supported the assertion that OIs are common causes of death of patients under ART.

TB is presumed to be a prominent cause of death among people living with HIV/AIDS. The result of our study showed that patients co-infected with TB had roughly twice the rate of death than uninfected patients. A study conducted in mainland China and Guinea-Bissau (West Africa) associated TB co-infection with high mortality of HIV patients (Xueyan *et al*., 2008 and Norrgren *et al.*, 2010, respectively). Moreover, a study undertaken in North Carolina indicated prevailing significant association between TB co-infection and high risk of death of patients under treatment (Gadkowski *et al*., 2009). The possible reason for these findings is that TB is not only the most common opportunistic infection (OI), it also increases risk of contracting other opportunistic infections.

Another important marker of the severity of the disease is the functional status of patients. The reason thereof is a person can work smoothly if his/her faculty is normal or not deteriorated. In other words, an individual who is in bed or at state of ambulatory is unable to earn his/her daily income and can suffer a hardship in life that in turn facilitate his/her morbidity and mortality rate. This study shows that patients in working and ambulatory functional status experienced a better survival time than those in bedridden functional status. Similar to the findings of this study Seage *et al*. (1997) stated functional status as a potential predictor of survival time.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Despite some limitations, the study has shown the feasibility of the rapid ART scale up service in reducing mortality. The study has employed survival statistical analysis to find out factors that associated with high mortality of HIV patients on ART in AFGTH. In this study, the overall mean estimated survival time of patients under the study was 72.0 months. The study has shown a high mortality of the cohort in the earlier months of treatment. For instance, nearly 32.6%, 50% and 70.9% deaths occurred within three months, six months and twelve months of ART initiation, respectively. The study, via the multivariable proportional hazards Cox regression model, revealed that advanced WHO clinical stages (III and IV), lower CD4 cell count (<50 cells/mm^3), TB co-infection, being employed, history of OIs, being bedridden and ambulatory functional status were strongly related to mortality. However, there were variables that were significant at univariable stage of analysis but not at multivariable analysis stage. These were number of rooms, the prescribed regimen, weight, educational level, age and household size. Moreover, variables that were significant neither at univariable nor multivariable analyses were gender, marital status, religion, risk behavior and substance use.

5.2 Recommendations

- Since lower CD4 cell count, being bedridden and WHO clinical stage are markers of the progression of the disease, patients should be informed about the same. Thus, it will be a warning bell for patients to start the treatment early.
- The high early mortality has to be addressed by increasing the availability of early HIV diagnosis and treatment services.
- Increased screening for HIV and routine health maintenance for those infected are urgently needed in order to facilitate management of both opportunistic infections and the secondary prevention of HIV infection.
- Prompt initiation of TB treatment in order to reduce patient mortality thereof.
- Creating favorable working conditions of the patients so that being engaged in work should improve survival time of the patients; for instance, reducing long time separation of the army from their beloved one.
- Eventually, health workers and data clerks whose duties are to handle information related to patients should be provided with necessary training that will enables them to secure quality data for stakeholders at all level. Moreover, causes of death rather than HIV/AIDS shall be better recorded separately for patients on ART.

REFERENCES

1. Aalen, O.O. (1975). Statistical inference for a family of counting process. Ph.D. Thesis, University of California, Berkeley.
2. Aalen, O.O. (1978). Non parametric inference for a family of counting processes. *Annals of Statistics, 6: 701 - 726.*
3. Alan Whiteside, Alex de Waal and Tsadkan Gebre-Tensae (2006). AIDS, security and the military in Africa: Sober appraisal. *African Affairs, 105(419): 201–218.*
4. Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Math Bioscience, 6: 1 – 11.*

5. Andinet Worku Alemu and Miguel San Sebastian (2010). Determinants of survival in adult HIV patients on antiretroviral therapy in Oromiyaa, Ethiopia. *Global Health Action; 3: 5398.*
6. Andrew Boulle, Peter Bock, Meg Osler, Karen Cohen, Liezl Channing, Katherine Hilderbrand, Eula Mothibi, Virginia Zweigenthal, Neviline Slingers, Keith Cloete and Fareed Abdullah (2008). Antiretroviral therapy and early mortality in South Africa. *Bulletin of the World Health Organization: 86:678–687.*
7. Arthur M. Smith (1992). MilitaryM edicine: Not the Same as Practicing Medicine in the Military. *Armed Forces & Society 18(4):576-594.*
8. Asgeir Johannessen, Ezra Naman, Bernard J Ngowi, Leiv Sandvik, Mecky I Matee, Henry E Aglen, Svein G Gundersen and Johan N Bruun (2008). Predictors of mortality in HIV-infected patients starting antiretroviral therapy in a rural hospital in Tanzania. *BMC Infectious Diseases, 8:52.*
9. Assefa Seme, Damen Haile Mariam and Alemayehu Worku (2005).The association between substance abuse and HIV infection among people visiting HIV counseling and testing centers in Addis Ababa, Ethiopia. *Ethiop.J.Health Dev; 19(2):116-125.*
10. Azusa Sato (2008). Is HIV/AIDS a threat to security in fragile states? Aids, security and conflict initiative (ASCI) research report no.10.
11. Ba,Oumar, O'Regan, Christopher, Nachega, Jean, Cooper, Curtis, Anema, Aranka, Rachlis, Beth and Mills and Edward J. (2008). HIV/AIDS in African militaries: an ecological analysis. *Medicine, Conflict and Survival, 24(2): 88 — 100.*
12. Bayeh Abera, Fisseha walle, Tsehaye Tewabe, Atnaf Alem and Mohammed Yessin (2010). ART-naive HIV patients at Feleg-Hiwot Referral Hospital Northwest, Ethiopia. *Ethiop. J. Health Dev; 24(1):3-8.*
13. BN Kong, JI Harwell, P Suos, L Lynen, S Mohiuddin, S Reinert and D Pugatch (2007). Opportunistic infections and HIV clinical disease stage among patients presenting for care in Phnom Penh, Cambodia. *Southeast Asian J Trop Med Public Health; 38(1).*
14. C Delpierre, L Cuzin, V Lauwers-Cances, G D Datta, L Berkman, and T Lang (2008). Unemployment as a risk factor for AIDS and death for HIV-infected patients in the era of highly active antiretroviral therapy. *Sex Transm Infect; 84:183–186.*

15. Centers for Disease Control and Prevention (CDC) (1993): revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Morb Mortal Wkly Rep; 41:1–19.*
16. Christian Laurent, Ndeye Fatou Ngom Gueye, Cheikh Tidiane Ndour, Pape Mandoumbe´ Gueye, Martin Diouf, Ndella Diakhate´, Ndeye Coumba Toure´ Kane, Isabelle Lanie`ce, Adama Ndir, Laurence Vergne, Ibrahima Ndoye, Souleymane Mboup, Pape Salif Sow and Eric Delaporte (2005). Long-Term Benefits of Highly Active Antiretroviral Therapy in Senegalese HIV-1–Infected Adults. *J Acquir Immune Defic Syndr; 38:14–17.*
17. Collett, D. (2003). Modelling survival data in medical research (Second edition). Chapman and Hall/CRC, London.
18. Cornell M., Myer L., Kaplan R., Bekker L.G. and Wood R. (2009). The impact of gender and income on survival and retention in a South Africa antiretroviral therapy programme. *Trop Med Int Health; 14(7): 722–731.*
19. Cox, D.R. (1972). "Regression models and life Tables (with Discussion)." *Journal of the Royal statistical society, 34: 187-220.*
20. David G. Klein Baum and W. Klein (2005). Survival Analysis a self learning text, second edition. Springer Science+Business Media, Inc. , New York.
21. De Waal, A. (2005). Issue Paper 1: HIV/AIDS and the Military. AIDS, Security and Democracy: Expert Seminar and Policy Conference, Clingendael Institute, The Hague.
22. Degu Jerene, Are Næss and Bernt Lindtjørn (2006[a]). Antiretroviral therapy at a district hospital in Ethiopia prevents death and tuberculosis in a cohort of HIV patients. *AIDS Research and Therapy; 3(10).*
23. Degu Jerene, Aschalew Endale, Yewubnesh Hailu and Bernt Lindtjørn (2006[b]). Predictors of early death in a cohort of Ethiopian patients treated with HAART. *Biomedical central Infectious Diseases; 6:136.*
24. Ethiopia Demographic and Health Survey (2005).
25. Ferradini L., Jeannin A., Pinoges L., Izopet J., Odhiambo D., Mankhambo L., Karungi G.,Szumilin E., Balandine S., Fedida G., Carrieri M. P., Spire B., Ford N., Tassie J.M., Guerin P.J. and Brasher.C (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. *Lancet; 367: 1335–42.*

26. Fikremarkos Merso. Women & Girls and HIV/AIDS in Ethiopia. An assessment of the policy and legal framework protecting the rights of women and girls and reducing their vulnerability to HIV/AIDS. *http://ethiopia.unfpa.org/drive/WomenandGirlsandHIV-AIDSinEthiopia.pdf.*
27. George R. Seage, Constantine Gatsonis, Joel S. Weissman, Jennifer S. Haas, Paul D. Cleary, Floyd J. Fowler,Michael P Massagli, Valerie E. Stone,Donald E. Craven, Harvey Makadon, Joan Goldberg,Kathryn Coltin,Kimberly S. Levin and Arnold M. Epstein (1997). The Boston AIDS Survival Score (BASS): A Multidimensional AIDS Severity Instrument. *American Journal of Public Health; 87(4).*
28. Grambsch P.M and Therneau T.M (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrik; 81(3): 515-526.*
29. Hans Norrgren, Sana Bamba, Zacarias J. Da Silva, Tuija Koivula, Sören Andersson (2010). Higher mortality in HIV-2/HTLV-1 co-infected patients with pulmonary tuberculosis in Guinea-Bissau, West Africa, compared to HIV-2-positive HTLV-1-negative patients. *International journal of infectious disease; 14:142-147.*
30. Helmut Kloos, Damen Haile Mariam and Bernt Lindtjørn(2007). The AIDS Epidemic in a Low-Income Country: Ethiopia. *Human Ecology Review, 14(1).*
31. Hosmer, D.W. and Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York.
32. Isidore Sieleunou, Mohamadou Souleymanou, Anne-Marie Schonenberger, Joris Menten and Marleen Boelaert(2009). Determinants of survival in AIDS patients on antiretroviral therapy in a rural centre in the Far-North Province, Cameroon. *Tropical Medicine and International Health, 14(1): 36–43.*
33. Jane Leserman, Eric D. Jackson, Johb M. Petitto, Robert N. Golden, Susan G. Silva, Diana O. Perkins, Jianwen Cai, James D. Folds, and Dwight L. Evans (1999). Progression to AIDS: The effects of stress, depressive symptoms, and social support. *Psychosomatic medicine 61:397–406.*
34. Jean-François Etard, Ibrahima Ndiaye, Marion Thierry-Mieg, Nde`ye Fatou Ngom Gue`ye, Pape Mandoumbe´ Gue`ye,Isabelle Lanie`ce, Alle´ Baba Dieng, Assane Diouf, Christian Laurent, Souleymane Mboup, Papa Salif Sow and Eric Delaporte (2006). Mortality and

causes of death in adults receiving highly active antiretroviral therapy in Senegal: a 7-year cohort study. *AIDS; 20: 1181-1189.*

35. John Kemoli Sagala (2008). HIV/AIDS prevention strategies in the Armed Forces in Sub-Saharan Africa. A critical review. *Armed Forces & Society, 34(2): 292-313.*
36. Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc 1958; 53:457-81.*
37. Katie Taylor-Smith, Hannock Tweya, Anthony Harries, Erik Schoutene and Andreas Jahn (2010). Gender differences in retention and survival on antiretroviral therapy of HIV-1 infected adults in Malawi. *Malawi Medical Journal; 22(2): 49-56.*
38. Kleinbaum, D.G. and Klein, W. (2005). Survival Analysis a self learning text, second edition. Springer Science+Business Media, Inc., New York.
39. Kloos H., Lindtjorn B. and Damen Haile Mariam (2007). The AIDS Epidemic in a Low-Income Country: Ethiopia. *Human Ecology Review, 14 (1): 39-42.*
40. KY Chow, LW Ang, I Verghesse, SK Chew and YS Leo (2005). Measurable predictive factors for progression to AIDS among HIV-infected patients in Singapore. *Ann Acad Med Singapore; 34(1):84-9.*
41. L. Beth Gadkowski, Carol D. Hamilton, Myra Allen, Ellen R. Fortenberry, Julie Luffman, Elizabeth Zeringue and Jason E. Stout (2009). HIV-Specific Health Care Utilization and Mortality among Tuberculosis HIV Co-infected Persons. *AIDS patient care and and STDs; 23(10).*
42. Ladjane Santos Wolmer de Melo, Heloisa Ramos Lacerda, Eduardo Campelo, Emanuel Moraes and Ricardo Arraes de Alencar Ximenes (2008). Survival of AIDS Patients and Characteristics of Those Who Died Over Eight Years of Highly Active Antiretroviral Therapy, at a Referral Center in Northeast Brazil. *The Brazilian Journal of Infectious Diseases; 12 (4):269-277.*
43. Lawn SD, Myer L, Orrell C, Bekker LG and Wood R (2005). Early mortality among adults accessing a community-based antiretroviral service in South Africa: implications for programme design. *AIDS*; 19 (18):2141-8.
44. Leigh F. Johnson and Rob E. Dorrington (2006). Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demographic research, 14(22):541-574.*

45. Magdalena A. Recsky, Zabrina L Brumme, Keith J. Chan, et al. (2004). Antiretroviral resistance among HIV-infected persons who have died in British Columbia, in the era of modern antiretroviral therapy. *The Journal of Infectious Diseases; 190:285-92.*
46. Mehret, M., L. Khodakevich and D. Zewdie (l990). HIV-1 infection and related risk factors among female sex workers in urban areas in Ethiopia. *Ethiopian Journal of Health Development; 4(2): 163-170.*
47. MoH. ART information tool Kit, Ethiopia. Addis Ababa (2006).
48. N. Balakrishnan and C.R. Rao (2004). Handbook of Statistics, Vol. 23.
49. National AIDS Control Council, Office of the President, Kenya (2008). UNGASS 2008 Country Report for Kenya. NACC, Nairobi.
50. Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of quality technology, 1: 27 – 52.*
51. Nelson, W. (1972). Theory and application of hazard plotting for censored failure data. *Technometrics, 14: 945 – 965.*
52. Nicole Fraser, Marelize Görgens-Albino and John Nkongolo (2008). Rapid analysis of HIV epidemiological and HIV response data about vulnerable populations in the Great Lakes Region of Africa. *Great Lakes Initiative on HIV/AIDS (GLIA).*
53. Office of AIDS Research Advisory Council (OARAC) (2009). Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents.
54. Petros Isaakidis, Marie-Eve Raguenaud, Vantha Te, Chhraing S Tray, Kazumi Akao, Varun Kumar, Sopheak Ngin, Eric Nerrienet and Rony Zachariah (2010). High survival and treatment success sustained after two and three years of first-line ART for children in Cambodia. *Journal of the International AIDS Society, 13:11.*
55. S J Gange, Y Barrón, R M Greenblatt, K Anastos, H Minkoff, M Young, A Kovacs, M Cohen, W A Meyer III and A Muñoz (2002). Effectiveness of highly active antiretroviral therapy among HIV-1 infected women. *J Epidemiol Community Health; 56:153–159.*
56. Schoenfeld, D. (1982). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika; 7: 145-153.*
57. Stephen D. Lawn, Anthony D. Harries, Xavier Anglaret, Landon Myer and Robin Wood (2008). Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa. *AIDS; 22:1897–1908.*

58. Stephen D. Lawn, Landon Myer, Guy Harling, Catherine Orrell, Linda-Gail Bekker, and Robin Wood (2006). Determinants of mortality and nondeath losses from an Antiretroviral Treatment Service in South Africa: implications for program evaluation. *Clinical Infectious Diseases; 43:770–6.*
59. Therneau, T.M., Grambsch, P.M., and Fleming, T.R. (1990). Martingale-based residuals for survival models. *Biometrika; 77: 147-160.*
60. UNAIDS (December, 2007), "Global AIDS epidemic update".
61. UNAIDS (May, 2006), "Report on Global AIDS Epidemic'' http://data.unaids.org/pub/GlobalReport/2006/2006_gr-executivesummary_en.pdf
62. UNAIDS, 2008 Report on the global AIDS epidemic.
63. UNAIDS, 2009. AIDS epidemic update. UNAIDS, Geneva.
64. UNDP (Sept., 2005), "Human Development Report." International cooperation at a crossroads aid, trade and security in unequal world. http://hdr.undp.org/en/media/HDR05_complete.pdf
65. Volmink J, Siegfried N, van der Merwe L and Brocklehurst P. (2007). Antiretroviral for reducing the risk of mother-to-child transmission of HIV infection. Cochrane database of Systematic Reviews, Issue 1.
66. Xueyan J., Hongzhou L., Yuexin Z., Zengquan Z., Hanhui Y., Qingxia Z., Hui W., Fuxiang W., Wangqian J., Hongqing S., Limin D., Yan S., Yun H., Wubuer M., Aibaidoula Y., Xiaofeng S., Yadong H., Wei T., Boping Z. and Cao Y. (2008). A Cross-Sectional Study of HIV and Tuberculosis Co infection Cases in Mainland China. *Southern Medical Association; 101(9):914-917.*
67. Yehenew Getachew, Befekadu Gashaw, Fentaw Abegaz, Paul Janssen, Ingrid Van Keilegom and Luc Duchateau (2009): Survival analysis. North-South-South project in Biostatistics Series VLIR-UOS, Belgium.

APPENDIX

Table 1: Socio-demographic characteristics of the study participants in AFGTH, AA, 2011.

Covariate/factor	Category	Censored (%)	Dead	Total
Gender	Female	164(88.2%)	22	186
	Male	484 (88.3%)	64	548
Marital status	Never married	233(86.9%)	35	268
	Married	317(88.3%)	42	359
	Others	97(91.5%)	9	106
Religion	Muslim	38(82.6%)	8	46
	Coptic orthodox	560(88.6%)	72	632
	Others	50 89.3%	6	56
Education level	No education	38(88.4%)	5	43
	Primary	267(86.1%)	43	310
	Secondary and above	343(90.0%)	38	381
Employment status	Employed	456(86.0%)	74	530
	Not employed	192(94.1%)	12	204
No of Rooms	Only one	360(87.2%)	53	413
	2 rooms and more	288(89.7%)	33	321
House hold size	Only one	217(85.1%)	38	255
	two and more	431(90.0%)	48	479
Substance use	No	452(87.8%)	63	515
	Yes	196(89.5%)	23	219
Functional status	Working	261(95.6%)	12	273
	Ambulatory	237(89.4%)	28	265
	Bedridden	150(76.5%)	46	196
Risk behavior	Regular	24(85.7%)	4	28
	Casual or both regular and casual	624(88.4%)	82	706

Table 2: Clinical characteristics of the study participants on initiation of ART in AFGTH, AA, 2011.

Covariate/factor	Category	Censored (%)	Dead	Total
TB co-infection	No	311(94.0%)	20	331
	Yes	337(83.6%)	66	403
	D4T-based	352(87.3%)	51	403

ART regimen	AZT-based	296(89.4%)	35	331
	No	85(98.8%)	1	86
OIs	Yes	563(86.9%)	85	648
	Stage I and II	199(99.0%)	2	201
WHO clinical stage	Stage III	286(88.5%)	37	323
	Stage IV	163(77.6%)	47	210

Table 3: Summary statistics of continuous variables included in the study of HIV patients under ART in AFGTH, AA, 2011.

Status of patient	Continuous variable	Mean	Std. Deviation	Minimum	Maximum	Median	Q1	Q3
censored	Time	37.10	22.395	1	82	42.00	13	55
	Age	34.93	7.725	20	60	34	29	40
	CD4	100.32	69.937	1	353	89	40.25	148
	Weight	53.85	9.376	28	92	54	48	59
uncensored	Time	11.38	13.784	1	68	6.50	2.75	15.25
	Age	36.76	8.434	22	62	36	30.00	42.00
	CD4	59.13	52.897	1	241	44.50	16.75	111
	Weight	50.99	8.370	35	74	50	45	55
Overall	Time	34.09	23.092	1	82	38.50	10.75	53.00
	Age	35.14	7.827	20	62	34	29	40
	CD4	95.50	69.408	1	353	83	36	139.25
	Weight	53.51	9.304	28	92	53	48	59

Table 4: Kaplan-Meier analyses of survival times for patients on antiretroviral treatment according to important socio-demographic and clinical characteristics of HIV patients, in AFGTH, AA, 2011.

Covariate / Factor	Category	Estimate	Std. Error	95% confidence interval	
Gender	Male	71.8	1.2	69.4	74.2
	Female	72.0	1.0	70.0	74.0
Marital status	Married	71.5	1.4	68.8	74.2

	Never married	69.2	1.9	65.5	72.8
	Others	75.3	2.1	71.2	79.5
Religion	Muslim	64.0	4.5	55.2	72.7
	Coptic orthodox	72.2	1.1	70.0	74.3
	Others	72.8	3.1	66.7	79.0
Education level	No education	62.7	3.8	55.3	70.2
	Primary	68.8	1.7	65.5	72.2
	Secondary and above	73.9	1.3	71.4	76.3
Employment status	employed	70.1	1.3	67.6	72.6
	not employed	73.4	1.3	70.8	75.9
No. of room	Only one	69.7	1.5	66.9	72.6
	2 rooms and more	73.7	1.4	71.0	76.4
Risk behavior	Regular	54.2	3.9	46.5	61.8
	Casual or both regular and casual	72.1	1.0	70.1	74.1
Substance use	No	70.8	1.2	68.4	73.2
	Yes	73.0	1.8	69.5	76.4
WHO stage	Stage I/II	75.3	0.5	74.3	76.3
	Stage III	70.7	1.4	67.8	73.5
	Stage IV	62.7	2.4	57.9	67.5
OIs	No	79.1	0.9	77.3	80.9
	Yes	70.8	1.1	68.6	73.0
TB co-infection	No	76.0	1.1	73.9	78.1
	Yes	67.9	1.6	64.8	71.0
ART Regimen	D4T-based	70.2	1.8	66.7	3.7
	AZT-based	72.4	1.4	69.8	75.1
HHS	only one	68.0	1.9	64.3	71.8
	2 hhs and more	73.6	1.2	71.3	75.9
Functional status	[9]Working	75.7	0.9	73.8	77.5
	[10]Ambulatory	72.1	1.6	69.0	75.2
	[11]Bedridden	60.0	2.8	54.5	65.4
Overall survival time		72.0	1.0	70.0	74.0

Table 5: Results of the Log-rank test for the categorical variables of HIV patients under ART in AFGTH, AA, 2011.

Covariate / factor	DF	Chi-square	P-Value
Gender	1	0.010	0.921
Marital status	2	2.573	0.276
Religion	2	2.078	0.354

[9] ***Working: An individual able to perform usual work in and out of the house, harvest, go to school for children, normal activities or playing.***

[10] ***Ambulatory: An individual able to perform activities for daily living.***

[11] ***Bedridden: An individual unable to perform activities for daily living.***

Level of education	2	4.718	0.095
No of Rooms	1	2.070	0.150
Risk behavior	1	0.237	0.626
ART regimen	1	1.621	0.203
Substance use	1	0.345	0.557
Employment status	1	9.334	0.002
House hold size	1	5.918	0.015
Functional status	2	51.603	0.000
TB co-infection	1	19.6689	0.000
OIs	1	9.781	0.002
WHO clinical stage	2	49.083	0.000

Table 6: Results of the univariable proportional hazards Cox regression model of HIV patients under ART in AFGTH, AA, 2011.

Variable	B	SE	Wald	DF	Sig.	Exp(B)	LR(Sig.)	-2log L	Score
Gender	0.025	0.247	0.010	1	0.921	1.025	0.921	1087.051	0.921
Marital status			2.511	2	0.285		0.259	1084.362	0.279
Married	0.578	0.374	2.385	1	0.122	1.782			
Never married	0.383	0.367	1.086	1	0.297	1.466			
Religion			2.020	2	0.364		0.405	1085.253	0.357
Muslim	0.674	0.541	1.554	1	0.213	1.963			
Coptic orthodox	0.186	0.426	0.192	1	0.662	1.205			
Education			4.600	2	0.100		0.099	1082.445	0.096
Primary	0.183	0.473	0.151	1	0.698	1.201			
Secondary and above secondary	-0.295	0.476	0.385	1	0.535	0.744			
Employment	0.916	0.311	8.646	1	0.003	2.499	0.001	1076.404	0.002
Room	0.317	0.222	2.038	1	0.153	1.373	0.149	1084.980	0.152
HHS	0.521	0.218	5.743	1	0.017	1.684	0.018	1081.490	0.015
Risk	0.248	0.512	0.234	1	0.628	1.282	0.641	1086.843	0.627
Substance	0.143	0.244	0.342	1	0.559	1.153	0.554	1086.712	0.558
Functional status			41.507	2	0.000		0.000	1040.340	0.000
Ambulatory	0.974	0.345	7.962	1	0.005	2.650			
Bedridden	1.939	0.325	35.586	1	0.000	6.952			
WHO clinical stage			29.146	2	0.000		0.000	1029.514	0.000
Stage III	2.541	0.726	12.241	1	0.000	12.688			
Stage IV	3.304	0.723	20.905	1	0.000	27.217			
OIs	2.463	1.006	5.996	1	0.014	11.740	0.000	1071.749	0.002
TB	1.076	0.255	17.748	1	0.000	2.932	0.000	1066.280	0.000
Regimen	0.279	0.220	1.599	1	0.206	1.321	0.203	1085.439	0.205
Age	0.020	0.013	2.302	1	0.129	1.020	0.136	1084.833	0.129
Weight	-0.042	0.013	10.942	1	0.001	0.959	0.001	1075.546	0.001
CD4 count	-0.011	0.002	27.116	1	0.000	0.989	0.000	1052.277	0.000

Remark: The value of -2LL for the null model is 1087.061.

Table 7: Results of the multivariable proportional hazards Cox regression model containing the variables significant at 20 - 25% level in the univariable proportional hazards Cox regression model of HIV patients under ART in AFGTH, AA, 2011.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	7.073	1	0.008	8.465	0.004
Functional status	10.296	2	0.006	11.157	0.004
WHO clinical stage	16.413	2	0.000	24.894	0.000
OIs	4.981	1	0.026	11.231	0.001
TB	3.706	1	0.054	3.982	0.046
CD4	5.900	1	0.015	6.499	0.011
Room	0.001	1	0.971	0.001	0.971
ART	0.226	1	0.635	0.227	0.634
Wght	0.743	1	0.389	0.754	0.385
Education	1.987	2	0.370	1.994	0.369
Age	2.122	1	0.145	2.070	0.150
HHS	1.935	1	0.164	1.947	0.163

Note that the value of -2LL for the model containing the covariates in this table is 959.683

Table 8: Results of the multivariable proportional hazards Cox regression model after eliminating the variable number of rooms from the multivariable proportional hazards Cox regression model in Table 7.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	7.173	1	0.007	8.633	0.003
Functional status	10.516	2	0.005	11.399	0.003
WHO clinical stage	16.416	2	0.000	24.902	0.000
OIs	4.980	1	0.026	11.230	0.001
TB	3.709	1	0.054	3.985	0.046
CD4	5.982	1	0.014	6.593	0.010
ART	0.226	1	0.634	0.227	0.633
Wght	0.742	1	0.389	0.753	0.385
Education	2.017	2	0.365	2.024	0.364
Age	2.141	1	0.143	2.089	0.148
HHS	2.701	1	0.100	2.653	0.103

Table 9: Results of the multivariable proportional hazards Cox regression model after eliminating the variable ART regimen from the multivariable proportional hazards Cox regression model in Table 8.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	7.115	1	0.008	8.559	0.003
Functional status	10.873	2	0.004	11.779	0.003
WHO clinical stage	16.326	2	0.000	24.750	0.000
OIs	4.968	1	0.026	11.197	0.001
TB	4.027	1	0.045	4.351	0.037
CD4	5.924	1	0.015	6.532	0.011
Weight	0.733	1	0.392	0.743	0.389
Education	1.985	2	0.371	1.992	0.369
Age	2.084	1	0.149	2.033	0.154
HHS	2.929	1	0.087	2.875	0.090

Table 10: Results of the multivariable proportional hazards Cox regression model after eliminating the variable weight from the multivariable proportional hazards Cox regression model in Table 9.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	6.985	1	0.008	8.391	0.004
Functional status	11.858	2	0.003	13.097	0.001
WHO clinical stage	17.172	2	0.000	25.895	0.000
OIs	4.979	1	0.026	11.223	0.001
TB	4.166	1	0.041	4.507	0.034
CD4	6.341	1	0.012	7.039	0.008
Education	2.477	2	0.290	2.481	0.289
Age	1.836	1	0.175	1.798	0.180
HHS	2.980	1	0.084	2.924	0.087

Table 11: Results of the multivariable proportional hazards Cox regression model after eliminating the variable education from the multivariable proportional hazards Cox regression model in Table 10.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	6.934	1	0.008	8.336	0.004
Functional status	12.248	2	0.002	13.736	0.001
WHO clinical stage	17.501	2	0.000	26.937	0.000
OIs	5.105	1	0.024	11.602	0.001
TB	4.237	1	0.040	4.589	0.032
CD4	5.579	1	0.018	6.175	0.013
Age	1.552	1	0.213	1.524	0.217
HHS	3.972	1	0.046	3.884	0.049

Table 12: Results of the multivariable proportional hazards Cox regression model after eliminating the variable age from the multivariable proportional hazards Cox regression model in Table 11.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	7.050	1	0.008	8.482	0.004
Functional status	12.357	2	0.002	13.920	0.001
WHO clinical stage	17.024	2	0.000	26.348	0.000
OIs	5.128	1	0.024	11.647	0.001
TB	4.381	1	0.036	4.749	0.029
CD4	5.726	1	0.017	6.336	0.012
HHS	3.388	1	0.066	3.318	0.069

Table 13: Results of the multivariable proportional hazards Cox regression model after eliminating the variable HHS from the multivariable proportional hazards Cox regression model in Table 12.

Covariates/ Factors	Wald χ^2	DF	Wald(sig.)	LR χ^2	LR(sig.)
Employment status	7.111	1	0.008	8.556	0.003

Functional status	13.823	2	0.001	15.520	0.000
WHO clinical stage	16.448	2	0.000	25.726	0.000
OIs	4.722	1	0.030	10.431	0.001
TB	4.437	1	0.035	4.817	0.028
CD4	5.697	1	0.017	6.325	0.012

Table 14: Percentage changes in the coefficients of the variables included in Table 13 when the variables that were not significant in the univariable proportional hazards Cox regression model are added one at a time.

	Gender	Marital status	Religion	Risk	Substance Use
Employment status	-10.9608	4.410112	2.409987	-7.38346	-0.45095
Ambulatory	-3.0584	0.927853	-4.68442	-2.33182	-0.48615
Bedridden	-5.83438	0.721309	-2.33123	0.878061	-0.64147
Stage III	-1.1886	0.041443	1.795512	1.250209	-0.96156
Stage IV	-0.89794	0.235524	0.459417	-0.05743	-0.41445
OIs	0.837308	0.957817	0.483508	-0.81568	0.564797
TB	-1.79226	-2.39555	-4.40921	-10.4412	0.863125
CD4	3.631462	2.501722	1.967764	-2.56312	2.092166

Table 15: Percentage changes in the coefficients of the variables included in Table 13 when the variables that were not significant in the multivariate proportional hazards Cox regression model are added one at a time.

	No room	ART	Weight	Education	Age	HHS
Employment status	3.932959	-0.57808	-1.05342	-0.31788	1.295392	0.520052
Ambulatory	0.209887	-0.29111	9.107513	7.769191	0.3107	1.420076
Bedridden	2.949248	0.931096	4.832963	3.009561	0.192556	4.491644
Stage III	0.730317	-0.37016	0.513295	1.62917	-0.02347	0.575337
Stage IV	0.027884	-0.36821	1.406268	1.425006	-0.39098	-0.45255
OIs	-1.41406	-0.19297	0.321434	1.293067	0.605166	-4.1497
TB	-1.81594	3.839111	1.976968	1.267486	0.431997	0.274671
CD4	-5.07724	-0.38887	2.001823	-6.80719	1.18801	-1.32402

Table 16: Wald statistics and corresponding p-values of possible interaction terms, added one at a time, to the variables included in the model in Table 13.

Interaction between	Covariates/ Factors	Wald χ^2	DF	P-value
Employment status	Functional status	2.464	2	0.292
	WHO clinical stage	0.141	2	0.932
	OIs	0.141	2	0.932
	TB	0.509	1	0.476
	CD4 count	0.000	1	0.998
Functional status	WHO clinical stage	0.613	4	0.962
	OIs	0.094	2	0.954
	TB	1.552	2	0.460
	CD4 count	2.705	2	0.259
WHO clinical stage	OIs	0.050	2	0.975
	TB	1.622	2	0.444
	CD count	4 4.674	2	0.097
OIs	TB	0.022	1	0.882
	CD4 count	0.384	1	0.535
TB	CD4 count	0.666	1	0.415

Table 17: Results of the multivariable proportional hazards Cox regression model containing the variables in Table 4.1 and their interaction with log time (in months).

Covariates interacted with Log time	Parameter estimate	Standard Error	Wald χ^2	DF	P-value	Hazard Ratio
Employment	-0.720	0.623	1.336	1	0.248	0.487
Functional			1.872	2	0.392	
Ambulatory	-0.908	0.695	1.707	1	0.191	0.403
Bedridden	-0.501	0.680	0.544	1	0.461	0.606
WHO stage			0.111	2	0.776	
Stage III	0.01784	0.66925	0.0007	1	0.9787	1.018
Stage IV	0.11572	0.66857	0.0300	1	0.8626	1.123
OIs	-2.476	1.872	1.748	1	0.186	0.084
TB	-0.471	0.573	0.676	1	0.411	0.624
CD4 cell count	-0.007	0.004	2.631	1	0.105	0.993

Overall test	3.0739	8	0.9296

Table 18: The five highest differences in the parameter estimates of the variables included in the model in Table 4.1 when the data value for each patient is in turn deleted from the model.

Covariates	Deleted Observation(i)	$\Delta_{j(-i)} = \hat{\beta}_j - \beta_{j(-i)}$	$\lvert\Delta_{j(-i)} = \hat{\beta}_j - \beta_{j(-i)}\rvert$
Employment status			
	396	0.08884	0.08884
	89	0.08761	0.08761
	311	0.08579	0.08579
	531	0.08356	0.08356
	434	0.08115	0.08115
Ambulatory	396	0.09222	0.09222
	419	0.08601	0.08601
	166	0.085	0.085
	89	0.08373	0.08373
	339	0.08297	0.08297
Bedridden	339	0.09504	0.09504
	419	0.09297	0.09297
	396	0.09218	0.09218
	166	0.09098	0.09098
	697	0.08723	0.08723
Stage III	532	0.46935	0.46935
	596	0.05018	0.05018
	396	-0.04757	0.04757
	5	-0.04691	0.04691
	334	-0.03813	0.03813
Stage IV	583	0.46644	0.46644
	656	0.05582	0.05582
	593	-0.05564	0.05564
	359	-0.05472	0.05472
	354	-0.04304	0.04304
OIs	244	0.3256	0.3256
	382	-0.07733	0.07733
	262	-0.06102	0.06102

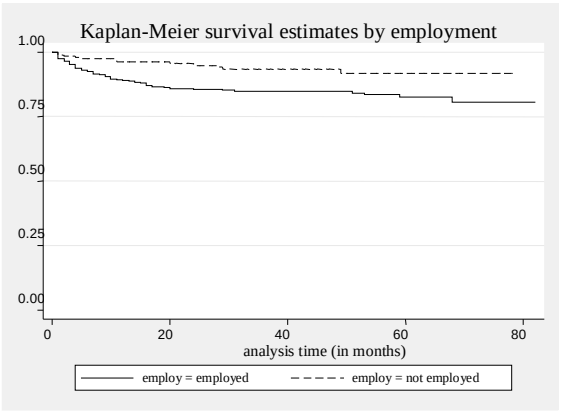
	508	-0.04845	0.04845
	679	-0.0459	0.0459
TB	422	0.05754	0.05754
	464	0.05531	0.05531
	595	0.05052	0.05052
	297	0.05033	0.05033
	521	0.04935	0.04935
CD4	695	-0.00063	0.00063
	319	-0.00058	0.00058
	129	-0.00051	0.00051
	626	-0.00051	0.00051
	152	-0.00037	0.00037

Table 19: Results of the Likelihood ratio, Score and Wald tests for testing the global null hypothesis: BETA=0

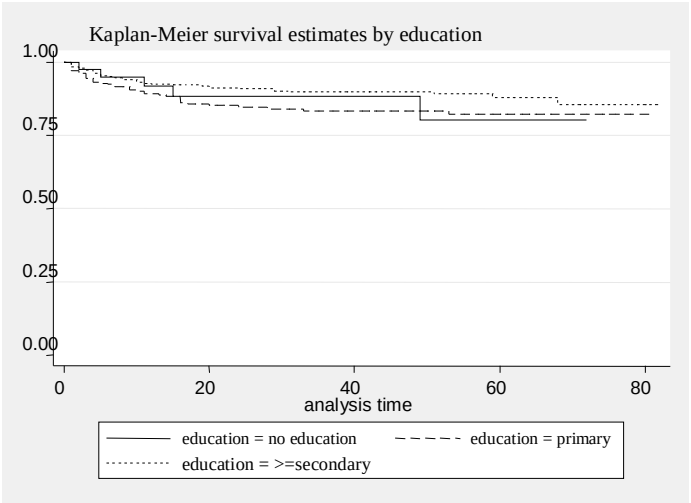
Test	χ^2	DF	P-value
Likelihood Ratio	119.0826	8	<0.0001
Score	107.1392	8	<0.0001
Wald	74.3775	8	<0.0001

Figure 1.1 (a – n): Plots of Kaplan-Meier survivor functions based on different factors, of HIV patients under ART in AFGTH, AA, 2011.

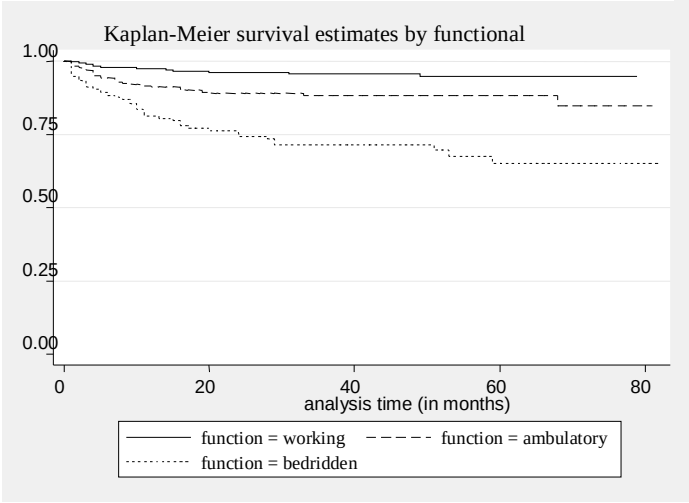
a.


Kaplan-Meier survival estimates by employment
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time (in months)
employ = employed
employ = not employed

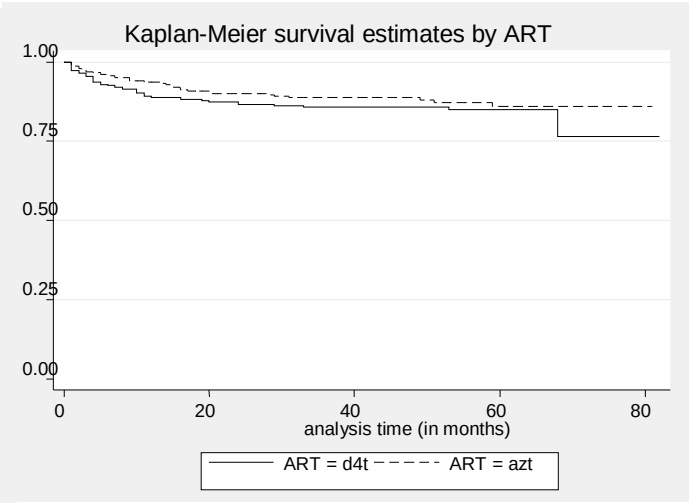
b.


Kaplan-Meier survival estimates by education
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
education = no education
education = primary
education = >=secondary

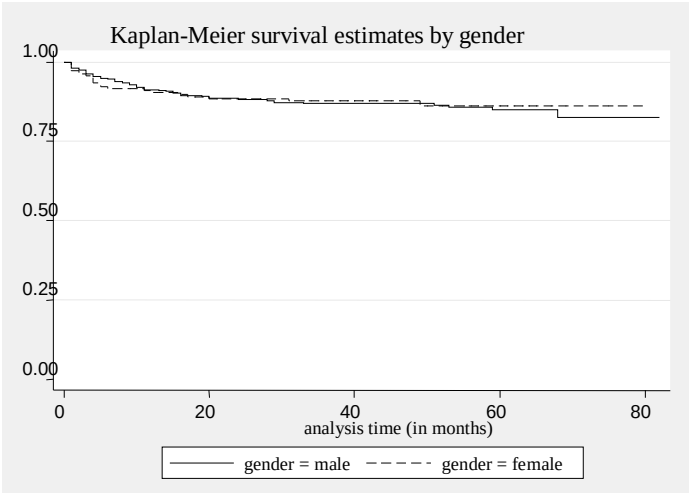
c.


Kaplan-Meier survival estimates by functional
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time (in months)
function = working
function = ambulatory
function = bedridden

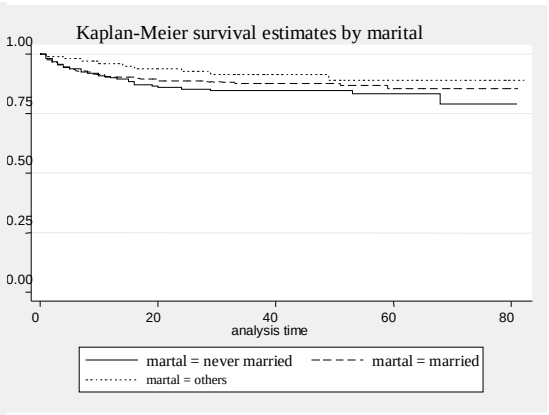
d.


Kaplan-Meier survival estimates by ART
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time (in months)
ART = d4t
ART = azt

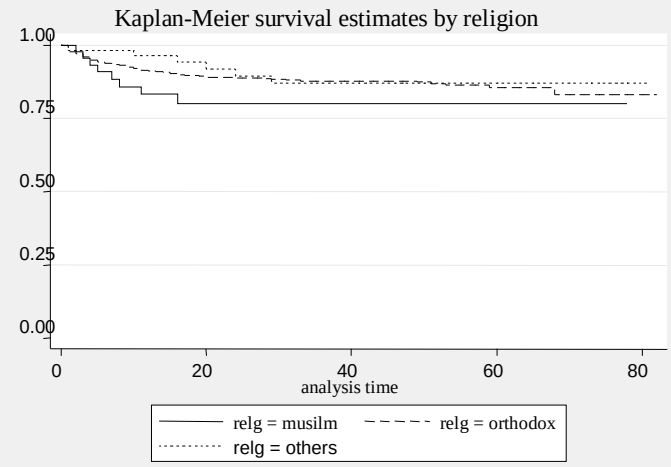
e.


Kaplan-Meier survival estimates by gender
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time (in months)
gender = male
gender = female

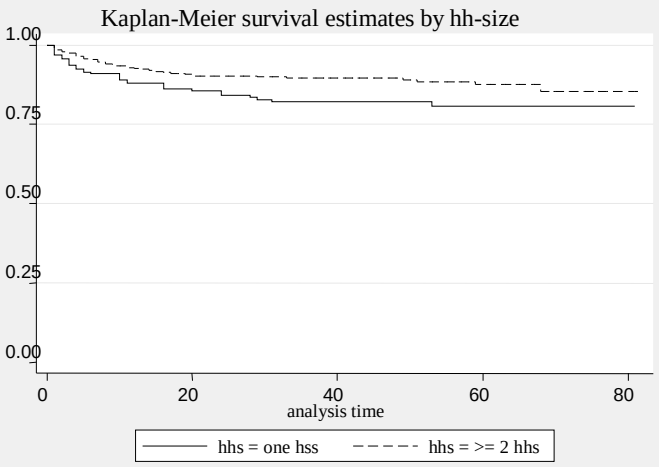
f.


Kaplan-Meier survival estimates by marital
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
martal = never married
martal = married
martal = others

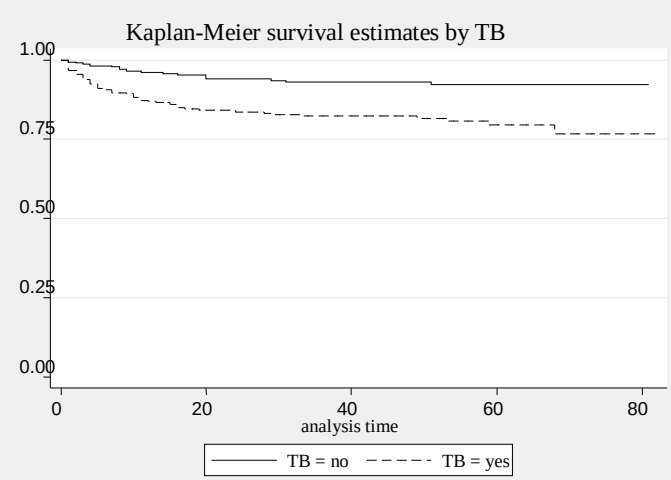
g.


Kaplan-Meier survival estimates by religion
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
relg = musilm
relg = orthodox
relg = others

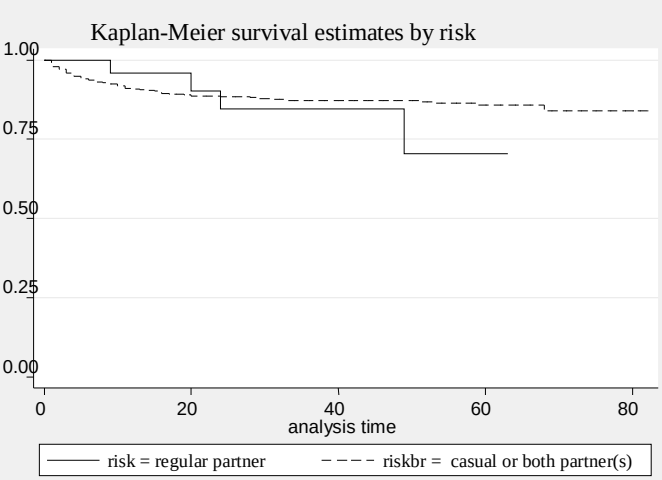
h.


Kaplan-Meier survival estimates by hh-size
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
hhs = one hss
hhs = >= 2 hhs

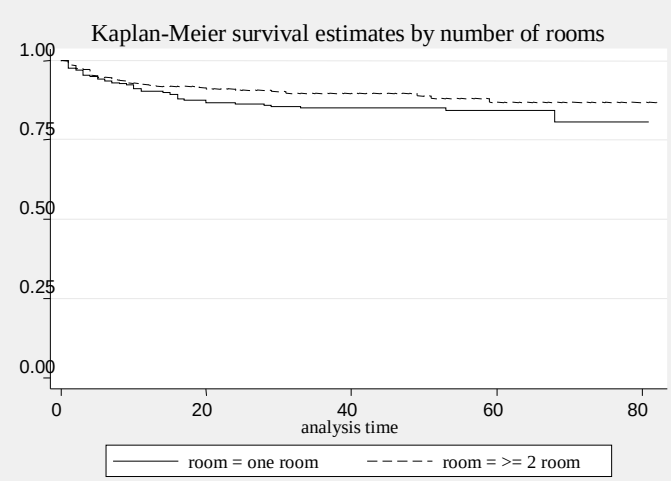
i.


Kaplan-Meier survival estimates by TB
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
TB = no
TB = yes

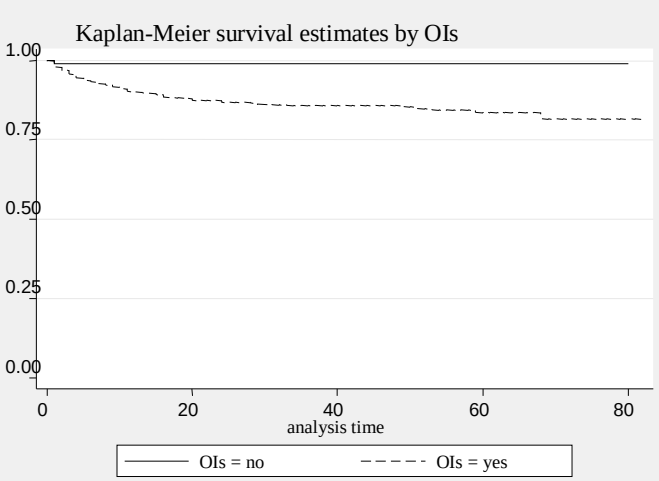
j.


Kaplan-Meier survival estimates by risk
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
risk = regular partner
riskbr = casual or both partner(s)

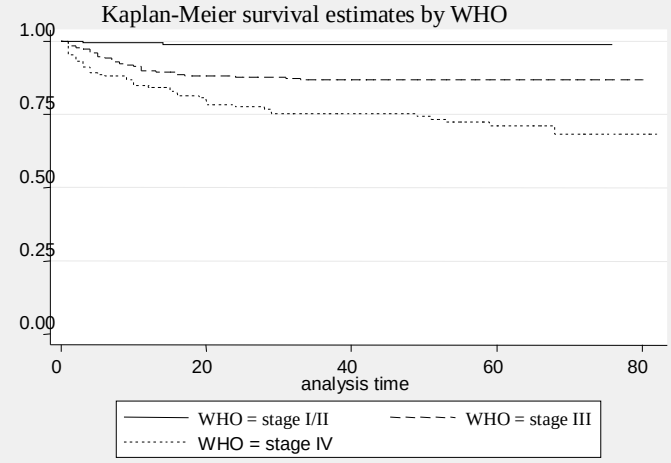
k.


Kaplan-Meier survival estimates by number of rooms
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
room = one room
room = >= 2 room

l.


Kaplan-Meier survival estimates by OIs
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
OIs = no
OIs = yes

m.



n.

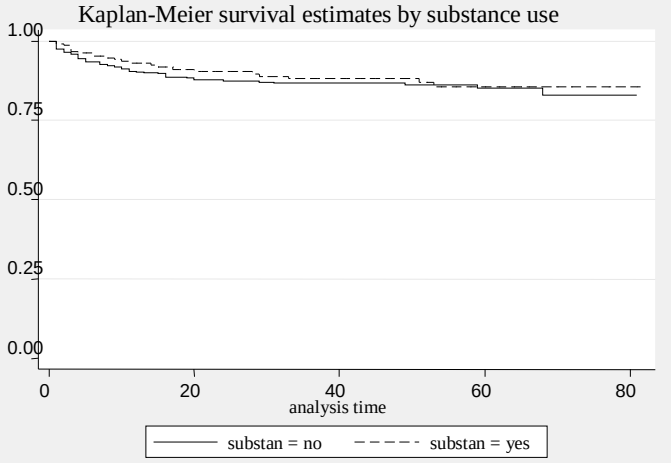
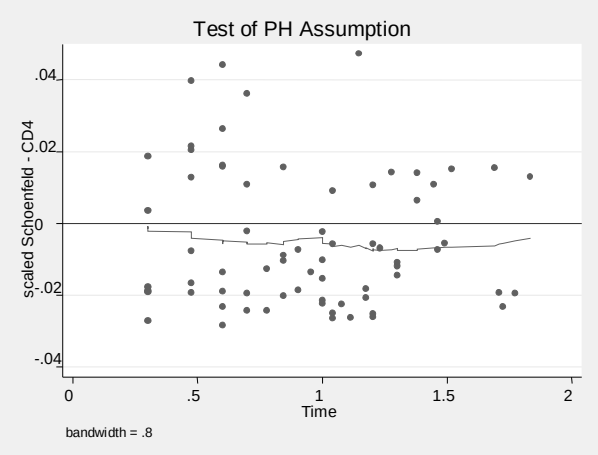
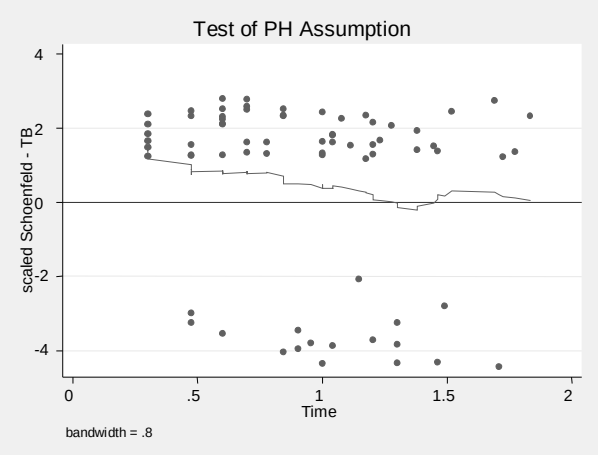


Figure 1.2 (a – h): Graphs of the scaled Schoenfeld residuals and their lowess smooth obtained from the model in Table 4.1 for the covariates CD4 count, TB, OIs, employment status, ambulatory, bedridden, clinical stage III and stage IV. The line that passes through zero is a reference line.

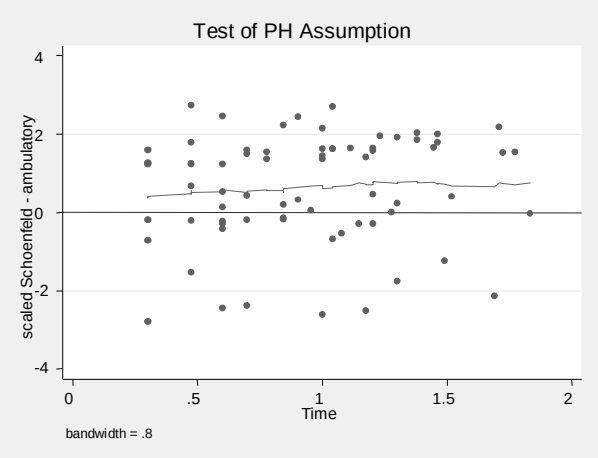
a.



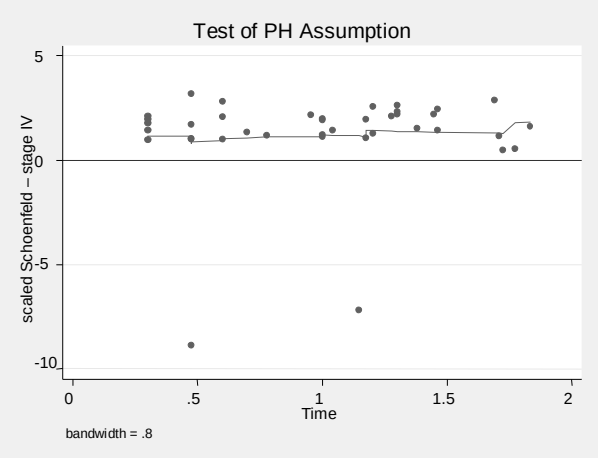
b.



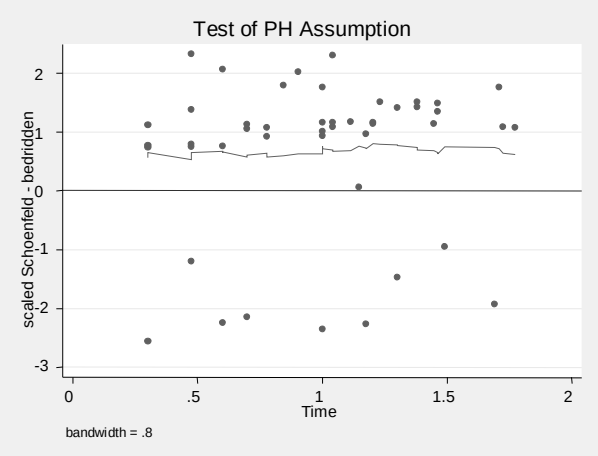
c.


Test of PH Assumption
scaled Schoenfeld - ambulatory
4
2
0
-2
-4
0
.5
1
1.5
2
Time
bandwidth = .8

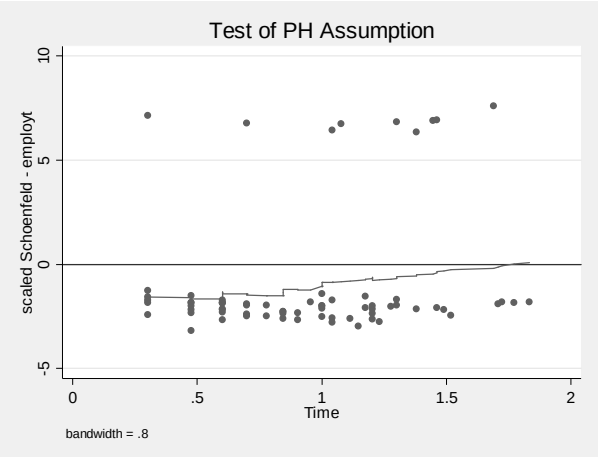
d.


Test of PH Assumption
scaled Schoenfeld – stage IV
5
0
-5
-10
0
.5
1
1.5
2
Time
bandwidth = .8

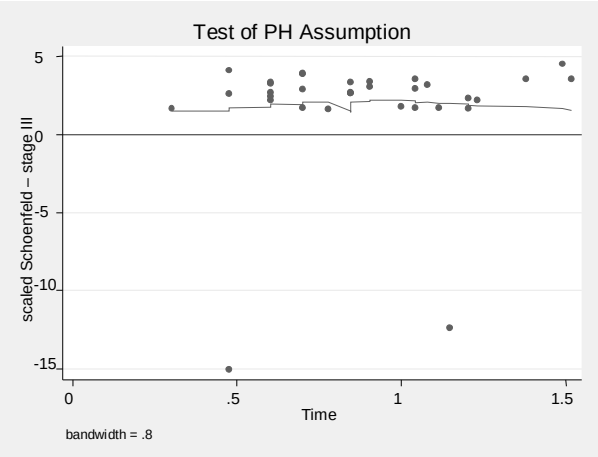
e.


Test of PH Assumption
scaled Schoenfeld - bedridden
2
1
0
-1
-2
-3
0
.5
1
1.5
2
Time
bandwidth = .8

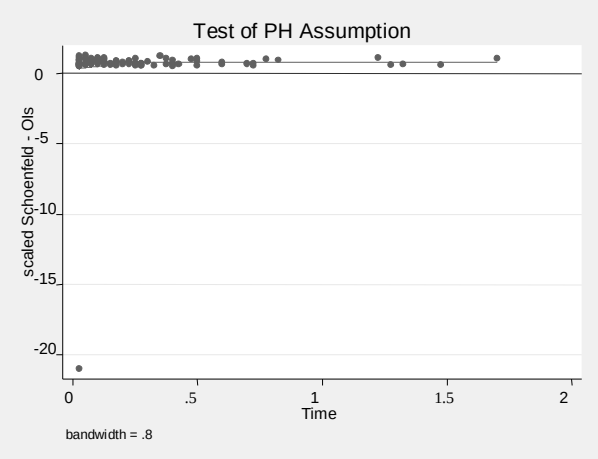
f.


Test of PH Assumption
scaled Schoenfeld - employt
10
5
0
-5
0
.5
1
1.5
2
Time
bandwidth = .8

g.


Test of PH Assumption
scaled Schoenfeld – stage III
5
0
-5
-10
-15
0
.5
1
1.5
Time
bandwidth = .8

h.


Test of PH Assumption
scaled Schoenfeld - OIs
0
-5
-10
-15
-20
0
.5
1
1.5
2
Time
bandwidth = .8

DECLARATION

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

Name: Ketema Kebebew

Signature: ……..

Date: ……

Place: College of Natural Science, Addis Ababa University

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

Name: Professor Eshetu Wencheko

Signature: ……..

Date: ……..

Place: College of Natural Science, Addis Ababa University