UNIVERSITY FOR DEVELOPMENT STUDIES

TAMALE

JOINT LONGITUDINAL AND SURVIVAL MODELLING OF HIV IN THE UPPER WEST REGION

ABDUL RAHAMAN ABDUL BASIT

2015

UNIVERSITY FOR DEVELOPMENT STUDIES

JOINT LONGITUDINAL AND SURVIVAL MODELLING OF HIV IN THE UPPER WEST REGION

BY

ABDUL RAHAMAN ABDUL BASIT (BSc. Mathematical Science (statistics option))

(UDS/MAS/0027/13)

THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, FACULTY OF MATHEMATICAL SCIENCES, UNIVERSITY FOR DEVELOPMENT STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN APPLIED STATISTICS

JULY, 2015

DECLARATION

I hereby declare that this dissertation is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere. Related works by others which served as a source of knowledge has been duly referenced.

Abdul Rahaman Abdul Basit …… ……

(Candidate) Signature Date

We hereby declare that the preparation and presentation of the dissertation was supervised in accordance with the guidelines on supervision of dissertation laid down by the University for Development Studies.

Dr. Albert Luguterah …… ……

(Supervisor)

Mr. Suleman Nasiru …… ……

(Co-Supervisor) Signature Date

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my supervisor Bishop Albert Luguterah (Phd) and Mr. Suleman Nasiru who, through his co-supervision of this research work, open my mind into practical aspect of my study area and other area of studies.

I am very much thankful to the Upper West Regional Health Information Officer, Mr. Nani Tengey. Mr. Yakubu Tifere of Jirapa Hospital and Mr. Emmanuel Wondoh of Upper West Regional Hospital for their guidance and support during my data collection for this research work.

I wish to register my sincere gratitude to my parents Mr. Abdul Rahman Husein and Madam Aminat Froko for their prayers towards my success. I also recognized the efforts of my brother and sisters including my dear wife, Aminat kunatteh. The brotherly support offered me by the Internal Auditor of kasena Nankana Municipal Assembly, Mr. Abdul Rahman Nuhu is highly appreciated.

My friends around and far are not left out; I thank them so much for the physical and moral support they gave me during my study period.

DEDICATION

I dedicate this piece of work to my parents Mr. Abdul Rahman Husein and Madam Aminat Froko.

ABSTRACT

The main aim of this research was to develop a joint longitudinal and survival model for HIV patients in the Upper West Region. The research covers a period of 8years starting from January 2006 to December 2014 with a total population of 119 HIV patients who were on two different ART regimens. Ninety-one (76.5%) of the population were females and 28 (23.5%) were males, of which 29 (24.4%) were alive, 78 (65.5%) were lost to follow-up and 12 (10.1%) experienced the event (death). Multivariate test used to study the pattern of change of CD4 count showed that the pattern of change of mean CD4 count for alcohol drinking patient and non-drinking patients were insignificantly different as well as the case for cigarette smokers and non-smokers. Pattern of change of mean CD4 count for educated and non-educated patients were also not significant. However, the pattern of change of mean CD4 count for married and unmarried patients were significantly different as well as that of males and females. The study revealed that AZT/3TC/NVP contribute better to patients survival time and CD4 count growth using weibull model and linear mixed effects model respectively. The two models further unveiled that factors such as preCD4 count, gender, and duration of treatment (months) significantly determines HIV patient's CD4 count, whilst drug regimen, age and preCD4 determine the survival of patient. Results from the joint model indicated a significant association between the repeated CD4 count measurement and survival time of HIV patients.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ABC	Abstinence, be faithful, Condom use
AFT	Accelerated Failure Time
AIC	Akaike's information criterion
AIDS	Acquired Immune Deficiency Syndrome
ART	Anti-retroviral Therapy
ARV	Anti-retroviral
AZT	Zidovudine
BIC	Bayesian Information Criterion
CI	Confidence Interval
EFV	Efavirenze
FDA	Food and Drug board Association
GHO	Global Health Observatory
HAART	Highly Active Anti-retroviral Therapy
HIV	Human Immune deficiency virus
NVP	Nevirapine
PH	Proportional Hazard
REML	Restricted maximized log-likelihood

SIV	Simian Immune deficiency virus
STIs	Sexual Transmitted Infections
TB	Tuberclosis
WHO	World health Organisation
3TC	Lamivodine

CHAPTER ONE

INTRODUCTION

1.0 Background to the study

In medical studies, periodically measured disease markers are used to monitor the progression of the disease so as to enable health care givers to prescribe the appropriate treatment for a patient at a particular period. One of such area broadly researched into is the dynamics of CD4 count in Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome (HIV/AIDS) patient. The Human Immunodeficiency Virus (HIV) pandemic has caused a resurgence of TB, resulting in increased morbidity and mortality world-wide. This accentuate the progression of each other as a result of the synergistic complex relationship that exists between them, this further convolutes the challenge each of these singularly presents to the world and health professionals in their management (www.HIV/Tuberculosis co-infection, 2014).

Human Immunodeficiency Virus (HIV) is the leading cause of death worldwide and to that effect first December is declared as worlds AIDS day (www.HIV/Tuberculosis co-infection, 2014). HIV is found in the body fluids of an infected person (semen and vaginal fluids, blood and breast milk). The virus is passed from one person to another through blood-to-blood and sexual contact. In addition, infected pregnant women can pass HIV to their babies during pregnancy, delivering and through breast feeding. Acquired Immune Deficiency Syndrome (AIDS) is the advance stage of HIV. With AIDS the virus has advance,

causing significant loss of CD4 count (white blood cells). This is achieved when the HIV infect the vital cells in the human immune system such as the helper T cells called CD4+ T cells. CD4 cells or T-cells play a major role in protecting the body from infection. They send signals to activate your body's immune response when they detect "intruders," like viruses or bacteria (Louise, 2011).

Once a person is infected with HIV, the virus begins to attack and destroy the CD4 cells of the person's immune system. These T-cells are critical component of immune mechanism involved in protection against infection, including tuberculosis. The alteration of such immune system makes an infected person much more vulnerable to opportunistic infections and other diseases. This susceptibility increases as the disease progresses. Nelson (2007), said we cannot win the battle against AIDS if we do not also fight Tuberculosis, Tuberculosis is too often a death sentence for people with AIDS; TB hampers the effort of HIV treatment. Thus, quick declining of CD4 counts is the resultant effect of TB on HIV fast progression into AIDS (Rosas-Taraco *et al.*, 2006). HIV uses the machinery of the CD4 cells to multiply (make copies of itself) and spread throughout the body; this process is called the HIV life cycle (Louise, 2011).

Monitoring the course of HIV infected person essentially depends on the measurement of the CD4+ T lymphocytes. The absolute number of CD4 cells per micro-liter of blood represents the CD4 count in adults whilst that of children is expressed as the percentage of total lymphocyte or total T lymphocytes and this

has enormous prognostic and therapeutic implication for most HIV treatment decisions (Rodriguez *et al.,* 2005). It is an important indicator that tells how well your immune system is working. The CD4 count of a healthy adult/adolescent ranges from 500 cells/mm^3 to 1,200 cells/mm^3. A patient with CD4 counts less than 200 cells/mm^3 is one of the ways to determine whether a person living with HIV has progressed to a different stage of the infection.

The T cell count is used by HIV care provider to decide when to put a patient on antiretroviral therapy (ART). ART involves taking a combination of HIV medicines otherwise called an HIV regimen every day. This treatment regimen prevents HIV from multiplying and destroying your infection-fighting CD4 cells. The regimens however can't cure HIV, it aids patient live a longer, healthier life and reduce the risk of HIV transmission. When the amount of HIV in your blood is lowered by ART, it allows the CD4 cells to reproduce and increase in number. Higher CD4 count gives patient the ability to fight HIV and other infections.

Currently, HIV treatment guidelines recommend that care provider request a CD4 count test every 3 to 6 months after a patient is put on ART to see how well a patient is responding to treatment. Depending on the progress of patient health status, care provider may switch to every 6 to 12 months once treatment has increased your CD4 levels to higher levels with a suppressed viral load. When CD4 count reaches normal levels (500 cells/mm^3 to 1500 cells/mm^3) and viral load remains suppressed, care giver may not check CD4 count unless there is

deterioration in health or increase in viral load of patient. CD4 count can vary depending on the time of day your blood is drawn and on whether you have other infections or illnesses, like the flu or sexually transmitted infections (STIs). The trend of CD4 count whether it is rising or falling over time is what is really important not an individual test result. CD4 is an important biomarker repeatedly-measured on each patient and this observed biomarker series are frequently important health indicators that represent the progression of HIV. Therefore longitudinal and survival data are simultaneously collected to study the trajectory of patient health statue. These longitudinal and survival characteristics have usually been analyzed separately with well-established methods but with their correlation ignored. In order to take of care of the ignored latent relationship between such longitudinal and time to event data as a results of these separate analysis this piece seeks to jointly model the two instead.

1.1 Problem Statement

There are several circumstances in which both repeatedly measured biomarker and time to event data are collected simultaneously on individual subject. The progression of the disease is usually monitored on dynamics surrounding the essential biomarker that support the immune system against such a disease. In this case, CD4+ is the essential biomarker always considered in the treatment and management of HIV/AIDS patient, Thus CD4 count is regarded as the

fundamental indicator for prognostic information and a guide as to whether HIV infected individual is qualified for antiretroviral therapy (Phillips *et al.,* 2007). The efficacy of antiretroviral therapy reflects on the initial viral decay rate and the increment of CD4 count (Ding and Wu, 2001). A normal CD4 count is from five hundred to about thousand five hundred cell per cubic millimeter of blood and It is therefore important to pay attention to the pattern of results of this marker, as the decline of CD4 count is an indicator of a significant progression of the disease and the consequent effect will be poor survival time. These repeated measured biomarker and time to event data collected simultaneously have usually been analysed separately ignoring their relationship.

As results, there is the need to investigate how the rate of change of CD4 count affects survival time of patient. Hence joint modeling is most appropriate method to establish the relationship between repeated measurements of CD4 count and patients' survival time.

1.2 Research Questions

i. What is the pattern of change of CD4 counts of HIV/AIDS patients on treatment?
ii. What is the association of CD4 count measurement on time to event?
iii. What model will best fit the significant impact of the rate of change of CD4 count on the survival time of patient on treatment?

iv. Which prognostic factors are predictive of time to event occurrence?

1.3.0 Objectives of the study

1.3.1 Main objective

The main aim of this research is to develop a joint longitudinal and survival model for HIV patients in the Upper West Region.

1.3.2 Specific objectives

In line with the general objective, the following specific objectives were considered;

i. To investigate within subject patterns of change of CD4 count of HIV patients on treatment.
ii. To examine the effect of different anti-retroviral regimen on CD4 lymphocyte count.
iii. To determine the prognostic factors of HIV infection persons.
iv. To model the significant impact of rate of change of CD4 counts on the survival time of patients on treatment.

1.4 Significance of the study

The essential biomarker always considered in the treatment and management of HIV/AIDS patient is the CD4+. It is regarded as the fundamental indicator for prognostic information and a guide as to whether HIV infected individual is

qualified for antiretroviral therapy (Phillips *et al.*, 2007). Therefore, modeling the significant impact of the rate of change of CD4 count on the survival time of HIV/AIDS patient will be helpful to policy makers, non-governmental organisation and health institutions.

It will also guide patients and their guardians to put measures into place in order to boost the CD4 counts of patient. The study will further open many pathways for other researchers in various areas of study.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter reviewed related works on HIV, survival modeling, longitudinal modeling and joint modeling of longitudinal and survival data.

2.1 The Genesis of HIV

The most interesting lent virus in terms of research into the origin of HIV is the Simian Immune deficiency virus (SIV) that affects Monkeys, which is believed to have existed for about 32,000 years from today (www.avert.org/origin-aids-hiv.htm, 2014). The most commonly accepted theory is that of the 'Hunter'. It is believed that the SIV was transferred to human as a result of chimps being killed and eaten or their blood getting into contact with the cuts or wounds of the Hunter. This theory could be supported as there were several early strains of HIV, each with a quite different genetic make-up thus anytime it is transferred to man, it would have developed in a slightly different way (www.avert.org/origin-aids-hiv.htm, 2014).

In a survey of 1099 sample, it was revealed that ten people were infected with Simian Foamy virus (SIV), which was previously only associated with primates. As a result of such discoveries, there has been a call to place a band on the hunting of bush meat to prevent the transfer of Simian viruses

(www.avert.org/origin-aids-hiv.htm, 2014). Moreover, the Hunters' body would have fought off the SIV, but on few occasions it adapted itself within a new Human host and subsequently became HIV-1. Subsequently, two types of HIV have been characterized: HIV-type 1 and HIV-type 2. Each appears to package their RNA differently; HIV-type 1 will bind to any appropriate RNA, whereas HIV-2 will preferentially bind to the mRNA that is used to create the Gag protein itself. This may mean that HIV-type 1 is more able to mutate (HIV-type 1 infection progresses to AIDS faster than HIV-type 2 infection and is responsible for the majority of global infections). HIV-type 1 is the virus that was initially discovered and it is more virulent, more infective, and is the cause of the majority of HIV infections globally. The lower infectivity of HIV-type 2 compared to HIV-type 1 implies that fewer of those exposed to HIV-2 will be infected per exposure. Because of its relatively poor capacity for transmission, HIV-type 2 is largely confined to West Africa (www.avert.org/origin-aids-hiv.htm, 2014)

HIV infects vital cells in the human immune system such as helper T cells (specifically $CD4^+$ T cells), macrophages, and Dendritic cells. HIV infection leads to low levels of $CD4^+$ T cells through a number of mechanisms, including apoptosis of uninfected bystander cells, direct viral killing of infected cells, and killing of infected $CD4^+$ T cells by CD8 cytotoxic lymphocytes that recognize infected cells. When $CD4^+$ T cell numbers decline below a critical level, cell-mediated immunity is lost, and the body becomes progressively more susceptible to opportunistic infections (www.wikipaedia.htm, 2014).

2.2 The State of HIV/AIDS

Coupled with the uneven perpetual spread HIV/AIDS pandemic around the world and the recent furore in Ghana regarding HIV/AIDS ambassador personally confession of her HIV status as negative, is not only an indictment to the government of Ghana but an indication that leaders of the world still today remain in a limbo concerning the status of HIV epidemics in most other areas due to the poor and inadequate data on the prevalence of HIV (www.citifmonline.com, 2014). The outbreak of the HIV epidemic has since recorded nearly 78 million people infected with the HIV virus and claimed about 39 million people lives. At the end of 2013 about 35 million worldwide have been recorded living with the virus. Most importantly, about 0.8% of the working group globally is with HIV virus, although the constraint it imposes significantly varies between countries and regions (WHO/ GHO, 2015).

Centrally to Africa, although diverse efforts are being amassed to counter the HIV epidemic, the situation remains persistent and progressively complicated. For instance, in sub-Saharan Africa, six in every ten adult men, eight in ten adult women and nine of every ten children are living with HIV virus (UNAID, 2011). The hybridized progressive pattern of the epidemic with varied intensity and velocity can be recognized in antenatal clinics of most cities in Southern Africa, as about 45 percent of women tested during pregnancy carry the HIV virus, this rate is about ten or more times in urban antenatal clinics in several countries in Central or West Africa. Research work conducted in this regard suggest migratory

and behavioral pattern may result these differences, however the pattern of interconnection through a convoluted social and sexual network must be closely researched into so as to unravel the variations thereof (UNAID, 2011). In spite of the positive impact of the interventions had on sub-Sahara Africa and the Caribean, these regions still contribute about 71 percent of newly infected persons in 2001 (UNAIDS report, 2012).

The situation in Ghana is dynamic as the early days of the disease present dire consequences ranging from economic to social development. In 1999, two hundred and ten people were estimated to be infected with the virus as 380,000 were officially reported to have been infected with the virus in 1998. Cumulatively, since the beginning of the epidemic in Ghana the estimated number of deaths is about 114,000 by 1999 and one million additional deaths were estimated over the ensuing period (1999-2014) (www.wikipaedia.org/HIV/AIDS in Ghana, 2012). However, the trend of event changed regarding the spread of the virus infection and death of infected patients as government and non-governmental organizations put efforts together to arrest the situation.

Though globally, HIV prevalence is expected to increase within the next 10 years (Yehia and Frank, 2011), Fred Nana Poku, 2014 unwrapped that the national prevalence rate in Ghana had stabilize at 1.5 percent as at 2010 and further declined in 2011 and 2012 to 1.3 percent. Whilst suggesting behavioral change he

added that the prevalence of HIV among sex workers had reduced from 35 percent in 2006 to 25 percent in 2009 and further to 11 percent in 2012.

2.3 Interventions to control the epidemic

The US Food and Drug Association (FDA) approval of the Nucleoside Reverse Transcriptase Inhibitor (NRTI) zidovudine (AZT) in 1987 against HIV and the subsequent development of more combination therapies brought a sigh of relieve to the world. This drug does not only delay the progression to AIDS but made the infected persons quite productive and thereby enhanced their quality of life (Broder, 2010). Arguably, patients still die out of AIDS as the drug is unable to subdue the virus for long period of time (Management of HIV/AIDS, 2015). To ensure strong effective control of the AIDS pandemic and possible eradication of the disease, a coordinated multifaceted approach was developed by introducing highly active anti-retroviral therapy (HAART). Whilst Hammer *et al.,* (1997) underscores the significant benefit of combining two NRTIs with a new class of anti-retroviral, quantitatively, it was revealed that there has been a telling benefit of about 60%-80% decline in the progression of infected person to AIDS, death and hospitalization (Management of HIV/AIDS, 2015). The early initiation of treatment regimen can successfully reduce HIV transmission by 96% (Cohen *et al.,* 2011). Also mother-to-child transmission in low and middle income countries can be prevented using anti-HIV drugs (Siegfried *et al.,* 2011; Tudor Car *et al.,* 2011; Santos *et al.,* 2012). The cost of the drug presented a great challenge in the

early days of discovery to the world especially developing countries. However, world health organization responded to this challenge stating that the ARV regimens are now available, very affordable, more efficient, safer and simpler than before to all countries even the most poorest (WHO, 2013).

Additionally, behavioral change has been recognized as an effective step to fighting the spreading of HIV. The ABC (Abstinence, Be faithful, and Condom use) method emphasize on fidelity, fewer sexual partners and a later age of sexual debut. The promotion of the ABC method gave a ten percent drop of HIV cases in Uganda in the period of 1990-2001 (www.medwiser.org/hiv/aids, 2015). Normally, developing countries record large number of late HIV diagnosis which affects treatment as the virus might develop into AIDS. The associated opportunistic infection and dementia seldom not require additional treatment, these treatment are often not available. Moreover, personal and family shame, blame and stigma from society coupled with inadequate education on HIV/AIDS shield individuals going for treatment (Steward *et al.*, 2012).

2.4 Relevance of CD4 Cells in HIV

CD4+ is an essential biomarker always considered in the treatment and management of HIV/AIDS patient. The dynamics therein calls for various research works on the CD4+ count to enhance the quality of lives of patients. CD4 cells, sometimes called T-cells are type of lymphocytes (white blood cells).

These cells are of two types; T-4 cells otherwise called CD4+ ('helper' cells) and T-8 cells normally referred to as CD8+ cell ('suppressor' cells). The CD4+ leads the attack against infections whilst the CD8+ ends immune response and kills cancer cells and cells infected with virus. The T-4 cells are normally referred to as CD4 because they have CD4 molecules on their surface and it is an important biomarker of HIV disease progression. The CD4 cells are essentially important component of the human immune system which begins to deplete as the viral load progress high. Thus CD4+ count is regarded as the fundamental indicator for prognostic information and a guide as to whether HIV infected individual is qualified for antiretroviral therapy. Moreover, the efficacy of antiretroviral therapy reflects on the initial viral decay rate and the increment of CD4 count (Ding and Wu, 2001).

HIV usually, infects the CD4 cells and become part of the cells, when these cells tend to fight an infection, they multiply and make copies of HIV. As results an infected person with no treatment begins to lower the CD4 count thus immune system becomes weak for opportunistic infections such as TB to set in. The restoration of the immune system is largely dependent on how well the CD4 cells are counted, as study has revealed that the disease progression is delayed significantly when the CD4 increases in response to treatment (Nelson *et al.*, 2007).

WHO, (2010) recommended patient with CD4 count below 350 $cellmm^3$ be initiated into ART. Several factors among fatigue, time of day, and stress as well as vaccination can cause lots of changes in the number of CD4 cells. For instance, when the body fights an infection the number of white blood cells increases implying CD4 count goes up. This particular behavior of CD4 cell affects treatment of HIV especially when it is infected with TB. To improve the survival and quality of life effectively, it is crucial to identify the factors influencing the disease progression for early intervention and to encourage more follow up for treatment (Nelson *et al.,* 2007). Maintaining higher CD4 count and complete cessation of smoking may reduce the risk of non-AIDS-defining cancer among patients (Krishnan *et al.,* 2011). A multivariate regression conducted to evaluate predictors of CD4+ count and HIV-1 RNA levels in a multicenter AIDS cohort study revealed that old men of 50yeasr and above on Highly Anti Active Retroviral Therapy (HAART) had their mean CD4+ adjusted upwards per microliter and similar pattern but quite higher adjustment of man CD4+ count was found in younger men (Xiuhong *et al.,* 2011).

2.5 Empirical Works on Longitudinal Analysis

Longitudinal analysis has been applied in several spheres of life ranging from medical, clinical and social investigations. Longitudinal analysis conducted on 51 HIV-negative pregnant individuals and 25 non pregnant individuals indicated that the average white blood cells, CD4 and CD8 of pregnant individuals progress

upwards than non-pregnant individuals and these increments included the parturition period through the pregnancy. However, significantly lowered absolute lymphocyte count and, percentage and absolute CD4 count is experienced by pregnant women though the absolute mean CD8 count appear not affected (Towers *et al.,* 2010).

In addition a research conducted using longitudinal analysis on different category of groups on HAART shows that amongst patients who started and remain on HAART, late presentation is related with lower rate of suppression, with increased blunted CD4 count compared with late starters. Differences existing between these two categories suggest that other factors may be contributing to the adverse treatment outcomes in late presenters (Waters *et al.*, 2011).

A multivariate analysis conducted on 110 men and 42 women with a median interval of 12.9 months visit using longitudinal model, evaluated the relationship between CD4 count, viral load and HAART with changes in trunk and extremity composition, revealed that alterations in trunk fat, extremity fat and lean mass are predicted by baseline viral load and change in CD4l count (McDermott *et al.*, 2005)

To determine the related impacts of cannabis on the risk behaviors among HIV-infected person, a longitudinal regression analysis was conducted on Russian HIV infected risky drinkers. Results obtained were needle sharing and several sexual episodes. Besides, drug injection and others were realized and finally, cannabis usage was significantly associated with needle sharing. as a results, it leads to

high rate of HIV transmission as it may increase risky sex behavior and other drug usage (Tyurina *et al.,* 2013).

2.6 Empirical Works on Survival Analysis

Having originated centuries ago from mortality tables, difficulties presented to military personnel during the Second World War as results of military equipment reliability brings new dimensions of survival analysis. Successes obtained as a results, promoted it in usage in the industries as the demand for safe and reliable product increased including medical research and clinical trials. As the interest in survival analyses increased, further research was conducted and that gave birth to non-parametric and semi-parametric approach from parametric approach. Consequently, some challenges were encountered as reported by Peto *et al.,* (1977) that using sloping lines to join Kaplan Meier survival curves present incorrect estimates. The development and improvement of statistical methods and procedure including survival models marginally improved between 1970 and 1990. It is within this period that Fleming *et al.,* (1991) improved mathematical theory of survival analysis based on martingale theory and counting stochastic process. The dominance won by survival analysis in medical research and clinical trials makes it universally accepted for data analysis in several medical fields. Following that, Cox (1972) develops the Cox model for analysis of survival to help identify differences in survival due to treatment and prognostic factors.

As the development progresses with challenges, Singh and Mukhopadhyay (2011) revealed that the logistic model ignores the survival time and censoring information and hence the Cox model is preferred as it gives way for the baseline hazard function and the survival curves can be estimated. Meanwhile, the proportional hazard (PH) models and its extension are largely to ascertain the effect of an intervention in the presence of covariates. The PH may not hold where the covariate are to accelerate the event occurrence, as a results, the accelerated failure time (AFT) model is used on 1236 tuberculosis patients admitted in a randomized control clinical trial. The AFT model present better prediction than Cox model (Ponnuraja and Venkatesan, 2010). Although there are still challenges in survival data analysis and its presentation, the progressive development and improvement of survival analysis is as a results of the fact that it's' application in research fields has increased over the recent years (Altman, 1991; Anderson, 1991). Jakperic and kpakpo (2013) studied the Effects of Prognostic Factors in Recovery of TB Patients in the Upper West Region using Kaplan-Meier estimator and the Logistic Regression model. A total of 400 patients were studied over period, 256 were males whiles 144 were females. Their result revealed that male patients were more than female patients and that goes to support WHO (2002) report, which said in most settings, tuberculosis incidence rates are higher in males at all ages than in females.

Using Cox regression and Product-Limit estimator, it was observed that the survival rates for males and females tuberculosis patients were 85.71% and

88.46% respectively. Whilst the age of patient at diagnosis, category, and type of patient were crucial determinants of treatment outcome, the study reported a median recovery time of 22 weeks in the Region. The risk of relapse and death were found to be related to age (Dioggban and Michael, 2012).

2.7 Empirical Works on Joint Modeling

Hitherto, repeated measured and time to event data have always been analyzed separately, and these separate models are overly simplified as the association between component is not always accounted for and also, failing to rope in all available information in an integrated manner, may provide invalid inference.

To address the problems of separate modeling, a two stage modeling is proposed in which the repeated measurements are modeled using a random effect in the first stage to estimate the subject-specific value of the covariate. In the second stage, the random effect modeled value is substituted into the survival model as time varying covariate. Morrell *et al.*, (2000) in their study revealed that aside, the two stage method has the ability to arrest measurement error, it also allows the covariate for each patient to be estimated but leaves parameter estimates still bias and inefficient. In addition, the two-stage has failed to address the association that exist between longitudinal and survival process. As a results Wulfsohn *et al.*,(1997) propose joint modeling and its extensions.

Moreover, the joint modeling approach has ability to ensure the effect of a covariate on the longitudinal process is separated from its effect on survival. And also, covariate effect can be described and visualized (Hyun *et al.,* 2013). In furtherance to the development of this approach, a two-step procedure is considered by using a modified EM algorithm for estimation, a growth curve random component model is assumed for true CD4 count with normal errors. Estimates of this model is then substituted into the Cox proportional hazards model to obtain estimates of the survival parameters (Wulfsohn *et al*., 1997). In trying to further develop the two-step procedure, Henderson *et al*., (2002) in their study assumed that the longitudinal and survival are conditionally independent given the one process covariates, linked the two process by considering the use of an unobserved or latent bivariate Gaussian process.

Combining a Linear Gaussian random effect sub model for the repeated CD4+ count measurement and Cox or Weibull survival sub model, linked through their shared dependence on the latent variable. Its revealed that that the hazard rate of death depended on the longitudinal progression of CD4+ counts, that is a patient's baseline CD4+ count and the rate of change in CD4+ counts significantly impact on his or her survival time (Hyun *et al.*, 2013).

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter looks at the source of data and the statistical methodologies that were used in the entire study to achieve the desired objectives.

3.1 Study area

The Upper West Region is one of the ten regions of Ghana. It is located at the North Western corner of Ghana with latitude 9.8°- 11.0° North and longitude 1.6°- 3.0 west, bordered by Upper East region to the east, Northern region to the south, and Burkina Faso to the west and north.

It covers a geographical area of 18,476 square kilometers which represents 12.7% of the total land area of Ghana. It is the seventh largest region in Ghana in total area, and it is made up of eleven (11) districts. The region has a total population (2010 census) of 576,583, of which 276,445 (47.9%) were males and 300,138 (52.1%) females. The major economic activity in the region is agriculture and this region host the famous Hippopotamus Sanctuary located South West of Wa, along the Black Volta River in the Wa West District (Wechau). The Gwollu Wall in the Sissala District also serves as the hometown of one of Ghana's past presidents - Dr. Hilla Limann (www. bits.wikimedia.org, 2014).

3.2 Study and Target population

The study population considered was HIV patients on ART whose CD4+ counts are monitored from January 2006 to December 2014. The target population consists of HIV patients on ART who were ten (10) years and above within the study period.

3.3 Source of data and data collection

Secondary data were obtained at the public health department specifically the HIV/AIDS unit at the various District Hospitals including the Regional Hospital, where enrolled client's CD4 counts were monitored regularly. Data collection was restricted to individuals infected with HIV/AIDS on treatment who had their therapy during January 2006 to December 2014 and were ten (10) years and above. Members under the study were censored if they dropped along the study period or failed to experience the event.

Both longitudinal and survival data were obtained from individual patient's folders. The two outcome variables considered in this study include CD4 cell count per mm^3 of blood measured repeatedly for approximately every six months interval visits and survival endpoint or time to event is death. A common measuring (observation) time limit is used for all patients where the maximum observation time considered to be on the 78^{th} month (such that $n_i \leq 12$). The survival response in months was obtained by subtracting the date of entry for

treatment from the date of the last visit. Nine (9) potential explanatory variables were considered in this study viz; Drugs, PreCD4, Alcohol, Smoking, Time, Education, Marital status, Gender, and Age.

3.4 Profile Analysis

Profile can be described as a broken line that graphically joins interdependent observations which are measured repeatedly on the same subject. These observations may be interdependent because they are taken over distinct time periods. Profile analysis is an extension of one-way MANOVA involving P response variable administered to individuals, an average profile consists of the average of the observations within each treatment level, and the Profile analysis therefore compares these average profiles from two or more treatment groups.

It is most commonly used in two scenarios thus comparing the same dependent variables between groups over several time-points and when there are several measures of the same dependent variable. The usual hypothesis is $H_0 : \mu_1 = \mu_2$ thus when differences between groups are constant across variables, then profiles are said to be parallel.

Consider the mean CD4 count at various time points of one level of a group of a sample of size $x_{11,}x_{12,}x_{13,}x_{14,} \dots x_{n_1}$ from the P-variate normal distribution with mean vector μ_1 and a covariance matrix and the other level of the group sample of

size $x_{21,}x_{22,}x_{23,}x_{24,}\ldots x_{n_2}$ from the P-variate normal distribution with mean vector μ_2 and a covariance matrix.

assuming

$$X_{1j}\sim \mathrm{N}(\mu_1,\Sigma) \quad \text{and} \quad X_{2j}\sim \mathrm{N}(\mu_2,\Sigma)$$

for j= $1 \ldots n_l$ and independent

we write our hypothesis as

$$H_0 : C\mu_1 = C\mu_2 \tag{3.1}$$

a contrast matrix $C_{(p-1)\times p}$ is a linear transformation on p variables to obtain a $(p-1)$ new variables as below `

$$C_{(p-1)\times p} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} \tag{3.2}$$

taking

$$CX = C\begin{bmatrix} X_1 \\ \vdots \\ .X_p \end{bmatrix} = \begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \\ \vdots \quad \vdots \\ X_{p-1} - X_p \end{bmatrix} \tag{3.3}$$

which are measured on the groups, this leads us to write the following random variables;

$$CX_{1j}\sim N_{p-1}(C\mu_1, C\Sigma C'),$$

$$CX_{2j} \sim N_{p-1}(C\mu_2, C\Sigma C').$$

now to obtain the covariance matrix Σ, since the estimates of $C\Sigma C' = CS_{pool}C'$, we calculate the pool variance,

$$S_{pool} = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2} \tag{3.4}$$

to test our hypothesis H_0, we make use of the Hoteling's T^2 test for two independence

samples by substituting the above into the Hoteling's T^2 equation.

Hoteling's T^2 test

In a case in which *p* variables are measured on each sampling unit we assume that the random sample is available from $N_p(\mu, \Sigma)$, but μ and Σ are unknown. Replace μ and Σ with $\bar{X}$ *and* S respectively.

Reject H_{01} if

$$T^2 = (\bar{X}_1 - \bar{X}_2)'C' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) CS_{pool}C'\right]^{-1} C(\bar{X}_1 - \bar{X}_2) > C^2 \tag{3.5}$$

where

$$C^2 = \frac{(n_1+n_2-2)(p-1)}{n_1+n_2-p} \mathcal{F}_{(p-1)(n_1+n_2-2)}(\alpha) \tag{3.6}$$

3.4.1 MANOVA

Multivariate Analysis of Variance (MANOVA) provides test to compare the mean vectors of K random samples for significant difference when the levels (g) of the grouping variable are two or more ($g \geq 2$). But for the case of two levels of the independent term (group) in a model, the Hotelling's T^2 described in equation (3.5) would be used to test the equality of the mean vectors between the two levels.

Consider g population mean vectors of model,

$$y_{ij} = \mu + \tau_i + +\varepsilon_{ij} \tag{3.7}$$

where;

for all $j = 1, \ldots, n_i$ cases per group, and $i = 1, \ldots, g$ groups.

y_{ij} = observation vector $p \times 1$

μ =overall mean vector $p \times 1$

$\tau_i = ith$ treatment effect vector $p \times 1$

ε_{ij} = residual for the ith group on the jth case $p \times 1$, $\varepsilon_{ij} \sim N_p(0, \Sigma)$

Each component of y_{ij} satisfies the one way ANOVA model, but now the model includes covariances among the components. These covariance are assumed to be

equal across the populations, hence the sum-of-squares and cross products (SSCP), total corrected squares and cross-product yields,

$$(x_{ij} - \bar{x})(x_{ij} - \bar{x})' = [(x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x})'][(x_{ij} - \bar{x}_i) + (x_i - \bar{x})']' \quad (3.8)$$

for squares and cross-product we have,

$$(x_{ij} - \bar{x})(x_{ij} - \bar{x})' + (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

cross-products subsequently gives us,

$$(x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})' + (\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i)'$$

for sum of squares,

$$\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x})' = \left(\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)\right)(\bar{x}_i - \bar{x})' \quad (3.9)$$

$$= \left(\left(\sum_{j=1}^{n_i} x_{ij}\right) - n_i\bar{x}_i\right)(\bar{x}_i - \bar{x})'$$

$$= n_i(\bar{x}_i - \bar{x}_i)(\bar{x}_i - \bar{x})' = 0 \quad (3.10)$$

Now summing the rest over j and i will give within groups SSCP

$$W = E = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \quad (3.11)$$

This will subsequently result to,

$$= (n_1 - 1)S_1 + (n_2 - 2)S_2 + \cdots + (n_g)S_g \quad (3.12)$$

where S_i is the sample covariance matrix for ith group.

Regarding between groups SSCP we have,

$$B = H = \sum_{i=1}^{g} n_i(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' = \sum_{i=1}^{g} n_i \hat{\tau}_i \hat{\tau}_i' \quad (3.13)$$

To test the hypothesis $H_o: \tau_1 + \tau_2 = \cdots = \tau_g = 0$ is true, then B(H) should be close to zero.

For $H_{0,}$ testing, we use the wilk's Lambda (likelihood ratio statistic),

$$\Lambda^* = \frac{|W|}{|W+B|} = \frac{|W|}{|T|} = \frac{\prod_{l=1}^{p} \lambda_l}{\prod_{l=1}^{p} \lambda_l^*} \quad (3.14)$$

where λ_l is the eigenvalues of W and λ_l^* eigenvalues for T if H_o is true then B is close to 0, impliying $T \approx W$ and $\lambda_l \approx \lambda_l^*$

However, if H_o is false then B is not close to 0, implying values on diagonals of T, which will be positive and hence large, implying $\lambda_l < \lambda_l^*$, Λ^* is small.

There are more than one way to combine the information in B and W such as;

Hotelling-Lawely Trace

Hotelling-Lawely statistic often called the generalized T^2- statistic,

$$trace(E^{-1}H) = tr(HE^{-1}) = \sum_{l=1}^{g} \bar{\lambda}_l \quad (3.15)$$

where $\bar{\lambda}_l$ is eigenvalue of HE^{-1}

Roy's largest root

Roy's largest root is also known as Hotelling's generalized T^2 statistic and it's taken as λ_l and somtimes

$$\theta = \frac{\bar{\lambda}_1}{1+\bar{\lambda}_1} \qquad (3.16)$$

bounded between zero and one

however, Wilk's Lambda is desirable because it can be converted exactly to an F statistic.

Pilai's Trace

Pillai's is an extension of Roy's statistic, equation (3.24). if the mean vectors do not lie in one dimension, the information in the additional terms $\frac{\lambda_l}{(1+\lambda_l)}$, $i = 2,3, \dots, s$ may be helpful in rejecting H_0

$$trace(B(B+W)^{-1}) = trace(H(H+E)^{-1}) = \sum_{i=1}^{p} \frac{\bar{\lambda}_l}{1+\bar{\lambda}_l} \qquad (3.17)$$

where $\bar{\lambda}_l$ is the eigenvalue of HE^{-1}.

3.5 Modeling Approach

3.5.1 Longitudinal modeling

In a longitudinal or follow up studies observations on individuals are repeatedly measured over time. As a result, modeling with simple linear regression will not be appropriate because of the assumption of independent observations. In that regards, linear mixed effect are designed in which the repeated measurements using linear regression model are fitted where parameters vary over subjects. Linear mixed effect model takes into accounts within and between sources of variation, they are flexible enough to account for the natural heterogeneity

population and they can handle any degree of missing drop out data in the longitudinal data. In addition, each subject has a subject-specific mean response over time with each i^{th} subject measured at times $S_{i1} \ldots S_{in_i}$ model as;

$$y_i = \mu_i(s) + W_{1i}(s) \tag{3.18}$$

$$= X_{1i}^T(s)\beta_1 + Z_{1i}^T(s)b_i + \varepsilon_{i,} \tag{3.19}$$

where

$$\mu_i(s) = X_{1i}^T(s)\beta_1 \tag{3.20}$$

$$W_{1i}(s) = Z_{1i}^T(s)b_i \tag{3.21}$$

$$\mathrm{b_i} \sim \mathrm{N_q}(0, \psi) \qquad \varepsilon_\mathrm{i} \sim \mathrm{N}_{n_i}(0, \sigma^2 I)$$

y is an n_i dimensional vector of observed responses.

β_1 is a p dimensional vector of fixed effects.

b_i is a q dimensional vector of random effects.

$X_{1i}^T(s)$ a matrix of size $(n \times p)$ fixed effects possibly time-varying covariates is the mean response.

$Z_{1i}^T(s)$ is a matrix of size $(n \times q)$ random effects covariates that incorporates random effects in the model.

3.5.2 Covariance Structure for the Mixed effect model

Longitudinal data analysis will be incomplete without the appropriate underlying covariance structure of the data. Most of these structures, have their covariance between two observations on the same subject depends only on the length of the time interval between measurements called the lag with constant variance over time. The covariances can be characterized in terms of the variances and the correlations expressed as a function of the lag. The heterogeneous versions of these covariance structures are a simple an extension; that is the variances, along the diagonal of the matrix, do not have to be the same. Note that this adds more parameters to be estimated, one for every measurement. Some of these structures are;

3.5.2.1 Variance Component

The variance covariance structure is known as the standard variance components, it specifies that there is no correlation between any pair of observations. The covariance structure is given as

$$corr\ (Y_i) = \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix} \quad (3.22$$

3.5.2.2 Compound Symmetric

The compound symmetric specifies that observations on the same patient have homogeneous covariance and homogeneous variance. The structure is given as

$$corr\ (Y_i) = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \ldots & 1 \end{bmatrix} \tag{3.23}$$

3.5.2.3 Autoregressive (1)

This structure indicates homogenous variance, it also specifies that covariance between observations on the same patient are not equal, but decrease toward zero with increasing distance. Two measurements that are right next to each other in time are more correlated than measurement far apart taking cognisance the value of ρ also characterizes this structure. The correlation due to the measurements distance is given by the exponential function. The structure is given below;

$$corr\ (Y_i) = \begin{bmatrix} 1 & \rho & \rho^2\rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ & \rho^3\rho^2 & \rho & 1 \end{bmatrix} \tag{3.24}$$

3.5.2.4 Unstructured

The unstructured structure specifies no patterns in the covariance matrix, and is completely general, however the generality brings the disadvantage of having a very large number of parameters. The structure is given below;

$$corr\ (Y_i) = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix} \tag{3.25}$$

3.6 Survival data analysis

Survival analysis is a powerful tool for studies aimed at analysing event times. In particular, clinical follow-up studies may be interested in analysing the time until an event occurs, normally understood as death contracting a disease or recurrence of a disease after treatment. These procedures allow the covariates effects on survival or hazards to be investigated. In this way, the dependent variable is the time until that event. The presence of censoring in survival data is what makes the difference, and consequently requiring specific methodologies, such as survival analysis.

Let T be the survival time and C the censoring time. Define the follow-up time $Y = min(T, C)$

and let $\delta = 1(T \leq C)$ denote the censoring indicator. Considering the probability density function of T, $f(t) = P(T = t)$ which represents the unconditional probabilities of death, the survival function is defined as the complement of the cumulative distribution function.

3.6.1 Log rank test

The log rank test was used to compare the difference in survival among groups such as Drugs regiment, Alcohol, Smoking, Education and Gender. This is a well-known and widely used test statistic. For k factor of groups, the log rank test the hypothesis that;

H_0: All survival curves are the same,

H_1: Not all survival curves are the same.

Log rank test approximates a chi-square test which compares the observed number of failures to the expected number of failure under the hypothesis. The test statistics is given by;

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3.26}$$

where $O_i$and E_i are the observed and expected number of death respectively. K-1 is the degree of freedom with K being the number of groups. A large chi-squared value will lead to the rejection of the null hypothesis in favour of the alternative.

3.6.2 Survival models

Both parametric and semi parametric models are available to model survival data and the commonly used models includes weibull, exponential, lognormal and log logistic including cox proportional hazards models. For parametric survival models, time is assumed to follow some distribution whose probability density

function $f(t)$ can be expressed in terms of unknown parameters. Once a probability density function is specified for survival time, the corresponding survival and hazard functions can be determined.

Survival function; $(t) = P(T \geq t)$ is thus obtained from

$$S(t) = \int_t^{\infty} f(u)du \tag{3.27}$$

and Hazard function $h(t)$ otherwise known as instantaneous rate also referred to as force of mortality or the age-specific failure rate is also ascertained from

$$h(t) = \frac{-d[S(t)]/dt}{S(t)} \tag{3.28}$$

this eventually would give Cumulative hazard function $\mathrm{H}(t)$;

$$\mathrm{H}(t) = \int_0^t h(u)du \tag{3.29}$$

equation (3.36) can further give the survival function in terms of hazards function by exponentiating the negative of the cumulative hazard function. The cumulative hazard function is the integral of the hazard function between integration limits of 0 and t;

$$S(t) = exp\left(-\int_0^t h(u)du\right) \tag{3.30}$$

finally the probability function can be expressed as the product of hazard function and survival function,

$$f(t) = h(t)S(t) \tag{3.31}$$

3.6.2.1 Exponential Distribution

This distribution is described as one parameter model because the hazard is constant over time. The risk of an event happening is flat over time.

Consider the density function, $f(t) = \lambda e^{-\lambda t} \; for \; t \geq 0$ (3.32)

but

$$S(t) = e^{-\lambda t}$$

from equation (3.39) $h(t) = \frac{f(t)}{S(t)}$

$h(t) = \lambda$ hence constant hazard. (3.33)

also from equation (3.36) the cumulative hazard function $\mathrm{H}(t)$ is determined,

$$H(t) = \lambda t \quad (3.34)$$

3.6.2.2 Weibull Distribution

This model is flexible as compared to the exponential model because its hazard rates are not constant. It is a two-parameter model i.e. λ and p where, λ is the scale parameter, p is the shape parameter. Thus, it informs whether the hazard is increasing, decreasing, or constant over time.

consider the survivor function

$$\mathrm{S}(t) = e^{-\lambda t^p} \quad (3.35)$$

$$f(t) = -\frac{d}{dt}S(t) = p\lambda t^{p-1}e^{-\lambda t^p} \quad (3.36)$$

where $h(t) = p\lambda t^{p-1}$

using equation (3.36) $H(t) = \lambda t^p$ (3.37)

The shape parameter can be interpreted as:

If $p < 1$, then the hazard is monotonically decreasing with time.

If $p > 1$, then the hazard is monotonically increasing with time.

If $p = 1$, then the hazard is flat and we have the exponential model.

3.6.2.3 Log-Logistic Model

The hazard function for Log-Logistic is defined as;

$$h(t, X) = \frac{\frac{\lambda t}{\gamma}\left[\frac{1}{\gamma} - 1\right]}{\gamma\left(1 + (\lambda t)^{1/\gamma}\right)} \quad (3.38)$$

where

$$\lambda_i = e^{-(X_i\beta)} \quad (3.39)$$

The log-logistic model have two parameters as the Weibull model, λ being the location parameter and γ as the shape parameter. The hazard for Log-logistic is not monotonic. The shape parameter is defined as:

If $\hat{\gamma} < 1$then the conditional hazard first rises, then falls.

If $\hat{\gamma} \geq 1$ then the hazard is declining

The survivor function for the log-logistic is

$$S(t) = \frac{1}{1 + (\lambda t)^{1/\gamma}} \quad (3.40)$$

the density function defined as;

$$f(t)=\frac{\frac{\lambda t}{\gamma}\left(\frac{1}{\gamma}-1\right)}{\left\{\gamma\left(1+(\lambda t)^{1/\gamma}\right)\right\}^2} \tag{3.41}$$

3.6.2.4 Log-Normal Model

The survivor function for log-normal model is:

$$S(t)=1-\Phi\left\{\frac{\ln(t)-\mu}{\sigma}\right\} \tag{3.42}$$

where Φ is the standard Normal Cumulative distribution function and $\sigma = \beta X$. The density function denoted as

$$f(t)=\frac{1}{\sigma t\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\ln(t)-u)^2\right\} \tag{3.43}$$

The hazard function for the Log-normal is given as:

$$h(t)=\frac{\frac{1}{\sigma t\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\ln(t)-u)^2\right\}}{1-\Phi\left\{\frac{\ln(t)-\mu}{\sigma}\right\}} \tag{3.44}$$

The hazard rate for this model is similar to the log-logistic that is, where $\hat{\gamma} < 1$ the hazard first rises and then falls.

3.6.2.5 The Cox Proportional Hazards Regression Model

The Cox model is used to determine the effects predictor variables have on the survival time. Cox (1972) proposed a semi-parametric model making it more robust to produce results that will closely approximate to a correct parametric model. This model is usually written in terms of the hazard model formula. It defines the hazard at time t for a patient and a number of explanatory variables represented by$K$. The variable K represents a collection of covariates that is modeled to predict the patient's hazard. This is defined by;

$$h(t,K) = h_0(t)\exp\left(\sum_{i=1}^{p}\beta_i K_i\right) \tag{3.45}$$

where, $h_0(t)$ is the baseline hazard function, K_i is the explanatory or the predictor variable and β_i is the regression coefficients.

3.6.3 Joint Model structure

Association between the longitudinal and survival processes can arise in two ways. One is through common explanatory variables and the other is through stochastic dependence between W_{1i} and W_{2i}. The joint model consists of two linked sub models, which is often refer to as the measurement model for the longitudinal process, and the intensity model for the survival process. This joint modeling strategy can connect models for longitudinal data and survival data with each other. When association between the two processes exists, we should obtain less biased and more efficient inferences by using this joint model.

3.6.3.1 Longitudinal sub model

Taking the linear mixed effects model that incorporates subject-specific variances as proposed by (Lyles *et al.*, 1999) as the sub model in equation (3.26),

$$y_i = \mu_i(s) + W_{1i}(s)$$

where $W_{1i}(s) = b_{1i}s + b_{0i}$

3.6.3.2 Survival sub model

The survival sub model usually takes the form of proportional hazard model,

$$h_i(t) = h_0(t)\exp(W_{2i}(s) + \emptyset \boldsymbol{v_i}) \quad (3.46)$$

where;

$h_0(t)$ Represent the base line hazard function.

$\boldsymbol{v_i} \in \cup_{\boldsymbol{i}}$ is a vector of base line covariates with corresponding log hazard ratios $\emptyset$

$W_2(s)$ is similar to $W_1(s)$ in equation (3.26) .

finally the joint model consists of the two linked sub models,

$$h_i(t) = h_0(t)\exp(\alpha W_1(s) + \emptyset \boldsymbol{v_i}) \quad (3.47)$$

The parameter α indicates the strength of association between the longitudinal biomarker and the time-to-event. If $\alpha = 0$ then the joint model is reduced to separate models and fitting a joint model will not yield any advantage.

3.6.3.3 Likelihood for the Joint model

Constructing the likelihood for the joint model,

$$\prod_{i=1}^{n}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{m_i} f\left(y_i(t_{ij})|b_i,\theta\right)\right)f(b_i|\theta)f(T_i,d_i|b_i,\theta)db_i\right] \quad (3.48)$$

where

$$f\left(y_i(t_{ij})|b_i,\theta\right) = (2\pi\sigma_e^2)^{-1/2}exp\left\{-\frac{y_i(t_{ij})-W_i(s)}{2\sigma_e^2}\right\}, \quad (3.49)$$

$$f(b_i|\theta) = (2\pi|V|)^{-1/2}exp\left\{-\frac{b_i{}'V^{-1}b_i}{2}\right\}, \quad (3.50)$$

The likelihood component under the Weibull or exponential sub model can be express as;

$$f(T_i,d_i|b_i,\theta) = [h_o(T_i)exp(\alpha W_i(s)+\emptyset v_i)]^{d_i}exp\left\{-\int_0^{T_i}h_0(u)exp(\alpha W_i(u)+\emptyset v_i)du\right\} \quad (3.51)$$

3.6.1.5 Model selection criteria

Akaike's information criterion (AIC) and Bayesian Information Criterion (BIC) are indices of relative goodness-of-fit and was used to compare models with the same fixed effects but different covariance structures. Both of these criteria apply rather generally for purposes of model selection and hypothesis testing. Smaller AIC or BIC values show a better fit. However, the BIC is preferred if the distribution have a sufficiently large sample size because it penalises more severely than the AIC does. In that regard, the two criteria will not always agree on the choice of best model or hypothesis since our objective is parsimony, hence BIC is more reliable than the AIC criterion.

$$AIC = L(\hat{\theta}) - q \quad (3.52)$$

$$BIC = L(\hat{\theta}) - (q + 2)\, log(N^{*}) \tag{3.53}$$

where;

$L(\hat{\theta})$ is the maximized log-likelihood(ML) or restricted maximized log-likelihood (REML).

q is the number of parameters in the covariance matrix,

P is the number of fixed effect parameters.

N is the number of subjects.

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.0 Introduction

This chapter of the research presents preliminary analysis of the data, further analysis of separate and joint model and discussion of results.

4.1 Preliminary Analysis

4.1.1 General Descriptive Statistics

This research covers a period of 8years starting from January 2006 to December 2014 with a total population of 119 HIV patients who are on two different ART regimens viz; AZT/3TC/NVP and AZT/3TC/EFV. At the end of the studies, 91 (76.5%) of the population were females and 28 (23.5%) were males, of which 29 (24.4%) were alive, 78 (65.5%) were lost to follow-up and 12 (10.1%) experienced the event (death). The study also revealed that the mean age recorded was 41.3years, with 22years and 67years being the minimum and maximum ages respectively. Regarding marital status, 73 (61.3%) were married and 46 (38.7%) not married. More than half of the patients thus 65 (54.6%) have not gotten any form of formal education and 54 (45.4%) had a feel of various levels of education. Regarding smoking and alcohol drinking, 17 (14.3%) smoke various types of cigarette whiles 52 (43.7%) drink various types of alcohol as shown in Table 4.1.

Table 4.1: General Descriptive Statistics

Variables	Count	Percentage
Gender		
F	91	76.47
M	28	23.52
Drug		
AZT/3TC/EFV	38	31.56
AZT/3TC/NVP	81	68.43
Outcome		
Alive	29	24.36
Lost	78	65.54
Dead	12	10.08
Education		
Yes	54	45.37
No	65	54.62
Married		
Yes	73	61.34
No	46	38.65
Smoking		
NO	102	85.71
YES	17	14.28
Alcohol		
NO	67	56.30
YES	52	43.69

Table 4.2 indicates the descriptive statistics of CD4 count of patients on ART in which males recorded a mean CD4 count of 417.4 with females recording a higher value of 539.6. Moreover, males patients recorded a minimum value of 7.0 with females recording a higher minimum value of 40.0. Also, male patients registered a maximum value of 1707 with female patients recording a value of 1885.0. Married patients had the least minimum CD4 count of seven (7) and the highest maximum CD4 count value of 1885 compared to non-married patients.

With regards to life styles, patients who take alcohol have a mean CD4count value of 527.7 higher than patients who do not take alcohol recording 486.8 and

their corresponding minimum and maximum value following the same trend but with different values. However, a mean CD4 count of 431.4 is registered for patients who smoke and 517.1 for non-smokers.

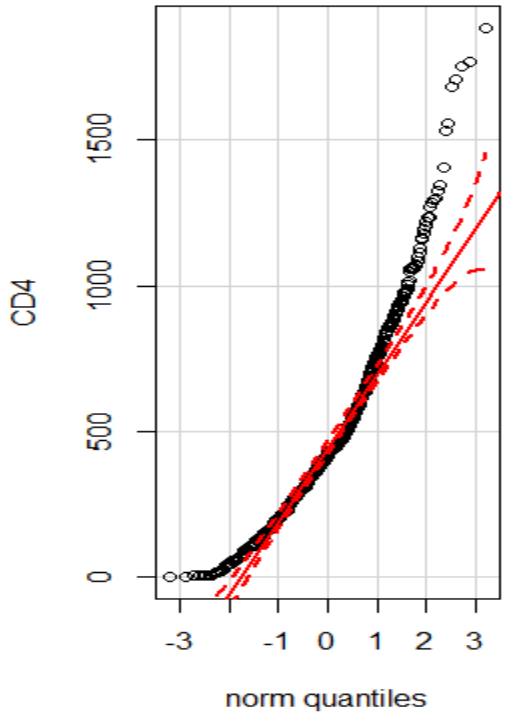
Meanwhile, patients who do not smoke recorded a minimum CD4 count value of seven while patients who smoke recorded minimum CD4count of 58.0. However, patients who smoke recorded a maximum CD4 count of 1238.0 and patients who do not smoke recorded 1885.0.

Table 4.2: Descriptive Statistics of CD4 Count ($cell/\mu L$) on Covariates

Variables	Mean	SE Mean	StDev	Minimum	Median	Maximum
Gender						
F	539.6	12.6	283.6	40.0	492.0	1885.0
M	417.4	21.5	282.8	7.00	377.0	1707.0
Drug						
AZT/3TC/EFV	474.1	16.5	241.7	7.00	439.0	1672.0
AZT/3TC/NVP	524.2	14.2	306.2	40.0	444.5	1885.0
Education						
Yes	488.4	14.4	276.2	7.000	432.0	1755.0
No	532.2	17.1	300.5	40.0	470.0	1885.0
Married						
Yes	481.2	14.5	289.5	7.0	429.5	1885.0
No	546.6	16.8	282.5	46.0	472.0	1707.0
Smoking						
No	517.1	11.9	292.8	7.0	453.0	1885.0
Yes	431.4	27.8	231.3	58.0	374.0	1238.0
Alcohol						
No	486.8	14.8	265.3	40.0	438.0	1672.0
Yes	527.7	16.2	306.3	7.0	448.0	1885.0

To ensure that the longitudinal data meets the normality assumption, a Q-Q plot of the CD4count was plotted and it was realised that most of the values deviated

from the normal line as indicated in Figure 4.1a. As a result a square root transformation was performed on the CD4count and after which it was observed that the data has attained normality as shown in Figure 4.1b.



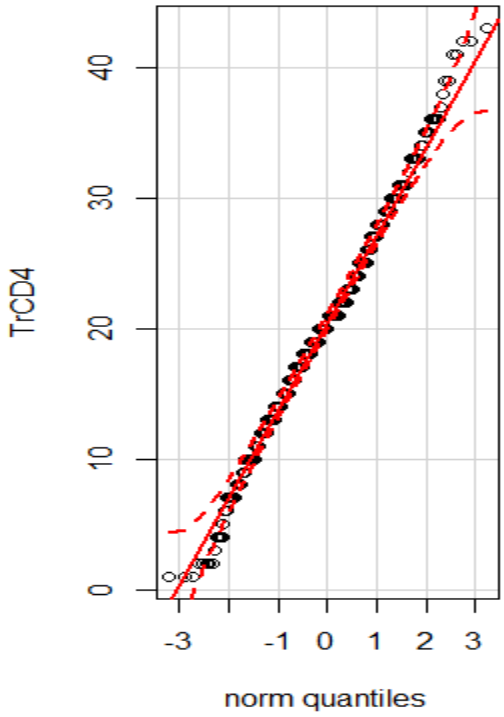


Figure 4.1a Q-Q plot of CD4 count **Figure 4.1b Q-Q plot of $\sqrt{\mathbf{CD4}}$ count**

4.1.2 Exploring the general mean structure of CD4 count

Following the data transformation, we need to ascertain the possible relationship of mean CD4 count over the space of time and as shown in Figure A1 in the appendix with a fitted trend line. The appropriate trend line was determined using AIC and BIC values after fitting the various trend line. Hence, we considered the polynomial model of order 3 since it has the least AIC and BIC value.

Table 4.3: Information criteria for trend models

Model	AIC	BIC
Linear	82.47128	84.97092
Exponential	82.21864	84.71828
Logarithmic	81.77224	84.27188
Polynomial (order=2)	84.26567	87.59852
Polynomial (order=3)	**66.57995**	**70.74601**
Polynomial (order=4)	67.91535	72.91463

4.1.3 Profile analysis

Figure A1 in the appendix indicated that the mean value of CD4 count for males and females were changing over time. The mean CD4 count of males precipitously declined immediately only after performing better than their female counterpart in the twelve month of treatment. Moreover, the profiles of females generally outperformed their males' counterpart with CD4 count during the treatment period.

Regarding alcohol, there was a change of mean CD4 count over time with those who drink alcohol and those who do not drink. Further investigation revealed that those who drink alcohol had a better initial CD4 count than those who do not drink from the beginning of the treatment. However, the mean CD4 count of those who drink declined at the 42nd month where non-drinkers performed better and precipitously inclined at the 66th month as shown in Figure A2 in the appendix. However, Figure A3 in the appendix revealed that there was a change of mean CD4 count over time with those who smoke and those do not smoke. Smokers performed better at the beginning of treatment but sharply declined immediately

after the 6th month and continuous to decrease but start to increase at a decreasing rate after the 30th month, however, non-smokers performance is increasing at a decreasing rate.

There is a conspicuous increment of CD4 count as results of two the drug regimens as indicated in Figure A4 in the appendix. The graph further showed that there is a mean change in CD4 count overtime and the pattern of the change is almost the same for AZT/3TC/NVP and AZT/3TC/EFV. Figure A5 in the appendix indicated that the mean values of CD4 count of educated and non-educated were changing over time. However, the profiles showed non-educated patients generally outperformed their educated counterpart with CD4count during the treatment period. The mean values of CD4 count of married and non-married patients were changing over time. But, the profiles showed that non-married patients generally outperformed their married counterpart on the average CD4count during the treatment period as shown in Figure A6 in the appendix.

4.1.4 MANOVA TEST

From Table 4.4 the multivariate analysis of variance conducted on the various groups revealed that the mean CD4 count for males and females were significantly different as well as married and non-married patients. The test further revealed that the mean CD4 count for patients in each group, drug regimens and Educational status were not significantly different. Regarding

Alcohol and Smoking, the mean CD4 count for patients in each group are also not significantly different.

Table 4.4: MANOVA Test for Groups

Parameter/Criterion	Test Statistic	F	N. DF	D. DF	P
Drug					
Wilks'	0.99888	0.752	1	671	0.386
Lawley-Hotelling	0.00112	0.752	1	671	0.386
Pillai's	0.00112	0.752	1	671	0.386
Roy's	0.00112				
Gender					
Wilks'	0.97506	17.161	1	671	0.000
Lawley-Hotelling	0.02557	17.161	1	671	0.000
Pillai's	0.02494	17.161	1	671	0.000
Roy's	0.02557				
Education					
Wilks'	0.99455	3.676	1	671	0.056
Lawley-Hotelling	0.00548	3.676	1	671	0.056
Pillai's	0.00545	3.676	1	671	0.056
Roy's	0.00548				
Married					
Wilks'	0.99051	6.426	1	671	0.011
Lawley-Hotelling	0.00958	6.426	1	671	0.011
Pillai's	0.00949	6.426	1	671	0.011
Roy's	0.00958				
Smoking					
Wilks'	0.99542	3.088	1	671	0.079
Lawley-Hotelling	0.00460	3.088	1	671	0.079
Pillai's	0.00458	3.088	1	671	0.079
Roy's	0.00460				
Alcohol					
Wilks'	0.99815	1.245	1	671	0.265
Lawley-Hotelling	0.00185	1.245	1	671	0.265
Pillai's	0.00185	1.245	1	671	0.265
Roy’s	0.00185				

D-Denominator, N-Numerator

4.2 Further analysis

4.2.1 Longitudinal Analysis

From Table 4.5, using AIC and BIC we compared the intercept model and intercept and slope model. It revealed that the model with intercept and slope has the least values of AIC and BIC and as a result we proceeded with a model with random intercept and random slope.

Table 4.5: AIC and BIC values of different linear mixed effect models

MODEL	AIC	BIC
Intercept	4680.9	4756.3
Intercept and Slope	**4670.9**	**4746.9**

We determined the underlying covariance structure appropriate for the data by exploring some covariance structures. Using the AIC and BIC values as indicated in Table 4.6, the unstructured covariance structure has the least values of AIC and BIC and hence considered as the appropriate underlying covariance structure.

Table 4.6: Statistic for the Covariance structures

Covariance Structure	**AIC**	**BIC**
Variance Component	3975.1	4055.7
First Autoregressive AR (1)	4089.1	4167.0
Compound Symmetry	4089.1	4167.0
Unstructured	**3973.0**	**4055.3**

Table 4.7 shows that the initial CD4 count, gender and time of treatment (in months) were significant determinants of change in CD4 counts of patients on ART. Male patients have about 2.21 units decrease of CD4 counts less than their female counterparts for every CD4 count. The advantage of early diagnosis and

treatment was shown in the PreCD4 count, thus when patient starts treatment early, the average change gained in CD4 counts is expected to be 0.016 units for every CD4 count. The rate of change in CD4 count is 0.048 counts per unit increase in time, suggesting the rate of change of CD4 count increase with time while a patient CD4 count decrease by about 0.03 units for every additional year of a patient's age holding other factors constant. Other covariates used in the study but were not statistically significant includes alcohol drinking, cigarette smoking, education, married, and drug. Though cigarette smoking and alcohol drinking were not significant, patients in these categories have a positive change in CD4 count by 0.52 units and 1.08 units respectively. However, the parameter estimates of cigarette smoking and alcohol drinking suggest that patients in the categories have CD4 higher than non-smokers and non-drinkers respectively. In addition, the study revealed that educated patients have a non-statistical significant decrease of CD4 count of about 0.35 units less than their non-educated counterparts and married patients also have about 0.65 units lower than their unmarried counterparts. Drug regimen of NVP components would have an insignificant increase of 0.11 units more than regimen of EFV component.

Table 4.7: Parameter Estimate of Mixed Effect of full model

Parametres	Value	Std.Error	DF	t value	p value
Intercept	18.415667	2.0194988	553	9.118929	0.0000
Compared with EFV					
Drug (NVP)	0.105318	0.7207710	553	0.146119	0.8839
Compared with Female					
Gender (Male)	-2.216603	0.9605134	115	-2.307728	0.0228
Compared with (NO)					
Education (YES)	-0.34869	0.8054122	553	-0.432944	0.6652
Compared with (NO)					
Alcohol (YES)	1.081634	0.8386667	553	1.289706	0.1977
Compared with (NO)					
Smoking (YES)	0.523765	1.1737440	115	0.446235	0.6563
Time	0.047982	0.0122346	553	3.921831	0.0001
preCD4	0.016444	0.0026337	115	6.243499	0.0000
Married (YES)	-0.648003	0.8356824	553	-0.775417	0.4384
age	-0.028564	0.0380705	553	-0.750301	0.4534

Random effects	StdDev	
Intercept	3.9513331	(Intr)
Time	0.0632366	-0.365
Residual	3.6553211	

From Table 4.7 the full linear mixed effect model is given as;

$$\boldsymbol{E(CD4) = 18.42 + 0.11 * NVP - 2.22 * Gender(Male) - 0.35 * Educated + 1.08 * Alcohol0.52 * Smoking + 0.05 * Time + 0.02 * PreCD4 - 0.65 * Married - 0.03 * Age} \quad (4.0)$$

The stepwise method was used in the selection of the reduced model guided by AIC and BIC model selection criterion. Following that the only covariates

included in the final model were PreCD4 count, Gender, Drug and Alcohol as shown in Table 4.8.

Table 4.8: Model Selection for prediction

Model	AIC	BIC
Drug, PreCD4, Time, Gender, Education, Maritalstatus, Alcohol, Smoking, Age	3974.69	4072.556
Drug, PreCD4, time, Gender, Education, Alcohol, Smoking, Age	3964.10	4110.263
PreCD4, Time, Gender, Education, Alcohol, Smoking, Age	3958.04	4043.907
PreCD4, Time, Gender, Alcohol, Smoking, Age,	3953.53	4016.795
PreCD4, Time, Gender, Alcohol, Age	3947.56	3992.754
PreCD4, Time, Gender, Alcohol	3946.24	3986.908
PreCD4, Time, Gender	**3945.70**	**3981.851**

The reduce model further revealed that gender, time and preCD4 count were significant in the rate of change of CD4 count as indicated in Table 4.9.

Table 4.9: Estimates of Reduced Linear Mixed-effect Model

Pararmeters	Value	Std.Error	DF	t value	p-value
Intercept	17.242413	0.7444227	558	23.162127	0.0000
Gender (Male)	-2.405326	0.9201100	116	-2.614173	0.0101
Time	0.047306	0.0118133	558	4.004455	0.0001
PreCD4	0.016489	0.0026335	116	6.261213	0.0000
Random effects	StdDev				
Intercept	4.06727742	(Intr)			
Time	0.06143932	-0.425			
Residual	3.66074861				

From Table 4.8 the reduced linear mixed effect model is given as;

$$\boldsymbol{E(CD4) = 17.24 + 0.02 * preCD4 + 0.05 * Time - 2.41 * Gender(male)} \quad (4.1)$$

4.3 Survival Analysis

4.3.1 Log-rank test for equality of Survivor function by groups

The significant difference among groups of the covariates was determined using the log rank test of equality as shown in Table 4.10. The test showed a significant difference of survival among the groups; smoking, alcohol and drug regimen. However, covariates such as education, married, and gender were not significantly different.

Table 4.10: Log-rank test for equality of Survivor function by groups

GROUP	Chisq	Df	P-value
Drug	7.20	1.00	0.01
Gender	0.30	1.00	0.56
Alcohol	4.10	1.00	0.04
Smoking	4.80	1.00	0.03
Married	1.80	1.00	0.18
Education	0.00	1.00	0.96

In other to choose the appropriate survival model for this research work; Cox PH models and four parametric survival models; Exponential, Loglogistic, Weibull and Lognormal were explored and then compared. Using AIC and BIC as shown in Table 4.11, Weibull model appeared the best since it has the least values of AIC and BIC.

Table 4.11: Model comparison

Criterion	Cox model	Weibull	Exponential	Llogistic	Lnormal
AIC	7626.4	**573.05**	761.768	1455.8	901.871
BIC	7673.057	**690.55**	879.265	1568.875	1019.370

4.3.2: The Weibull Model

The Weibull model as shown in Table 4.12 revealed that drug regimen of NVP component will lower the hazard of patient by 2.54 ($p = 6.95\text{e} - 04$) than EFV component of drug regimen. However, the results revealed there was an increase in hazard with an increase in age by hazard of 1.03 significant at $p = 1.43\text{e} - 02$. Patient with high preCD4 count lowers the hazards by 1.00 which is statistical significant at $p = 2.34\text{e} - 04$. However, the rest of the covariates were all non-statistical significant to the survival time of patient. To determine the prognostic factors and the reduced model for prediction, the AIC stepwise selection criterion was used. The reduced model indicated that covariates such as PreCD4 count, Age and Drug are significant determinants of patient's survival time as shown in Table 4.13.

Table 4.12: Weibull model of all covariates

Parameter	Value	StdError	z	p
Intercept	3.24842	0.61861	5.25	1.51e-07
Compared with EFV				
Drug (NVP)	0.93058	0.27439	3.39	6.95e-04
Compared with (NO)				
Alcohol (YES)	-0.46628	0.26321	-1.77	7.65e-02
Compared with (NO)				
Smoking (YES)	-1.12953	0.201633	-5.6019	2.12e-01
Compared with (NO)				
Married (YES)	0.61088	0.28377	2.15	3.13e-01
Compared with (NO)				
Education (YES)	0.39391	0.26645	1.48	1.39e-01
Compared with Female				
Gender (Male)	-0.42460	0.30209	-1.41	1.60e-01
PreCD4	0.00396	0.00107	3.68	2.34e-04
Age	0.03295	0.01345	2.45	1.43e-02
Log(scale)	-0.62320	0.103190	-6.0394	1.55e-09
Scale= 0.536				

From Table 4.12 the full Weibull model is given as;

$$\boldsymbol{h(t) = \rho t^{(\rho-1)} \exp(0.93 * Drugs + 0.0032 * preCD4 + \ 0.03 * Age - 1.13 * smoking + 0.61 * married(yes) - 0.47 * alcohol(yes) + 0.4 * education(yes) - 0.42 * gender(male))} \quad (4.2)$$

Table 4.13: Stepwise Covariates selection

Covariates	AIC	BIC
Drug, preCD4, Religion, Education, Married, Gender, Alcohol, Smoking, Age	217.505	263.567
Drug, preCD4, Religion, Education, Gender, Alcohol, Smoking, Age	214.170	249.394
Drug, preCD4, Religion, Education, Gender, Smoking, Age	212.399	244.914
Drug preCD4 Religion Education Smoking Age	211.264	241.069
Drug, preCD4, Age	**210.851**	**232.527**

Table 4.14 is the weibull reduced model in which all the parameters contribute significantly to the model. The NVP regiment which lowers the hazard by 2.34 is the common regiment patient used by patients

Table 4.14: Estimate of reduced model

	Value	Std Error	z	p
(Intercept)	3.90825	0.56445	6.92	4.39e-12
Drug (NVP)	0.84828	0.24895	3.41	6.56e-04
precd4	0.00316	0.00106	2.98	2.87e-03
Age	0.03015	0.01322	2.28	2.25e-02
Log(scale)	-0.49536	0.107221	-4.62	3.84e-06
Scale= 0.609				

From Table 4.14 the reduced Weibull model is written as;

$$h(t) = \rho t^{(\rho-1)} \exp(0.85 * Drugs + 0.0032 * preCD4 + 0.03 * Age) \quad (4.3)$$

4.4 Joint Model

Because convergence failure precluded accurate computation of posterior model summaries, we used a Weibull model with *r* = 1 (i.e., an exponential) resulting in Table 4.15. The results revealed that the time of treatment has statistically significant positive effect on the average CD4 and thus indicating that there is increase of 0.0000103 units of CD4 count per units increase in time (95% CI: 4.46e-06, 0.000016). The results in addition revealed that PreCD4 count or initial CD4 count also positively affect the CD4 insignificantly, thus indicating that with early diagnosis and treatment the expected average increment in CD4 count was 0.0159314 units (95% CI: 0.0108126, 0.0210502). But male patient has a significant reduction on the average CD4 count of 2.873938 units (95% CI: -4.707743, -1.040132) as compared to their female counterparts.

Treatment NVP regimen component effect lowers the hazard of death by -1.16353 (95% CI:-2.208942, -0.1181169) than EFV regimen component. However, the other survival covariates that are non-statistically significant include age and preCD4 count. The estimated association parameter in the joint model was highly negative with a value of -0.3317109 (95% CI:-0.5826281, -0.0807937) and it is

statistically significant in the weibull model ($p = 0.010$). This indicates that there is strong evidence of association between the two sub-models.

Further investigation of this association showed that both initial level and slope of CD4 count were negatively associated with the hazard of death. This finding is clinically predictable since patients with more drastic decline in CD4 count would have poorer survival. Under this joint model, a patient's survival is associated with his/her longitudinal data pattern of the rate at which CD4 count increase.

Table 4.15: Parameter Estimates of Joint Model

Parameters	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Longitudinal						
Time	0.0000103	2.95e-06	3.47	0.001	4.46e-06	0.000016
Gender	-2.873938	0.9356323	-3.07	0.002	-4.707743	-1.040132
preCD4	0.0159314	0.0026117	6.10	0.000	0.0108126	0.0210502
Int	20.71174	1.342994	15.42	0.000	18.07952	23.34396
Survival						
Assoc:value ln_lambda	**-0.331711**	0.1280213	-2.59	0.010	-0.5826281	-0.0807937
Age	-0.0350959	0.0309258	-1.13	0.256	-0.0957094	0.0255176
preCD4	0.0040346	0.0031685	1.27	0.203	-0.0021755	0.0102447
Drug (NVP)	-1.16353	0.5333836	-2.18	0.029	-2.208942	-0.1181169
Int	3.273715	2.52749	1.30	0.195	-1.680074	8.227505

Random Effect	EStimates	Std. Err	[95% Conf. Interval]	
Time	0.0000144	4.34e-06	8.02e-06	0.000026
Int	4.104442	0.3520551	3.469309	4.855849
corr(Time*Int)	-0.6864479	0.1326006	-0.8699059	-0.3361284
Residual	3.620334	0.1211074	3.390582	3.865655

The longitudinal sub-model and survival sub-model was obtained from Table 4.15;

thus longitudinal sub-model is written as,

$$\boldsymbol{E(CD4) = 20.71 + 0.02 * preCD4 + 0.0000103 * Time}$$

$$\boldsymbol{-2.87 * Gender(male)} \quad (4.4)$$

and the survival sub-model in the absence of random effects,

$$\boldsymbol{\log\big(h(t)\big) = 3.27 - 1.2 * Drugs + 0.004 * preCD4 - 0.04 * Age} \quad (4.5)$$

The incorporation of stochastic variable from the linear mixed effect model into the survival sub model gave us the joint model,

$$\boldsymbol{\log\big(h(t)\big) = -1.164 * Drugs + 0.004 * preCD4 + -0.035 * Age - 0.332 * (4.104442 + 0.0000144 * Time)} \quad (4.6)$$

4.5 Discussion

The exploratory analysis revealed that the total number of patients considered for this study was 119, of which 76% were females and 23% males. The higher number of females as compared to males may be as a result of government intervention programme, the free Antenatal care program within the country in which all pregnant women who passed through the free health care delivery

system will be screened of HIV. Regarding CD4 counts, female patients recorded the highest average CD4 count and the highest minimum and maximum CD4 count as well as the highest variability. The exploratory analysis also revealed that about 44% of patients consume various kinds of alcohols and about 14% smoke various types of Cigarettes. The MANOVA test on the mean profiles of smokers and non-smokers indicated that there is no statistical significant difference in the pattern of change of mean CD4 count at 5% significance level. The same test conducted on patients who drink and those who do not drink proves no significant difference in the pattern of change of their mean profiles.

The MANOVA test also showed that there is no statistical difference in the mean profiles of CD4 count of educated and non-educated patients. In addition, Majority of patients (68%) were on AZT/3TC/NVP regiment whilst 32% of them were on AZT/3TC/NVP. Figure A4 in the appendix showed that the mean profiles for the two drug regimens had an increasing pattern of CD4 count overtime and the MANOVA test applied to the mean profiles showed that there were no significant drug regimen differences on the pattern of change of patient mean CD4 count at the 5% significance level. The MANOVA test also revealed that there were no significant educational status differences in the pattern of change of patients' mean CD4 count at the 5% significance level, this was further demonstrated in Figure A5 in the appendix as the mean profiles seems to move in the same pattern. However, the profiles showed that the non-educated patients generally outperform their educated counterparts.

However, consistent with Adams and Luguterah (2013), the MANOVA test revealed that there was significant gender difference in the pattern of change of mean CD4 count. Figure A1 further indicated that the profiles of females generally outperformed their males' counterpart with CD4 count during the treatment period. The MANOVA test showed that there was a significant marital status difference in the pattern of change of mean CD4 count of patients. This was demonstrated pictorially in Figure A6 in the appendix which indicated that as non-married patients showed an increasing pattern of mean CD4 count, married patients showed a decreasing pattern of mean CD4 counts.

The linear mixed effects model employed to cater for the longitudinal data revealed that a patient initial CD4 count significantly determined his/her current CD4 count suggesting that a patient with higher initial CD4 count will have 0.016 units of CD4 better than patients with lower initial CD4 count. The time of treatment significantly affect patients CD4 count positively, showing that there was positive relationship between time of treatment and CD4 count growth, as per every six months there was 0.05 units increment in patient CD4 count. Gender in addition significantly affects CD4 count as male patients have 2.22 units of CD4 count lower than their female counterpart and this is contrary to Adams and Luguterah (2013), but consistent with Mair *et al.* (2008). This might be as a result of various negative life-styles indulged by these male patients. Moreover, the effect of marital status has non-statistical significance on patients CD4 count though married patients have 0.65 units decrease in CD4 count compared with

their unmarried counterparts, this might be as a results multiple reinfection and unstable mind due complications in marriage. Alcohol and cigarette also have non-statistical significant effects on CD4 count though the results proved that patients of these categories have higher growth in CD4 counts than their respective counterparts. Drug (AZT/3TC/NVP) has a non-statistical significant effect on CD4 count, but cause an increase in CD4 by 0.11units than AZT/3TC/EFV regimen.

On the equality of survival, the log rank test conducted revealed that patients' survival functions as a result of the two drug regimens differed significantly. The test also revealed that there was significant differences between the survival functions of patients who smoke various types of cigarettes and those who do not smoke as well as those who drink alcohol and those who do not drink alcohol. However, there was no significant gender difference in the survival functions of patients. Marital status and educational status were also not significantly different in their respective survival functions.

Various models such as cox, lognormal, log logistic, exponential and weibull models were compared using AIC and BIC and weibull model had the least AIC and BIC values. The weibull model indicated that age and Drug were the significant determinant of patient survival time whilst the rest of the covariates were insignificant. The reduced weibull model however indicated the preCD4 count was also significant determinant of patient survival time.

Furthermore, to assess the relationship between the trajectory of CD4+ counts over time and the risk of death among HIV patients, we used joint modelling. The association between the longitudinal and survival process can arise in two ways: through common explanatory variables or through stochastic dependence between subject-specific random effects (Guo *et al.*, 2004).

Allowing the hazard of death to depend on the longitudinal process of an intercept and a slope, thus a patient's baseline CD4+ count and the impact of the rate of change in CD4+ counts on a patients' survival time will results in the joint model. In order to avoid convergence failure that prevents computation of posterior summaries we used a weibull model of $P = 1$ (i.e., an exponential). In addition, the general pattern of the CD4 count follows a polynomial trend of order three and as such three was used for power transformations of the time variable included in the longitudinal sub-model as random effects. The the joint model results differed much more from the separate models results, for instance, the random parameter estimates for the joint model have significantly smaller values than the separate models, Suggesting longer patient survival time or lower hazard rate of death (Ping *et al.*, 2012). The results from the joint model further indicated that the male patients have a much more significant lower CD4 count values than their female counterparts but drug has a reduced hazard rate less than the case of the separate model.

Consistent with Hyun *et al*. (2013) the estimated association parameter of the joint model is negative and highly significant suggesting a strong relationship between the two sub-models. The negative value of the association parameter implies the rate of change of CD4 count significantly lowers the hazard of death. Furthermore, the association showed that both the patient's baseline CD4 count and the rate of change in CD4+ counts (slope) were negatively associated with the hazard of death. This is clinically not doubtable since patients with more drastic decline in CD4 count would have poorer survival.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.0 INTRODUCTION

This chapter presents the conclusion and recommendations based on the analysis of the obtained data and literature reviewed.

5.1 Conclusion

We observed that the general pattern of the mean CD4 count followed a polynomial trend of order three. That pattern of change of mean CD4 count among gender and marital status differed significantly; non-married patients had a better increasing pattern of average change of CD4 count than married patients and female patients also had a better increasing pattern of mean change of CD4 count.

Drug regimen AZT/3TC/NVP contributed to the mean CD4 count increment better than AZT/3TC/EFV and reduced the hazard rate of death better than AZT/3TC/EFV.

Factors such as preCD4 count, gender, and duration of treatment (months) significantly determines HIV patient's CD4 count, whilst drug regimens, age and preCD4 determine the survival of patient.

We observed that there was a significant association between the repeated measured CD4 count and patients survival time. Hence the rate of change of CD4 count has a significant effect on the survival time of patient.

5.2 Recommendations

Joint modelling of CD4+ count progression and survival time should be performed to obtain a clear picture of the effect of specific covariate.

Stakeholders in the health sector and guardians of patients should guide, encourage and monitor patients to continue to adhere to treatment as recommended.

A research should be conducted to investigate into the reasons male patients are associated with low CD4 count recovery rate.

More public education should be done to encourage people to go for voluntary counselling and testing so that infected persons can be put on early treatment.

The measurement of CD4 count can be affected by some factors within patients and environmental factors. As results viral load should rather be measured so as to effectively monitor patients.

The impact of the rate of change of insulin on the survival of diabetic patients with the development of cardiovascular disease as time-to-event should studied using joint modelling.

REFERENCE

Adams, M., and Luguterah, A. (2013). Longitudinal Analysis of Change in CD4+ Cell Counts of HIV Patients on Antiretroviral Therapy (ART) in the Builsa District. *European Scientific Journal*, **9**(33**)**: 1857 – 7881.

Altman, D.G. (1991). Development in the 1980's. *Statistics in Medicine*, **10**: 1897-1913.

Anderson, P.K. (1991). Survival analysis 1982-91: the second decade of proportional hazard regression model. *Statistics in Medicine*, **10**: 1931-1941.

Behavioral change is an effective way of reducing the spreading of HIV (2015). Available at www.medwiser.org/hiv-aids/around-the-world/aids-in-africa. Accessed on 13th March, 2015.

Broder S. (2010). The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral Research*, **85**(1): 1–18.

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., Hakim, J. G., Kumwenda, J., Grinsztejn, B., Pilotto, J.

H. S. *et al.* (2011). Prevention of HIV-1 infection with early antiretroviral therapy. *New England Journal of Medicine,* **365**(1): 493–505.

Cox, D. R. (1972): Regression models and life tables. *Journal of the Royal Statistical* Society*; Series B,* **34**:184-220.

Ding, A. and Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics,* **2**(1): 13–29.

Dioggban Jakperik, and Bernard Kpakpo Acquaye (2013). Assessing the Effects of Prognostic Factors in Recovery of Tuberculosis Patients in the Upper West Region. *Mathematical Theory and Modeling*, **3**(11): ISSN 2224-5804.

Fred Nana Poku (2014). HIV prevalence rate stabilized in Ghana. Available at www.ghanaweb.com/GhanaHomePage/health/artikel.php?ID=306338. Accessed on 16th, April 2014.

Fleming, T. R. & D. P. Harrington (1991). Counting process and survival analysis. *John Wiley and Sons,* pp429.

Guo, X. and Carlin, B.P. (2004). Separate and Joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, **58**(1): 16-24

Hammer, S.M., Squires, K.E., Hughes, M.D., Grimes, J.M., Demeter, L.M., Currier, J.S., Eron, J.J. Jr., Feinberg, J.E., Balfour, H.H. Jr., Deyton, L.R., Chodakewitz, J.A., 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. *New England Journal of Medicine*, **337** (11): 725-33.

Henderson, R., Diggle, P., Dobson A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics*, **1**(4): 465–480.

HIV/AIDS epidemic in Ghana (2013). Available at www.wikipaedia.org/HIV/AIDS in Ghana. Accessed on 13th February, 2015.

HIV/Tuberculosis co-infection (2014). Available at www.modernghana.com. Accessed On November 26, 2014.

Hyun J. L., Prosanta M. and Stuart S. (2013). Joint modeling of longitudinal and event time data; Application to HIV study. *Journal of Medical Statistics and Informatics.* ISSN 2053-7662.

I never had HIV – Ghanaian HIV/AIDS ambassador confesses. Retrieved from www.citifmonline.com/2014/12/15/i-never-had-hiv-ghanaian-hivaids-ambassador-confesses.

Jakperik, D., and Ozoje, M. (2012): Survival analysis of average recovery time of tuberculosis patients in Northern region, Ghana: *International Journal of Current Research*, **4**:123-125.

Krishnan, S., Schouten, J. T., Jacobson, D. L., Benson, C. A., Collier, A. C., Koletar, S. L., Mitsuyasu, R. (2011). Incidence of Non-AIDS-Defining Cancer in Antiretroviral Treatment-Naïve Subjects after Antiretroviral Treatment Initiation: An ACTG Longitudinal Linked Randomized Trials Analysis. *Oncology*, **80**(1): 42–49.

Louise, C. (2011). History of HIV/AIDS. Available at www.webmd.com/hiv-aids/ss/slideshowaids-retrospective. Accessed on 19th March, 2014.

Lyles, R. H., Munz, A., Xu, J., Taylor, J. M. G. and Chmiel, J. S. (1999). Adjusting for measurement error to assess health effects of variability in biomarkers. *Stat.Med*., 18, 1069-1086.

Management of HIV/AIDS (2015). Available at www.wikipedia.org/w/index.php?title=Management_of_HIV/AIDS Accessed on 19th July, 2014.

Mair, C., Hawes, S. E., Agne, H. D., Sow, P. S., N'doye, I., Manhart, L. E., Gottlieb,G.S. and Kiviat, N. B. (2008). Factors associated with CD4+ lymphocyte counts in HIV-negative Senegalese individuals. *Clin Exp Immunol*, **151**(3): 432–440

McDermott, A. Y., Terrin, N., Wanke, C., Skinner, S., Tchetgen, E., & Shevitz, A. H. (2005). CD4+ Cell Count, Viral Load, and Highly Active Antiretroviral Therapy use are independent predictors of body Composition alterations in HIV-infected adults: A longitudinal study. *Clinical Infectious Diseases*, **41**(11): 1662–1670.

Medical News Today (2014). What is tuberculosis? What causes tuberculosis? Available at www.medicalnewstoday.com/categories/tuberculosis. Accessed on 22nd Oct, 2014.

Morrell, C.H. and Brant, L.J. (2000). A two-stage linear mixed-effects and Cox model for longitudinal data with measurement error and survival. *American Statistical Association*, 198-203.

Nan, M., Laird, James H. (1982). Random-Effects Models for Longitudinal Data. *International Biometric Society*, **38(**4): 963-974.

Natural History of Opportunistic Disease in an HIV-Infected Urban Clinical Cohort. *Annals of Internal Medicine*, **124** (7): 633.

Nelson, M. Fatkenheuer, G. Konourina, I. *et al*. (2007). ART-experienced patients infected with CCR5-tropic; 14th Conference on Retroviruses and Opportunistic Infections. *Los Angeles*, CA, USA.

Nelson M. (2007). We cannot win the battle against AIDS if we do not also fight Tuberculosis. Available at www.action.org/resources/item/tb-hiv-co-infection. Accessed on 12th October, 2014.

Origin of HIV and AIDS (2014). Available at www.avert.org/origin-aids-hiv.htm. Accessed on 1st March, 2015.

Peto, R. Pike, M.c., Armitage p., Breslow, N.E. et al. (1977). Design and analysis of randomized clinical trials requiring prolong observation of each patient II. Analysis and examples. *British Journal of cancer*, **35**(1): 1-39.

Phillips, A.N., Gazzard, B.G., Clumeck, N., *et al*. (2007). When should antiretroviral therapy for HIV be started? *British Medical Journal*, **334**:76-78.

Ping, W., Wei Shen, and Mark Ernest B. (2012). Joint modeling of longitudinal outcomes and survival using latent growth modeling approach in a mesothelioma trial. *Health Services* & *Outcomes* Research *Methodology*, **12**(2-3): 182–199.

Ponnuraja and Venkatesan (2010). Survival model for TB. Indian Society for Education and Environment (iSee); *Indian Journal of Science and Technology*, **3**(7): 0974- 6846

Rodriguez, W. R., Nicolaos, C., Pierre, N.F., Susan, G., Sanghamitra, M., Meredith, D., Mina, H. and Trevor, P. (2005). A microchip CD4+

counting method for HIV monitoring in resource-poor settings**.** *Peer-Reviewed Open-Access Journal Medicine***, 2(7):** e182.

Rosas-Taraco, A.G., Arce-Mendoza, A.Y., Caballero-Olin, G. and Carmona, M.C. (2006). Gyrase Mycobacterium tuberculosis upregulates co-receptors CCR5 and CXCR4 while HIV modulates CD1 4 favoring concurrent infection. *Gyrase AIDS Res. Hum. Retroviruses*, **22**(1): 45-51.

Santos, S.P., Amado, C.A. (2012). Assessing the efficiency of mother-to-child HIV prevention in low and middle-income countries using data envelopment analysis. *Santos MF Health Care Management Science*, **15**(3):206-22.

Siegfried, N., van der Merwe, L., Brocklehurst, P., Sint, T.T. (2011). Antiretroviral for reducing the risk of mother-to-child transmission of HIV infection. *Cochrane database of systematic reviews*, **6**(7): CD003510.

Singh R, Mukhopadhyay K. (2011). Survival analysis in clinical trials: Basics and must know areas. *Perspect Clin Res*, **2**:145-8.

Steward, W.T., Bharat, S., Ramakrishna, J., Heylen, E., Ekstrand, M, L. (2013). Stigma is associated with delays in seeking care among HIV-infected people in India. *Journal of International Association Provide Care*, **12**(2): 103-9.

Towers, C. V., Rumney, P. J., & Ghamsary, M. G. (2010). Longitudinal study of CD4+ *AIDS* cell counts in HIV-negative pregnant patients. *Journal of Maternal-Fetal & Neonatal Medicine*, **23**(10): 1091–1096.

Tudor Car, L., van-Velthoven, M.H., Brusamento, S., Elmoniry,H., Car, J., Majeed, A., Atun, R. (2011). Integrating prevention of mother-to-child HIV transmission (PMTCT) programmes with other health services for preventing HIV infection and improving HIV outcomes in developing countries. *Cochrane database of systematic reviews*, **15**(6):CD008741.

Tyurina, A., Krupitsky, E., Cheng, D. M., Coleman, S. M., Walley, A. Y., Bridden, C., Samet, J. H. (2013). Is cannabis use associated with HIV drug and sex risk behaviors among Russian HIV-infected risky drinkers? *Drug & Alcohol Dependence*. **132**(1/2): 74–80.

UNAID AIDS Report on Africa (2011). Available at www.unaid.org. Accessed on 3rd August, 2014.

UNAIDS World AIDS Day Report (2012). The composite budget of the Builsa district assembly for the 2012 fiscal year. Available at www.ghanadistrict.com. Accessed on 5th March, 2015.

Waters, L., Fisher, M., Anderson, J., Wood, C., Delpech, V., Hill, T., Sabin, C. (2011). Responses to highly active antiretroviral therapy and clinical events in patients with a low CD4 cell count; late presenters vs. late starters. *HIV Medicine,* **12**(5): 289–298.

World health organization (2010). Antiretroviral therapy for HIV Infection in Adults and Adolescent: Recommendation for a public health Approach Available at www.who.int/hiv/topics/treatment/en/index.html. Accessed on July 10, 2014.

World health organization (2013). Consolidated guidelines on the use of antiretroviral drugs for Treating and preventing HIV infection. Available at www.who.int/hiv/pub/guidelines/arv2013. Accessed on July 15, 2014.

World health organization (2015). Global Health Observatory (GHO) data. Available at www.who.int/sysmedia/media/resources/favicon.ico. Accessed on May 10, 2015.

World health organization (2002). Global tuberculosis report. Available at www.who.int/tb. Accessed on July 10, 2014.

Wulfsohn, M.S., Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics,* **53**:330–339.

Xiuhong, L., Margolick, J. B, Jamieson ,B. D., Rinaldo, C. R., John, P. Phair, J. P and Lisa, P. Jacobson, L. P. (2011). CD4+ T-Cell Counts and Plasma HIV-1 RNA Levels Beyond 5 Years of Highly Active Antiretroviral Therapy. *Journal of Acquired Immune Deficiency Syndromes,* **57**:421–428.

Yehia, B., Frank, I. (2011). Battling AIDS in America: an evaluation of the National HIV/AIDS Strategy. *American Journal of Public Health,* **101**(9): e4–e8.

APPENDIX

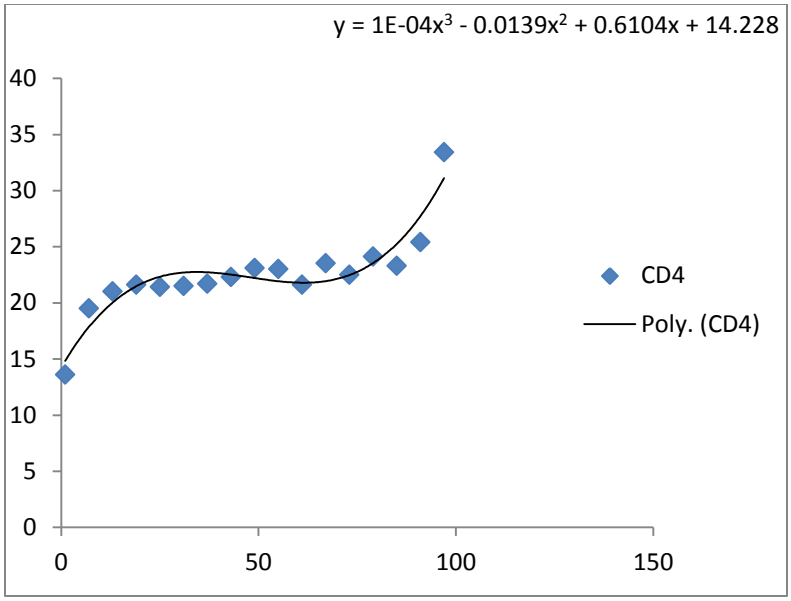


Figure A1: General pattern of CD4 count over time

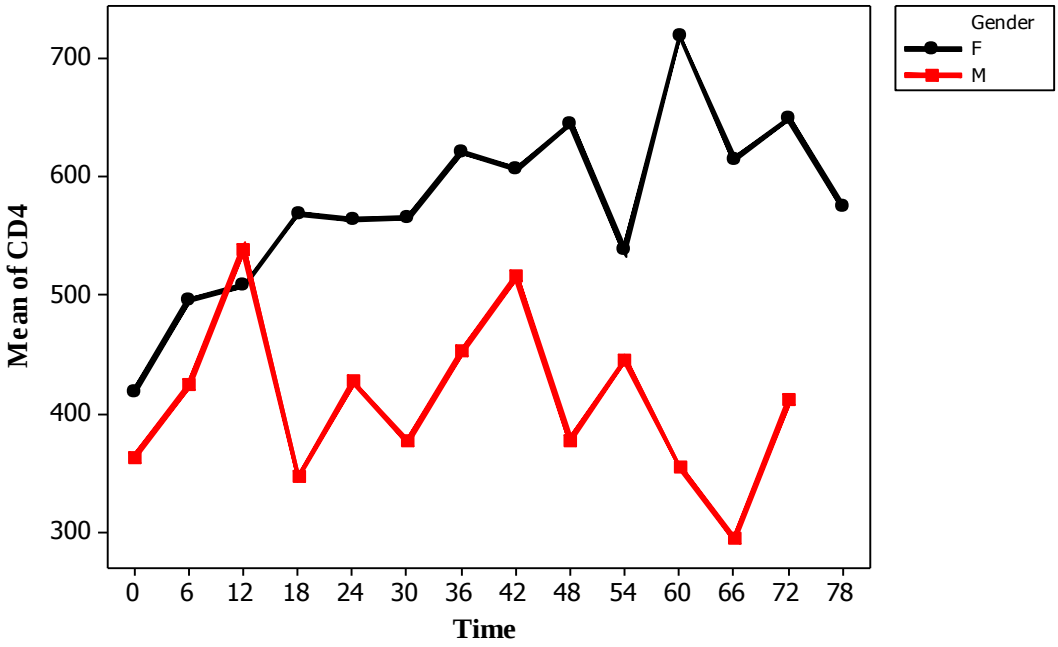


Figure A2: Profile plot of CD4 count by Gender

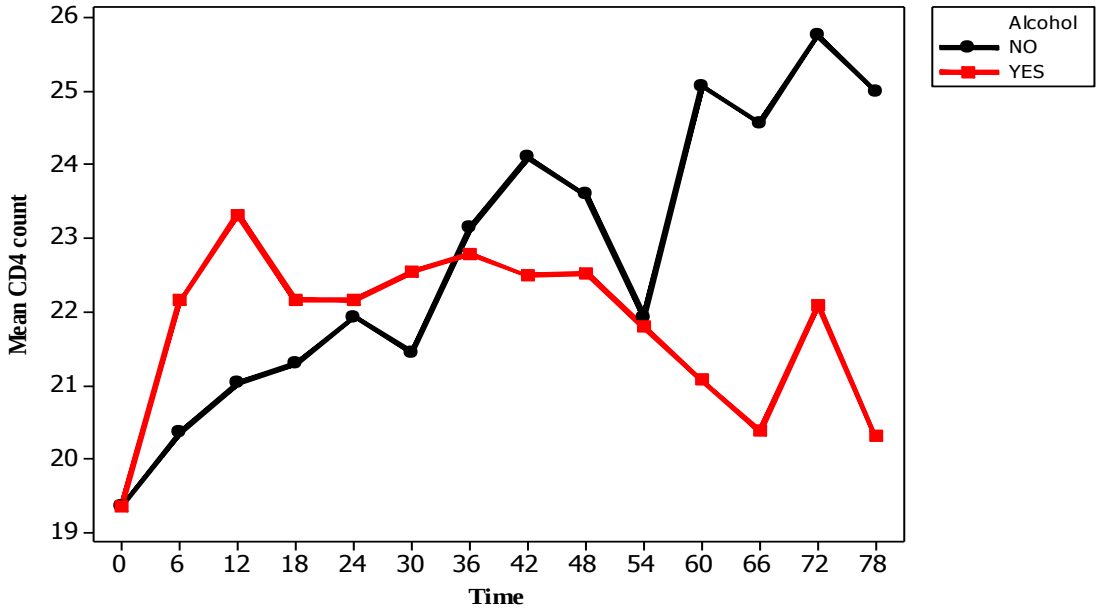


Figure A3: Profile plot of CD4 count by Alcohol

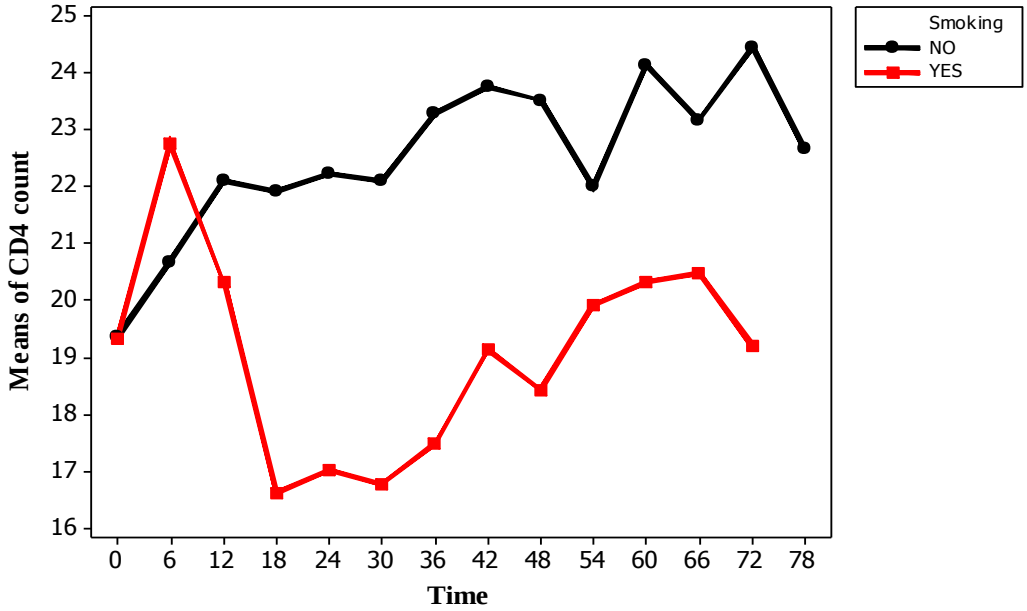


Figure A4: Profile plot of CD4 count by smoking

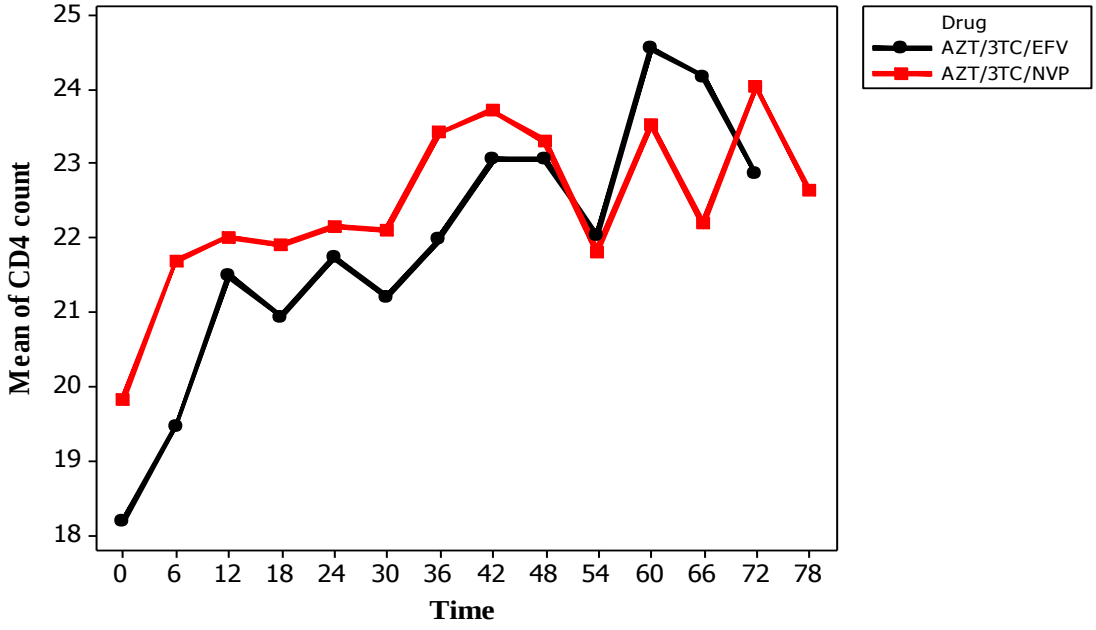


Figure A5: Profile plot of CD4 count by Drug

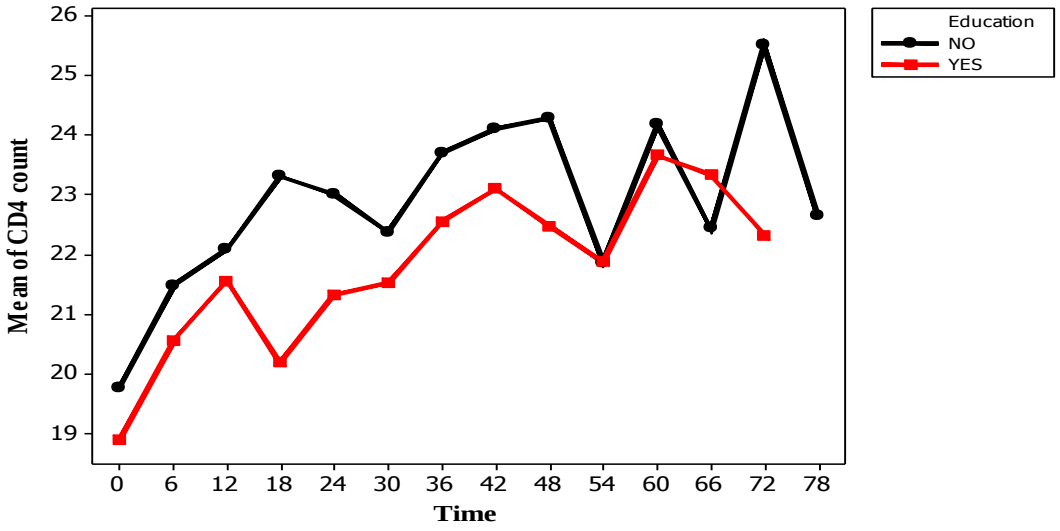


Figure A6: Profile of CD4 count by Education

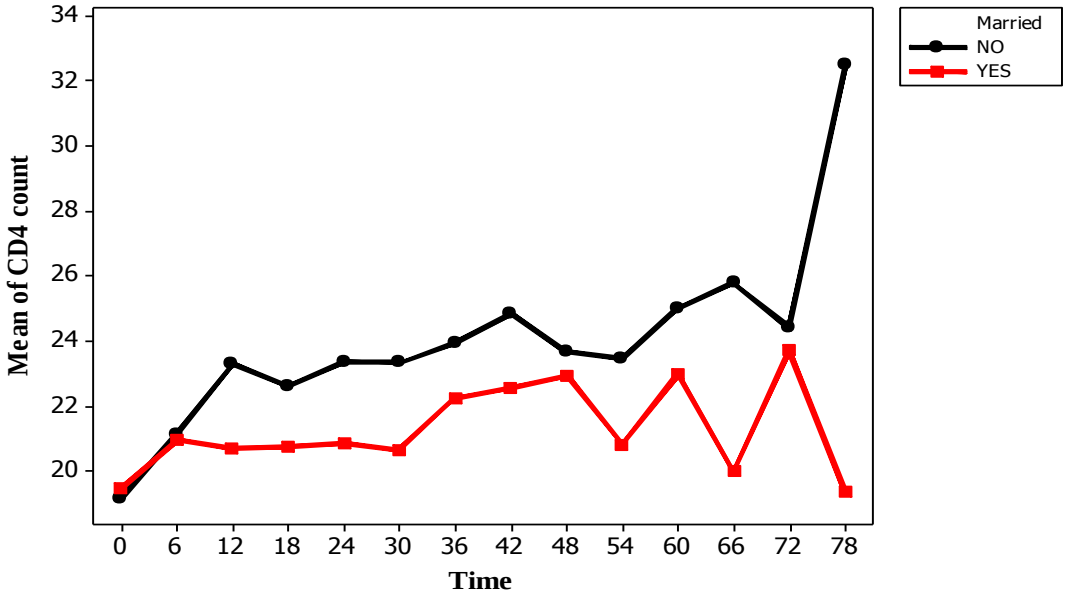


Figure A7: Profile plot of CD4 count by Married

.