ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAMME
DEPARTMENT OF STATISTICS

ASSESSMENT OF FACTORS ASSOCIATED WITH HIGH RISK OF MORTALITY OF HIV PATIENTS TREATED WITH HIGHLY ACTIVE ANTIRETROVAIRAL THERAPY IN JIMMA ZONE, SOUTH WESTERN ETHIOPIA: APPLICATION OF SURVIVAL ANALYSIS METHODS

By
Birtukan Tsehaineh

A thesis submitted to the Office of Graduate Studies of Addis Ababa University in partial fulfilment of the requirements for the Degree of Masters of Science in Statistics (Biostatistics)

September 2010

ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

DEPARTMENT OF STATISTICS

ASSESSMENT OF FACTORS ASSOCIATED WITH HIGH RISK OF MORTALITY OF HIV PATIENTS TREATED WITH HIGHLY ACTIVE ANTIRETROVAIRAL THERAPY IN JIMMA ZONE, SOUTH WESTERN ETHIOPIA: APPLICATION OF SURVIVAL ANALYSIS METHODS

By
Birtukan Tsehaineh

Approved by the board of examiners:

Yilma Tefera (Lecturer)_____ Signature_____
Chairman, Department graduate committee

Mekonnen Taddesse (Asst. Prof.)__ Signature_____
Examiner

Girma Taye (Ph.D.)_____ Signature_____
Examiner

ACKNOWLEDGMENTS

I am sincerely grateful to my advisor Prof. Eshetu Wencheko for his invaluable help, assistance and guidance in the completion of this work. Had it not been for his dedication, this paper would not have its present shape. Thank you Prof.!

My sincere acknowledgment also goes to my friends and colleagues who were there for me every step of the way especially I want to say thank you to Shimelis, Aster, Melat and friends, and my brother Eshetu.

Finally I would like to thank Genet and Yared; data clerks of Jimma University Specialized Hospital ART clinic for helping me collect the data for this study.

'Thank you all!!'

TABLE OF CONTENTS

Pages

List of tables

List of figures

ACRONYMS

3TC	Lamivudine
ADC	AIDS Defining Complex
AIDS	Acquired Immunodeficiency Syndrome
ANOVA	Analysis of Variance
ART	Antiretroviral therapy
ARV	Antiretroviral
BMI	Body Mass Index
CI	Confidence Interval
D4T	Stavudine
EFV	Efavirenz
HAART	Highly Active Antiretroviral Therapy
HCT	HIV Counselling and Testing
HGB	Haemoglobin
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
IDU	Injection Drug Use/User
JUSH	Jimma University Specialized Hospital
LR	Likelihood Ratio
MLE	Maximum Likelihood Estimate/Estimator
NEV	Neverapine
NNRTI	Non-Nucleotide Reverse Transcriptase Inhibitor
NRTI	Nucleotide Reverse Transcriptase Inhibitor
OI	Opportunistic Infection
PI	Protease Inhibitors
PLWHA	People Living with HIV/AIDS
PMTCT	Prevention of Mother to Child Transmission
PYO	Person Years of Observation
RTI	Reverse Transcriptase Inhibitors
TB	Tuberculosis
TLC	Total Lymphocyte Count
VCT	Voluntary Counselling and Testing
WHO	World Health Organization
ZDV	Zidovudine

ABSTRACT

Background: The introduction of HAART has brought about a significant reduction in the morbidity and mortality of patients with HIV infection. However, the mortality rate of patients treated with HAART is still very high in resource-poor settings. Factors contributing to this high mortality rate are poorly understood.

Objective: The objective of this study is to identify the determinant factors of HIV associated mortality in a cohort of HIV infected patients treated with HAART.

Method: The study has reviewed patient forms and follow up cards of 832 patients treated with HAART in Jimma University Specialized Hospital from 2003-2007. The minimum follow up time was 1 month and the maximum was 78 months. Kaplan-Meier survival curves and Log-Rank test were used to compare the survival experience of different groups of patients and proportional hazards Cox model was used to explore the factors associated with increased risk of mortality.

Results: Some 144 patients died during the follow up time of which 48.6% and 68.8% deaths occurred within three and six months of HAART initiation, respectively. The overall mean estimated survival time of patients was 63.7 months. Factors/covariates associated with increased risk of mortality were older age, (HR=1.03, 95% CI: 1.01-1.051), low CD4 count at baseline (HR=0.994, 95% CI: 0.992-0.996), low weight at baseline (HR for a 5kg change=0.902, 95% CI: 0.816-0.996), bedridden and ambulatory functional status (HR=6.904, 95% CI: 4.005-11.902) and (HR=2.877, 95% CI: 1.899-4.360), respectively, co-infection with TB (HR=1.906, 95% CI: 1.305-2.784) and substance use (HR=1.42, 95% CI: 1.016-1.985).

Conclusion: The mortality rate of patients was high at the earlier time of treatment. Moreover, laboratory and clinical factors are associated with increased risk of mortality. Thus, those patients with poor laboratory/clinical characteristics should be identified and treated early before they progress to advanced disease stages.

CHAPTER ONE
INTRODUCTION

1.1 Background

AIDS is an acronym for acquired immune deficiency syndrome, the final disease manifestation from infection by human immunodeficiency virus (HIV). For the first time AIDS was recognized among homosexual men in the United States of America in 1981 and it was only in 1983 that the human immunodeficiency virus was defined as the primary cause of AIDS **(1-3)**.

Today it is one of the largest public health crises endangering the human race. In almost three decades since its first cases were recognised, it has claimed the lives of millions of people making it one of the most devastating epidemics. In the countries most heavily affected, HIV has reduced life expectancy by more than 20 years, slowed economic growth, and deepened household poverty. It has already caused an estimated 25 million deaths worldwide and has generated profound demographic changes. In Sub-Saharan Africa alone, the epidemic has orphaned nearly 12 million children aged less than 18 years. The natural age distribution in many national populations in Sub-Saharan Africa has been dramatically skewed by HIV, with potentially perilous consequences for the transfer of knowledge and values from one generation to the next **(4)**.

The total number of people living with HIV worldwide continued to grow in 2008, reaching an estimated 33.4 million which was more than 20% higher than the number in 2000, and the prevalence was roughly threefold higher than in 1990. Moreover, 2.7 million became newly infected and 2 million died due to AIDS in 2008 alone. Sub-Saharan Africa remains the most heavily affected region, accounting for 71% of all new HIV infections and AIDS related deaths globally in 2008 **(5)**.

There is currently no cure for HIV infection, and viable vaccine candidates are years away. Yet the development of life-saving antiretroviral drugs has brought new hope. In high-income countries, combination antiretroviral therapy has extended and improved life for large numbers of people living with HIV/AIDS and transformed perceptions of HIV/AIDS from a fatal disease to a manageable, chronic illness. In the poorer parts of the world – precisely the regions where HIV/AIDS has spread most rapidly – this transformation has not yet happened **(6)**.

Highly active antiretroviral therapy (HAART) was a breakthrough in care and treatment of people living with HIV, leading to a reduction in mortality and an improvement in the quality of life. Antiretroviral drugs significantly lowered the rate of HIV transmission from mother to child, and antiretroviral therapy (ART) has become an integral part of the continuum of HIV care **(7)**.

The earliest evidence of HIV infection in Ethiopia was found in 1984, with the first case reported in 1986. The major avenue of transmission of HIV infection in Ethiopia is heterosexual intercourse and the practice of multiple sexual partnerships, particularly in urban areas. Illegal medical practices and harmful traditional practices are also potential routes of transmission. The prevalence of HIV was relatively low in the 1980's but it increased rapidly in the 1990's **(8, 9)**. Since then it has become a major public health and development problem and was declared an emergency in 2002 by the Ethiopian government **(8)**.

Adult HIV prevalence in 2009 was estimated to be between 1.4% and 2.8% in the country. The last single point estimate exercise which was done in 2007, estimated urban and rural prevalence of 7.7% and 0.9%, respectively, for 2009. Prevalence was 1.8% for males and 2.8% for females, and women accounted for 59% of the HIV-positive population. According to the 2007 single point estimate, there were an estimated 1,116,216 people living with HIV in 2009, of which 336,160 were eligible for ART. There were an estimated 131,145 new HIV infections and 44,751 AIDS-related deaths of which females accounted for 57% of the total infections and deaths. The total estimated number of HIV-positive pregnant women and annual HIV positive births in the same year were 84,189 and 14,140, respectively. There were an estimated 72,945 children less than 15 years old living with HIV, out of which 20,522 needed ART. Due to the combined effect of poverty and AIDS, more than 5.4 million children under the age of 18 years were orphaned out of which 855,720 (16%) lost at least one parent due to AIDS **(10)**.

As things stand today in the country, the provision of free antiretroviral drugs is taken as one of the treatment options to reduce the effects/consequences of the epidemic. In this regard the HIV Counselling and Testing (HCT) program has shown considerable improvement both in terms of service expansion as well as utilization. The number of health facilities providing ART reached 512 in December 2009. A total of 5.8 million people (53% male) received HIV counselling and testing in 2008/09, this is a 22% increase from the previous year. ART coverage increased from 46% in 2008 to 53% in 2009. As of end of 2009 there were a total of

241,236 people ever started ART and 176,644 currently on ART. Females accounted for 57.9% of ART clients. A total of 11,000 children were ever started ART, including 8,761 currently on ART as of December 2009. ART coverage for children was 43%. However, lost to follow up to ART service was 28% by the end of 2008 which is a major challenge **(10)**.

Almost 30 years have now elapsed. Thirty years, in which HIV infection has changed from a fatal condition to a manageable chronic illness. Thirty years, in which the development of ART has been one of the dramatic advances in the history of medicine. However, for the vast majority of people living with HIV/AIDS, ART is still light years away, largely inaccessible in resource- poor countries where HIV continues to devastate families, communities and societies, especially the poor and the socially marginalized **(11)**. Of the estimated 9.5 million people in need of treatment in 2008 in low- and middle-income countries, only 42% had access. Further more all indications point to the number of people needing treatment rising dramatically over the next few years **(5)**.Thus, without rapid access to properly managed treatment, these millions of women, children and men will die **(6)**.

This human toll and the accompanying social and economic devastation can be averted. The delivery of ART in resource-poor settings, once thought impossible, has been shown to be feasible. The prices of antiretroviral drugs, which until recently put them far beyond the reach of low-income countries, have dropped sharply. A growing worldwide political mobilization, led by people living with HIV/AIDS, has educated communities and governments, affirming treatment as a human right. The World Bank has channelled increased funding into HIV/AIDS. New institutions such as the Global Fund to Fight AIDS, Tuberculosis and Malaria and ambitious bilateral programmes, including the United States Presidential Emergency Plan for AIDS Relief, have been launched, reflecting an exceptional level of political will and unprecedented resources for the HIV/AIDS battle. However, there still remain a lot that needs to be done in this regard **(6)**.

1.2 Statement of the problem

It is a well known fact that Ethiopia is one of the most countries hit by the HIV/AIDS epidemic. The epidemic has claimed the lives of the country's adult that would otherwise have contributed immensely to the country's development in many respects. Ever since the first infections were identified in the country, a lot has been done both by the government and other stake holders to curb the effect of the disease specially to suppress the risk of transmission.

Numerous researches have also been conducted that tried to address many of the issues that arise in connection with the HIV epidemic. One thing that should, however, be noted is that many of these research works mainly focused on the assessment of the prevalence and the study of the numerous prevention measures that should be undertaken to stop or reduce the spread of the epidemic. Little attention seems to have been given to study the situation of those living with the virus. Especially when seen in light of the current situation in the country where we have over 1 million individuals living with the virus and over 1 billion birr allocated for HIV mobilization in 2008/2009 only **(10)**, it is important that research works focus on the study of PLWHA situation. This is more so when we see the large number of HIV positive people who are following ART.

Thus, it should be stated at this juncture that researches should be conducted to evaluate the effectiveness of the free antiretroviral drugs treatment option. Though research work is not non-existent in this area, many of them focused on the study of drug adherence. Save for a few studies, many of the factors that can possibly affect the survival of people who are in ART follow up remain unstudied. And in most cases these studies were limited to urban areas mainly Addis Ababa.

The effectiveness of HAART could vary from region to region (this variation is also generally a reflection of the variation that exists between and within countries and regions as regards HIV prevalence and its epidemiological patterns) because of the difference in background disease burden (such as tuberculosis or intestinal parasites), viral subtypes, and possible genetic differences in drug metabolism. However, such arguments are based on little data from the resource-limited settings **(5, 48-50)**.

Given this as a back drop, this thesis will focus on the consideration of some of the possible factors/variables that may possibly influence the survival status of people who are following ART in South West Ethiopia. The study of these factors will provide information to put in place efficient ART system that will take into account the huge cost that is involved in the system.

1.2 Objectives of the study

The general objective of the study is:

To explore the determinant factors of HIV associated mortality in a cohort of HIV infected patients treated with HAART in Jimma University Specialized Hospital using survival analysis based on non-parametric and semi-parametric methods.

The specific objectives of the study are:

1. To estimate the survival time of HIV patients treated with HAART.
2. To compare survival time among the different groups of HIV patients treated with HAART.
3. To determine the factors and/or covariates that affect the survival of HIV patients treated with HAART.
4. To demonstrate the application of survival analysis techniques to the data.

1.4 Applications of the result

1. The findings from this study will be expected to give information for public health practitioners and stakeholders who are working in the areas of giving care, support and treatment for HIV/AIDS patients.
2. The findings of this study will provide empirical evidence for program planners, decision makers and ART program implementers at different levels by enabling them to access a baseline data on predictors of survival on PLWHA after the advent of HAART.
3. The study will also aid HIV patients by providing information about those socio-demographic factors and baseline clinical/laboratory characteristics, which have influence on the survival of patients, so that they can improve their quality of life accordingly.
4. It is also believed that results of the research will be used as a basis for further study in the area.

1.5 Limitation of the study

1. The study is conducted based on secondary data which might have incomplete and biased information.
2. The study was restricted to adults, and results might not be applicable to infants and children.
3. Underestimation of mortality due to lost to follow up patients included in the study. In addition to this all deaths are assumed to be caused by AIDS.
4. The study includes only baseline values of the variables. i.e. CD4 cell count stability or improvement, weight lose or gain, treatment adherence, treatment switches or substitution, number of missed appointments, which are associated to mortality of AIDS patients, are not included in the study.
5. The study used CD4 cell count as indirect serogate indicator instead of viral load. Also the study used weight, which might be affected by height, instead of BMI.

CHAPTER TWO
LITERATURE REVIEW

2.1. General overview on the use of ART

Antiretroviral medications are designed to inhibit the reproduction of HIV in the body. The main effect of antiretroviral treatment is to suppress viral replication, allowing the individual's immune system to recover and protect him/her from the development of AIDS and death. In other words what this means is that if ART is effective, the deterioration of the immune system and the onset of AIDS can be delayed for years thereby improving the quality of life of the victims. Standard ART (also known as HAART) consists of the use of at least three ARV drugs to maximally suppress the HIV virus and stop the progression of HIV disease **(12, 13)**.

ART changes the natural history of HIV infection. Results from medical studies on HAART have been extremely impressive. Since its introduction in 1996, mortality and morbidity rates in HIV-infected individuals in countries with widespread access to HAART have plummeted. The use of antiretroviral medicines dramatically reduced AIDS related illnesses and death in countries where these drugs are widely accessible **(4, 14)**. Although the treatments are not a cure and continue to present new challenges with respect to side-effects and drug resistance ART as disease modifying therapy for established HIV infection has produced dramatic effects on morbidity and mortality among HIV-infected patients. As a result of the widespread use of ART, the HIV/AIDS pandemic which was once regarded as an infectious disease with an almost universal fatal outcome has been transformed into a manageable chronic infectious disease **(15, 16)**.

A study in the US conducted from January 1994 through June 1997 which evaluated 1255 patients who were using antiretrovirals revealed that mortality among the patients declined from 29.4 per 100 person-years in 1995 to 8.8 per 100 person-years in the second quarter of 1997. The incidence of major opportunistic infections (OI) also declined from 21.9 per 100 person-years in 1994 to 3.7 per 100 person-years by mid-1997. There were also reductions in mortality and incidence of opportunistic infections regardless of sex, race, age, and risk factors for transmission of HIV **(14)**.

Another multinational study conducted in Europe also demonstrated that mortality rates for HIV-infected persons have become much closer to general mortality rates since the introduction of HAART. According to this study, persons infected sexually with HIV now appear to experience mortality rates similar to those of the general population in the first five years following infection though a mortality excess remains as duration of HIV infection lengthens **(17)**. However, in developing countries where we have most of the 33.4 million people currently living with HIV/AIDS the opportunities created by ART have not been shared. For instance, in 2003 the WHO estimated that by the end of the year there were close to 6 million people in developing countries were in immediate need of life sustaining ART. However, it was only about 400,000 persons who were being treated **(6)**. This situation does not seem to have changed much though there are noticeable developments. Of the estimated 9.5 million people in need of treatment in 2008 in low- and middle-income countries, only 42% had access, up from 33% in 2007 **(5)**.

Moreover, for resource limited countries like Ethiopia the WHO uses a standardized approach of public health including ARV drugs for the purpose of reaching as many patients as possible **(16)**. As a result the most common first line drug regimen combinations given to those patients in such countries usually consists 2NRTIs (either D4T + 3TC or ZDV +3TC) + 1NNRTI (either NVP or EFV). The above formula gives the following four possible combinations: D4T + 3TC + NVP, D4T + 3TC + EFV, ZDV +3TC + NVP, ZDV +3TC + EFV **(7)**. This gives a better option for patients to change and substitute one regimen by another in the same class during side effects, drug-drug interactions, toxicity, pregnancy etc. For example, it is possible to substitute NVP in place of EFV to prevent vertical transmission of HIV and avoid teratogenicity during pregnancy.

Despite problems associated with access to the drugs, lack of trained manpower and laboratory facilities, numerous researches conducted in resource limited settings on the provision of ART have shown positive outcomes. For example, in an article that reviewed 28 other articles involving studies conducted in 14 African countries, positive health outcomes were reported in addition to high level treatment adherence comparable to those in the industrialised world showing that provision of ART in these countries is feasible **(18)**. A similar outcome was also reported in a cross-sectional study conducted on 137 patients who were under ART in Uganda. When HAART was initiated, it was reported that 93% of these patients were unable to hold a job. However, 85 % of them reported feeling “good” to “excellent” after at least 12 weeks of

treatment, and 96% of them reported better performance at home or work **(19)**. The WHO also reported that the sustained progress in expanding access to HIV prevention and treatment and care services in low- and middle-income countries has increasingly shown positive effects of scaling up treatment on mortality **(20)**.

A final point that needs to be highlighted is the fact that despite its advantages ART is not an easy process. For its optimal use the treatment needs highly qualified personnel and sophisticated laboratory facilities in addition to high level of commitment on the side of patients, which makes it difficult as it is a treatment that will have to be taken for life. In controlled clinical trials, it was shown that six different combinations of ARV drugs have reduced the viral load to less than 500 copies/ml in 60 to 90% of patients **(21-23)**. However, since ART involves complicated dosing schedules, side effects in a significant number of patients, and a high degree of adherence (almost more than 95%) to maintain viral suppression, numerous studies outside clinical trials have shown that between 10–50% of patients who start HAART fail to achieve adequate virological control **(24, 25)**. This in the long run reduces the effectiveness of the drugs and leads to the development of drug resistant strains **(26-27)**. And thus, failure to achieve adequate virological control will further result in the recurrences of illness and with it the inevitability of decreased quality of life and premature death.

This being generally the case it should however be stated that most African studies relate only to short term follow up which in most cases is 12 months unlike the case in the industrialized world where HAART management gets more difficult with time due to increasing toxicity, decreasing adherence, treatment failure & switches and emergence of drug resistance **(28)**.

2.2. Predictors of Mortality/Survival

A study was conducted in Tanzania with the objective of assessing mortality and to identify predictors of mortality in HIV-infected patients starting ART in one rural hospital. The study used Kaplan-Meier models to estimate mortality and Cox proportional hazards models to identify predictors of mortality and it was found that male sex, severe malnutrition and WHO stage IV were associated to progression to death and no such associations were found for age, religion, education level and active TB in the univariable analysis **(29)**. A similar retrospective cohort study in northern Thailand, which used the same method of analysis, also revealed that

sex, age group, registered year, clinical status, CD4 group, and ARV drug group were all significantly related to death in the univariable analysis, but the association with sex and age group was not significant in the multivariate analysis **(45)**.

To assess the effect of adherence to HAART on survival **(30)** has conducted a retrospective cohort study on 897 Ugandan patients who are treated with HAART between May 2004 and December 2006. Kaplan-Meier curves and Cox proportional hazards regression model were used in the analysis. In this study, which found out that non-adherence to HAART was significantly associated with mortality, it was shown that patients with a CD4 count of less than 50 cells/mm^3 had a higher risk of mortality (HR = 4.3; 95% CI: 2.22–5.56) compared to patients with a CD4 count equal to or greater than 50 cells/mm^3 (HR = 2.4; 95% CI: 1.79–2.38).

A collaborative analysis of prospective studies which includes 13 cohort studies from Europe and North America assessed the Prognosis of 12,574 adult HIV-1-infected patients who were on HAART. The end points considered were progression to a new AIDS-defining disease or death and to death alone. The prognostic model that generalised best was a Weibull model, stratified by baseline CD4 cell count and transmission group. Baseline CD4 cell count was strongly associated with the probability of progression to AIDS or death: compared with patients starting HAART with less than 50 CD4 cells/µl, adjusted HRs were 0·74 (95% CI: 0·62–0·89) for 50–99 cells/µl, 0·52 (95% CI: 0·44–0·63) for 100–199 cells/µl, 0·24 (95% CI :0·20–0·30) for 200–349 cells/µl, and 0·18 (95% CI :0·14–0·22) for 350 or more CD4 cells/µl. Other independent predictor of poorer outcome was advanced age **(31)**.

Another study which included 1,308 patients on HAART was conducted in rural Malawi. The study used Product-limit estimates (Kaplan-Meier) to asses the probability of survival and Cox model to asses the independent predictors of death after initiation of HAART. Low BMI, WHO stage IV, male sex, and baseline CD4 count lower than 50 cells/µl were independent determinants of death in the first 6 months. However, age at HAART initiation and regimen types are not significantly associated with early death **(32)**.

To assess the incidence of mortality, clinical characteristics and outcome of co-infection with HIV and TB, **(33)** conducted a cross-sectional study of co-infected cases reported from nine domestic hospitals throughout mainland China. Treatments for TB and HIV were provided to

241 patients and mortality attributable to co-infection was reported for 15.8% of the cases. Immune function among most patients was suppressed based on reduced CD4 cell counts. HIV/TB co-infection was related to high mortality even when HAART and/or drug therapy for TB was provided. It was found that fever, fatigue, and weight loss were experienced by the majority of HIV/TB patients (87.6%, 61.0% and 60.2%, respectively), making these the leading clinical symptoms among co-infected patients. Another study conducted in Baltimore USA also revealed that, accounting for potential confounders, including CD4 cell count and viral load, the hazard of AIDS-related death was 2.4 times more for the person-time with TB compared to the person-time without TB **(34)**.

A cohort study in Abidjan in which 792 adults started ART, with a median CD4 cell count of 252 cells/ml and were followed for a median of 8 months, showed that patients who experienced severe morbidity had higher risks of mortality, virological failure and immunological failure. Other independent risk factors for mortality and/or severe morbidity were at baseline: high viral load, advanced clinical stage, past history of tuberculosis, low BMI, low haemoglobin and low CD4 cell count; and during follow: low CD4 cell count and persistently detectable viral load **(35)**.

To evaluate sex differences in HIV disease progression before (pre-1997) and after (post-1997) the introduction of HAART, a study which investigated 6,923 seroconverters from a collaboration of 23 HIV seroconverter cohort studies from Europe, Australia, and Canada was conducted. Within a competing risk framework, the authors used Cox proportional hazards models allowing for late entry to evaluate sex differences in time from HIV seroconversion to death, to AIDS, and to each first AIDS-defining disease and death without AIDS. While no significant sex differences were found before 1997, from 1997 onward, women had a lower risk of AIDS and death than men did. Compared with men, women also had lower risks of AIDS dementia complex, tuberculosis, Kaposi's sarcoma, lymphomas, and death without AIDS. According to this research sex differences in HIV disease progression have become larger and statistically significant in the era of HAART, supporting a stronger impact of health interventions among women **(36)**.

Although the consensus is that gender does not influence HIV progression, its relevance may depend on the setting. In a Spanish study, conducted to study gender differences in HIV progression to AIDS and death in a cohort of HIV infected injection drug users (IDUs),

Kaplan-Meier methods and Cox regression adjusting for gender, age, and calendar period fitted as time dependent covariates were employed. Of the 929 IDU, 24.7% were women. ART was initiated in 44% of women and 34% of men. Risk of AIDS was lower in women in both univariate and multivariate analyses although it is not statistically significant. In general, the study concluded that HIV progression was lower in female IDUs before and after the introduction of HAART and their uptake of ART was higher than male IDUs **(38)**.

Population-based, prospective cohort study was conducted on 3,116 antiretroviral naive HIV-infected patients in a province-wide HIV/AIDS treatment program in British Columbia, Canada, to compare survival rates among HIV-infected patients initiating HAART with and without a history of injection drug use. The main outcome measure of the study is all-cause mortality. Of the 3,116 patients, 915 were IDUs (29.4%), 579 were female (18.6%), and the median age was 39.4 years. Treatment with HAART was initiated between August 1, 1996, and June 30, 2006. The median duration of follow-up was 5.3 years for IDUs and 4.3 years for non-IDUs. Patients were followed up until June 30, 2007. At 84 months after the initiation of HAART, the product limit estimate of the cumulative all-cause mortality rate was similar between the IDUs (26.5%) and non-IDUs (21.6%) (Wilcoxon P=0.47). The time-updated Cox regression also showed that, the hazard ratio of mortality was similar between IDUs and non-IDUs after adjustment for baseline CD4 cell count, adherence, and physician experience. In general injection drug use was not associated with decreased survival among HIV-infected patients initiating HAART in this study **(37)**.

In Uganda a retrospective cohort of 427 HIV-1 patients who were initiated on ART was studied to establish the effect of AIDS defining complexes (ADCs) on immunological recovery among patients initiated on ART. Kaplan-Meier survival curves were employed to estimate median times, log rank test to compare different categories and Cox proportional hazard models were used at multivariate analysis. The median time to gaining 50 CD4 cells/µl from the baseline value after ART initiation was longer in the ADC group (9.3 months) compared to the non-ADC group (6.9 months) (log rank test, $p = 0.027$). At multivariate analysis after adjusting for age, sex, baseline CD4 count, baseline HIV viral load, total lymphocyte count and adherence level, factors that shortened the median time to immunological recovery after ART initiation were belonging to the non-ADC group (HR = 1.31, $p = 0.028$), adherence to ART of $\geq 95\%$ (HR = 2.22, $p = 0.001$) and a total lymphocyte count ≥ 1200 cells/mm^3 (HR = 1.84, $p = 0.003$). A low baseline CD4 count of ≤ 200 cells/µl (HR = 0.52, $p = 0.001$) was associated with

a longer time to immunological recovery. There was no interaction between low CD4 counts and ADC group. Patients with ADCs take longer to regain their CD4 counts due to the defect in the immune system. This may prolong their risk of morbidity and mortality **(39)**.

A prospective cohort study conducted to determine the long-term incidence of TB and associated risk factors among individuals receiving HAART in South Africa found that risk of TB was independently associated with CD4 cell count < 100 cells/ml (P = 0.04),WHO stage III or IV disease (P = 0.01) and age < 33 years (P = 0.01). Risk of TB was not independently associated with plasma viral load, previous history of TB, low socioeconomic status or sex. Despite similar virological responses to HAART, blood CD4 cell count increases were much smaller among patients who developed TB than among those who remained free of TB **(40)**.

In Peru 564 initially antiretroviral-naive HIV-infected persons who received combination ART were followed over time to asses the prolonged effectiveness of ART in a developing country. Comparison of categorical data between groups was evaluated by the χ^2 test, and comparison of continuous data between groups was evaluated by the Student's t-test and Wilcoxon test. Survival after the initiation of ART was estimated using the method of Kaplan-Meier. The incidence of opportunistic infections (OIs) was estimated using cumulative incidence estimates. Statistical comparison for time-to-event data was performed using the likelihood ratio test from Cox regression models. The overall survival rate was 96% at year 2, 94% at year 4, and 91% at year 5. Among persons who initiated therapy with CD4 counts <100 cells/mm^3, the overall survival rate at 3 years was 95%. OIs while on ART occurred in 20% of persons. Patients who received 2 RTI plus a PI had slightly better survival rates and less opportunistic disease in the first year of therapy as compared with those receiving 2 RTI and a NNRTI or 3 RTI **(41)**.

To determine the relationship between mortality risk and the CD4 cell response to ART, a cohort of 2,423 patients on ART and who had a median baseline CD4 105 cells/µl were observed for up to 5 years of ART in South Africa. Kaplan-Meier analyses were used to estimate cumulative mortality, univariate and multivariate mixed effect Poisson regression models were used to estimate the association between baseline risk factors, updated CD4 cell counts and viral load and the incidence of mortality. In addition Wilcoxon rank-sum tests were used to compare medians. Older age, WHO clinical stage 4, updated CD4 cell counts and

detectable updated viral load measurements were all significantly associated with mortality risk in both the crude analyses and the multivariate model with the exception of male sex which is significant in the crude analysis only. However, updated CD4 cell counts were the variable most strongly associated with death **(42)**.

Defaulting diminishes the immunological benefit of ART and increases AIDS-related morbidity, mortality and hospitalizations. In resource-constrained settings where the health care services are not well developed, poor adherence to treatment and defaulting from treatment are the two major challenges that ART programs face **(51)**. To determine the prevalence and factors associated with defaulting ART in Jimma, Ethiopia, unmatched case control study was conducted. Taking hard drugs (cocaine, cannabis and IV drugs), excessive alcohol consumption, being bedridden, living outside Jimma town and having HIV negative or unknown HIV status partner were independently associated with defaulting ART in the logistic regression analysis **(43)**. A French study conducted to assess adherence to HAART in a cohort of patients infected by HIV through IDU has also shown that non-adherence was associated with younger age, alcohol consumption, frequency of negative life-events during the prior 6 months and active drug use after adjustment by logistic regression **(44)**.

To identify prognostic indicators of survival at different CD4+ cell levels, independent of HAART, among IDUs a study was conducted in Baltimore. The outcome of interest was HIV-related death, defined as death after an AIDS diagnosis or a diagnosis of another infectious disease such as sepsis or endocarditis. Analysis using the Cox proportional hazards model showed that prior hospitalization and AIDS diagnosis were independent consistent risk factors for HIV mortality, whereas outpatient/emergency room (ER) visits, sepsis or endocarditis and alcohol use were significantly associated with higher mortality, specifically for the low CD4 count group. It was also shown that survival among HIV-infected IDUs improved since the introduction of HAART, although utilization of HAART was incomplete **(46)**.

In the era of HAART, maximizing health-related quality of life has become a high priority of long-term management of HIV-infected individuals. To identify the predictors for lower quality of life among HAART-using HIV infected homosexual and bisexual men participants in 4 cities of USA, a longitudinal cohort study was conducted and predictors were assessed using the random effects model. Quality of life before HAART initiation was a strong predictor of quality of life subsequent to HAART initiation. Older age, lower socioeconomic status, less

male sexual partners, no alcohol drinking, and more advanced HIV disease stage were significant predictors for lower physical health summary score. In addition, more outpatient visits, depression, amprenavir use, antiretroviral drug interruption, recreational drug use, and less social support were significantly associated with lower mental health summary score **(47)**.

In south Ethiopia a series of studies were conducted on HIV infected patients to investigate the risk factors for HIV disease progression among untreated patients, to assess the effect of ART on patient mortality and tuberculosis incidence rate and to identify risk factors for mortality in patients treated with ART. Survival analysis methods such as log rank test, Kaplan Mayer and Cox regression models were employed in the study and the main outcome measures were time to death and time to diagnosis of tuberculosis. Among 207 untreated patients followed for a median duration of 19 months the mortality rate was 46 per 100 PYO, mortality increased with advanced disease stage and diarrhoea, oral thrush and low total lymphocyte count were significant markers of mortality. In addition among 152 patients treated with HAART the overall mortality was 16.7 per 100 PYO[1]. The highest death rate occurred in the first month of treatment. Being in WHO clinical stage IV and having TLC<=750/mcL were independent predictors of death but BMI <=18.5 kg/m^2 at baseline was associated with death in univariate analysis. More over on the follow up data; decline in TLC, HGB and BMI was associated with death in univariate analysis only **(48-50)**.

[1] Person year of observation (PYO) = number of observation multiplied by the number of years.

CHAPTER THREE
DATA AND METHODOLOGY

3.1. The Data

The data for this study were obtained from the ART clinic in Jimma University Specialized Hospital (JUSH) in Jimma town, Ethiopia. The hospital serves as a teaching and referral centre for the population of Jimma zone and other adjacent zones and regions in South-West Ethiopia. It has a separate unit for the ART program. JUSH launched ART treatment service in July 2003 for PLWHA who are financially strong and have the capability to pay for ART according to the national treatment guidelines. Later by the year 2005 the hospital started to register and give free ART services for those who fulfil standard diagnostic criteria to start ART.

In addition to adherence counselling and providing ART for HIV patients, the clinic also provides screening, follow up and referral services for TB patients. All diagnosed HIV positive individuals from VCT, PMTCT or inpatient services were sent to the ART centre and registered. For those HIV patients, who were eligible to start ART according to the national treatment guideline, a unique ART number was given to each patient.

This study reviews patient's intake forms and follow-up cards of HIV patients taking HAART in JUSH ART clinic. The patients' chart include the patient intake forms and follow-up cards, which are prepared by Federal Ministry of Health (FMOH) to be uniformly used by clinicians to simply identify and document clinical and laboratory variables. A total of 832 patients in the clinic, on whom HAART was initiated between the times October 2003 and August 2007, were included in the study and they were followed up until March 30, 2010. The data were collected by two data clerks working in the clinic and coded and analyzed using the statistical packages SPSS 13, SAS 9.2 and Stata 9.2.

3.2. Variables included in the study

3.2.1. The response (dependent) Variable:

The response/outcome variable is the survival time of HIV patients, the length of time from ART start date until the date of death/censor measured in months.

3.2.2. Predictor (independent) Variables:

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

- Age in completed years
- Sex (Male, Female)
- Marital status (Never Married, Married, Others)
- Level of Education (No education, Primary, Secondary and above)
- Religion (Muslim, Orthodox, Others)
- Employment status (yes, no)
- Residence (Jimma, Out of Jimma)
- Past opportunistic infection (yes, no)
- TB incidence (yes, no)
- HIV disclosure (yes, no)
- Number of rooms (one, two or more)
- Risk behaviour (Regular sexual partner, Casual or both sexual partner)
- Substance (Tobacco, Alcohol, Soft drugs) use (yes, no)
- Functional status (Working, Ambulatory and Bedridden)
- WHO clinical stage (Stage I or Stage II, Stage III and Stage IV)
- Baseline CD4 cells count
- Baseline weight
- Regimen (AZT-based, D4T-based)

3.3. The methodology: Survival analysis

Survival analysis is the phrase used to describe the analysis of data in the form of a well defined time origin until the occurrence of some particular event or end point. Generally, survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. The problem of analyzing time-to-event data arises in a number of applied fields such as medicine, biology, public health, epidemiology, engineering, economics and demography. In medical research, the time origin will often correspond to the recruitment of an individual into an experimental study, such as a clinical trial to compare two or more treatments. This in turn may coincide with the diagnosis of a particular condition, the commencement of a treatment regimen, or the occurrence of some adverse event. If the end point is the death of a patient, the resulting data are literally survival times. However, data of a similar form can be obtained when the end point is not fatal, such as

the relief of a pain, or the recurrence of symptoms. In this case the observations are often referred to as time-to-event data.

One of the most important differences between the outcome variables modelled via linear and logistic regression analyses and the time variable in the survival data is the fact that we may only observe the survival time partially. The variable time actually records two different things. For those subjects who experienced the event (most of the time death), it is the outcome variable of interest, the actual survival time. However, for subjects who were alive at the end of the study, or for subjects who were lost to follow-up, time indicates the length of follow-up (which is a partial or incomplete observation of survival time). These incomplete observations are referred to as being censored. Most survival analyses consider a key analytical problem called censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly.

Survival data are not in general amenable to standard statistical procedures, such as mean, standard deviation and ANOVA, used in statistical analysis and mostly they are positively skewed.

3.3.1. Descriptive methods for survival data

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. Such summaries may be of interest in their own right or as a precursor of a more detailed analysis of the data. Routine applications of standard formulas for measures of central tendency and variability will not yield estimates of the desired parameters when the data include censored observations. In summarizing survival data there are two functions of central interest namely the survivor function and the hazard function.

The survivor function S (t):

Let T be a random variable associated with the survival times, t be the realization of the random variable T and $f(t)$ be the underlying probability density function of the survival time t. The cumulative distribution function $F(t)$, which represents the probability that a

subject selected at random will have a survival time less than some stated value t, is then given by:

$$F(t) = P(T < t) = \int_0^t f(u)du \ , \quad t \geq 0 \qquad [1]$$

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

$$S(t) = P(T \geq t) = 1 - F(t) \ , \quad t \geq 0 \qquad [2]$$

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

From equations [1] and [2] the relationship between $f(t)$ and $S(t)$ can be derived as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}(1 - S(t)) = \frac{-d}{dt}S(t) \ , \ t \geq 0 \qquad [3]$$

The hazard function h(t):

The hazard function is widely used to express the risk of hazard of death at time t, and is obtained from the probability that an individual dies at time t, given that the individual has survived up to time t. It is also known as the conditional failure rate in reliability, the force of mortality in demography, the intensity function in stochastic process, the age specific failure rate in epidemiology, the inverse of the Mill's ratio in economics or simply the hazard rate. It gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t.

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

$$h(t) = \lim_{\Delta t \to 0} \frac{p\{an\ individual\ fails in the time\ \text{int}erval\ (t, t+\Delta t) \setminus it\ survives until time\ t\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{p\{t \leq T \leq t + \Delta t \setminus T \geq t\}}{\Delta t}$$

By applying the theory of conditional probability and the relationship in equation [3], the hazard function can be expressed in terms of the underlying probability density function and the survivor function as follows

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d}{dt} \ln S(t) \qquad [4]$$

A related quantity is the cumulative hazard function $H(t)$ defined by

$$H(t) = \int_0^t h(u)du = -\ln S(t) \qquad [5]$$

Thus $S(t) = \exp(-H(t))$ consequently $f(t) = h(t)\exp(-H(t))$ [6]

The Kaplan-Meier, Nelson-Aalen and Life Tables are the most widely used to estimate the survival and hazard functions. These methods are said to be non parametric or distribution-free since they do not require specific assumption to be made about the underlying distribution of the survival times.

The Kaplan-Meier (KM) estimator, also known as the product limit estimator, is the estimator used by most software packages. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t . The technique becomes a little more complicated but still manageable when censored times are included.

Suppose t_1, t_2,…, t_n be the survival times of n independent observations and $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(m)}$, $m \leq n$ be the m distinct ordered death times. The formula for the KM estimator of the survival function at failure time t is given by **(55)**:

$$\hat{S}_{KM}(t) = \prod_{i=1}^{k} \left(\frac{n_i - d_i}{n_i} \right) \quad for\ t_{(k)} \leq t < t_{(k+1)}\ \ k = 1,2,\cdots,m \qquad [7]$$

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

In the above formula n_i is the number of individuals who are at risk of dying at time t_i and d_i is the number of individuals who failed (died) at time t_i .

From equation [6] the KM estimator of the cumulative hazard function can be estimated as:

$$\hat{H}_{KM}(t) = -\ln\left(\hat{S}_{KM}(t)\right) \qquad [8]$$

The variance of the KM survival estimator which is also known as the Greenwood's formula is **(55)**:

$$Var\left(\widehat{S}_{KM}(t)\right) = \left(\widehat{S}_{KM}(t)\right)^2 \sum_{i=1}^{k} \frac{d_i}{n_i\left(n_i - d_i\right)} \quad , \quad for\, t_{(k)} \le t < t_{(k+1)} \qquad [9]$$

Another alternative estimator of the survival function and the corresponding commutative hazard function at time t due to Nelson and Aalen as stated in **(53, 55)**, which is based on the individual failure times is given by

$$\widehat{S}_{NA}(t) = \prod_{i=1}^{k} \exp\left(-\frac{d_i}{n_i}\right) \quad and \quad \widehat{H}_{NA}(t) = \sum_{i=1}^{k} \frac{d_i}{n_i} \, , \quad for\ t_{(k)} \le t < t_{(k+1)} \qquad [10]$$

Although the Nelson-Aalen estimate of the survivor function has been shown to perform better than the KM estimate in small samples, in many circumstances, the two estimates will be very similar, particularly at the earlier survival times. Moreover, the KM estimate is regarded as an approximation to the Nelson-Aalen estimate.

Comparison of Survival Curves

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

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

These tests, which have the same generalized form or algebraic presentation, include the Log Rank test, Generalized Wilcoxon test, Tarone and Ware test, Peto-Peto-Prentice test and Harrington-Fleming test. The calculation of each test is based on a contingency table of groups by status at each observed survival time. The general form of these test statistics for the comparison of survival functions between two groups can be defined as follows:

$$Q = \frac{\left[\sum_{i=1}^{m} w_i \left(d_{1i} - \hat{e}_{1i}\right)\right]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \qquad [11]$$

where:

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

n_{1i} is the number of individuals at risk in group 1 just prior to failure time t_i

n_{2i} is the number of individuals at risk in group 2 just prior to failure time t_i

n_i is the number of individuals at risk in both groups 1 and 2 just prior to failure time t_i

d_{1i} is the observed number of failure (death) in group 1 at failure time t_i

$\hat{e}_{1i} = \frac{n_{1i} \times d_i}{n_i}$ is the expected number of failures corresponding in group 1 at 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 failures in group 1 at time t_i

w_i is the weight for censor adjustment at failure time t_i.

The test statistic Q has χ^2 distribution with 1 degree of freedom under the null hypothesis that the two survivorship functions are the same when the total number of observed events and sum of expected number of events are large and assuming that the censoring experience is independent of group **(54)**. The statistic Q can be extended for comparing more than two groups of survival experience. [See details in **54** and **55**].

The most frequently used tests, the log rank test and the generalized Wilcoxon test, use weight $w_i = 1$ and $w_i = n_i$, respectively. Of the two tests, the log-rank test is the more suitable (powerful) when there is no difference between the survival times of the two groups is tested against the alternative that the hazard of death in any given time for an individual in one group is proportional to the hazard at that time for a similar individual in the other group. For other types of departures from the null hypothesis that there is no difference in survival times of the two groups, the Wilcoxon test is more appropriate than the log-rank test. Specifically the

Wilcoxon test is more powerful when the failure times have a lognormal distribution, with equal variance in both groups but a different mean. This is the assumption of accelerated failure time model. Both tests will lack power if the survival or hazard curves cross each other; however that does not necessarily make them invalid **(**The details are available in **53**, **54** and **55)**. One approach that can be used when the survival or hazard curves cross each other is using clinically plausible stratification variables, factors that may confound the analysis, and testing group differences with in the strata.

3.3.2. Modelling of survival data

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

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

Suppose $x = (x_1, x_2, \cdots, x_p)^{'}$ is a vector of explanatory variables associated with an individual and $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^{'}$ is the vector of regression coefficient associated with the explanatory variables. Two general classes of models have been used to relate covariates effects to survival

namely, the family of multiplicative hazard model and the family of additive hazard rate models **(53)**.

In the family of multiplicative hazard rate models, the conditional hazard rate of an individual is a product of a baseline hazard rate ($h_0(t)$) and a non-negative function of the explanatory variables ($r(\beta, x)$). That is,

$$h(t \mid x) = h_0(t) r(\beta, x) . \qquad [12]$$

A second class of models for the hazard rate is the family of additive hazard rate models. Here, the conditional hazard function is modelled as

$$h(t \mid x) = h_0(t) + \sum_{j=1}^{p} x_j(t) \beta_j(t) . \qquad [13]$$

The regression coefficients of this model are functions of time so that the effect of a given covariate on survival is allowed to vary over time and their values may be positive or negative, but their values are restricted so that $h(t \mid x)$ must be positive.

The Cox regression model

Let $h(t, x, \beta)$ be the hazard rate at time *t* for an individual with covariate vector *x*. The Cox regression model is defined as **(57)**:

$$h(t, x, \beta) = h_0(t) \exp(\beta' x) \qquad [14]$$

where, $h_0(t)$ is an arbitrary baseline hazard rate, β and x are as defined above.

The Cox model can also be regarded as a linear model for the logarithm of the relative hazard, that is,

$$\ln\left[\frac{h(t, x, \beta)}{h_0(t)}\right] = \beta' x . \qquad [15]$$

When no explanatory variables are included in the Cox model, it reduces to the baseline hazard. This property of the Cox model is the reason why $h_0(t)$ is called the baseline hazard function. Moreover, the Cox model is a multiplicative hazards model and it is semi-parametric, often used to describe survival time in a comparative sense, since a parametric form is assumed only for the covariate effect and the baseline hazard is treated non-parametrically.

The Cox model is often called a proportional hazards model, the main assumption of the model, because the hazard rates of two individuals with distinct values of covariates are proportional **(52)**. This can be verified by considering the hazard ratios of different individuals. The logarithm of the hazard ratio for two individuals having two distinct covariate values x_j and x_i can be expressed as

$$\ln\left[\frac{h(t,x_j,\beta)}{h(t,x_i,\beta)}\right] = \ln\left[\frac{h_0(t)\exp(\beta' x_j)}{h_0(t)\exp(\beta' x_i)}\right] = \beta'(x_j - x_i). \quad [16]$$

Clearly the above ratio is independent of time which means that the log hazard ratio is constant at any given time. Moreover, the hazard ratio does not depend on the value of the covariate; rather it depends on the difference between the covariate values.

By applying the method of calculus and the relationship in equation [6], the cumulative hazard function and the survivor function for the Cox model are, respectively, given as:

$$H(t,x,\beta) = H_0(t)\exp(\beta' x) \quad [17]$$

$$S(t,x,\beta) = [S_0(t)]^{\exp(\beta' x)} \quad [18]$$

where $H_0(t)$ and $S_0(t)$ are the baseline cumulative hazard and survivor functions, respectively.

The Cox proportional hazards model is a "robust" model, so that the results from using the Cox model will closely approximate the results for the correct parametric model. In other words, the 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.

The explanatory variables in the Cox model, in equation [14], are time-independent. It is possible, nevertheless, to consider these as time-dependent variables. If time dependent variables are considered, the Cox model form may still be used, but such a model no longer satisfies the proportional hazards assumption, since the hazard ratio of two individuals varies with time, and is called the extended Cox model. The extended Cox model can be expressed as

$$h(t,x,\beta) = h_0(t)\exp\left(\beta' x + \delta' x(t)\right) \quad [19]$$

where, the vectors ***x*** and ***x(t)*** are the vectors of time independent and time dependent covariates, respectively, and δ is the vector of coefficients for the time dependent covariates.

Fitting the Cox model

Suppose the survival data based on *n* independent observations are denoted by the triplet (t_i, x_i, c_i) *for* $i = 1,2,\cdots,n$

where,

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

$c_i = \begin{cases} 1 \text{ for failed observations} \\ 0 \text{ for censored observations} \end{cases}$ is event indicator.

$x_i = (x_{1i}, x_{2i}, \cdots, x_{pi})^t$ is the value of the covariate vector for subject i.

Fitting the Cox model to observed survival data entails estimating the unknown regression coefficients. Also, the baseline hazard function must be estimated. It turns out that these two components of the model can be estimated separately. The coefficients should be estimated first and the estimates are then used to construct an estimate of the baseline hazard function. The regression coefficients in the proportional hazards Cox model, which are the unknown parameters in the model, can be estimated using the method of maximum likelihood.

The likelihood for survival data is constructed by considering the censored observations and the failed observations separately. In the case of failed observations the survival time is exactly t_i, thus the contribution to the likelihood is the probability that the subject fails at time t_i, i.e. $f(t_i, \beta, x_i)$. And, for censored observations the contribution for the likelihood is that the probability that a subject survives at least t_i time units, i.e. $S(t_i, \beta, x_i)$ **(54)**.

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

$$f(t_i, \beta, x_i)^{c_i} \, S(t_i, \beta, x_i)^{1-c_i}.$$

Consequently, the likelihood function which holds for any censored survival data with generalized hazard function $h(t_i, \beta, x_i)$, which may not assume proportional hazards, will be

$$L(\beta) = \prod_{i=1}^{n} f(t_i, \beta, x_i)^{c_i} \; S(t_i, \beta, x_i)^{1-c_i} = \prod_{i=1}^{n} h(t_i, \beta, x_i)^{c_i} \; S(t_i, \beta, x_i).$$

However, Cox showed that the relevant likelihood function which considers the baseline hazard rate as a nuisance parameter; he called it a partial likelihood function, for the proportional hazards model assuming no tied survival times is given in **(57)**:

$$L_p(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta' x_i)}{\sum_{j \varepsilon R(ti)} \exp(\beta' x_j)} \right]^{c_i} \qquad [20]$$

where, $R(t_i)$ represents the risk set just prior to time t_i.

Thus, the corresponding log-partial likelihood function is:

$$LL_p(\beta) = \sum_{i=1}^{n} c_i \left[\beta' x_i - \ln \sum_{j \varepsilon R(ti)} \exp(\beta' x_j) \right] \qquad [21]$$

From equation [21] above, the vector of efficient scores and the observed information matrix can be found, respectively, as follows:

$$U(\beta) = \left(\frac{\partial LL_p(\beta)}{\partial \beta_1}, ..., \frac{\partial LL_p(\beta)}{\partial \beta_p} \right)' \qquad [22]$$

$$I(\beta)_{pxp} = \begin{Bmatrix} -\frac{\partial^2 LL_p(\beta)}{\partial \beta_1^{\,2}} & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_1 \partial \beta_2} & \cdots & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_1 \partial \beta_p} \\ -\frac{\partial^2 LL_p(\beta)}{\partial \beta_2 \partial \beta_1} & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_2^{\,2}} & \cdots & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 LL_p(\beta)}{\partial \beta_p \partial \beta_1} & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_p \partial \beta_2} & \cdots & -\frac{\partial^2 LL_p(\beta)}{\partial \beta_p^{\,2}} \end{Bmatrix}. \qquad [23]$$

The maximum likelihood estimate (MLE) of the coefficients and their estimated variances in the Cox proportional hazards model can be found by maximizing the log-partial likelihood function with respect to the parameters using iterative numerical analysis techniques such as Newton-Raphson which make use of the efficient scores and the observed information matrix.

According to the Newton-Raphson procedure an estimate of β at the (s+1)th of the iterative procedure, $\hat{\beta}_{s+1}$,is

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}(\hat{\beta}_s)U(\hat{\beta}_s), s = 0,1,2,...$$

where $U(\hat{\beta}_s)$ is the vector of efficient scores and $I^{-1}(\hat{\beta}_s)$ is the inverse of the observed information matrix, both evaluated at $\hat{\beta}_s$ .The process can be started by taking $\hat{\beta}_0 = (0,0,...,0)'$ and continue until the change in the likelihood function is sufficiently low.

Thus, after getting the MLE, $\hat{\beta}$, the covariance matrix of $\hat{\beta}$ can be approximated by the inverse of the information matrix, evaluated at $\hat{\beta}$, that is

$$\text{var}(\hat{\beta}) = I^{-1}(\hat{\beta}) . \qquad [24]$$

Although the proportional hazards model for survival data assumes that the hazard function is continuous and, under this assumption, tied survival times are not possible, tied survival times can arise as a result of rounding process. Alternate methods have been suggested by Cox (modification of 1972) and by Breslow and Efron when there are ties. All of them reduce to the original partial likelihood suggested by Cox in the absence of ties. Despite its very complicated form, the appropriate likelihood function when there are ties is given by Kalbfleisch and Prentice. [See the details in **(53)** and **(55)]**.

Assessment of Regression Coefficients

Naturally the next steps following the fit of a regression model are the assessment of the regression coefficients and the formation of confidence intervals for the parameters and related quantities. Under the assumption of proportional hazards, there are three different tests for model assessment (the significance of the coefficients): the partial likelihood ratio test, the Wald test and the score test. These tests are presented below as discussed in **(53)** and **(54)**.

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

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

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

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

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

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

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

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

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

Under the null hypothesis $H_0 : \beta_j = 0$ the statistic $Z \sim N(0,1)$ at a significance level α,

Consequently the 100(1-α)% Wald statistic based confidence interval for β_j is

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

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

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

$$Q_s = U'(\beta_q, \hat{\beta}_{p-q}) I^{-1}(\beta_q, \hat{\beta}_{p-q}) U(\beta_q, \hat{\beta}_{p-q})$$

where: $U(\beta_q, \hat{\beta}_{p-q})$ and $I^{-1}(\beta_q, \hat{\beta}_{p-q})$ are the score vectors and inverse of the observed information matrix evaluated at the hypothesized value of β_q and the restricted partial maximum likelihood estimator of β_{p-q}. Under the above null hypothesis and for large sample $Q_s \sim \chi^2(q)$, at α level of significance.

Model Building/ Variable Selection Procedures

In many settings a variety of explanatory variables are measured and a major question in analysing such data sets is how to incorporate these variables in the modelling procedure. An initial step in the model selection procedure is to identify a set of explanatory variables that have the potential for being included in the linear component of the proportional hazards model. Once a set of potential variables has been isolated, the combination of variables that has to be used in modelling the hazard function has to be determined. An important principle in statistical modelling is the hierarchic principle, and means that interactions should not be fitted unless the corresponding main effects are present.

When the number of variables is relatively large, it can be computationally expensive to fit all possible models. In this situation, automatic routines for variable selection that are available in many software packages might seem an attractive prospect. These routines are based on forward selection, backward elimination or the combination of the two known as the stepwise procedure.

The model selection strategy depends to some extent on the purpose of the study. In a situation where the aim is to identify variables upon which the hazard function depends, instead of using the automatic variable selection procedures, the following procedure is recommended in **(55)**.

1. The first step is fitting a univariable model for each of explanatory variables and identifying the variables that are significant at some level of significance. (A significance level from 20% to 25% is recommended in Hosmer and Lemeshow (1999)).

2. The variables that appear to be important from step 1 are then fitted together in a multivariable model. In the presence of certain variables others may cease to be important. Consequently, backward elimination is used to omit nonsignificant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

3. Variables, which were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to

the model from step 2, with forward selection method. This process may result in terms in the model determined at step 2 ceasing to be significant.

4. A final check is made to ensure that neither no significant variable is eliminated from the model nor nonsignificant variable is included in the model. At this stage the interactions between any of the main effects currently in the model can be considered for inclusion if the inclusion significantly modifies the model.
For steps 2, 3 and 4 a level of significance around 10% is recommended in **(55).**

ASSESSMENT OF MODEL ADEQUACY

Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid, the fitted model must provide an adequate summary of the data upon which it is based. A complete and thorough examination of a model's fit and adherence to model assumptions is just as important as careful model development. Because methods used in assessing the adequacy of survival models have to cope with the occurrence of censored survival times, they are a little more complicated than the corresponding methods used in linear regression modelling. Many model checking procedures are based on quantities known as residuals. Residuals are values that can be calculated for each observation and have the feature that their behaviour is known, at least approximately, when the fitted model is satisfactory. The following residuals have been proposed for use by different authors in connection with the Cox regression model.

Cox-Snell residuals are residuals most widely used in the analysis of survival data **(55)**. The Cox-Snell residual for the i^{th} individual is given by

$r_i = \hat{H}_i(t) = -\ln \hat{S}_i(t)$, where $\hat{H}_i(t)$ *and* $\hat{S}_i(t)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} individual at time t.

If T is the random variable associated with the survival time of an individual and $S(t)$ is the corresponding survivor function, then the random variable $-\ln S(t)$ has an exponential distribution with unit mean irrespective of the functional form of $S(t)$.

Martingale residuals are modified Cox-Snell residuals and, defined as

$rm_i = c_i - r_i$, where c_i is censoring indicator and r_i is the Cox-Snell residual.

It can also be shown that these residuals sum to zero. Moreover, they are uncorrelated and their expected value is zero in large samples. In this respect they have properties similar to those possessed by residuals encountered in linear regression analysis **(55)**.

Deviance residuals: Although the martingale residuals share many of the properties possessed by residuals encountered in other situations, such as linear regression analysis, they are not symmetrically distributed about zero, even when the fitted model is correct. The deviance residuals which were introduced by **(56)**, are much more symmetrically distributed about zero and they are defined as

$$rd_i = sign(rm_i)\{-2[rm_i + c_i \ln(c_i - rm_i)]\}^{1/2},$$

where, rm_i is the martingale residual and $sign(rm_i)$ is the sign of the martingale residual for the i^{th} observation.

The deviance residuals are expected to be symmetrically distributed about zero if the fitted model is appropriate although they do not necessarily sum to zero.

Schoenfeld residuals: The above three residuals have the disadvantages that they depend heavily on the observed survival time and require an estimate of the cumulative hazard function. Schoenfeld proposed residuals that overcome these disadvantages **(59)**. These residuals are calculated for each individual and for each covariate. Thus, the Schoenfeld residual for the i^{th} observation in the k^{th} covariate, the i^{th} component in the k^{th} score vector, is given by

$$rs_{ik} = c_i \left\{ x_{ik} - \frac{\sum_{j \varepsilon R(ti)} x_{jk} \exp(\hat{\beta}^t x_j)}{\sum_{j \varepsilon R(ti)} \exp(\hat{\beta}^t x_j)} \right\},$$

where, the terms are as defined in equation [20].

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

Scaling a vector of Schoenfeld residuals by an estimator of its variance is more effective in detecting departures from the assumed model **(58)**. The vector of the scaled Schoenfeld residuals is then given by

$$rs_i^* = [\text{var}(rs_i)]^{-1} rs_i \approx m\,\text{var}(\hat{\beta}) rs_i,$$

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

Most of the model diagnostics in survival data are based on the residuals stated above.

I. Testing for linearity of covariates

After identifying a particular set of explanatory variables on which the hazard function depends, it is important to check that the correct functional form has been adopted for the continuous covariates. An improvement for the fit of a model may be obtained by using some transformation of the values of a variable instead of the original one. 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. LOESS smoothed curve can be superimposed on the scatter plots to give interpretation. If the functional form suggested (observed) 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. A straight line would then confirm that the appropriate transformation has been used to the covariate.

II. Subject-wise diagnostic measures

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 undue 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 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. It is therefore of particular importance to examine the influence of each observation on these estimates. 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 detected using 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.

Suppose that $\widehat{\beta}_j$ is the j^{th} parameter estimate and let $LL_p(\widehat{\beta})$ be the maximized log partial likelihood of the model containing all the *n* observations, $\widehat{\beta}_{j(-i)}$ be the j^{th} parameter estimate and $LL(\beta_{-i})$ be the maximized log partial likelihood of the model containing only the *n*-1 observations after deleting the i^{th} observation, respectively. Then, the statistic $\Delta_i\widehat{\beta}_j = \widehat{\beta}_j - \widehat{\beta}_{j(-i)}$, which is known as DFBETA, can be used as a measure of how the j^{th} parameter estimate would change, if the i^{th} observation was deleted from the data set. Also, the statistic $LD_i = 2[LL(\beta) - LL(\beta_{-i})]$, which is known as the likelihood displacement statistic, can be used as a measure of how the maximized partial log likelihood changes, if the i^{th} observation was deleted from the data set. Observations that influence a particular parameter estimate have a large absolute value of DFBETA than for other observations in the data set and observations that do influence the overall fit of the model are those which have large values of likelihood displacement statistics than the other observations in the data set. Index (patient identification number) plot of the DFBETAs for each explanatory variable in the model will then reveal whether there are observations that have an undue impact on the parameter estimate for any particular explanatory variable **(55)**.

III. Methods for the Assessment of Proportional Hazards

A crucial assumption made while using the Cox regression is that the proportional hazards assumption. However, there are various reasons that the model may not have proportional hazards. If hazards are not proportional, this means that the linear component of the fitted model varies with time in some manner. As a result a large number of tests and procedures have been proposed that can be used in advance of model fitting and after fitting the model. As can be seen in equation [16], the hazard ratio for two individuals in the proportional hazards model is independent of time and is constant over time. Thus the plot of the logarithm of the Kaplan-Meier cumulative hazards function based on different factors may help in assessing the proportional hazards assumption before fitting a Cox model. If this assumption is met, then the

plots will be more or less parallel. But, this method will not give any clue if the plots for different categories of covariates cross each other. The other method, which could be used after the fit of the model, is extending the proportional hazards model by defining several product terms involving each time independent variable with some function of time. That is, 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. The extended Cox model that simultaneously considers all time-independent variables of interest can be expressed as:

$$h(t, x, \beta) = h_0(t) \exp\left(\sum_{j=1}^{p} \beta_j x_j + \sum_{j=1}^{p} \delta_j x_j g_j(t) \right)$$

To check the proportional hazards assumption using a statistical test, we consider the null hypothesis that all the δ terms, which are coefficients of the $x_j \times g_j(t)$ product terms in the model, are zero. Usually the function $g_j(t)$ is chosen to be the logarithm of survival time i.e. $g_j(t)=ln(t)$. Under the null hypothesis that all the δ terms are zero, the model reduces to the proportional hazards model **(52, 54)**.

Moreover, the plot of scaled Schoenfeld residuals of each covariate against the logarithm of analysis time may give additional insight whether there is some departure from proportional hazards. The plots will be randomly distributed about zero if the proportional hazards assumption is met **(54, 55)**.

IV. **Overall Goodness of Fit**

If the fitted model is satisfactory (appropriate), the Cox-Snell residuals will behave as *n* observations from a unit exponential distribution. Thus, the plot of the estimated hazard rate of the Cox- Snell residuals ($\widehat{H}_r(r_i)$), versus r_i will give a straight line through the origin with slope unity if the fitted model is satisfactory. However, the drawback is that they do not indicate the particular departure from the model fitted, if there is any. An index plot of deviance residuals also highlight individuals whose survival time is not well fitted by the model. 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. Potential outliers have deviance residuals whose absolute values are too large.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1. GENERAL CHARACTERISTICS OF THE COHORT

The study included 832 HIV patients, who started HAART in Jimma University Specialized Hospital (JUSH) between the years 2005 and 2007, about whom complete information related to the study covariates is available. Among the patients, 686 started their treatment on D4T-based drug regimens. Of the total of 832 patients included, 475 were females, 167 were never married, and 219 were living in Jimma town. When HAART was initiated, 197 patients had clinical AIDS (WHO stage IV) and 58 patients had bedridden functional status due to the severe progression of the disease. TB and opportunistic infections (OI) were prevalent among 428 and 512 patients, respectively at baseline. Most of the patients (728) disclosed their HIV sero-status at least for one member of their family, 285 of the patients had casual sexual partners and 276 of the patients were substance users (Table 4.1.1 Appendix). The median age, CD4 count and weight of patients when HAART was initiated are 31 years (interquartile range 37-27 years), 135 cells/mm^3 (interquartile range 200- 79 cells/mm^3) and 50 kg (interquartile range 55-45 kg), respectively (Table 4.1.2 Appendix).

4.2. DESCRIPTIVE SURVIVAL ANALYSES

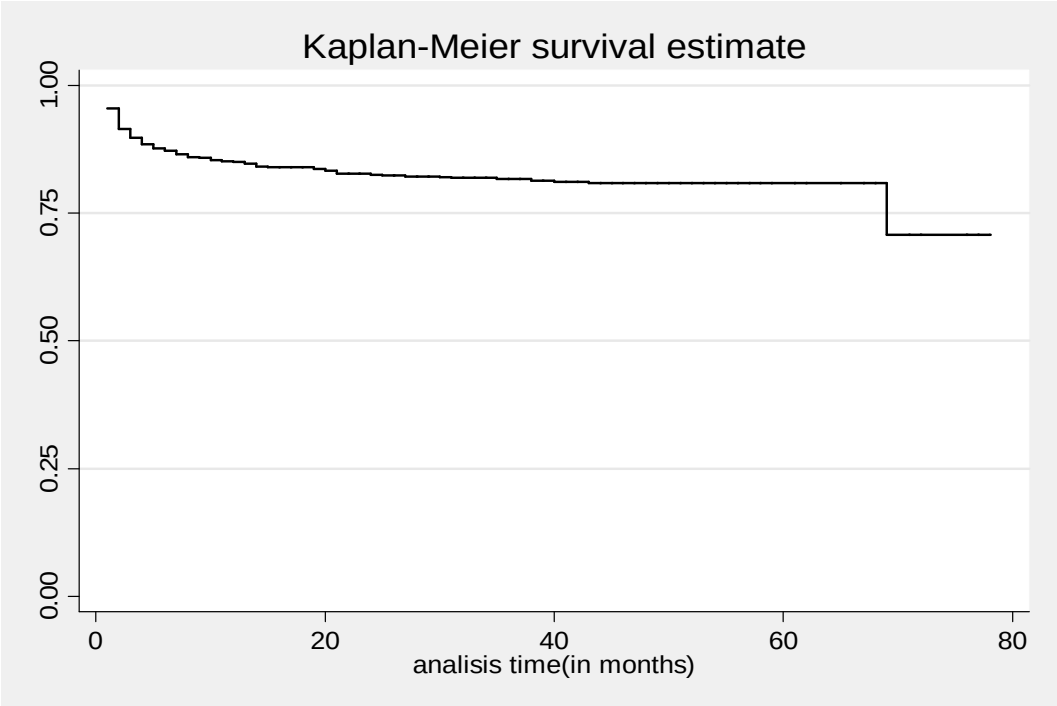
The patients were followed up for a median of 40 months. The minimum follow up time was 1 month and the maximum was 78 months. Some 144 patients died during the follow up time of whom 70 (48.6%) and 99 (68.8%) deaths occurred within three months and six months of HAART initiation, respectively. The overall mean estimated survival time of patients under the study was 63.7 (95% CI: 61.1- 66.3) months. Females have relatively lower survival time (59.1 months) than males (63.8 months). Patients with younger age (40 years or less) had survived for about 64.4 months while the mean survival time for older patients was 57.6 months. Table 4.2.1 presents the mean survival time of patients based on different socio-demographic and clinical characteristics.

Table 4.2.1: Kaplan-Meier analyses of survival times for patients on antiretroviral treatment according to important socio-demographic and clinical characteristics of HIV patients treated with HAART in HUSH, Jimma, 2010.

Covariate / Factor	Category	Estimate	Std. Error	95% confidence interval	
Sex	Female	59.1	1.5	56.2	61.9
	Male	63.8	1.6	60.7	67.0
Age group	Below 40	64.4	1.5	61.6	67.3
	40 and above	57.6	2.6	52.4	62.7
Marital status	Never married	55.1	2.0	51.1	59.0
	Married	63.8	1.7	60.5	67.0
	Others	56.6	1.5	53.6	59.6
Religion	Muslim	62.4	2.3	57.9	67.0
	Orthodox	63.3	1.3	60.7	65.9
	Others	61.8	2.1	57.8	65.9
Level of education	No education	62.0	2.4	57.3	66.8
	Primary	60.5	1.7	57.3	63.8
	Sec. or above	63.3	1.5	60.3	66.3
Employment status	Employed	64.1	2.2	59.9	68.3
	Non Employed	62.7	1.2	60.4	65.0
Place of residence	Jimma	63.7	2.1	59.7	67.8
	Others	62.2	1.5	59.2	65.1
No of Rooms	Only One	58.6	1.4	55.9	61.2
	Two or more	64.1	1.5	61.1	67.0
Disclosure status	Disclosed	63.6	1.4	60.9	66.3
	Not disclosed	49.0	2.0	45.0	52.9
Risk behaviour	Regular	66.2	1.2	63.8	68.5
	Casual or Both	54.6	1.7	51.3	57.9
Substance use	Yes	58.6	2.1	54.5	62.6
	No	66.3	1.2	64.0	68.6
Functional status	Working	71.3	1.0	69.4	73.1
	Ambulatory	57.1	1.8	53.6	60.6
	Bed ridden	39.2	5.1	29.1	49.2
WHO clinical stage	Stage 1 or 2	58.7	1.8	55.2	62.2
	Stage 3	65.2	1.2	62.8	67.7
	Stage 4	56.7	2.7	51.4	62.1
Past OI	Yes	62.2	1.5	59.2	65.1
	No	65.1	1.8	61.6	68.5
TB co-infection	Yes	60.9	1.4	58.2	63.6
	No	67.9	1.9	64.1	71.7
Drug regimen	D4T-based	63.3	1.4	60.6	66.1
	AZT-based	52.2	1.8	48.7	55.7
Over all survival time		63.7	1.3	61.1	66.3

The graph of the estimate of overall Kaplan-Meier survivor function showed that most of the deaths occurred in the earlier months of HAART initiation and it declined in the later months of follow up. Separate graphs of the estimates of the Kaplan-Meier survivor functions, for different factors, have also been constructed in order to asses whether there is difference in survival experience between different groups of individuals. Most of the graphs did not show differences between different categories. However, relatively larger gaps are observed in covariates such as substance use, WHO clinical stage, functional status and TB co-infection. The graphs of Kaplan-Meier survival estimates based on different categories of factors are presented in the displays of Figure 4.2.1.

Figure 4.1.1: The plot of the overall estimate of Kaplan-Meier survivor function of HIV patients treated with HAART in JUSH, Jimma, 2010.



Log-rank test was performed to investigate the significance of the observed difference in the Kaplan-Meier estimates of the survivor functions among different categories of the factors. According to the log-rank test, there was no significant difference in survival experience between the various categories of sex, marital status, religion, level of education, employment status, place of residence, number of rooms, disclosure status and drug regimen. However, the p-values of the log-rank test showed that the survival experience of patients in different

categories of risk behaviour, substance use, functional status, who clinical stage, past history of OI and TB co-infection differ significantly. Thus, those patients who had regular partner, who were not using substance (mostly khat), had working functional status, were in WHO clinical stage I or II, had no TB co-infection and who did not suffer from opportunistic infections had better survival experience. The log rank test result is shown in Table 4.2.2.

4.3. RESULTS OF THE COX PROPORTIONAL HAZARDS MODEL

In statistical modelling, when the number of variables is relatively large, it can be computationally expensive to fit all possible models. Thus, one of the options is fitting a multivariable model containing the variables that are significant at a modest level of significance in a univariable analysis. As a result, eighteen univariable Cox proportional hazards models, each containing one explanatory variable, were fitted. Table 4.3.1 presents the eighteen univariable models along with the necessary statistics.

The analysis of the data on survival times of HIV patients suggested that not all of the 18 explanatory variables are needed in a proportional hazards model. As a result, the most appropriate subset of these covariates to be included in the multivariable model will be selected based on their contribution to the maximized log partial likelihood of the model ($-2LL(\hat{\beta})$). From among the explanatory variables, functional status leads to the highest reduction in $-2LL(\hat{\beta})$, reducing its value from 1875.8 to 1797.418. This reduction of 78.382 is highly significant (p-value < 0.0001) when compared with percentage points of the χ^2 distribution on 1 degree of freedom. The next highest significant change in $-2LL(\hat{\beta})$ is obtained when the variable CD4 count is added to the null model in its continuous form, which reduced the statistic by 50.498. The reduction in $-2LL(\hat{\beta})$ on adding baseline weight, TB, WHO clinical stage, age, past history of OI, substance use and risk behaviour to the null model one at a time is 22.749, 13.889, 12.006, 8.01, 5.885, 5.76 and 5.208 respectively, making all of them significant at 5% level.

The next step is fitting a Cox Proportional hazards model that contains the variables functional status, baseline CD4 count, baseline weight, TB co-infection, WHO clinical stage, age at HAART initiation, past history of OI, substance use and risk behaviour, which leads to a value

of $-2LL(\hat{\beta})$ to 1730.482, and checking the effect of omitting each variable from the model. The increase in $-2LL(\hat{\beta})$ and p-values resulted from omitting each of the above variables in turn from this model are shown in Table 4.3.2. In particular, when past history of OI is eliminated, the increase in $-2LL(\hat{\beta})$ is almost 0. Thus, the variable past history of OI does not appear to be included in the model , in the presence of the other eight variables and as a result it became the first variable to be eliminated from the multivariable model.

Following the elimination of past history of OI, a model which considered the remaining eight variables is fitted and the effect of eliminating each variable from the model is assessed. The increase in $-2LL(\hat{\beta})$ and p-values resulted from omitting each of the eight variables are shown in Table 4.3.3. Thus, the minimum insignificant increase in $-2LL(\hat{\beta})$ of 1.073 (p-value 0.5848) is obtained when WHO clinical stage is removed from the model. As a result, WHO clinical stage can be excluded from the model which contained the eight variables and another model containing the remaining seven variables will be fitted.

Examining the reduction in the log partial likelihood and the p-values associated with it resulted from modelling the seven variables in Table 4.3.4, the minimum insignificant change is obtained when risk behaviour is removed from the model (LR χ^2=2.536, p-value 0.1113). As a result, the next step is fitting a model which ignores risk behaviour.

Using the likelihood ratio test p-values which compares the model containing the variables that were significant in the previous step and a model which eliminated one variable at a time, all the changes are found to be significant. Thus, all the variables functional status, baseline CD4 count, baseline weight, TB co-infection, age at HAART initiation and substance use should be retained in the model, which leads to a value of $-2LL(\hat{\beta})$ of 1734.091. Table 4.3.5 presents the comparison of the model containing all the six variables and the model with one variable eliminated.

Finally, the importance of the variables which failed to be significant, in the univariable analyses, as predictor or useful confounder of survival experience of patients should be assessed in the presence of others. Thus, those variables are added one at a time in the model containing the variables functional status, baseline CD4 count, baseline weight, TB co-

infection, age at HAART initiation and substance use, and both the improvement in $-2LL(\hat{\beta})$ and the change in the coefficients of the significant variables were examined (Table 4.3.6). However, none of those variables were found to be important and none of them need to be retained in the model. Therefore, the most appropriate main effects model is the one containing functional status, baseline CD4 count, baseline weight, TB co-infection, age at HAART initiation and substance use.

Although the model-based approach to the analysis of survival data identified the 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 continuous covariates. Thus, the martingale residuals obtained from a model excluding the covariate of interest are plotted against the excluded covariate to ensure the correct functional form required for the covariate. The resulting plots with a LOESS smoothed curve superimposed to aid in their interpretation are shown in Figures 4.3.1, 4.3.2 and 4.3.3. In all the three plots the martingale residuals are random showing no systematic pattern and the LOESS smoothed curves are straight lines. Thus, the three plots confirm that linear terms are required for age at HAART initiation, baseline CD4 and baseline weight in the model. The slope of the plot for age in Figure 4.3.1 is positive; corresponding to the positive coefficient of age in the fitted model, while the plots for baseline CD4 count and baseline weight in Figures 4.3.2 and 4.3.3 have negative slopes.

The final step in model development strategy is consideration of interaction terms that may be useful in the improvement of the model. Thus, all the possible interactions of the variables under the hierarchic principle are formed and the significance of adding each of the interactions in the main effects model, one at a time, is checked using the Wald test. However, the Wald test p-values in Table 4.3.7 indicate that none of them were found to be significant. Thus, the last model will be the one which contains only the main effects in Table 4.3.5. The parameter estimates and hazard ratios of the covariates are shown in Table 4.3.8. However, in order to make use of these estimates, assessment of the important assumptions associated with the proportional hazards Cox regression model must be met.

Table 4.3.8: 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 treated with HAART in JUSH, Jimma, 2010.

Covariates / Factors	DF	Parameter Estimate	Standard Error	Wald χ^2	P-Value	Hazard Ratio	95.0% CI for the Hazard Ratio	
Age	1	0.0299	0.0101	8.7765	0.0031	1.03	1.010	1.051
Baseline CD4 count	1	-0.0060	0.0012	23.4349	<.0001	0.994	.992	.996
Baseline weight	1	-0.0207	0.0102	4.1105	0.0426	0.979	.960	.999
Substance use(versus no)	1	0.3505	0.1710	4.2030	0.0404	1.42	1.016	1.985
Functional status(versus working)	2			49.4557	<.0001			
Ambulatory	1	1.0569	0.2120	24.8474	<.0001	2.877	1.899	4.360
Bedridden	1	1.9321	0.2779	48.3454	<.0001	6.904	4.005	11.902
Tuberculosis co-infection (versus no)	1	0.6449	0.1934	11.1210	0.0009	1.906	1.305	2.784

4.4. MODEL DIAGNOSTICS

After a model has been fitted, there are a number of aspects of the fit of a model that need to be studied. In this section we shall use a series of regression diagnostics for the final proportional hazards model.

4.4.1. CHECKING THE ADEQUACY OF THE PROPORTIONAL HAZARDS ASSUMPTION

The validity of Cox's regression analysis relies heavily on the assumption of proportionality of the hazard rates of individuals with distinct values of a covariate. If this assumption is not valid, then one may be appreciably misled by the results of the analyses. In this study, extended Cox model is used to test this assumption, a graphical check is also used to provide any additional insight into any departure from proportionality. Thus, all interactions of covariates with the logarithm of survival times are modelled together with the main effects, and then formal tests are applied on the coefficients of those interaction terms using the Wald test.

Table 4.4.1: Results of the multivariable proportional hazards Cox regression model containing the variables in Table 4.3.8 and their interaction with log time (in months)

Covariates / Factors	DF	Parameter Estimate	Standard Error	Wald χ^2	P-value	Hazard Ratio
Age	1	0.0471	0.0147	10.2478	0.0014	1.048
Baseline CD4 count	1	-0.0041	0.0019	4.7093	0.0300	.996
Baseline weight	1	-0.0411	0.0153	7.2300	0.0072	.960
Substance use	1	0.3668	0.2571	2.0351	0.1537	1.443
Functional status	2			30.5386	<.0000	
Ambulatory	1	1.5818	0.3791	17.4103	<.0000	4.864
Bedridden	1	2.4749	0.4482	30.4895	<.0000	11.880
Tuberculosis co-infection	1	0.8013	0.3101	6.6770	0.0098	2.229
Age	1	-0.0156	0.0095	2.6660	0.1025	.985
Baseline CD4 count	1	-0.0015	0.0011	1.6587	0.1978	.999
Baseline weight	1	0.0167	0.0088	3.6164	0.0572	1.017
Substance use	1	-0.0039	0.1559	.0006	0.9800	.996
Functional status	2			3.3411	0.1881	
Ambulatory	1	-0.3394	0.1939	3.0637	0.0801	.712
Bedridden	1	-0.3890	0.2729	2.0324	0.1540	.678
Tuberculosis co-infection	1	-0.1223	0.1715	.5088	0.4757	.885
Global test	7			12.05	0.0989	

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

As can be seen from Table 4.4.1, the Wald p-values confirm that none of the coefficients of the interaction terms are significant at the 5% level. Moreover, the inclusion of the interaction terms improved $-2LL(\hat{\beta})$ by 12.05 which is less than 14.07, the 5% percentile of the chi-squared distribution with 7 degrees of freedom. Therefore, there is no evidence against the null hypothesis that the coefficients of the time varying variables(interaction terms) are zero ascertaining the validity of the proportional hazards assumption for the data. The assumption of proportionality was also assessed graphically by plotting the scaled Schoenfeld residuals of each covariate against log time. Figures 4.4.1.1 - 7 show the plots of scaled Schoenfeld residuals of each covariate in the model against log of survival times along with the LOESS smoothed curves. The residuals are more or less random and LOESS smoothed curves have basically zero slope which is an indication of no evidence of non-proportionality.

4.4.2 CHECKING FOR INFLUENTIAL OBSERVATIONS

In the assessment of model adequacy, it is important to determine whether any particular observation, if any, has an undue impact (leverage) on inferences made on the basis of model fitted to an observed set of survival data. It is therefore of particular interest to examine the influence of each particular observation on these estimates. This is done by examining the extent to which the estimated parameters and the maximized likelihood in the fitted model are affected by omitting in turn the data record for each individual in the study. Thus, the DFBETA and the likelihood displacement statistics are used to examine the untoward effect of each observation on the j^{th} parameter estimate and the maximized log partial likelihood, respectively in the fitted Cox regression model. The five largest changes in the parameter estimates and in the log partial likelihood statistics are presented in Table 4.4.2.1 and Table 4.4.2.2, respectively.

The largest difference for age occurs for patient 641, but there are other differences with similar values. The change in the parameter estimate on omitting the data for this patient is -0.002661463. Therefore, omission of this patient increases the hazard of death relative to the baseline hazard. The standard error of the parameter estimate for age in the full data set is 0.0101, and so the maximum amount by which this estimate changed when one observation is deleted is about 26% of the standard error (less than one standard error). Thus, the change in age effect by deleting this patient can be considered as insignificant.

Omitting the data from patient 613 and patient 437 from the dataset brought the largest changes in the parameter estimates for the other two continuous variables; baseline CD4 count and baseline weight, respectively. The maximum change in the parameter estimates for CD4 count and weight when each observation is omitted in turn is 0.00038(32% of the standard error) and 0.00287 (28% of the standard error), respectively; both of them are within one standard error of the estimates. The effect of deleting these observations is decreasing the relative hazard of death, but again these decreases are not great. The differences in the parameter estimates for the levels of the categorical variables were also assessed. However, the largest differences are less than a quarter of the standard error of the corresponding estimate.

The largest value of the likelihood displacement statistic, the largest change in the value of the maximized log partial likelihood, is 0.28 corresponding to patient 641. This change is quite

very small. This means that the removal of any observation in the dataset does not bring considerable change in the maximized log partial likelihood value. Thus, at this point we can conclude that neither the estimates for each of the parameters nor the set of parameter estimates are affected by any of the observations in the dataset.

4.4.3 ASSESSMENT OF OVERALL GOODNESS OF FIT

After fitting the Cox model, its accuracy for predicting the survival of a given subject and the extent to which the fitted model provides an appropriate description of the observed data should be assessed. Thus, the Cox-Snell residuals are used to asses the overall goodness of fit of the model and the deviance residuals from the final Cox model are used to check if each observation is well fitted by the Cox model. The plot of the Nelson-Aalan estimate of the cumulative hazard function of the Cox-Snell residual against the Cox-Snell residuals is shown in Figure 4.4.3.1 below.

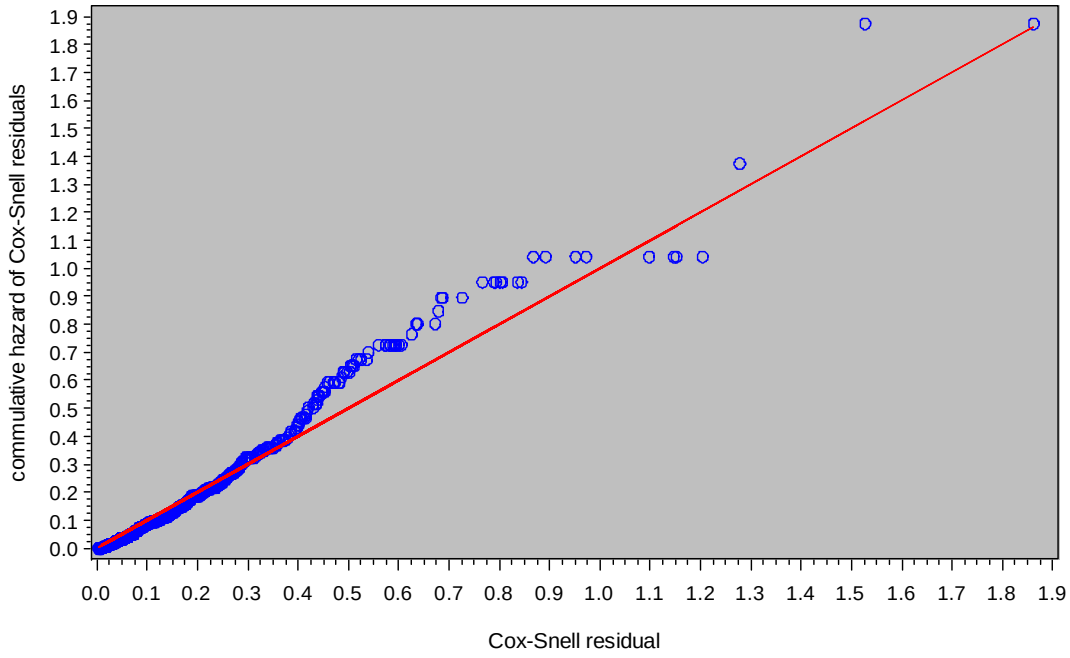


Figure 4.4.3.1: Cumulative hazard plot of the Cox-Snell residuals of the proportional hazards Cox regression model in table 4.3.8. The 45^0-striaght line through the origin is drawn for reference.

The plotted points in Figure 4.4.3.1 are fairly close to the straight line through the origin, which has unit slope. This suggests that the model fitted to the data is satisfactory. Also, the index plot of the deviance residuals in Figure 4.4.3.2 suggests that none of the observations are poorly fitted by the model. To identify the values of covariates which are poorly fit by the model, if any, the deviance residuals versus the values of each covariate was also plotted. The

plots in Figures 4.4.3.3 (a, b, c) are randomly distributed about zero with no systematic pattern suggesting all the values of the covariates are well fitted by the model. Moreover, the likelihood ratio test, the score test and the Wald test in table 4.4.3 suggest that the overall goodness of fit of the model is good.

4.5. INTERPRETATION AND DISCUSSION OF RESULTS

4.5.1 INTERPRETATION

When the multiple proportional hazards Cox model is used in the analysis of survival data, 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. On the other hand, the coefficient for a continuous explanatory variable is the estimated change in the logarithm of the hazard ratio for a unit increase in the value of the explanatory variable after adjustment for the remaining variables in the model. Thus, the interpretation of those variables that were significant in the final proportional hazards model of HIV patients treated with HAART in JUSH is as follows.

Age of patients at the start of HAART is found to have a significant effect on the mortality of patients (estimated HR=1.03, 95% CI: 1.01-1.051, p=0.0031). Since the estimated hazard ratio is greater than unity, other things being equal, the higher the age of a patient, the greater the hazard of death at any given time. In particular an increase of 1 year in age of patients increases the hazard rate of death by 3%. The 95% CI also indicates that the hazard rate could be as low as 1.01 or as large as 1.051. Baseline CD4 count is also another covariate which has a significant effect on the mortality of patients (estimated HR=0.994, 95% CI: 0.992-0.996, p=<0.0001). The estimated hazard ratio for a 50 cells/mm^3 increase in the baseline CD4 count is $\exp(-0.006 * 50) = 0.741$, and the corresponding 95 % CI for the hazard ratio is 0.659 to 0.833. The interpretation is that patients whose CD4 count is higher by 50 cells/mm^3 are dying at a hazard rate 26% lower than for patients with lower count. Also a 5 kilograms (kg) increase in the baseline value of weight of patients decreases the hazard rate of death by 10%. However, the decrease in the hazard rate could be as low as less than 1% or as high as 18% (the estimated HR for a 5kg change=0.902, 95% CI: 0.816-0.996).

Substance use, functional status and TB co-infection are the three categorical variables that are found to be significantly associated with the survival of patients in the fitted Cox regression

model. The hazard ratio for substance use is 1.42. Thus, patients who are abusing substance have a 42% higher risk rate of death than those that are not using substances. The confidence interval indicates that the risk of death for substance user patients could be higher by a quantity as large as 99% or as low as 1.6% than patients who are not using substance; p =0.0404.

The estimated relative risk (hazard ratio) of dying for patients co-infected with TB as compared to those who are not infected with TB is 1.906 (95% CI: 1.305-2.784). This means being co-infected with TB almost doubles the hazard rate of death. The 95 % confidence interval also suggests that the risk of death for TB co-infected patients is 1.306 times as low and 2.784 times as large as compared to patients who are not co-infected with TB after adjusting for other variables in the model.

The estimated risks of death for a patient with ambulatory and bedridden functional status compared to those patients in a working functional status are 2.877 (95% CI: 1.899-4.360) and 6.904 (95% CI: 4.005-11.902), respectively. This means that the hazard rate of death for ambulatory and bedridden patients is 2.8 times and 7 times the working patients, respectively. Moreover, the estimated hazard ratio of bedridden functional status compared to ambulatory functional status is $\exp(1.9321 - 1.0569) = 2.399$ (95% CI: 1.53 to 3.75). Since the confidence interval does not contain 1, an individual whose functional status is bedridden has a significantly higher hazard rate, at any given time, than patients with ambulatory functional status. Thus, compared to patients with ambulatory functional status the hazard rate of bedridden patients is 2.4 times the hazard rate of ambulatory patients.

4.5.2. DISCUSSION OF RESULTS

This study identified variables/factors that are significantly associated with increased risk of mortality. Identifying patients at a higher risk of death has the advantage that due attention will be given to the risk group during their follow up to minimize the risk of mortality while they are taking ART.

This historical cohort study found that the significant predictors of lower chance of survival in patients living with HIV/AIDS after initiation of ART were: older age, low CD4 count at baseline, low weight at baseline, substance use, being bedridden and ambulatory and co-

infection with TB. WHO clinical stage, OI at baseline and risk behaviour are only significantly associated with mortality in the univariable analysis but not in the multivariable analysis.

The mortality rate of patients in the earlier months of ART initiation was high and it declined in the later months of follow up. Studies conducted in southern Ethiopia and Tanzania also revealed that the mortality rates were high within the first three months of follow up **(29, 48-50)**. This may be attributable to the fact that most of the patients start HAART at the severe stage of the disease.

CD4+ cell count is the most important marker of HIV disease progression and a strong predictor of survival, independent of HIV viral load. The use of CD4+ cell level as the primary means to evaluate survival among HIV-infected individuals is also supported by the observation that duration of infection has less prognostic value than the CD4+ cell count at a given point in time. Survival difference based on baseline CD4 count has been revealed by many studies. A study from Thailand showed that patients with low baseline CD4 cell count are at higher risk of mortality **(45)**. A similar study in Uganda also found that the risk of mortally of patients having CD4 count of less than 50 cells/mm^3 is more than 4 times compared to those patients having a CD4 count greater than 50 cells/mm^3 **(30)**. Studies from Malawi, South Africa, and Ivory Coast also identified lower CD4 count as a predictor for lower survival and/or increased rate of mortality **(32, 35, 40)**. A multinational study in Europe and North America also showed that as CD4 count of patients increase the hazard of death is declining **(31)**. Our results are also consistent with the findings in the above studies.

TB continues to be the leading cause of death in people living with HIV/AIDS. This study showed that TB increased the rate of mortality. Patients co-infected with TB had nearly 2 times higher risk of dying while on ART compared to non-infected. Studies from Baltimore, USA and mainland China have also shown that TB is an independent predictor of mortality in patients on ART, after controlling for potential confounders, including CD4 cell count and viral load (**33, 34)**. In the Baltimore study the hazard of AIDS related death was 2.4 times (95% CI: 1.2–4.7) more for the person-time with TB compared with the person-time without incident TB. A cohort study in Abidjan, Ivory Coast also found out that TB is a risk factor for immunological and virological failure, which leads to severe morbidity and mortality in adult patients treated with ART **(35)**.

The functional status of patients can be seen as an indicator of the severity of the progression of the disease. Those patients who are in ambulatory and working functional status have the strength to work and engage themselves in household activities which may help them generate additional income and improve their quality of life. Even though patients who are staying in bed in hospitals are accessing medical care and support, cannot do anything by their own to create a stress-free environment for themselves. Thus, as expected, patients with ambulatory and bedridden status are at a higher risk of mortality than working patients.

Symptomatic disease (WHO stages III and IV) was associated with mortality in many studies **(29, 32, 40, 42, 48-50)**. However, our finding shows that WHO clinical stage is associated with mortality in the univariabe analysis only, possibly reflecting differences in the accuracy of clinical staging or homogeneity of cohorts with respect to this variable. Thus, in the univariable Cox regression, being WHO stage IV led to increased rate of mortality when compared to stage I and II.

Substance abuse is another significant predictor of death in our study. Substance abuse was associated with mortality, non adherence to medication and lower quality of life in many studies **(43, 44, 46, and 47)**. Substance abuse may impair judgment and the ability to adopt and maintain routine medication use. Therefore, screening for drug abuse and excessive alcohol use, and supportive counselling and treatment for drug abuse might help in promoting long term adherence to ART.

Even though different previous studies demonstrated body mass index (BMI) as one of the strong predictors of mortality **(32, 35 48-50)**, it was not possible to calculate BMI for patients since only the weight but not the height of the patients has been recorded at baseline. However, the study revealed that the weight of patients is significantly associated with mortality and those patients having lower weight are at higher risk of mortality.

Compared to younger patients, older patients were more likely to have greater risk of mortality according to this study. Likewise, older age was also associated with increased rate of mortality and low quality of life and defaulting, which leads to mortality **(42, 47)**. However in France a study demonstrated that younger age was associated with non adherence to ART **(44)**. A study in Tanzania also showed that age is not significantly associated with mortality **(29)**.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

In this study we tried to identify the factors that are associated with high risk of mortality in HIV patients treated with HAART in Jimma university specialized hospital using the methods of survival analysis. The Kaplan-Meier results showed that the general mean estimated survival time of patients after HAART initiation is 63.7 months. The mortality rate was very high in the earlier months of HAART initiation and tended to stabilize later. About 48.6 % and 68.8% of the deaths occurred within three and six months of HAART initiation, respectively. Moreover, the results of the multivariable proportional hazards Cox regression model showed that lower CD4 count at the start of HAART, lower weight, older age, TB co-infection, substance use and being of bedridden or ambulatory functional status are associated with higher risk of mortality. Having more than one sexual partner, experience of past OI and being in WHO clinical stage IV are associated with increased rate of mortality in the crude (univariable) analysis but not in the multivariable analysis. However, none of the socio-demographic variables have significant association with survival. Similarly disclosure status of patients and the type of drug regimen used at the start of treatment are not significantly associated with reduced survival of patients.

5.2. RECOMMENDATIONS

- Health workers and other ART clinic staff should plan for more frequent contacts with patients during the early phase of treatment in order to prevent the many deaths that occur during the early weeks of ART. Moreover, the factors that are causes for or associated with increased rate of mortality in the early stages of HAART initiation need to be explored.

- Having lower CD4 count, being bedridden and WHO clinical stage are indicators of the progression of the disease. Therefore, patients should be informed about the need for early diagnosis of HIV infection and starting treatment early is very important. Moreover, treatment of opportunistic infection parallel to the ART programme may reduce the risk of mortality.

- Having more than one partner may expose the patients as well as the general population to drug resistant virus. So, continuous advice and counselling should be provided for patients as part of health education so as to bring behavioural change.

- Even though lower weight is associated with increased rate of mortality, the cause of having lower weight in patients is still not known. In this regard further studies need to be undertaken to determine the cause for having lower weight in patients under HAART. Also, the high mortality risk associated to older patients need be investigated.

- Most of the patients in our cohort initiated their therapy with NVP-based regimens which are not recommended for patients co-infected with TB according to the ART guideline. Thus, EFV-based regimens should be provided to patients with TB to improve their survival. Further to this, special care should be given to TB/HIV patients.

- A separate treatment programme for drug user patients is important and careful monitoring of drug adherence should be made available, as the effect of the treatment is highly dependent on adherence. Also, the greater risk of mortality associated with drug use need be explored.

- Finally, health workers and peer educators and data clerks, working with patients under HAART, should be given special training to improve the quality of the data records of patients. Moreover, attempt should be made to investigate the causes of deaths that occurred out of hospitals, and mechanisms should be devised to trace patients lost to follow up.

References:

1. Barre-Sinoussi F., Chermann J.C., Rey F., Nugeyre M.T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vezinet-Brun F., Rouzioux C., Rozenbaum W., Montagnier L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for AIDS. Science; 220: 868-71.
2. Broder S. and Gallo R.C. (1984). A pathogenic retrovirus (HTLV-III) linked to AIDS. N Engl J Med; 311:1292-7.
3. Gallo R.C., Salahuddin S.Z., Popovic M. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science; 224: 500-3.
4. UNAIDS (2008). Report on the global AIDS epidemic. Geneva.
5. UNAIDS (2009). AIDS epidemic update 2009. Geneva.
6. WHO and UNAIDS (2003). Treating 3 million by 2005: Making it happen.
7. Federal HIV/AIDS Prevention and Control Office (July 2007). Guidelines for implementation of the antiretroviral therapy programme in Ethiopia. Ethiopia, Addis Ababa.
8. Ministry of Health (October 2002). AIDS in Ethiopia, disease prevention and control department, fourth edition, Addis Ababa, Ethiopia.
9. Derek P., Paul B. and Kathy A. (October 8 – 23, 2008). UNAIDS Second Independent Evaluation 2002-2008: Country Visit to Ethiopia, Summary Report.
10. Federal HIV/AIDS Prevention and Control Office (March 2010). Report on progress towards implementations of the UN declaration of commitment on HIV/AIDS. Ethiopia, Addis Ababa.
11. Hoffmann C., Rockstroh J.K., and Kamps B.S. (2007). HIV Medicine.15th edition, Flying Publisher.
12. http://www.globalhealthreporting.org/diseaseinfo.asp?id=262
13. www.who.org/int\
14. Palella F.J., Delaney K.M., Moorman A.C., Loveless M.O., Fuhrer J., Satten G.A., Aschman D.J., Holmberg S.D. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N Eng. J Med.; 338:853-860.
15. Palella FJ., Deloria-Knoll M., Chmiel J.S., Moorman A.C., Wood K.C., Greenberg A.E. and Holmberg S.D. (2003). Survival benefit of initiating antiretroviral therapy in HIV-infected persons in different CD4+ cell strata. Ann Intern Med.; 138(8):620-626.

16. WHO (2003). Scaling-up antiretroviral therapy in resource-limited settings. Treatment guideline for a public health approach. 2003 revision.
17. Bhaskaran K., Hamouda O. and Sannes M. (2008). Changes in the Risk of Death after HIV Seroconversion Compared with Mortality in the General Population. JAMA.; 300(1):51-59.
18. Akileswaran C., Lurie M.N., Flanigan T.P., and Mayer K.H. (2005). Lessons Learned from Use of Highly Active Antiretroviral Therapy in Africa. Clinical Infectious Diseases; 41:376–385
19. Spacek L.A., Shihab H.M., Kamya M.R., Mwesigire D., Ronald A., Mayanja H., Moore R.D., Bates M. and Quinn T.C. (2006). Response to antiretroviral therapy in HIV infected patients attending a public, urban clinic in Kampala, Uganda. Clinical Infectious Disease; 42: 252–9
20. WHO (2009). Towards universal access: the way forward.
21. Hammer S.M., Scquires K.E., Hughes M.D., Grimes J.M., Demeter L.M., Currier, J.S., Eron J.J., Feinberg J.E., Balfour H.H., Deyton L.R., Chodakewitz J.A. and Fischl, M.A. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. N Engl J Med.; 337:725-733.
22. Saag M., Gersten M., Chang Y. Long term virological and immunological effect of the HIV protease inhibitor Viracept (nelfinavir mesylate) in combination with zidovudine (AZT) and lamivudine (3TC). [Abstract]. The Infectious Disease Society of America 35th Annual Meeting, San Francisco, 13-16 September 1997:112.
23. Kahn J., Mayers D., Riddler S., Stein D, Bach M., Havlir D. (). Durable clinical anti-HIV-1 activity (60 weeks) and tolerability for efavirenz (DMP 266) in combination with indinavir (IDV) suppression to “less than 1copy/ml” (OD5 background) by Amplicor as a predictor of virologic treatment response [DMP 266-003, cohort IV] [Abstract]. 5th Conf Retrovir Oppor Infect.; 1-5 Feb 1998:208.
24. Paredes R., Mocroft A., Kirk O., Lazzarin A., Barton S.E., van Lunzen J., Katzenstein T.L., Antunes F., Lundgren J.D., Clotet B. (2000). Predictors of virological success and ensuing failure in HIV-positive patients starting highly active antiretroviral therapy in Europe: results from the Euro SIDA study. Arch Intern Med.; 160:1123-1132.
25. Ledergerber B., Egger M., Opravil M., Telenti A., Hirschel B., Battegay M., Vernazza P., Sudre P., Flepp M., Furrer H., Francioli P., Weber R. (1999). Clinical progression

and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Swiss HIV Cohort Study. Lancet; 353:863-868.

26. Powderly W.G., Saag M.S., Chapman S., Yu G., Quart B., Clendeninn N.J. (1999). Predictors of optimal virologic response to potent antiretroviral therapy. AIDS;13:1873-80
27. Raboud J.M., Harris M., Rae S., Montaner J.S. (2002). Impact of adherence on duration of virological suppression among patients receiving combination antiretroviral therapy. HIV Med; 3:118-24.
28. Frater A.J., Dunn D.T., Beardall A.J. (2002). Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. AIDS; 16:1139-1146.
29. Johannessen A., Naman E., Ngowi B.J., Sandvik L., Matee M.I., Aglen H.E., Gundersen S.G. and Bruun J.N. (2008). Predictors of mortality in HIV-1 infected patients starting antiretroviral therapy in a rural hospital in Tanzania. BMC Infectious Diseases, **8**:52
30. Abaasa A.M., Todd J., Ekoru K., Kalyango J.N., Levin J., Odeke E. and Karamagi C. A.S. (2008). Good adherence to HAART and improved survival in a community HIV/AIDS treatment and care programme: the experience of the AIDS support organization (TASO), Kampala, Uganda. BMC Health Services Research, 8:241
31. Egger M., May M., Chêne G., Phillips A.N., Ledergerber B., Dabis F., Costagliola D., Monforte A.D.A., Wolf F., Reiss P., Lundgren J.D, Justice A.C., Staszewski S., Leport C., Hogg R.S., Sabin C.A., Gill M.J., Salzberger B., Sterne J.A.C. (2002). HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet, 360: 119–29.
32. 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., Brasher.C (2006). Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. Lancet; 367: 1335–42
33. 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.

34. Lo´pez-Gatell H., Cole S.R., Margolick J.B., Witt M.D., Martinson J., Phair J.P. and Jacobson L.P. (2008). Effect of tuberculosis on the survival of HIV-infected men in a country with low tuberculosis incidence. AIDS, 22:1869–1873.
35. Moha R., Danela C., Messoua E., Ouassaa T., Gabillarda D., Anziana A., Aboa Y., Salamona R., Bissagnenea E., Seylera C., Eholie S. and Anglareta X. (2007). Incidence and determinants of mortality and morbidity following early antiretroviral therapy initiation in HIV-infected adults in West Africa. AIDS, 21:2483–2491.
36. Jarrin I., Geskus R., Bhaskaran K., Prins M., Perez- Hoyos S., Muga R., Herna´ndez-Aguado I., Meyer L., Porter K., Amo J. (2008). Gender Differences in HIV Progression to AIDS and Death in Industrialized Countries: Slower Disease Progression Following HIV Seroconversion in Women. Am J Epidemiol,168:532–540.
37. Wood E., Hogg R.S., Lima V.D. (2008). Highly Active Antiretroviral Therapy and Survival in HIV-Infected Injection Drug Users. JAMA, 300(5):550-554.
38. Hera M.G., Ferreros I., Amo J., Olalla P. G., Hoyos S.P., Muga R., Romero J., Guerrero R., Herna´ndez-Aguado I. (2004). Gender differences in progression to AIDS and death from HIV seroconversion in a cohort of injecting dug users from 1986 to 2001. J Epidemiol Community Health; 58:944–950.
39. Kigozi B.K., Sumba S., Mudyope P., Namuddu B., Kalyango J., Karamagi C., Odere M., Katabira E., Mugyenyi P. and Ssali F. (2009). The effect of AIDS defining conditions on immunological recovery among patients initiating antiretroviral therapy at Joint Clinical Research Centre, Uganda. AIDS Research and Therapy; 6:17.
40. Lawn S.D., Badri M. and Wood R. (2005). Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort. AIDS; 19:2109–2116.
41. Corey D.M., Kim H.W., Salazar R., Illescas R., Villena J., Gutierrez L., Sanchez J., and. Tabet S. R. (2007).Corey et al; Effectiveness of Combination Antiretroviral Therapy on Survival and Opportunistic Infections in a Developing World Setting: An Observational Cohort Study. J Acquir Immune Defic Syndr; 44:451–455.
42. Lawn S.D, Little F., Bekker L., Kaplan R., Campbel E., Orrell C. and Wood R. (2009). Changing mortality risk associated with CD4 cell response to antiretroviral therapy in South Africa. AIDS; 23:335–342.
43. Kebede D., Fikre H., Sibhatu B., Alemayehu A. and Biruk K.B. (2008). Defaulters from antiretroviral treatment in Jimma University Specialized Hospital, Southwest Ethiopia: Tropical Medicine and International Health; 13(3): 328–333.

44. Moattia J.P., Carrieria M.P., Spirea B., Gastauta J.A., Cassutod J. P., Moreaue J. (2000). Adherence to HAART in French HIV-infected injecting drug users: the contribution of buprenorphine drug maintenance treatment. AIDS; 14:151-155.
45. Pathipvanich P., Ariyoshi K., Rojanawiwat A., Wongchoosie S., Yingseree P., Yoshiike K., Warachit P., and Sawanpanyalert P. (2003). Survival Benefit from Non-Highly Active Antiretroviral Therapy in a Resource-Constrained Setting. JAIDS Journal of Acquired Immune Deficiency Syndromes; 32:157-160.
46. Galai N., Vlahov D., Bareta J.C., Wang C., Cohn S., and Sterling T.R. (2005). Prognostic Factors for Survival Differ According to CD4+ Cell Count Among HIV-Infected Injection Drug Users :Pre-HAART and HAART Eras. J Acquir Immune Defic Syndr; 38:74-81.
47. Liu C., Johnson L., Ostrow D., Silvestre A., Visscher B., and Jacobson L.P. (2006). Predictors for Lower Quality of Life in the HAART Era Among HIV-Infected Men. J Acquir Immune Defic Syndr; 42:470- 477.
48. Degu J. and Lindtjørn B. (2005).Disease Progression among Untreated HIV-Infected Patients in South Ethiopia: Implications for Patient Care. Medscape General Medicine; 7(3):66.
49. Degu J., Næss A and Lindtjørn B. (2006). Antiretroviral therapy at a district hospital in Ethiopia prevents death and tuberculosis in a cohort of HIV patients. AIDS Research and Therapy; 3:10.
50. Degu J., Endale A, Hailu Y and Lindtjørn B. (2006). Predictors of early death in a cohort of Ethiopian patients treated with HAART. BMC Infectious Diseases; 6:136.
51. Hogg R.S., Heath K., Bangsberg D. (2002). Intermittent use of triple-combination therapy is predictive of mortality at baseline and after 1 year of follow-up. AIDS; 16: 1051–1058.
52. Kleinbaum, D.G. and Klein, W. (2005). Survival Analysis a self learning text, second edition. Springer Science+Business Media, Inc. , New York.
53. Klein, J.P. and Moeschberger, M.L. (1997). Survival Analysis techniques for censured and truncated data. Springer-Verlag, New York.
54. Hosmer, D.W. and Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York.
55. Collett, D. (2003). Modelling survival data in medical research (Second edition). Chapman and Hall/CRC, London.

56. Therneau, T.M., Grambsch, P.M., and Fleming, T.R. (1990). Martingale-based residuals for survival models. Biometrika; 77: 147-160.
57. Cox, D.R. (1972). Regression models and life Tables (with Discussion). Journal of the Royal statistical society; series B 34:187-220.
58. Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika; 81:515-526.
59. Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. Biometrica; 69: 239-241.

APPENDIX

Table 4.1.1: The Distribution of Important Socio-Demographic and Clinical Characteristics of HIV patients treated with HAART in JUSH, Jimma, 2010.

Covariate / factor	Category	Censored	Dead (%)	Total
Sex	Female	394	81 (17.1)	475
	Male	294	63 (17.6)	357
Marital status	Never married	134	33 (19.8)	167
	Married	347	72 (17.2)	419
	Others	207	39 (15.9)	246
Religion	Muslim	180	42 (18.9)	222
	Orthodox	397	85 (17.6)	482
	Others	111	17 (13.3)	128
Level of education	No education	119	27 (18.5)	146
	Primary	252	43 (14.6)	295
	Secondary or above	317	74 (18.9)	391
Employment status	Employed	195	39 (16.7)	234
	Non Employed	493	105 (17.6)	598
Place of residence	Jimma	180	39 (17.8)	219
	Others	508	105 (17.1)	613
No of Rooms	One room	369	73 (16.5)	442
	Two or more rooms	319	71 (18.2)	390
Disclosure status	Disclosed	601	127 (17.4)	728
	Not disclosed	87	17 (16.3)	104
Risk behaviour	Regular	463	84 (15.4)	547
	Casual or Both	225	60 (21.1)	285
Substance use	Yes	216	60 (21.7)	276
	No	472	84 (15.1)	556
Functional status	Working	398	33 (7.7)	431
	Ambulatory	259	84 (24.5)	343
	Bedridden	31	27 (46.6)	58
WHO clinical stage	Stage I or II	130	19 (12.8)	149
	Stage III	411	75 (15.4)	486
	Stage IV	147	50 (25.4)	197
Past OI	Yes	340	88 (20.6)	428
	No	348	56 (13.9)	404
TB co-infection	Yes	404	108 (21.1)	512
	No	284	36 (11.3)	320
Drug regimen	D4T-based	563	123 (17.9)	686
	AZT-based	125	21 (14.4)	146

Table 4.1.2: Summary statistics of continuous variables included in the study of HIV patients treated with HAART in JUSH, Jimma, 2010.

Patient status	Continuous Variable	Mean	Std. Dev.	Minimum	Maximum	1st quartile	2nd quartile	3rd quartile
Censored	Follow up time	35.49	18.34	1	78	20	42	47
	Age	32.19	7.73	18	65	26	30	36
	CD4 count	153.79	81.69	20	500	88	146	209
	Weight	50.93	9	30	82	45	50	56
Dead	Follow up time	7.15	10.37	1	69	1	3	8
	Age	34.22	8.68	18	60	27	33	40
	CD4 count	105.32	82.02	3	372	38	87	144
	Weight	47.49	8.38	24	79	41	47	54
Overall	Follow up time	30.58	20.289	1	78	7	40	46
	Age	32.54	7.930	18	65	27	31	37
	CD4 count	145.40	83.729	3	500	79	137	200
	Weight	50.33	9.149	24	82	45	50	55

Table 4.2.2: Results of the Log-rank test for the categorical variables of HIV patients treated with HAART in JUSH, Jimma, 2010.

Covariate / factor	DF	Chi-square	P-Value
Sex	1	0.21	0.6466
Marital status	2	0.93	0.6287
Religion	2	2	0.3679
Level of education	2	2.15	0.3415
Employment status	1	0.13	0.7171
Residence	1	0.11	0.7369
Number of rooms	1	0.12	0.7282
Disclosure status	1	0.03	0.8519
Risk behaviour	**1**	**5.56**	**0.0184**
Substance use	**1**	**6.17**	**0.0130**
Functional status	**2**	**95.80**	**0.0000**
Who clinical stage	**2**	**13.48**	**0.0012**
Past OI	**1**	**5.97**	**0.0145**
TB co-infection	**1**	**13.38**	**0.0003**
Drug regimen	1	0.61	0.4335

Table 4.3.1: Results of the univariable proportional hazards Cox regression model of HIV patients treated with HAART in JUSH, Jimma, 2010.

Variables	B	SE	Wald	Df	Sig.	Exp(B)	LR Sig.	-2log L	Score
Sex	.076	.168	.205	1	.651	1.079	.651	1875.596	.651
Marital status			.903	2	.637		.641	1874.910	.636
Married	.224	.237	.896	1	.344	1.251			
Never married	.089	.199	.199	1	.655	1.093			
Religion			1.931	2	.381		.348	1873.688	.377
Orthodox	.372	.288	1.672	1	.196	1.451			
Muslim	.351	.266	1.743	1	.187	1.420			
Education			2.085	2	.353		.341	1873.648	.351
Primary	-.033	.225	.022	1	.882	.967			
Secondary or above	-.270	.192	1.989	1	.158	.763			
Employment status	.067	.188	.128	1	.720	1.070	.719	1875.671	.720
Residence	.062	.188	.110	1	.740	1.064	.741	1875.692	.74
Room	.057	.167	.118	1	.731	1.059	.731	1875.683	.731
Disclosure	.048	.258	.034	1	.854	1.049	.853	1875.766	.854
Risk	**.392**	**.169**	**5.353**	**1**	**.021**	**1.479**	**.022**	**1870.592**	**.020**
Substance	**.412**	**.169**	**5.933**	**1**	**.015**	**1.510**	**.016**	**1870.040**	**.014**
Functional status			**76.137**	**2**	**.000**		**.000**	**1797.418**	**.000**
Ambulatory	1.262	.202	39.113	1	.000	3.532			
Bedridden	2.222	.261	72.549	1	.000	9.229			
WHO clinical stage			**12.720**	**2**	**.002**		**.002**	**1863.794**	**.001**
Stage III	.203	.257	.626	1	.429	1.225			
Stage IV	.772	.270	8.185	1	.004	2.164			
OI	**.410**	**.171**	**5.747**	**1**	**.017**	**1.507**	**.015**	**1869.915**	**.016**
TB	**.683**	**.193**	**12.567**	**1**	**.000**	**1.980**	**.000**	**1861.911**	**.000**
Regimen	.182	.236	.597	1	.440	1.200	.430	1875.178	.439
Age	**.029**	**.010**	**8.448**	**1**	**.004**	**1.029**	**.005**	**1867.790**	**.004**
Weight	**-.046**	**.010**	**21.606**	**1**	**.000**	**.955**	**.000**	**1853.051**	**.000**
CD4 count	**-.008**	**.001**	**42.511**	**1**	**.000**	**.992**	**.000**	**1825.302**	**.000**

N.B the value of -2log L for the null model is 1875.8

Table 4.3.2: Results of the multivariable proportional hazards Cox regression model containing the variables significant at 20% level in the univariable proportional hazards Cox regression model of HIV patients treated with HAART in JUSH, Jimma, 2010.

Covariates/ Factors	DF	Wald χ^2	Wald P-value	LR χ^2	LR P-value
Age at HAART initiation	1	9.3634	0.0022	8.870	0.0029
Baseline CD4 count	1	23.0477	<.0001	26.535	<.0001
Baseline weight	1	3.2119	0.0731	3.302	0.0692
Risk behaviour	1	2.4527	0.1173	2.410	0.1206
Substance use	1	3.4013	0.0651	3.332	0.0680
Functional status	2	43.0526	<.0001	45.312	<.0001
WHO clinical stage	2	1.0774	0.5835	1.073	0.5848
Past opportunistic infection	**1**	**.0000**	**0.9963**	**.000**	**0.9963**
Tuberculosis co-infection	1	6.2385	0.0125	6.132	0.0133

Table 4.3.3: Results of the multivariable proportional hazards Cox regression model after eliminating the variable past opportunistic infection from the multivariable proportional hazards Cox regression model in Table 4.3.2.

Covariates/ Factors	DF	Wald χ^2	Wald P-value	LR χ^2	LR P-value
Age	1	9.3828	0.0022	8.890	0.0029
Baseline CD4 count	1	23.0826	<.0001	26.723	<.0001
Baseline weight	1	3.2186	0.0728	3.307	0.0690
Risk behaviour	1	2.4636	0.1165	2.420	0.1198
Substance use	1	3.4032	0.0651	3.334	0.0679
Functional status	2	43.1706	<.0001	45.406	<.0001
Who clinical stage	**2**	**1.0774**	**0.5835**	**1.073**	**0.5848**
Tuberculosis co-infection	1	10.8663	0.0010	11.930	0.0006

Table 4.3.4: Results of the multivariable proportional hazards Cox regression model after eliminating the variable WHO clinical stage from the multivariable proportional hazards Cox regression model in Table 4.3.3

Covariates/ Factors	DF	Wald χ^2	Wald P-value	LR χ^2	LR P-value
Age	1	9.3671	0.0022	8.869	0.0029
Baseline CD4 count	1	23.8494	<.0001	27.499	<.0001
Baseline weight	1	3.7914	0.0515	3.895	0.04844
Risk behaviour	**1**	**2.5827**	**0.1080**	**2.536**	**0.11125**
Substance use	1	3.7305	0.0534	3.648	0.05612
Functional status	2	45.9311	<.0001	47.083	<.0001
Tuberculosis co-infection	1	11.0256	0.0009	12.115	0.0005

Table 4.3.5: Results of the multivariable proportional hazards Cox regression model after eliminating the variable risk behaviour from the multivariable proportional hazards Cox regression model in Table 4.3.4

Covariates/ Factors	DF	Wald χ^2	Wald P-value	LR χ^2	LR P-value
Age	1	8.7765	0.0031	8.317	0.0039
Baseline CD4 count	1	23.4349	<.0001	27.078	<.0001
Baseline weight	1	4.1105	0.0426	4.223	0.0399
Substance use	1	4.2030	0.0404	4.103	0.0428
Functional status	2	49.4557	<.0001	50.299	<.0001
Tuberculosis co-infection	1	11.1210	0.0009	12.225	0.0005

Table 4.3.6: Likely hood ratio statistics and Percentage changes in the coefficients of the variables included in table 4.3.5, when the variables that were not significant in the univariable proportional hazards Cox regression model are added one at a time

Covariates / Factors	Sex	Marital status	Religion	Education	Employment status	Residence	Number of rooms	Disclosure status	Regimen
Age	2.26	-11.2	0.07	-4.57	1.03	-0.22	-3.77	0.42	1.11
Baseline CD4 count	0.92	3.11	0.11	2.18	-0.27	0.26	-0.25	0.12	-0.25
Baseline weight	-6.81	-8.74	0.72	-19.7	-7.57	-0.01	-0.32	-0.6	-1.46
Substance use	2.74	1.05	1	9.25	1.95	-0.63	3.74	-0.41	1.64
Functional status									
Ambulatory	-0.27	-0.85	0.14	-0.05	-1.39	-0.55	-0.22	-0.1	0.46
Bedridden	0.43	0.61	0.39	1.57	-0.01	-0.16	-0.55	0.31	0.52
Tuberculosis co-infection	0.30	1.82	-0.06	-0.03	0.02	-0.21	0.07	-0.34	0.18
LR χ^2	0.13	1.61	0.06	2.69	0.72	0.12	0.37	0.05	0.14
LR p-value	0.72	0.45	0.97	0.26	0.40	0.73	0.54	0.82	0.71

Table 4.3.7: Wald statistics and corresponding p-values of possible interaction terms, added one at a time, to the variables included in the model in Table 4.3.5

Interaction between	Covariates/ Factors	DF	Wald χ^2	P-value
Age	Baseline CD4 count	1	2.8879	0.0892
	Baseline weight	1	0.1412	0.7071
	Substance use	1	0.4844	0.4864
	Functional status	2	4.0906	0.1293
	Tuberculosis co-infection	1	1.1069	0.2927
Baseline CD4 count	Baseline weight	1	0.2247	0.6355
	Substance use	1	0.2152	0.6427
	Functional status	2	1.0331	0.5966
	Tuberculosis co-infection	1	0.4841	0.4866
Baseline weight	Substance use	1	0.4817	0.4876
	Functional status	2	0.3948	0.8209
	Tuberculosis co-infection	1	0.0803	0.7769
Substance use	Functional status	2	3.9192	0.1409
	Tuberculosis co-infection	1	0.2444	0.6211
Functional status	Tuberculosis co-infection	2	0.7669	0.6815

Table 4.4.2.1: The five highest differences in the parameter estimates of the variables included in the model in Table 4.3.8 when the data value for each patient is in turn deleted from the model.

Covariate / factor	Deleted observation (i)	$\Delta_{j(-i)} = \widehat{\beta}_j - \widehat{\beta}_{j(-i)}$	$\left\|\Delta_{j(-i)} = \widehat{\beta}_j - \widehat{\beta}_{j(-i)}\right\|$
Age	641	-0.002661463	0.002661463
	784	0.001767119	0.001767119
	654	-0.001553907	0.001553907
	785	-0.001531866	0.001531866
	677	0.001426637	0.001426637
CD4	613	0.000384307	0.000384307
	815	0.000359062	0.000359062
	786	0.000347438	0.000347438
	141	0.000286337	0.000286337
	794	0.000283782	0.000283782
Weight	437	0.002868645	0.002868645
	641	0.002462941	0.002462941
	704	-0.002023444	0.002023444
	639	-0.001848839	0.001848839
	261	0.001724296	0.001724296
Substance use (yes)	641	-0.032551331	0.032551331
	437	0.019443765	0.019443765
	705	0.019340439	0.019340439
	807	0.019177044	0.019177044
	424	0.019173742	0.019173742
Functional status (Ambulatory)	614	-0.035966530	0.035966530
	583	-0.034821143	0.034821143
	108	-0.034534823	0.034534823
	655	-0.033928999	0.033928999
	421	-0.033737400	0.033737400
Functional status (Bedridden)	641	-0.059978009	0.059978009
	218	-0.043881392	0.043881392
	421	-0.043660919	0.043660919
	655	-0.043121168	0.043121168
	614	-0.042935320	0.042935320
TB (Co-infected)	81	-0.028700960	0.028700960
	832	-0.028451000	0.028451000
	75	-0.028447769	0.028447769
	146	-0.028439996	0.028439996
	493	-0.027779846	0.027779846

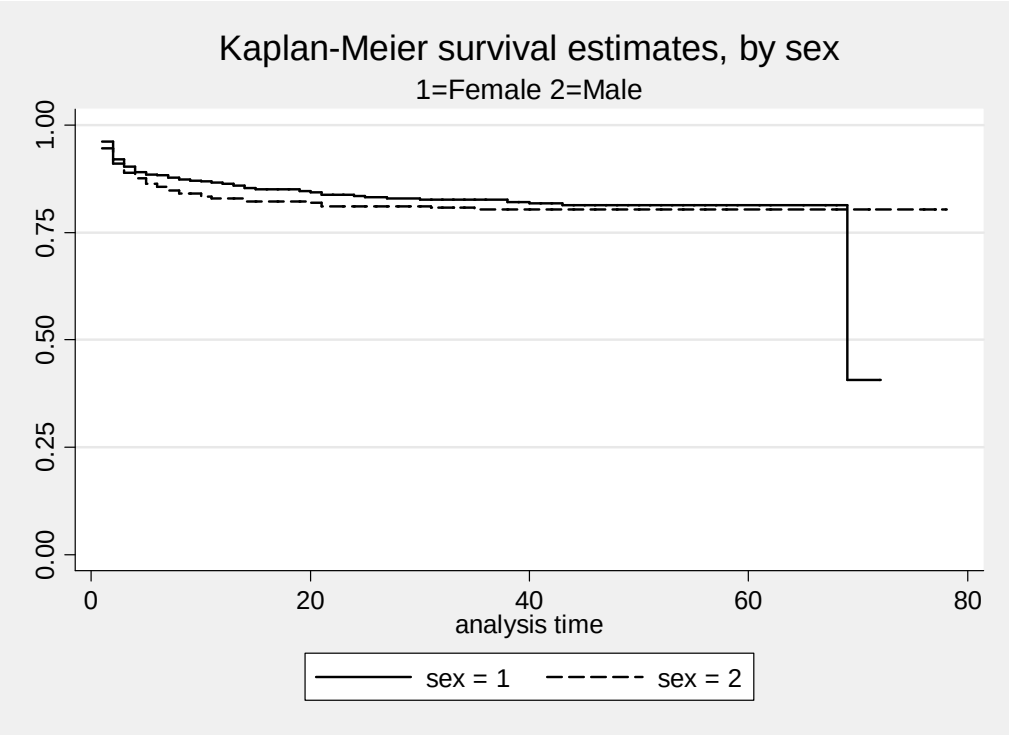
Table 4.4.2.2: The five highest likelihood displacement values when each observation is in turn deleted from the model in Table 4.3.8

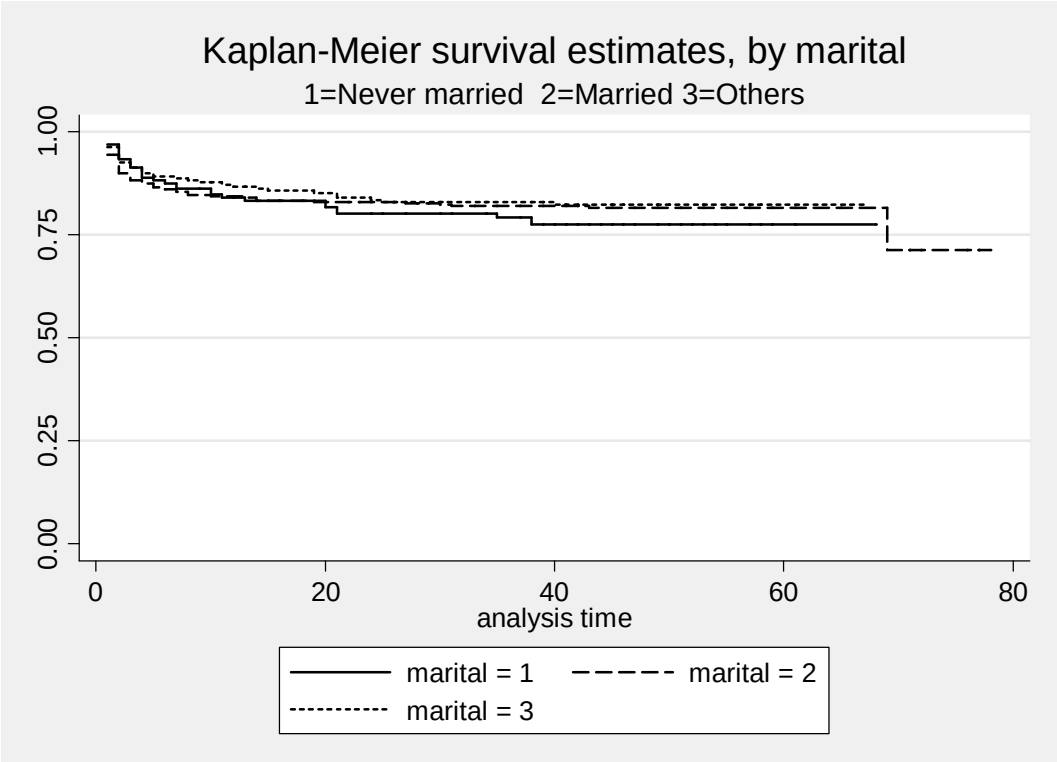
Deleted observation (i)	$LD_i = 2\left[LL(\beta) - LL(\beta_{-i})\right]$
641	0.277715
437	0.143899
613	0.132838
786	0.110308
815	0.100449

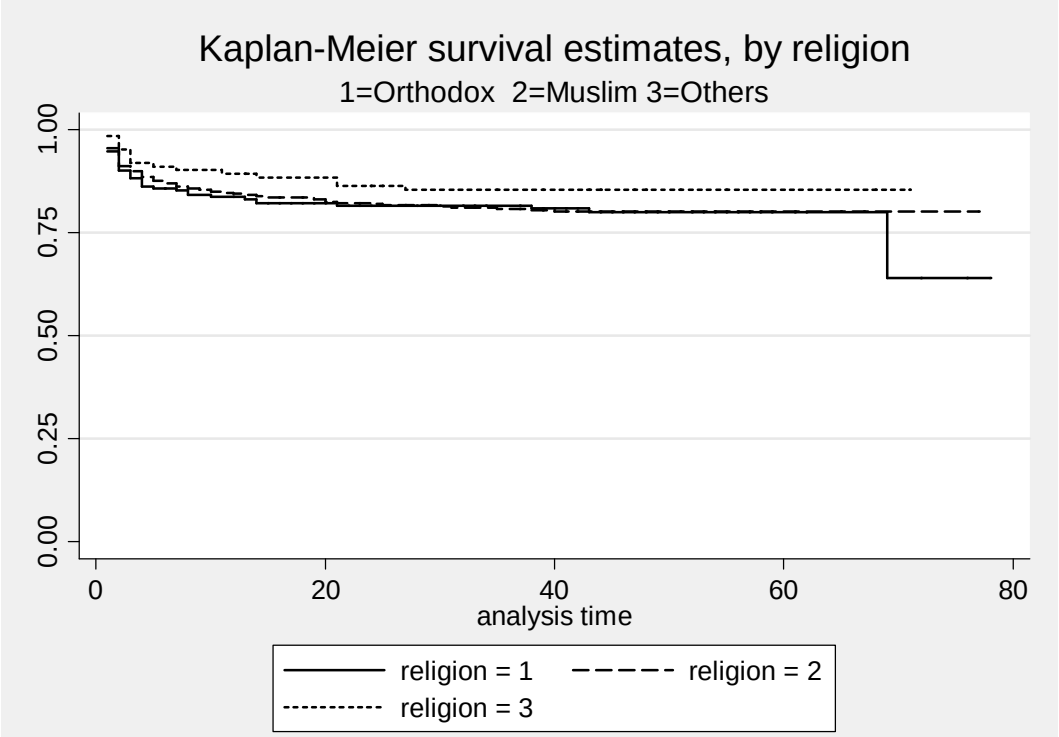
Table 4.4.3: Results of the Likelihood ratio, Score and Wald tests for testing the global null hypothesis: BETA=0

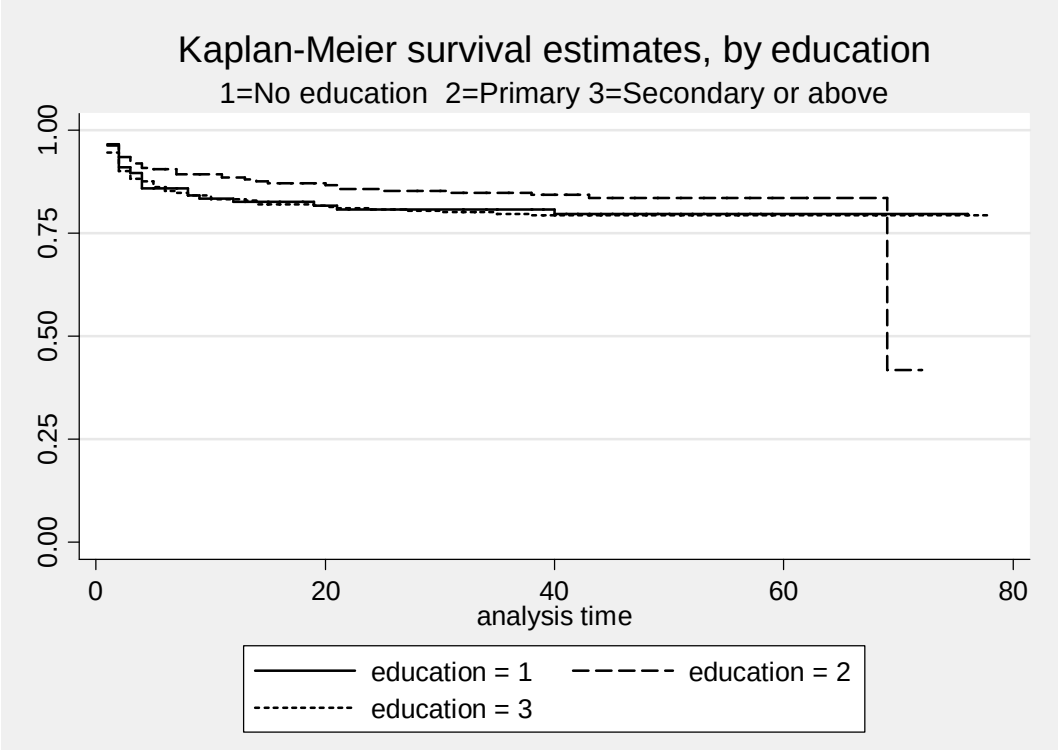
Test	DF	χ^2	P-Value
Likelihood Ratio	7	141.7097	<.0001
Score	7	154.4658	<.0001
Wald	7	129.0057	<.0001

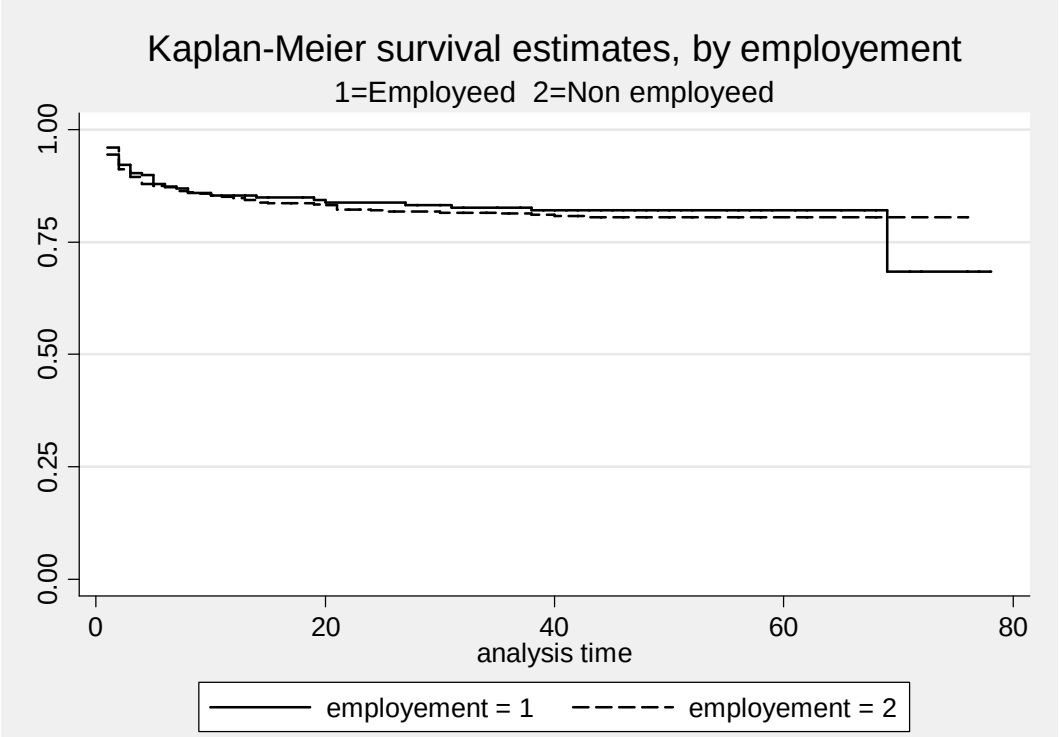
Figures 4.1.2: Plots of Kaplan-Meier survivor functions, based on different factors, of HIV patients treated with HAART in JUSH, Jimma, 2010.

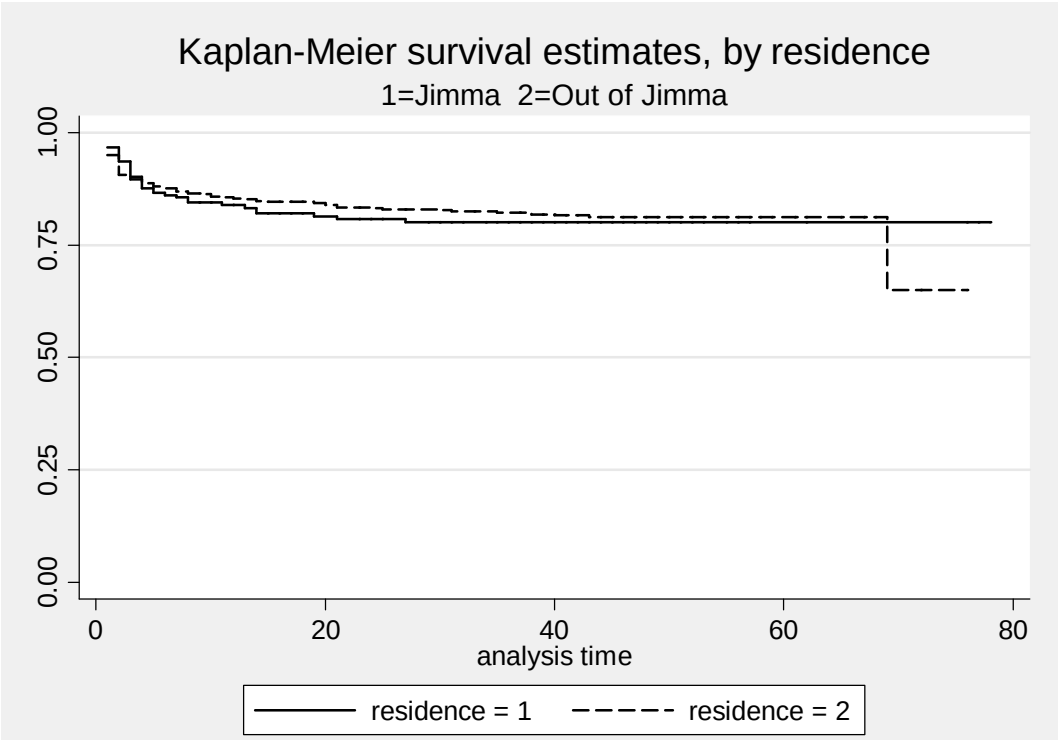


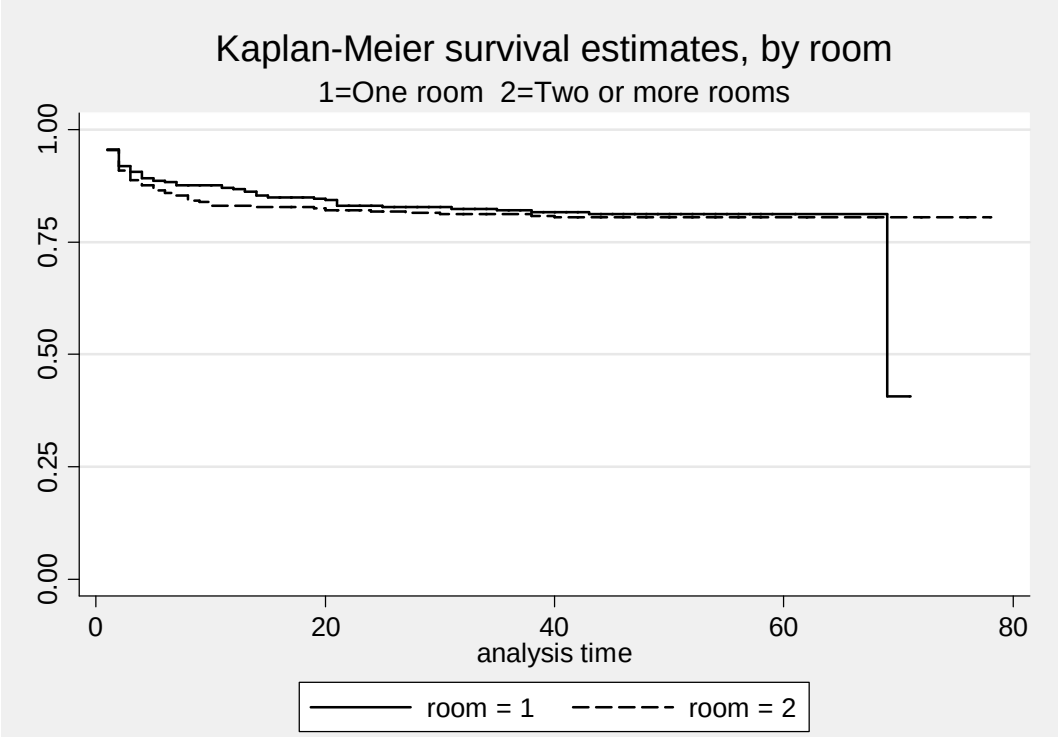


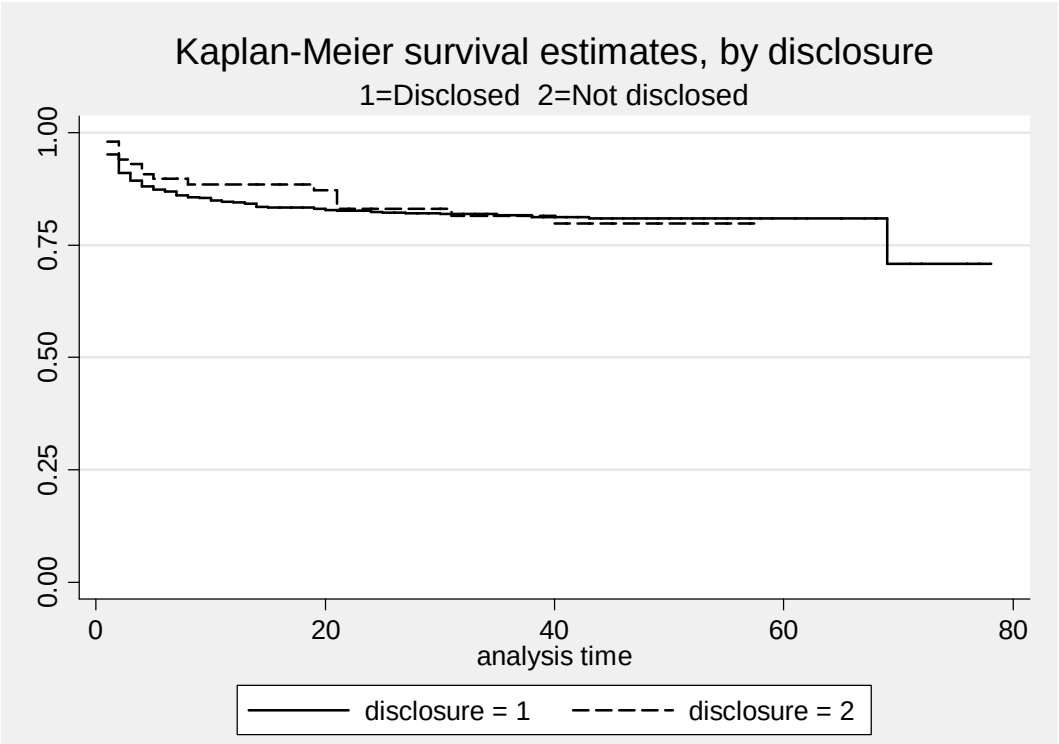

Kaplan-Meier survival estimates, by religion
1=Orthodox 2=Muslim 3=Others
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
religion = 1
religion = 2
religion = 3

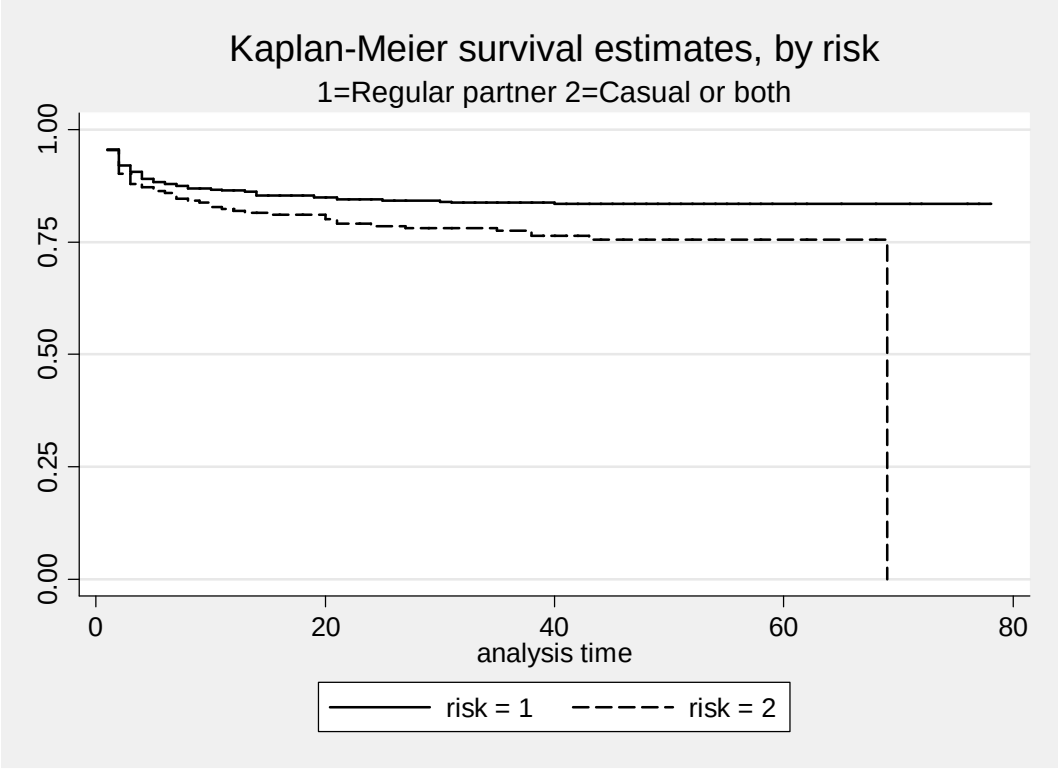

Kaplan-Meier survival estimates, by education
1=No education 2=Primary 3=Secondary or above
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
education = 1
education = 2
education = 3

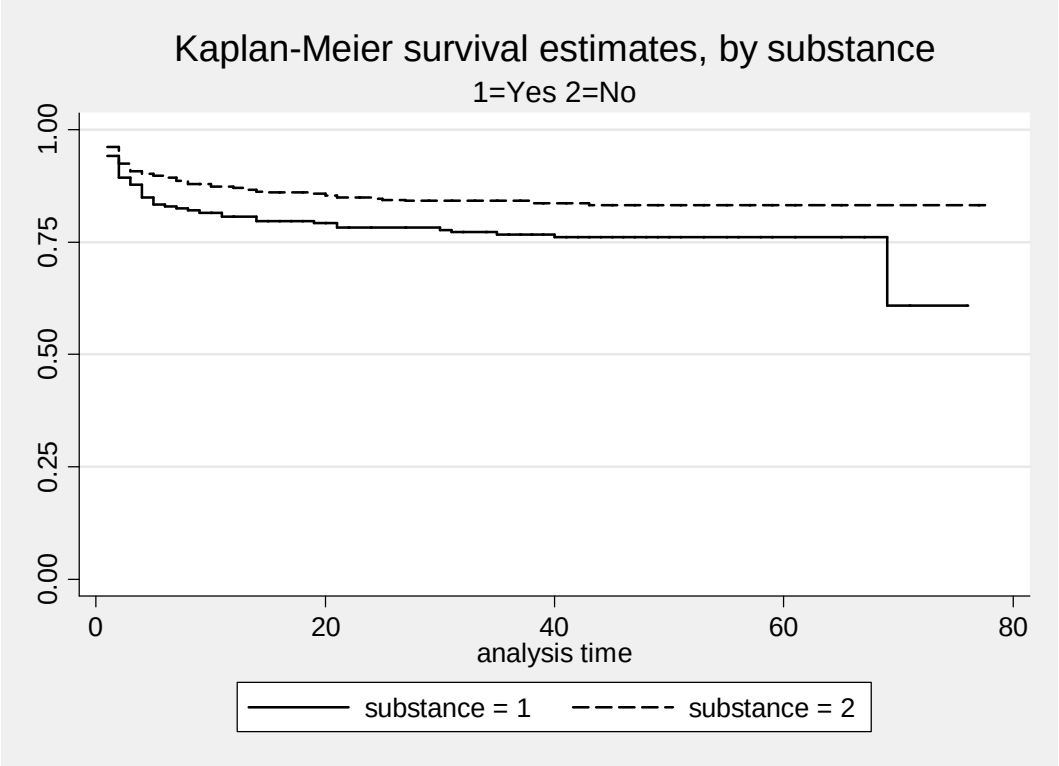

Kaplan-Meier survival estimates, by employement
1=Employeed 2=Non employeed
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
employement = 1
employement = 2

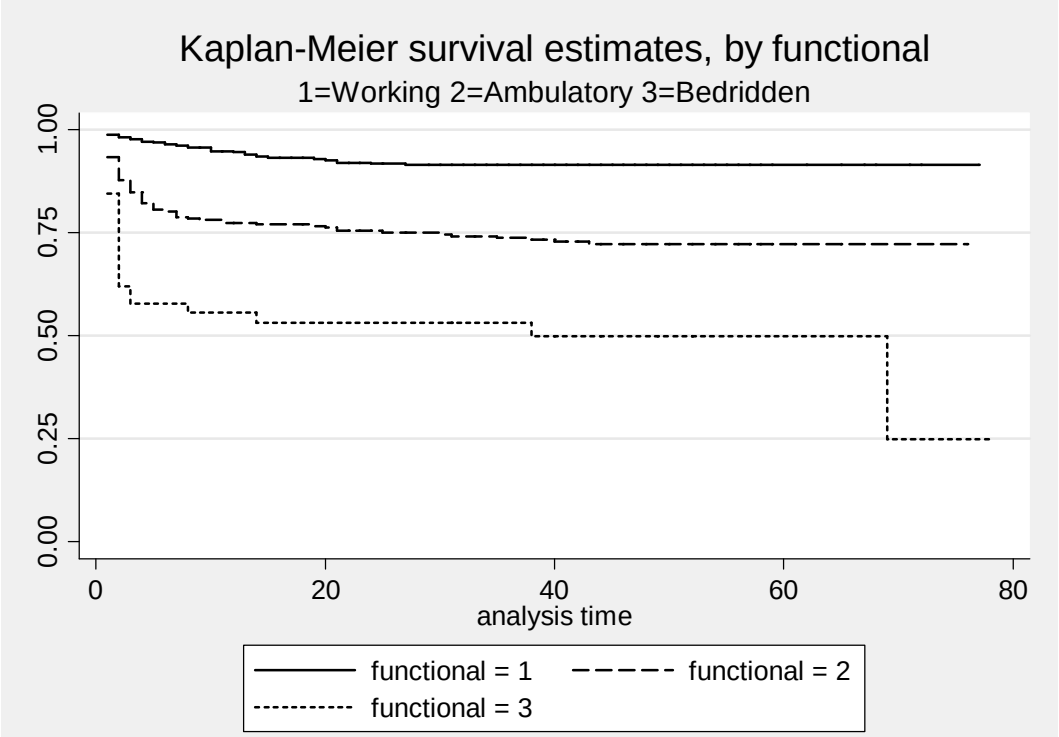

Kaplan-Meier survival estimates, by residence
1=Jimma 2=Out of Jimma
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
residence = 1
residence = 2

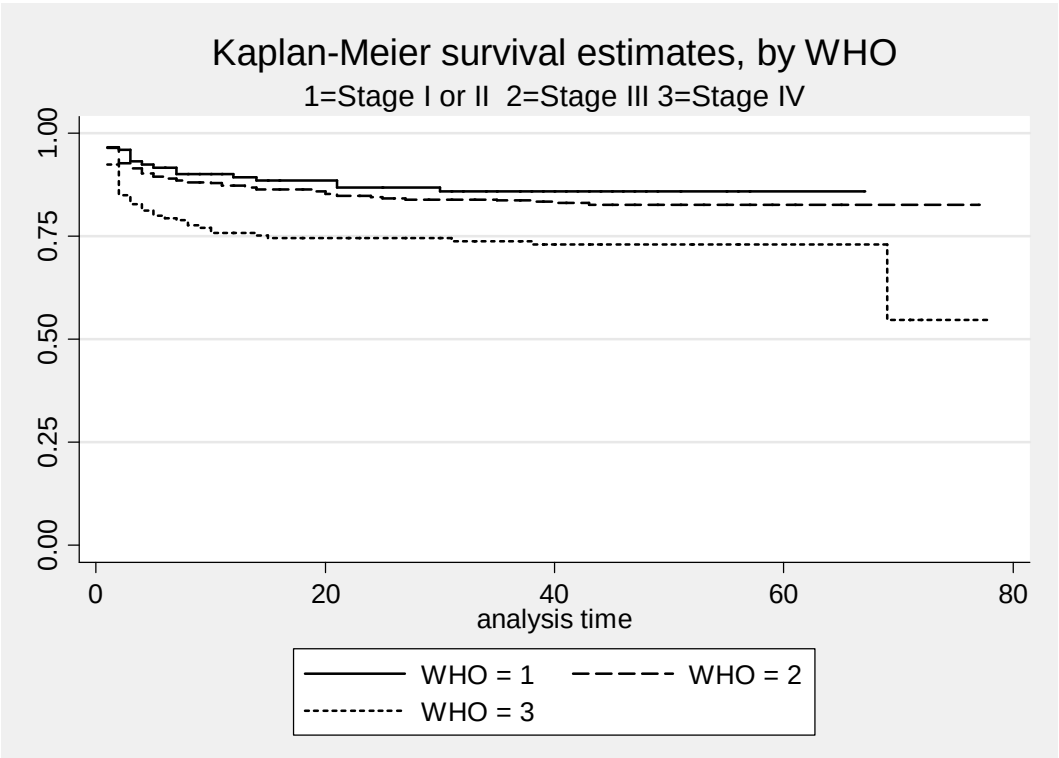

Kaplan-Meier survival estimates, by room
1=One room 2=Two or more rooms
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
room = 1
room = 2

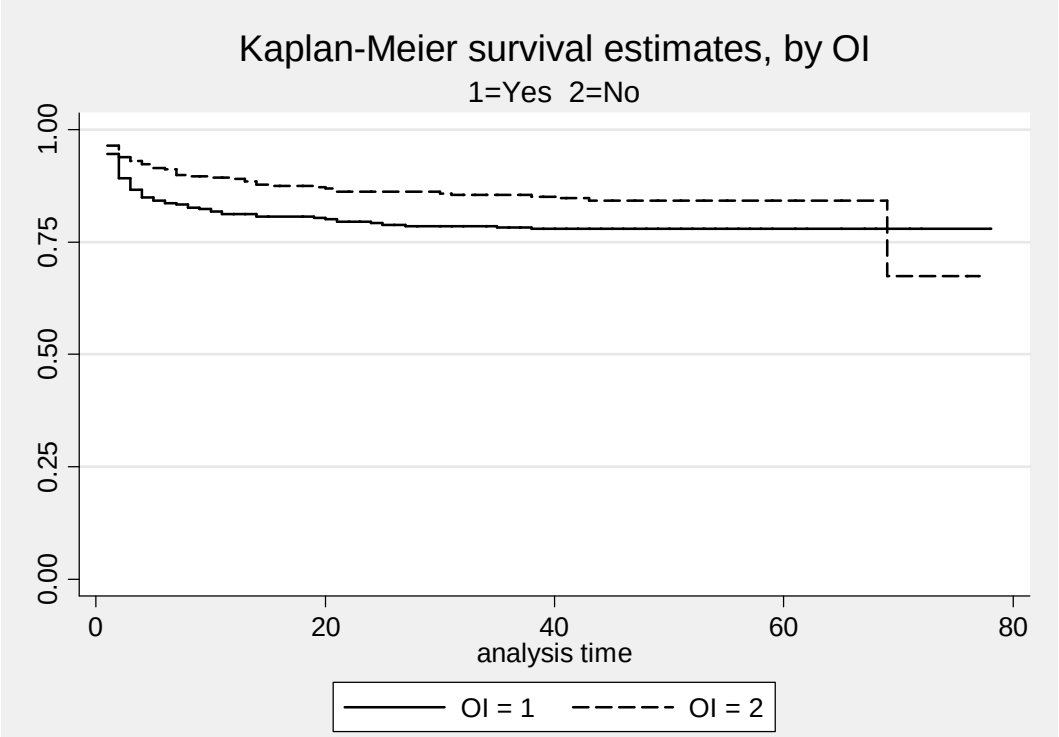

Kaplan-Meier survival estimates, by disclosure
1=Disclosed 2=Not disclosed
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
analysis time
disclosure = 1
disclosure = 2

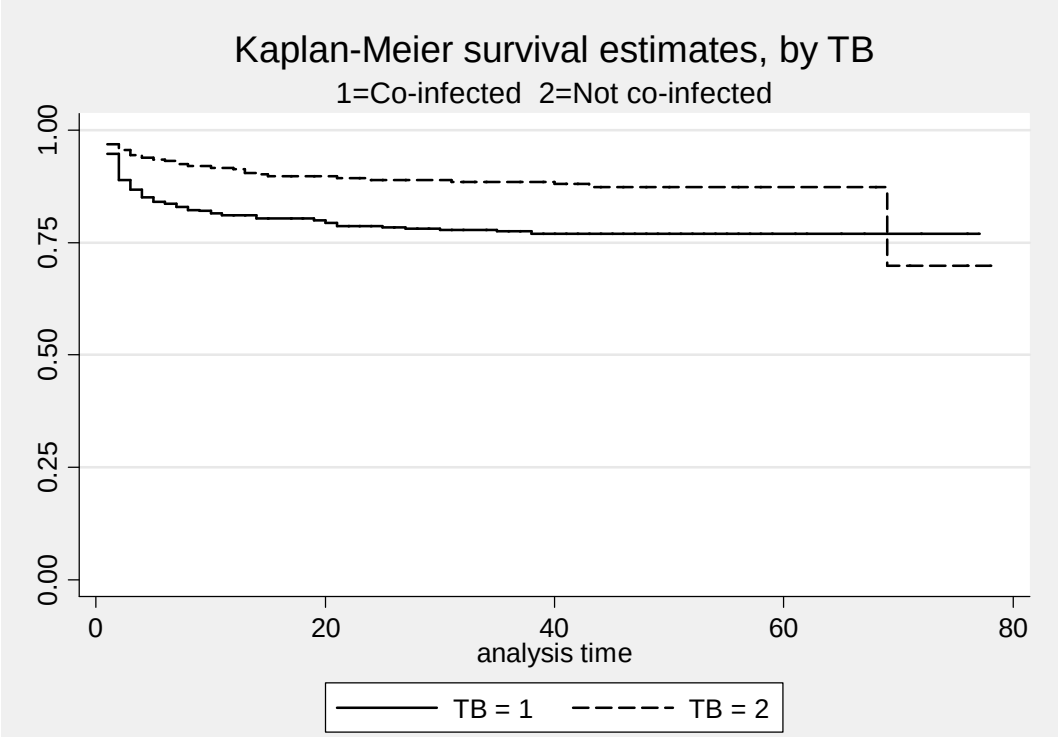

Kaplan-Meier survival estimates, by risk
1=Regular partner 2=Casual or both
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
risk = 1
risk = 2

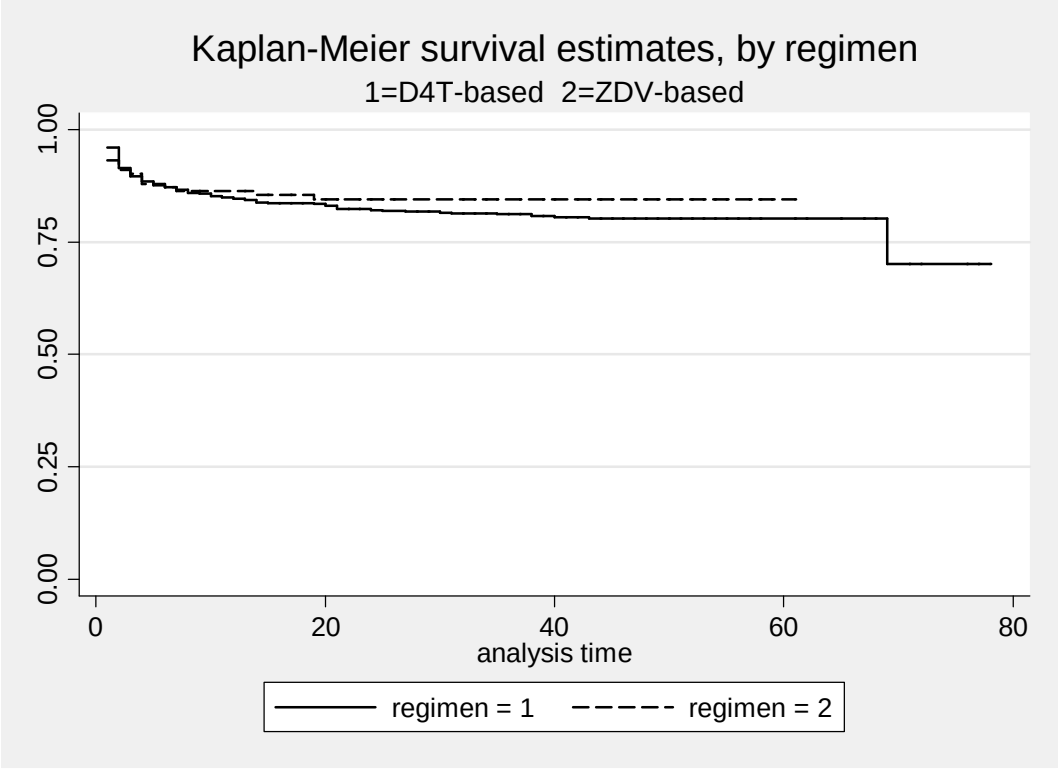

Kaplan-Meier survival estimates, by substance
1=Yes 2=No
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
substance = 1
substance = 2


Kaplan-Meier survival estimates, by functional
1=Working 2=Ambulatory 3=Bedridden
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
functional = 1
functional = 2
functional = 3


Kaplan-Meier survival estimates, by WHO
1=Stage I or II 2=Stage III 3=Stage IV
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
WHO = 1
WHO = 2
WHO = 3


Kaplan-Meier survival estimates, by OI
1=Yes 2=No
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
OI = 1
OI = 2


Kaplan-Meier survival estimates, by TB
1=Co-infected 2=Not co-infected
0.00
0.25
0.50
0.75
1.00
0
20
40
60
80
analysis time
TB = 1
TB = 2



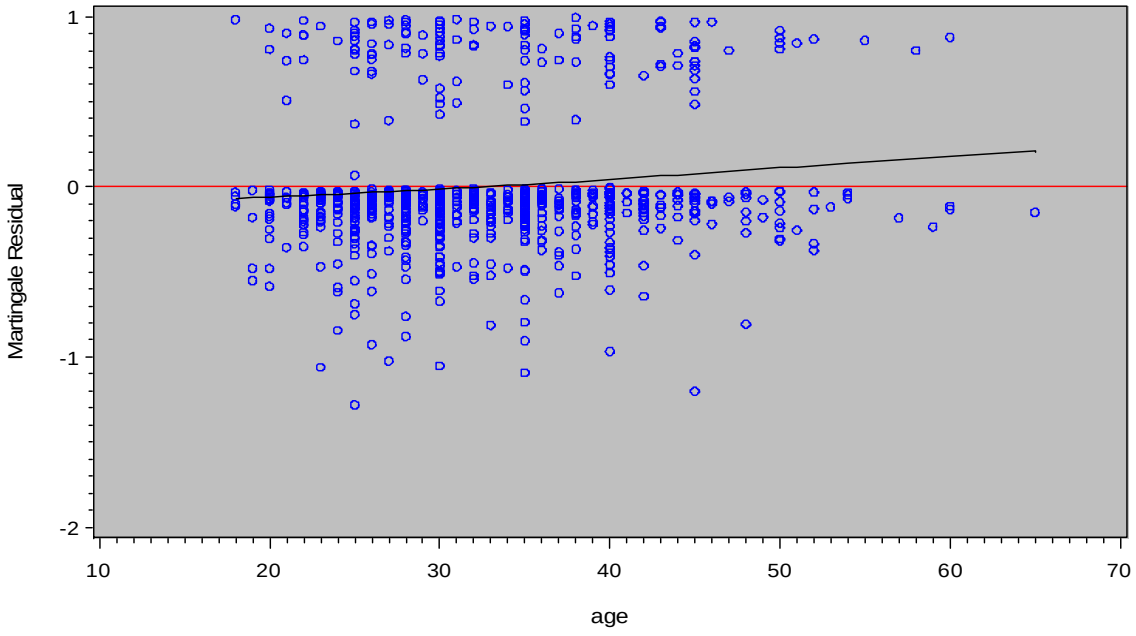


Figure 4.3.1: Plot of the martingale residuals for the model excluding age against the values of age, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

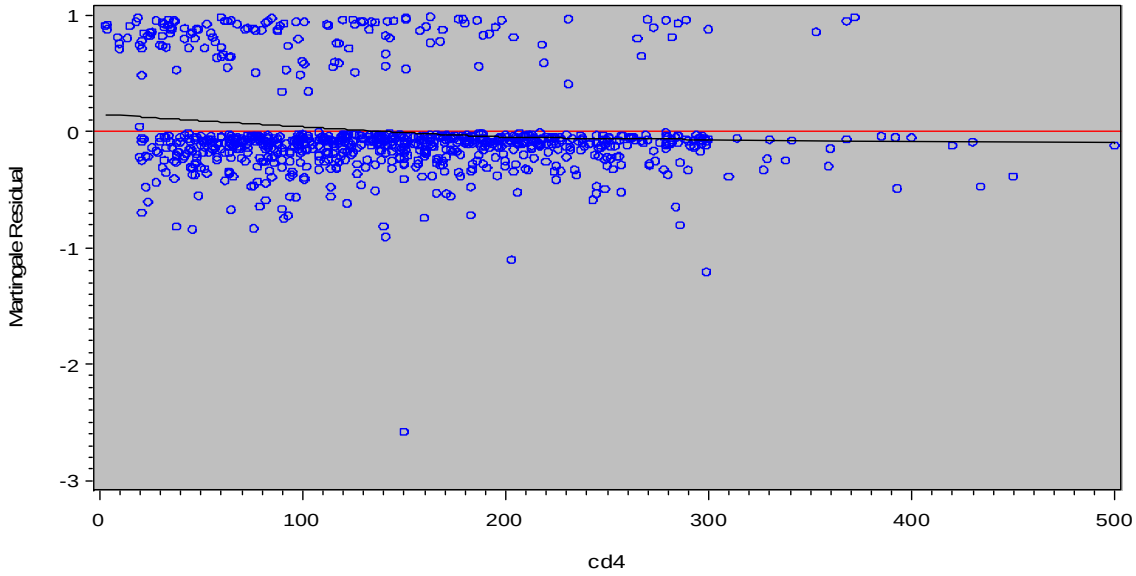


Figure 4.3.2: Plot of the martingale residuals for the model excluding baseline CD4 count against the values of baseline cd4 count, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

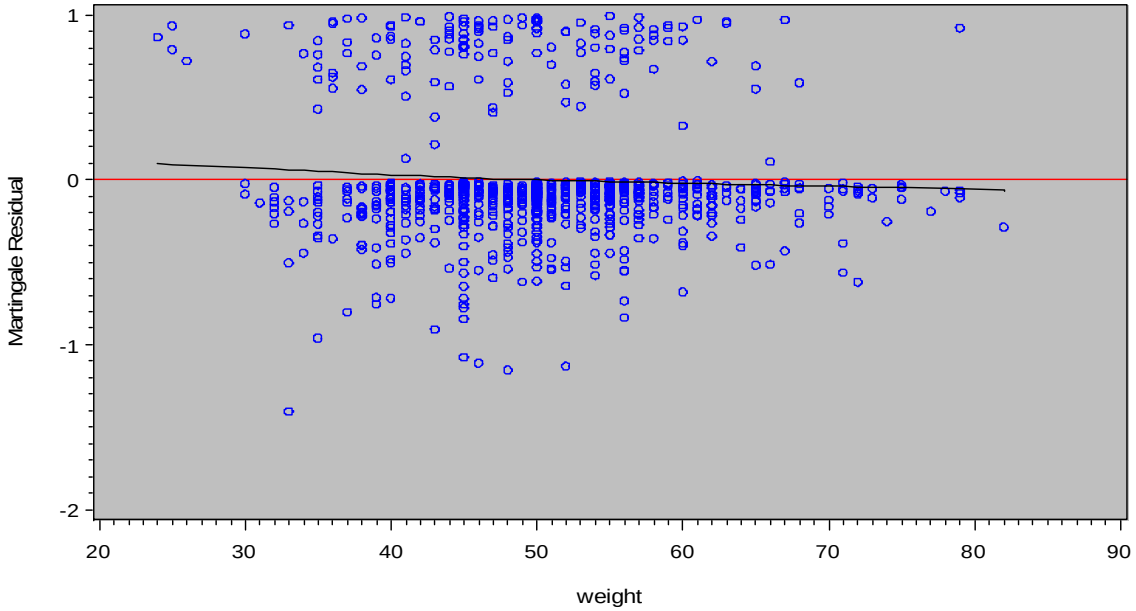


Figure 4.3.3: Plot of the martingale residuals for the model excluding baseline weight against the values of baseline weight, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

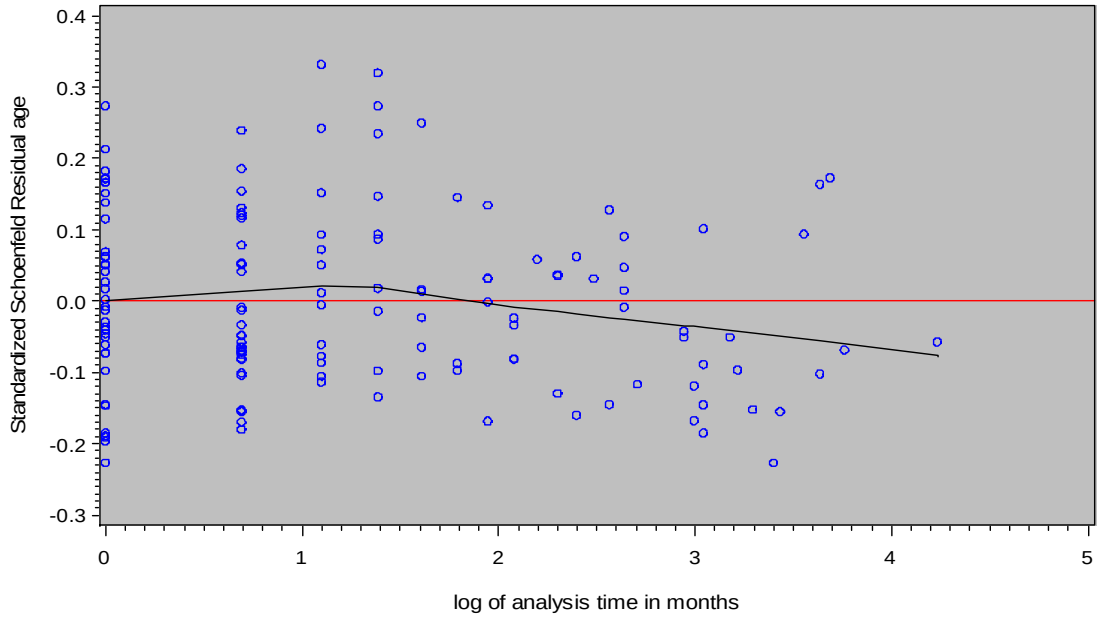


Figure 4.4.1.1: The plot of scaled Schoenfeld residuals for baseline age in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

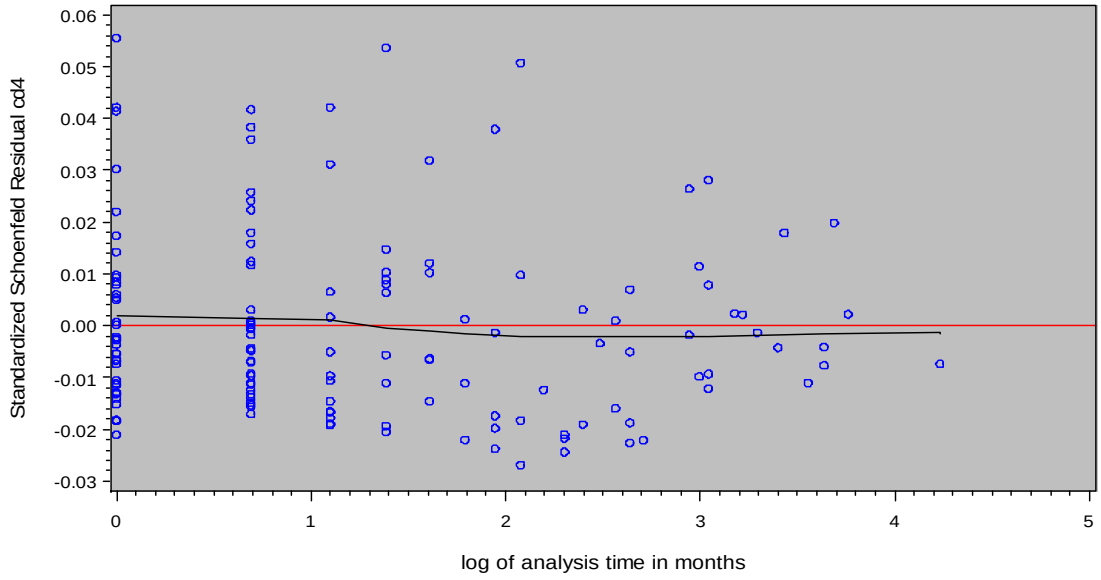


Figure 4.4.1.2: The plot of scaled Schoenfeld residuals for baseline CD4 count in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

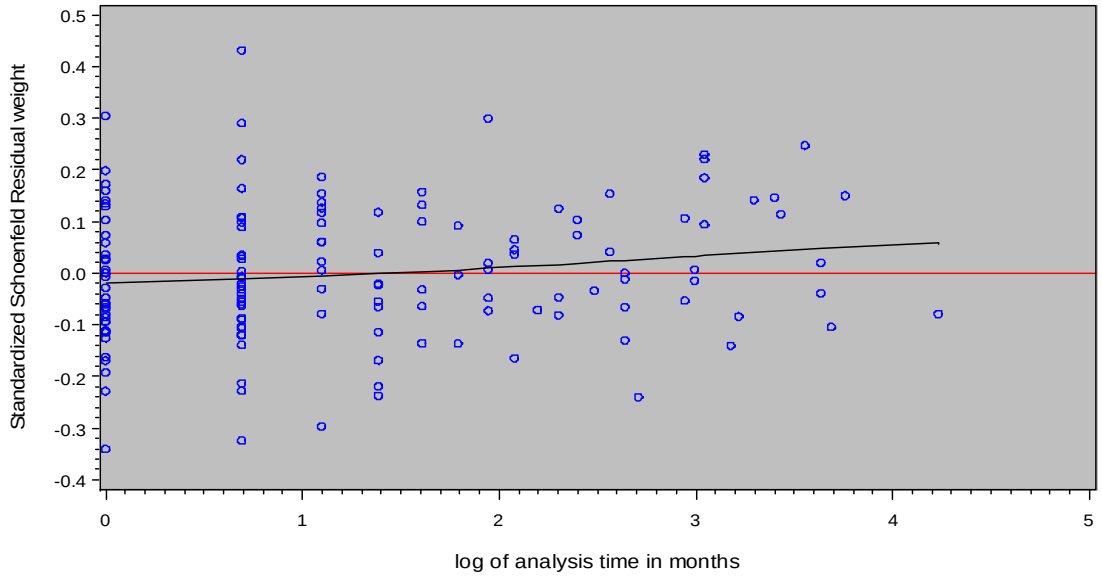


Figure 4.4.1.3: The plot of scaled Schoenfeld residuals for baseline weight in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

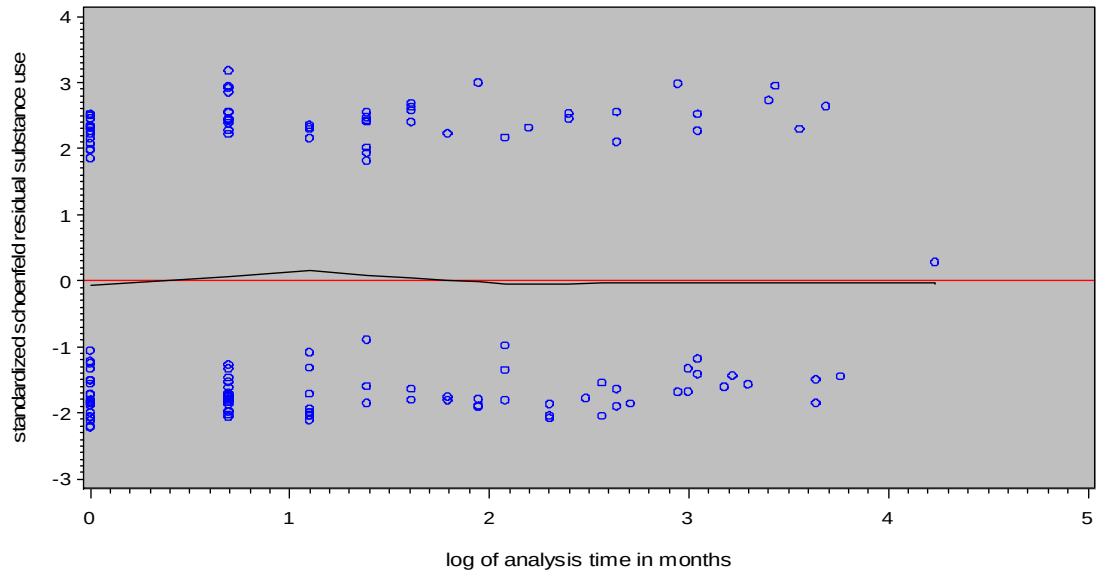


Figure 4.4.1.4: The plot of scaled Schoenfeld residuals for substance use in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

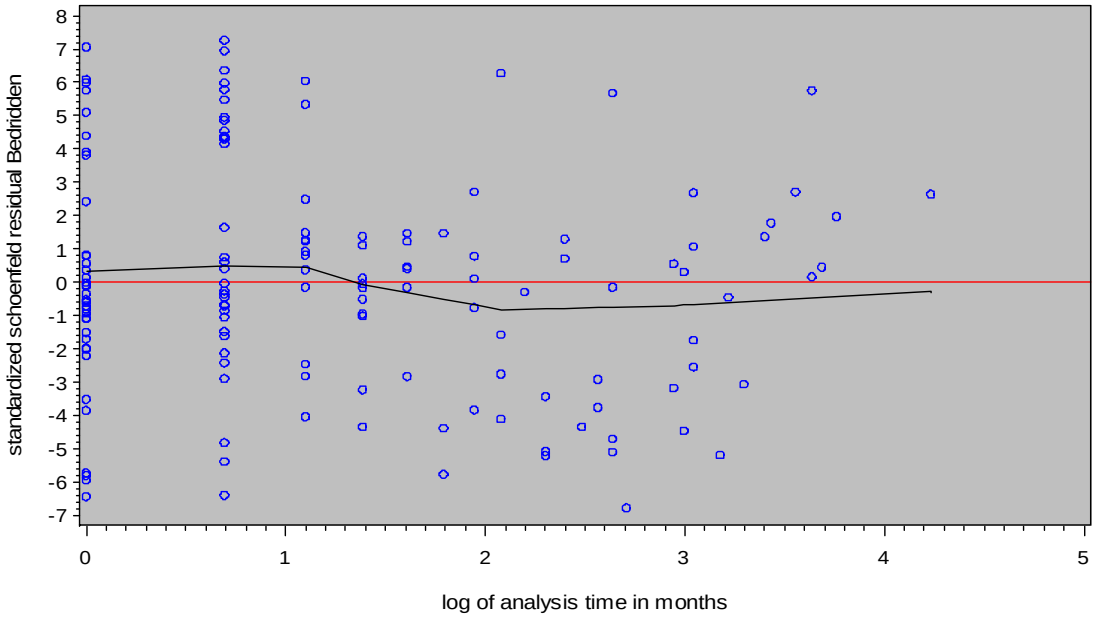


Figure 4.4.1.5: The plot of scaled Schoenfeld residuals for bedridden functional status in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

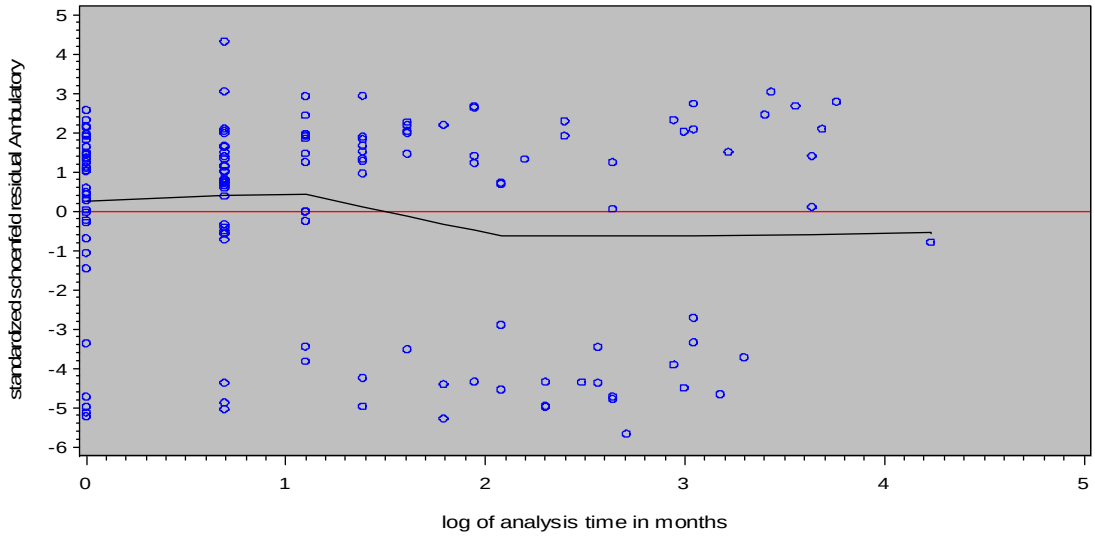


Figure 4.4.1.6: The plot of scaled Schoenfeld residuals for ambulatory functional status in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn

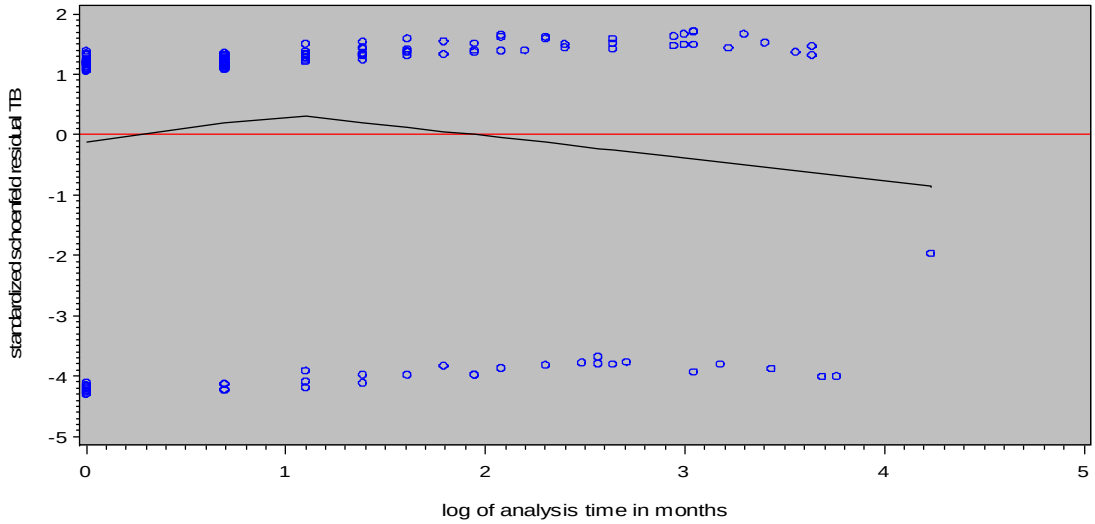


Figure 4.4.1.7: The plot of scaled Schoenfeld residuals for TB co-infection in the model in Table 4.3.8 against log of analysis time, with LOESS smoothed curve superimposed. The horizontal line through zero is drawn for reference.

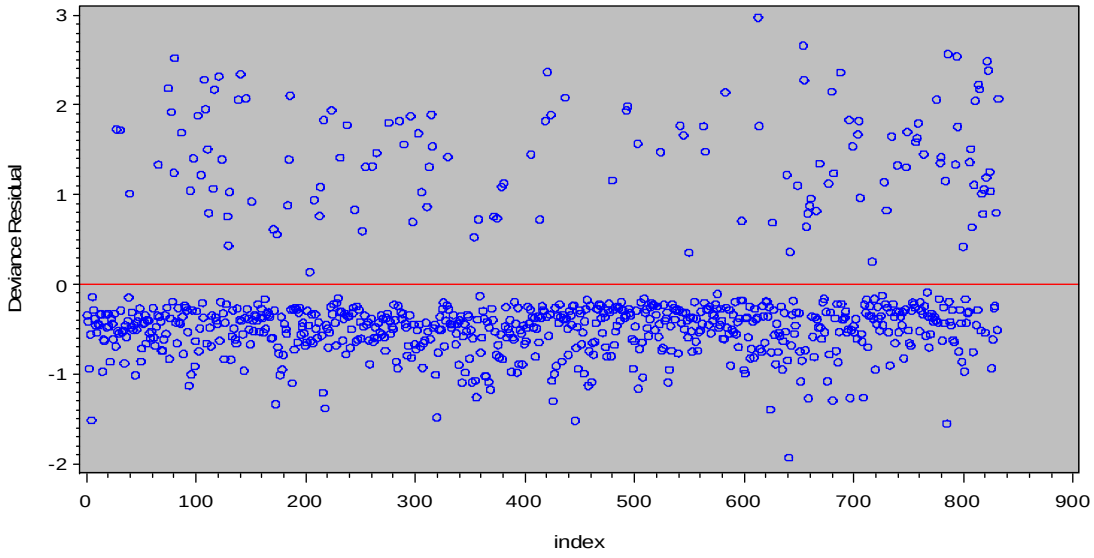


Figure 4.4.3.2: Index plot of the deviance residuals for the Cox proportional hazards model in Table 4.3.8. The horizontal line through zero is drawn for reference.

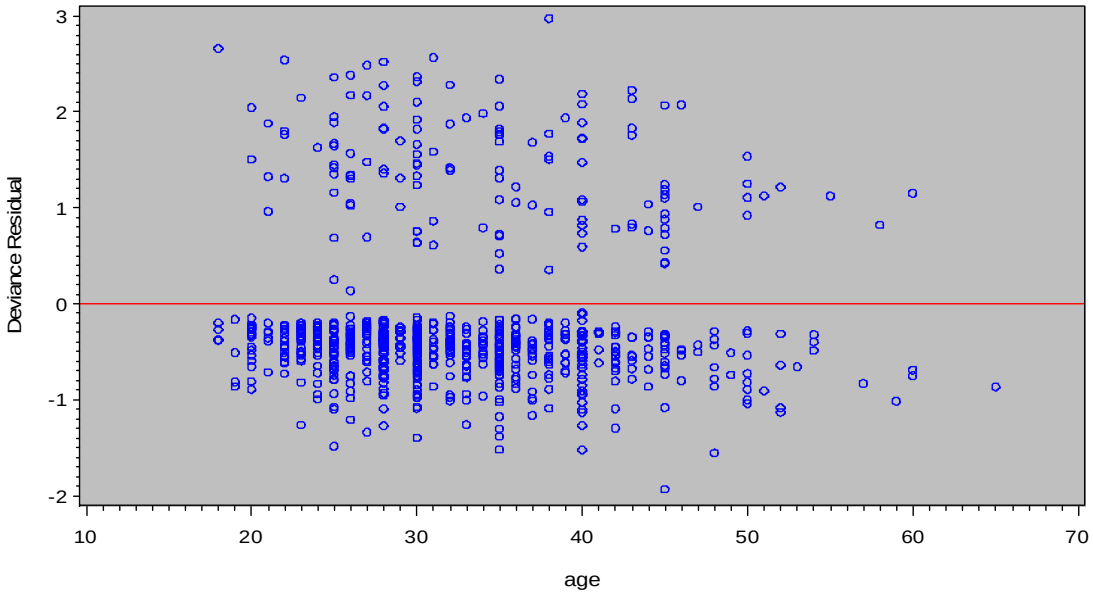


Figure 4.4.3.3 (a): Plot of the deviance residuals for the Cox proportional hazards model in Table 4.3.8 against age. The horizontal line through zero is drawn for reference.

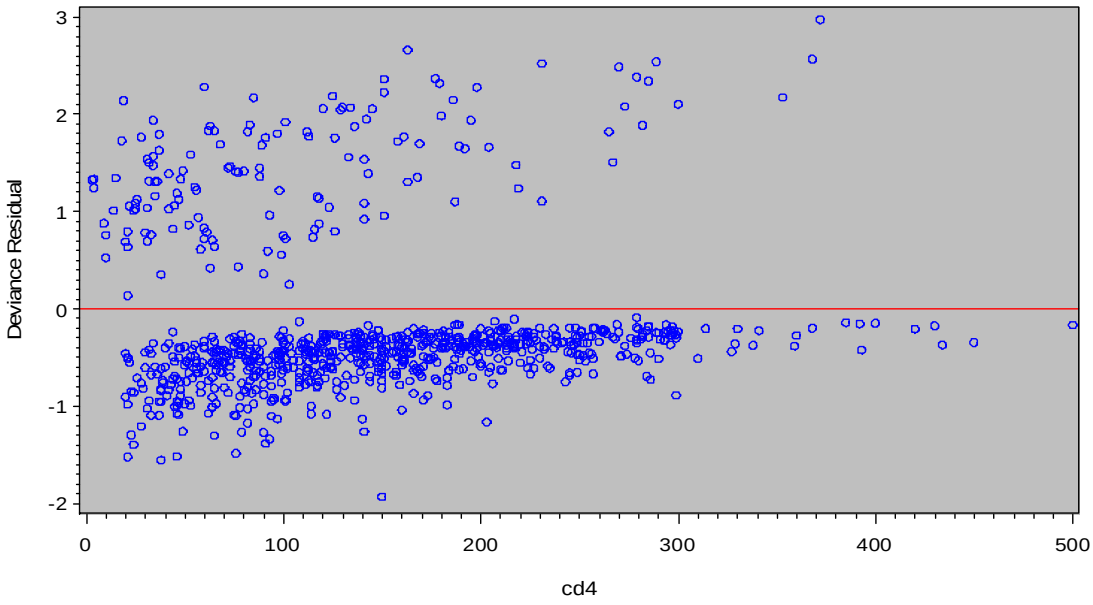


Figure 4.4.3.3 (b): Plot of the deviance residuals for the Cox proportional hazards model in Table 4.3.8 against baseline CD4 count. The horizontal line through zero is drawn for reference.

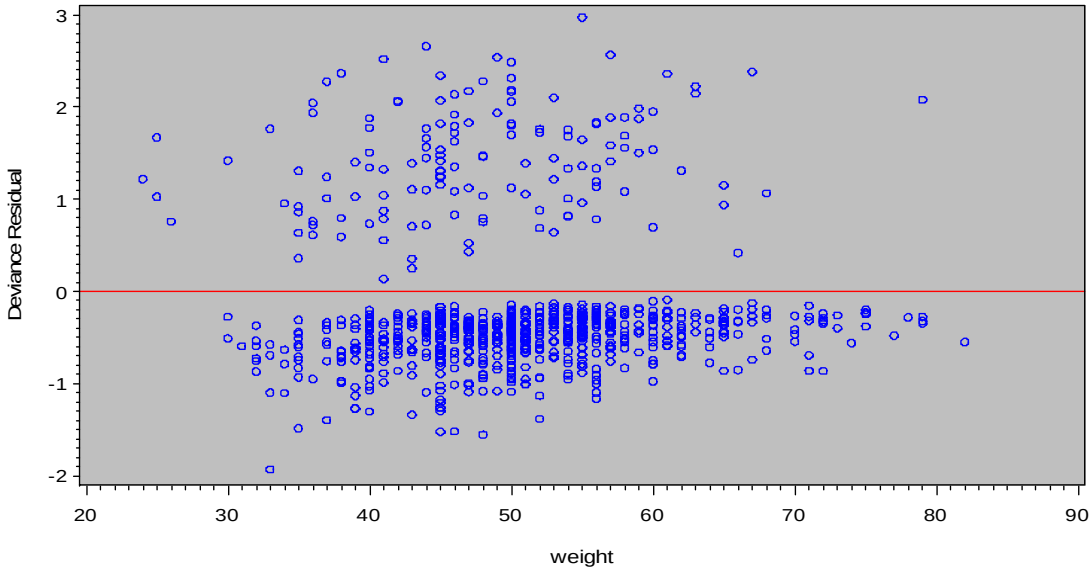


Figure 4.4.3.3 (c): Plot of the deviance residuals for the Cox proportional hazards model in Table 4.3.8 against baseline weight. The horizontal line through zero is drawn for reference.

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: Birtukan Tsehaineh Adamu

Signature: _____

Date: _____

Place: Faculty of 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: Faculty of Science, Addis Ababa University