SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

Application of Longitudinal Count Data Models to Progression of CD4 Count: A Case of Debre Markos Referral Hospital

By: Belay Desyebelew

Advisor: Birhanu Teshome (PhD)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS (BIOSTATISTICS)

June, 2017

Addis Ababa, Ethiopia

Decelaration

I, ***Belay Desyebelew***, do hereby declare that this thesis entitled: ***"Application of Longitudinal Count Data Models to Progression of CD4 Count: A Case of Debre Markos Referral Hospital"*** is entirely my own original work and has not been presented for higher degree at any other University or Institute anywhere for that award of any academic degree, diploma or certificate. All references made to works of other persons have been duly acknowledged.

Name: Belay Desyebelew Andualem

Signature: – – – – – – – – – – – – – – – – –

Place: Faculty of Science, Addis Ababa University

Date: June, 2017

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

– – – – – – – – – – – – – – – – –

Birhanu Teshome (PhD)

Approval

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF STATISTICS

This is to certify that, the thesis work prepared by ***Belay Desyebelew***, entitled: ***"Application of Longitudinal Count Data Models to Progression of CD4 Count: A Case of Debre Markos Referral Hospital"*** was carried out under strict supervision and has been approved for submission to the Graduate Programs of Addis Ababa University in partial fulfillment of the requirements for the award of the Degree of Master of Science in Statistics (Biostatistics) assembles with the regulations of the University and meets the accepted standards with respect to originality and quality.

Approved by the board of examiners:

Birhanu Teshome (PhD)	_____	_____
Advisor	Signature	Date
Shibru Temesgen (PhD)	_____	_____
Examiner	Signature	Date
Dejen Tesfaw (PhD)	_____	_____
Examiner	Signature	Date
_____	_____	_____
Chair of Department	Signature	Date

Acknowledgement

First, and foremost, I would like to thank the almighty God for being there with me in every step of my life. Next, I would like to gratefully and sincerely thank my advisor, Dr. Birhanu Teshome, for his support, guidance, encouragement and precious comments from the beginning till the completion of this work.

My special thanks goes to ART case unit information center staffs in Debre Markos Referral Hospital for their kind cooperation in providing me all the data. I would like to give my heartfelt thanks to my family for their love, support, encouragement in every phase of my life.

Finally, I would like to gratefully thank Addis Ababa University and Dire Dawa University for the chance they gave me to pursue my masters study.

Abstract

Application of Longitudinal Count Data Models to Progression of CD4 Count: A Case of Debre Markos Referral Hospital

Belay Desyebelew

Addis Ababa University, Ethiopia, 2017

Even though the world is fighting HIV disease in unity and patients are getting antiretroviral therapy treatment, HIV disease continues to be a serious health issue for parts of the world and large number of AIDS related deaths are being registered every year. A number of studies have been conducted to assess factors related with the progression of the disease using surrogate endpoints like CD4 cell count. The main objective of this study was to make use of appropriate statistical models to analyze CD4 cell counts data and identify associated risk factors affecting the CD4 cell progression of patients under ART tra etment in Debre Markos Refferal Hospital. In this longitudinal retrospective cohort based study, data was collected from 445 HIV patients registered for ART treatment between September, 2005 and August, 2014 in the Hospital. Poisson, Poisson-Gamma, Poisson-Normal, and Poisson-Gamma-Normal models were applied to account for overdispersion and correlation in the data. Poisson-Gamma-Normal model with random intercept was selected as a best model to fit the data based on different model selection criteria. The findings of the study revealed that time in months, sex of patients, baseline WHO stage and baseline CD4 cell count were found to be significant factors for progression of HIV patients' CD4 cell count. Patients who started ART at higher baseline CD4 counts evolved higher than those who started at lower CD4 counts. Therefore, patients should start ART treatment early to increase their CD4 cell count progression.

Keywords: CD4 count, Longitudinal data analysis, Poisson-Normal Model, Poisson-Gamma-Normal model, Antiretroviral therapy (ART), HIV/AIDS

Contents

Acronyms

AIDS	Acquired immune deficiency syndrome
AIC	Akaike information criterion
ART	Antiretroviral therapy
BIC	Bayesian information criterion
CD4	Cluster differentiation 4
CDC	Centers for diseases control and prevention
HIV	Human immunodeficiency virus
UNAIDS	Joint United Nations programme on HIV/AIDS
WHO	World health organization

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of the Study

HIV stands for human immunodeficiency virus and it gradually attacks the immune system, which is our body's natural defense against illness. The virus destroys a type of white blood cell called a T-helper cell and makes copies of it inside them. T-helper cells are also referred to as CD4 cells. CD4 count refers to the number of T-helper cells in a cubic millimeter of blood. When a persons CD4 count drops below 200 cells per milliliter of blood, they are said to have AIDS. If HIV left untreated, it can take around 10 to 15 years for AIDS to develop, which is when HIV severely damage the immune system. AIDS stands for acquired immune deficiency syndrome and is a syndrome caused by the HIV virus. It is when a persons immune system is too weak to fight off many infections, and develops when the HIV infection is very advanced. It is the last stage of HIV infection where the body can no longer defend itself. Currently, there is no cure for HIV or AIDS. However, with the right treatment and support, people can live longer and healthy lives with HIV. To do this, it is especially important to take treatment correctly and deal with any possible side-effects (AVERT 2016).

HIV disease continues to be a serious health issue for parts of the world. According to UNAIDS fact sheet 2016, there were about 2.1 million new cases of HIV in 2015 worldwide. About 36.7 million people were living with HIV around the world, and as of June 2016, 18.2 million people living with HIV were receiving medicines to treat HIV, called antiretroviral therapy (ART). In 2015, around 46% of all people living with HIV had access to treatment. An estimated 1.1 million people died from AIDS-related illnesses in 2015 and 35 million people have died from AIDS-related illnesses since the start of the epidemic (UNAIDS, 2016).

As can be seen from 2016 prevention gap report of UNAIDS, although Eastern and Southern Africa has only 6.2% of the world's population; it is home to half of the worlds people living with HIV. The region continued to be the hardest hit by the HIV epidemic, with 46% of the world's new HIV infections in 2015. Nearly 40% of new HIV infections in the region in 2015 were in South Africa, and another 50% occurred in eight countries: Ethiopia, Kenya, Malawi,

Mozambique, Uganda, the United Republic of Tanzania, Zambia and Zimbabwe.

According to the 2016 UNAIDS report, there were 19.0 million people living with HIV in 2015 in Eastern and Southern Africa and there were about 960,000 new HIV infections. In 2015, 10,252,400 people living with HIV were on antiretroviral treatment and 470,000 AIDS-related deaths were registered (UNAIDS, 2016).

It can be seen from the report that even though there were many people on antiretroviral treatment, there were also many AIDS-related deaths registered in 2015. This implies that further studies are needed to evaluate and identify factors related with poor response to the treatment. In this study, CD4 counts of HIV-infected patients undergoing ART treatment in Debre Markos referral Hospital were used to evaluate the disease response to the treatment over time and to identify factors that might affect CD4 count of HV-infected people who are on ART. The study employed generalized Poisson mixed and other models to fit CD4 count of HIV-infected people.

1.2 Statement of the Research Problem

Even though the number of people living with HIV on antiretroviral treatment is increasing from year to year, there are a number of AIDS-related deaths registered every year. This indicates that there are other factors affecting the progression of the disease and the survival status of HIV positive patients as well. Therefore, it is important to asses other factors those affect the effectiveness of the treatment over time. This can be done by measuring viral load and/or CD4 cell count over time. Most researches conducted in relation to HIV-infected patients are on the survival time of the patients after the initiation of ART (Chakravarty *et al*, 2014; Bhatta *et al*, 2013; Jemal *et al*, 2014; Ketema K. and Eshetu W., 2012).

Longitudinal data analysis should be used to analyze the progression of CD4 count over time of HIV-infected patients under ART treatment as there is correlation among CD4 counts of an individual measured at different time points. Although CD4 count is count data, some researchers considered the normal approximation and modeled CD4 count using linear mixed effects model (Zhou *et al*, 2010, Adams M. *et al*, 2013, Lubyayi *et al*, 2015; Lemma, 2016).

The distributions of count variables such as CD4 count are typically highly skewed. For these reasons, it may be inappropriate to use models that assume normally distributed errors (Allison, 2005). Analyzing count data as continuous normal leads to incorrect test results as count distribution is too skewed to satisfy normality, and normal model does not necessarily prevent negative estimated counts (Hedeker and Gibbons, 2006). Modeling such data using linear mixed models might give us inefficient estimates. So, we got it is better to use the natural and most appropriate longitudinal count models in order to find efficient and unbiased estimates in modeling CD4 count of HIV-infected patients who are under the treatment of antiretroviral therapy.

There were researches conducted in relation with survival time of the patients after the initiation of ART and level of ART adherence in Debre Markos Referral Hospital (Nurilgn *et al*, 2014; Mulugeta *et al*, 2014). But there was no enough evidence that researches have been conducted on the progression of CD4 counts of HIV-infected patients who are under the treatment of antiretroviral therapy in the Hospital. Therefore, this study examined the progression of the disease over time in HIV positive patients who are under ART treatment in the Hospital.

1.3 Objectives of the Study

1.3.1 General Objective

The main objective of the study is to fit and identify risk factors of CD4 count progression of HIV-infected patients who are under ART treatment in Debre Markos Referral Hospital.

1.3.2 Specific Objectives

The specific objectives that are planned to be addressed in this study are the following:

- To examine progression of CD4 count of HIV patients.
- To select appropriate model that can fit CD4 cell count of HIV patients.
- To identify factors that affect CD4 count of HIV patients.

1.4 Significance of the study

In practice, it is common to have count outcome variables in longitudinal data analysis. CD4 count is a good example of such outcome variables. Many researchers analysed progression of CD4 count over time using linear mixed model. But, using models which assume normal (multivariate) distribution for the error terms may give inappropriate estimates because the distribution of count variables is highly skewed.

This study used the natural and most appropriate models to study CD4 count over time of HIV-infected patients under ART in which most efficient estimates can be obtained. Therefore, the importance of this study is having a most efficient statistical model to fit CD4 count of HIV-infected patients under ART.

The scientific outputs obtained are important for both HIV-patients and ART service providers to give more attention and work on the factors that are responsible for change in CD4 count over time. It is important to identify the factors related to poor response of CD4 cell counts to the treatment.

Chapter 2

LITERATURE REVIEW

2.1 Role of CD4 Count In HIV Treatment

The CD4 T lymphocytes are coordinators of the body's immune response, e.g., providing help to B cells in the production of antibody, as well as unaugmenting cellular immune response to antigens. The CD or cluster of differentiation is a protein expressed on the surface of the cells of the hematopoietic system. CD4 T lymphocytes occupy the central position in regulating immune functions. CD4 T lymphocytes are the primary targets of HIV. The relentless destruction of CD4 T lymphocytes by HIV, either directly or indirectly, results in the loss of HIV-specific immune response, recall antibody response and, finally, non-specific immune response in the AIDS stage. (WHO, Regional Office for South-East Asia New Delhi, 2007).

The CD4 count measures how many CD4 cells one have in his or her blood. These are a type of white blood cell, called T-cells that move throughout ones body to find and destroy bacteria, viruses, and other invading germs. Test results help doctors to know how strong immune system of the patient is and guide HIV treatment choices. CD4 count can also be used to tell what stage of HIV or AIDS a patient has and what's likely to happen next. Keeping the patients CD4 count up can hold off symptoms and complications of HIV and help the patient live longer. A normal CD4 count is from 500 to 1,400 cells per cubic millimeter of blood. For some people, CD4 counts can drop dramatically, even going down to zero. But it is more important to pay attention to the pattern of CD4 counts than to any one test. In general, HIV infection is getting worse as CD4 count is going down. It means immune system is getting weaker and the patient is more likely to get sick. (WebMD Medical Reference Reviewed by Melinda Ratini, 2016).

According to WHO, Regional Office for South-East Asia New Delhi, (2007), within hours of exposure to HIV, CD4 T lymphocytes are found to be infected showing active viral replication. The infected CD4 cells release virions by budding through the cell membrane or by lysis of the infected cells. During the primary HIV infection, the number of CD4 T lymphocytes in the bloodstream decreases by 20% to 40%. HIV brings about the lysis of HIV infected cells as well as bystander uninfected cells using various mechanisms such as lysis of the cells infected with

HIV. Billions of CD4 T lymphocytes may be destroyed every day, eventually overwhelming the immune system's regenerative capacity. The impairment of HIV specific CD4 T lymphocytes function occurs very early in acute infection. Following acute primary HIV infection, one may remain free of HIV-related illnesses, often for years, despite ongoing replication of HIV in the lymphoid organs and relentless destruction of the immune system. However, during the period, the immune system remains sufficiently competent to provide immune surveillance and to prevent most infections. A number of assays such as cytokine induction, antigen induced proliferation, measurement of activation markers etc. can assess the functions of lymphocytes. However, the total CD4 T lymphocytes number still remains the most robust marker of immune competence.

The progressive loss of CD4 T lymphocytes eventually results in the loss of an ability to mount desirable immune response to any pathogen and vulnerability to opportunistic pathogens characteristic of AIDS. The estimation of peripheral CD4 T lymphocytes counts is relied upon for taking a decision on initiation of ART. The estimation of peripheral CD4 T lymphocytes counts has also been used as a tool for monitoring disease progression and the effectiveness of antiretroviral treatment (ART). The changes in the CD4 T lymphocytes counts are important indicators of the response to ART. HIV plasma virus load is a sensitive indicator of the progression of HIV disease. However, due to the relatively high cost of virus load estimation, the CD4 T lymphocytes count remains the most important key indicator for initiation and monitoring of ART and a measure of the effectiveness of the treatment in clinical trial evaluations. (WHO, Regional Office for South-East Asia New Delhi, 2007).

2.2 Review of Related Literatures

The scale-up of treatment is among the greatest successes of the global AIDS response to date. In the past two years the number of people living with HIV on antiretroviral therapy has increased by about a third, reaching 17.0 million people. These gains are largely responsible for a 26% decline in AIDS-related deaths globally since 2010, from an estimated 1.5 million in 2010 to 1.1 million in 2015. In the worlds most affected region, Eastern and Southern Africa, the number of people on treatment has more than doubled since 2010, reaching nearly 10.3 million people, and AIDS-related deaths have decreased by 36% since 2010 (UNAIDS, 2016).

The monitoring of CD4+ cell counts are a basis for assessing the effectiveness of most HIV treatments. Understanding the way CD4+ cells change over time could provide insight into the way Patients respond to treatment and how effective treatment is with time (Adams M., Luguterah A., 2013). These authors selected 139 HIV-infected patients enrolled for ART at HIV/AIDS monitoring program of the Builsa District hospital in the Upper East Region of Ghana. The CD4+ cell counts of the patients were initially taken on enrolment into the ART program and thereafter, every six months between January, 2008 and December, 2012. They modeled the change in CD4+ cell count over time using linear mixed model and obtained that the initial CD4+ cell count, age, gender and duration of treatment (in months) were significant determinants of the CD4+ cell counts of a patients on ART.

Jennifer *et al* 2010 stated that the extent of recovery of CD4 cells, once the patient has been placed on HAART, appears to be important predictor of treatment success; patients who achieve close to normal values could potentially have a normal lifespan. The study investigated that the baseline CD4 count was a significant predictor for HIV disease progression, survival and treatment outcome. Patients with lower CD4 counts are at risk for both AIDS and non-AIDS-related events. Serious non-AIDS-related diseases, such as liver, cardiovascular, renal and non-AIDS malignancies had contributed to the majority of morbidity and mortality among HIV-infected patients who are stable on HAART. Higher CD4 counts have already been shown to reduce these rates. (Jennifer *et al*, 2010).

A cohort study on factors associated with CD4 variations over time conducted by Montarroyos *et al*, (2014) on 1,870 HIV patients who were under ART treatment from July 2007 to December 2010 in the Correia Picanco and Oswaldo Cruz University Hospitals, State of Pernambuco, Brazil. The authors employed a multilevel modeling and identified age, smoking, hospital and changing doctors during outpatient follow up and use of drug were found to be significant factors for change in CD4 count (Montarroyos *et al*, 2014).

Grover *et al* (2015) applied genaralized poisson regression model to assess factors affecting the improvement of CD4 count of HIV patients undergoing ART treatment. The authors performed the analysis using Poisson,Negative binomial and generalized Poisson models. Depending on

the likelihood and AIC values they identified that generalized Poisson model fits the data well. Using generalized Poisson model the significant factors determining the rate of improvement in CD4 count were found to be gender, addiction to alcohol, WHO stage, body mass index, land status, opportunistic infections, difference in weight and occupation.

In a historical cohort study, Abbastabar H. *et al* (2016) employed Poisson regression model to determine factors that affect CD4 cell count of HIV patients. These authors analyzed a data from 1,565 HIV infected individuals and obtained that gender, unemployment, past or current addiction, positive imprisonment history, joint blade usage, HIV infection via injection or sex, longer HIV durations and age were associated with CD4 cell count.

Daniele de Brito Trindade *et al*, (2015) applied Poisson and Negative Binomial models using the multilevel (ML) approach and the generalized estimations equations (GEE) to model CD4 cell counts. The data for the study referred to 587 HIV seropositive patients in the city of Salvador, who were registered at the Laboratory Testing Control System (SISCEL, in Portuguese) of the Brazilian Ministry of Health between January 2002 and August 2012. Statistical modeling considered the number of CD4+ cells as the outcome, which was measured at different points in time after receiving the antiretroviral therapy. After analyzing the data they stated that the best marginal model to fit the data was the negative binomial (NB) with exchangeable correlation structure. Overdispersion was detected in the data, indicating NB as the most appropriate model.

Quasi-Poisson and negative binomial regression models were also applied to identify factors affecting initial CD4 cell count change due to antiretroviral therapy administered to HIV positive adults in NorthWest Ethiopia (Amhara region) by Awoke *et al*, (2016). The study was conducted among 792 HIV positive adult patients who already started antiretroviral therapy for one month of therapy. Depending on the information criteria (AIC and BIC) they declared that Quasi-Poisson model was preferable. Hence, parameter estimation and identification of predictors of initial CD4 cell were conducted using the Quasi-Poisson model. The authors obtained that adherence to antiretroviral therapy; initial CD4 cell count, household income, WHO stages, age, weight and owner of cell phone played a major role for the variation of CD4 cell count in their data.

Getachew *et al*, (2016) employed Poisson, Poisson-Gamma, Poisson-Normal, and Poisson-Normal-Gamma models to study CD4 count data collected from 222 HIV positive patients who were 15 years old and treated with ART drugs from September 2011 to May 2014 at Hossana District Queen Elleni Mohammed Memorial Hospital. A model comparison was made between the models and the Poisson-Normal-Gamma was chosen to be the best model for CD4+ cell counts data. In the study time since month of seroconversion, sex of the patient and age of patient were found to be potential risk factors for the change in the CD4+ cell counts of ART patients at HAART.

Chapter 3

DATA AND METHODOLOGY

3.1 Source of Data

The data of this study was collected from registered documents (patients' cards) of 445 HIV-positive patients who started antiretroviral treatment between September 30,2005 and August 21,2014 at ART program unit of Debre Markos referral Hospital. The CD4 count of the patients was collected at the initiation of the treatment and at different time points after the start of the treatment. Data on demographic and clinical characteristics of patients was collected as well.

3.2 Scope of the Study

The population of interest of this study was HIV-infected patients who were under antiretroviral therapy treatment in Debre Markos Referral Hospital that is located in the capital city of East Gojam zone, Debre Markos. Debre markos is located in Northwest of Ethiopia 299 kilo meters far from Addis Ababa.

3.3 Variables in the Study

3.3.1 Dependent Variable

The dependent variable of this study was the CD4 cell count per cubic millimeter of blood of HIV-infected patients who are under ART treatment.

3.3.2 Independent Variables

The independent variables considered in this study were selected based on related literatures of this area (Getachew T. *et al*, 2016; Awoke *et al*, 2016 and Adams M., Luguterah A., 2013).

- Sex of patients (male, female)
- Age of patients (age at the initiation of the treatment)
- Baseline CD4 count (the CD4 count of the patients at the start of the treatment)

- WHO clinical stage at baseline (stage I, stage II, stage III, stage IV)
- Marital status at baseline (never married, married,Divorced and window)
- Baseline weight
- Level of education at baseline (no education, primary, secondary and tertiary)
- Functional status at baseline(working, ambulatory and bedridden)
- TB status at baseline (negative and positive)
- Time in months

3.4 Methodology

3.4.1 Longitudinal Count Data Analysis

Longitudinal data is universal in a wide range of fields: medicine, public health, education, business, economics, psychology, biology and more. Longitudinal data analysis is a statistical analysis method in which the variable of interest (the dependent variable) is measured repeatedly over time and the repeated measurements taken from a subject over time are correlated. The distribution of the response variable in longitudinal data analysis may be Gaussian or non-Gaussian. The Gaussian longitudinal data are often analyzed by linear mixed model (Verbeke and Molenberghs, 2000) and generalized linear mixed model is the most frequently used mixed effects model in the context of discrete or non-Gaussian longitudinal data.

In practice, it is common to face response variables of count type like number of CD4 cells in a cubic milliliter of blood. Some data analysts treat CD4 cell count as continuous measure and apply linear mixed effects model. But that practice ignores two facts: the data are really discrete, and the distributions of count variables are typically highly skewed. For these reasons, it may be inappropriate to use models that assume normally (multivariate) distributed errors (Allison, 2005). In this study, a Poisson regression model with normal random effects and a model that accounts for both correlation between repeated measures and overdispersion simultaneously, combined (Poisson-Gamma-Normal) model were considered in line with Booth *et al* (2003) and Molenberghs *et al*, 2007 and 2010.

3.4.2 Data Exploration

Before directly modeling a given data, it is important to make exploratory analysis to observe the structure and pattern of the data. Data exploration gives some direction to select the appropriate statistical model for a given data. In this study, individual profile plot, mean profile plot and variance structure were considered as parts of data exploration.

Individual profile plots: It is a plot of the values of the outcome variable for each individual at different time points versus the time points at which the outcome variable is measured. It is useful to observe subject specific evolution over time and to decide on the random effects to be included in the model.

Average or mean profile plot: It is a plot of the average value of the outcome variable at each point of time versus the time points. It describes how the profile for a number of relevant subpopulations (the population as a whole) evolves over time. It is important to choose a fixed effects structure for the model to be considered.

Variance Structure: The evolution of the variance is important to build an appropriate longitudinal model.

3.4.3 Statistical modeling

The data of this study is longitudinal count data. As a result, subject specific (Poisson-Normal and Poisson-Gamma-Normal) models which includes subject specific random effects were employed in the study by starting from standard count data models(Poisson and Poisson-Gamma models).

Poisson-Normal-Model

The generalized linear mixed model is the most frequently used random-effects model for non-Gaussian repeated measurements. It is straight-forward extension of the generalized linear model by addition of random effects in the model.

In generalized linear mixed models, conditionally on q-dimensional random effects b_i, the outcomes Y_{ij} are assumed to be independent with exponential-family densities of the form:

$$f_{ij}(y_{ij}/b_i, \beta, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \Psi(\theta_{ij})] + C(y_{ij}, \phi)\} \tag{3.1}$$

where, the random effects b_i, often assumed to be drawn independently from the N(0, D).
D is the variance-covariance matrix of the random effects.
θ_{ij}= natural or canonical parameter
$\Psi(.)$ and C(.) are known functions.
Y_{ij}= the value of the outcome variable (CD4 count) for i^{th} individual at j^{th} time point.
β= a p-dimensional vector of unknown fixed regression coefficients.
b_i = a q-dimensional vector of unknown random regression coefficients for the i^{th}individual.
ϕ= a scale parameter

For μ_{ij} =E($Y_{ij}|b_i$) and known link function $\eta(.)$, the generalized linear mixed model can be expressed as:

$$\eta(\mu_{ij}) = \eta[E(Y_{ij}|b_i)] = X_{ij}^T\beta + Z_{ij}^T b_i \tag{3.2}$$

where, X_{ij} and Z_{ij} are p-dimensional and q-dimensional vectors of known covariate values, respectively (Molenberghs et al, 2007).

For the case of CD4 counts which is Poisson data (i.e. Y_{ij} ~Poisson(λ_{ij})), the generalized mixed Poisson model with normal random effects (Poisson-normal model) becomes:

$$\ln(\lambda_{ij}) = X_{ij}^T\beta + Z_{ij}^T bi \tag{3.3}$$

This model is said to be Poisson-Normal model because it assumes Poisson distribution for the CD4 counts and normal distribution for the random effects b_i.

Overdispersion

Overdispersion in Poisson models occurs when the response variance is greater than the mean. Overdispersion is caused by correlation between individual responses. Overdispersion also arises when there are violations in the distributional assumptions of the data, such as when the data

are clustered and thereby violate the likelihood independence of observations assumption. A model may be overdispersed if the value of the Pearson χ^2 statistic divided by the degrees of freedom is greater than 1.0 (Hilbe, 2011).

It is clear that the Poisson distribution forces equality between mean and variance. However, comparing the sample average with the sample variance might reveal that this assumption is not true for a particular set of data. Likelihood Ratio Test can also be used to test overdispersion in count data.

One way to account for overdispersion in count data is through a two-stage approach. A commonly encountered instance is by assuming that $Y_i|\theta_i \sim Poi(\theta_i\lambda_i)$. Where θ_i denote an independent and identically distributed (iid) sample of unit mean gamma random variables with shape parameter α (Booth *et al* (2003)). Conditional on θ_i, the CD4 count of the i^{th} patient, Y_i has a Poisson distribution with mean $\theta_i\lambda_i$.

The counts are then marginally independent negative binomial random variables (Y_i ~nb(α, λ_i)) with mean λ_i and variance $\lambda_i + \lambda_i^2/\alpha$. Hence, the parameter α quantifies the amount of overdispersion with $\alpha = \infty$ corresponding to no overdispersion with respect to the Poisson distribution. The mass function of the negative binomial random variables is given by

$$Pr(Y_i = y; \alpha, \lambda_i) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}(\frac{\alpha}{\lambda_i+\alpha})^{\alpha}(\frac{\lambda_i}{\lambda_i+\alpha})^{y} \tag{3.4}$$

The negative binomial model is given by $\log(\lambda_i) = X_i^T\beta$.

Poisson-Gamma-Normal Model

According to Molenberghs *et al*, (2007) and (2010) a model combining the ideas from the Poisson-Normal and overdispersion models for repeated Poisson data with overdispersion can be specified as follows $Y_{ij} \sim poi(\theta_{ij}\lambda_{ij})$

$$\lambda_{ij} = \exp(X_{ij}^T\beta + Z_{ij}^T b_i) \tag{3.5}$$

where θ_{ij} capture overdispersion and denote an independent and identically distributed (iid) sample of unit mean gamma random variables with shape parameter α and scale parameter β=$1/\alpha$.

$b_i \sim N(0, D)$ and $\theta_{ij} \sim Gamma(\alpha, \beta)$

This model is called Poisson-Gamma-Normal (combined) model because it includes both normal (b_i) and gamma (θ_{ij}) random effects to account for correlation and overdispersion, respectively.

3.4.4 Working Correlation Structures

In modeling longitudinal data it is important to specify the structure of correlation between the repeated measures of a subject. The most commonly used correlation structures include independence, exchangeable (compound symmetry), unstructured and autoregressive (Hilbe, 2011).

Independence correlation structure (IND): In this correlation structure observations are considered to be independent of one another. The structure assumes a zero correlation between subsequent measures of a subject within time points. This structure is useful if the size of panels is small and if there is evidently no time effect in the data. The scheme of this correlation structure can be shown as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Exchangeable correlation structure (Exch): The exchangeable correlation structure is the most commonly used structure. It is the default for several of the major commercial software implementations. The exchangeable correlation structure assumes that the correlations between measurements within time are the same, irrespective of any time interval. Any correlation value within the structure may be exchanged with any other hence the name exchangeable. The structure is as follows::

$$\begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

Unstructured correlation structure (UN): In the unstructured correlation structure all correlations are assumed to be different; correlations are freely estimated from the data. For our data with 7 time points, its structure is as follows:

$$\begin{pmatrix} 1 & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ C_1 & 1 & C_7 & C_8 & C_9 & C_{10} & C_{11} \\ C_2 & C_7 & 1 & C_{12} & C_{13} & C_{14} & C_{15} \\ C_3 & C_8 & C_{12} & 1 & C_{16} & C_{17} & C_{18} \\ C_4 & C_9 & C_{13} & C_{16} & 1 & C_{19} & C_{20} \\ C_5 & C_{10} & C_{14} & C_{17} & C_{19} & 1 & C_{21} \\ C_6 & C_{11} & C_{15} & C_{18} & C_{20} & C_{21} & 1 \end{pmatrix}$$

Autoregressive correlation structure (AR): The autoregressive (AR) correlation structure assumes that there is a marked decrease in correlation coefficient values with the corresponding increase in measurements within panel time intervals. Each off-diagonal from the main diagonal decreases by the square of the previous diagonal. For our data with 7 time points, the correlation structure can be expressed as:

$$\begin{pmatrix} 1 & C & C^2 & C^3 & C^4 & C^5 & C^6 \\ C & 1 & C & C^2 & C^3 & C^4 & C^5 \\ C^2 & C & 1 & C & C^2 & C^3 & C^4 \\ C^3 & C^2 & C & 1 & C & C^2 & C^3 \\ C^4 & C^3 & C^2 & C & 1 & C & C^2 \\ C^5 & C^4 & C^3 & C^2 & C & 1 & C \\ C^6 & C^5 & C^4 & C^3 & C^2 & C & 1 \end{pmatrix}$$

3.4.5 Methods of Parameter Estimation

Generalized linear mixed models (GLMMs) are a natural outgrowth of both linear mixed models and generalized linear models. GLMM's enable the accommodation of non-normally distributed responses and specification of a possibly nonlinear link between the mean of the response and the predictors, and they can model overdispersion and correlation by incorporating random effects. Although there were computational problems to numerically evaluate the high-dimensional integrals, maximum likelihood (ML) estimation was used in this study.

Random-effects models can be fitted by maximization of the marginal likelihood, obtained by integrating out the random effects from conditional densities of the form:

$$f_i(y_{ij}|b_i, \beta, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \Psi(\theta_{ij})] + C(y_{ij}, \phi)\}$$

Therefore, as Poisson distribution is a member of exponential families, the likelihood contribution of patient i for the Poisson-Normal model is:

$$f_i(y_i|\beta, D, \phi) = \int \prod f_{ij}(Y_{ij}|b_i, \beta, \phi) f(b_i|D) db_i \tag{3.6}$$

Where D is the variance-covariance matrix of the random effects.
This implies that the likelihood function can be given as

$$L(\beta, D, \phi) = \prod f_i(y_i|\beta, D, \phi) = \prod \int \prod f_{ij}(Y_{ij}|b_i, \beta, \phi) f(b_i|D) db_i \tag{3.7}$$

For the Poisson-Gamma-Normal or combined model the likelihood contribution of patient i and the likelihood function can be given respectively as

$$f_i(y_i|\beta, D, \alpha, \beta^*) = \int \prod f_{ij}(Y_{ij}|\beta, b_i, \theta_i) f(b_i|D) f(\theta_i|\alpha, \beta^*) db_i d\theta_i \tag{3.8}$$

$$L(\beta, D, \alpha, \beta^*) = \prod f_i(y_i|\beta, D, \alpha, \beta^*) = \prod \int \prod f_{ij}(Y_{ij}|\beta, b_i, \theta_i) f(b_i|D) f(\theta_i|\alpha, \beta^*) db_i d\theta_i \tag{3.9}$$

where β is a p-dimensional vector of unknown fixed regression coefficients and β^* is the scale parameter of the gamma distribution. The key problem in maximizing both likelihood functions (3.7 & 3.9) is the presence of n integrals over the random effects b_i and θ_i. To overcome this problem, different approximation methods to maximize the likelihood function have been proposed.

These include numerical integration and series expansion methods, including penalized quasi-likelihood and marginal quasi-likelihood, Laplace approximation, adaptive Gaussian quadrature approximation, etc. In this study we used glmer and glmer.nb R functions under packages MASS and lme4 with Laplace and adaptive Gaussian quadrature approximation methods.

3.4.6 Model and Variable Selection

For all models, to select the important variables, first the main effect, main effect by time interaction and main effect by main effect interactions were incorporated to the initial candidate models and, then the highly non-significant interaction effects were removed and the models were refitted again and so on. i.e. Unautomated back ward selection technique was employed to select significant factors to be included in the final model.

The best model that can fit the data was selected depending on different information criteria (AIC, BIC and -2loglikelihood). The model with smallest values of information criteria was selected as the best model to fit the data well.

Chapter 4

RESULTS AND DISCUSSION

4.1 Descriptive Analysis

In this section, data obtained from 445 HIV positive patients who were under ART treatment in Debre Markos Referral Hospital is summarized. These patients were enroled for antiretroviral treatment between September 30, 2005 to August 21, 2014 at ART program unit of Debre Markos referral Hospital. The data in this study indicated that majority of HIV patients (347 (77.98%)) started antiretroviral treatment with CD4 cell counts < 200 cells/mm^3. CD4 counts were taken from each patient in six month interval starting from baseline to the 36^{th} month. The minimum number of observations per subject was two and maximum of seven. Among the 445 patients 280 (62.9%) were females and the remaining 165 (37.1%) were males.

The summary of CD4 count at different time points is given in Table 4.1. As can be seen in Table 4.1, the mean CD4 count increased over time until the 24^{th} month, decreased at the 30^{th} month and then starts to increase after 30^{th} month. The same is true for the standard deviation of CD4 count. It increased until the 24^{th} month and comes down at month 30; and, starts to increase after month 30. The number of patients decreased at some point and increased at another which indicates that there is intermittent missingness in the data.

The average CD4 count of the patients at the start of the treatment was 155.3 CD4 cells/mm^3 of blood with standard variation of 102.25 CD4 cells/mm^3 of blood. The minimum and maximum baseline CD4 counts were three and 971 CD4 cells/mm^3 of blood, respectively.

Table 4.1: Summary of CD4 Count at Different Time Points

Time	0	6	12	18	24	30	36
n	445	372	320	283	271	279	261
Mean	156.28	294.05	323.58	356.37	385.59	376.78	398.08
Std	106.20	164.70	167.09	197.70	205.09	182.84	187.09

Data on demographic and clinical characteristics of the patients was collected at the start of antiretroviral treatment. The data of the continuous covariates included in this study (Age and

Weight) is summarized in Table 4.2. This study considered patients with age greater than 15 years of age. The mean baseline age of patients is 35.18 years with minimum and maximum age of 17 and 68 years, respectively. The average baseline age of female patients was 32.94 years and that of male patients was 38.74.

The minimum and maximum weight of the patients were 30 and 82 kilograms, respectively with mean baseline weight of 52.02 kilograms.

Table 4.2: Summary of Continuous Covariates at Baseline

Variable	Min	1^{st} Qu.	Median	Mean	$3^{rd}Qu.$	Max	Std
Age	17	28	33	35.18	42	68	10.5
Weight	30	45	52	52.02	58	82	9.26

As presented in Table 4.3, among the 445 patients included in this study, only 165 were males and 280 were females. The male patients have 134.84 mean baseline CD4 cell counts and female patients have 168.91 mean baseline CD4 count. On average, female patients started ART treatment at higher CD4 cell counts as compared with male patients.

The difference in mean CD4 cell count of the two groups increases as time increases. That means the average CD4 cell count of female 's seems higher than males at all time points and the difference increases over time.
Table 4.3 also showed that the average variability in CD4 cell count of females is higher than the average variability of males at all time points.

WHO stage III has high number of patients (282 (63.37%)) as compared with the other three stages. WHO stage II takes the second place in number of patients (77 (17.30%)) and WHO stage IV has the smallest number of patients (27 (6.07%)). As expected, patients with WHO stage 1 have higher CD4 cell counts at all time points as compared with patients of the other three stages of the disease.

Patients with working functional status have higher mean CD4 count at all time points than that of patients with ambulatory functional status. Among the 445 HIV patients included in this study, 346 (77.75%) were patients with working functional status and 99 (22.25%) were

Table 4.3: Summary of CD4 cell Progress for categorical Covariates

Covariates	Time	0	6	12	18	24	30	36
Sex								
Male	n	165	131	117	100	97	106	106
	Mean	134.84	263.29	278.22	310.07	322.27	324.72	339.45
	Std	91.56	147.75	137.98	157.59	172.40	151.31	168.81
Female	n	280	241	203	183	174	173	155
	Mean	168.91	310.76	349.72	381.68	420.90	408.68	438.05
	Std	112.19	171.20	176.85	212.68	213.66	193.30	188.93
WHO Stage								
Stage I	n	59	51	41	36	38	35	36
	Mean	230.54	378.25	390.24	505.97	462.39	398.23	475.39
	Std	159.03	211.53	188.69	282.77	212.41	196.87	205.71
Stage II	n	77	65	54	52	47	48	42
	Mean	139.39	252.88	298.06	285.42	319.51	366.50	338.83
	Std	79.22	136.39	165.44	147.63	173.05	169.14	155.14
Stage III	n	282	235	208	175	171	173	165
	Mean	145.39	285.81	320.65	346.84	384.11	376.36	394.65
	Std	85.87	153.787	160.36	177.60	203.85	185.30	184.775
Stage IV	n	27	21	17	20	15	23	18
	Mean	155.89	308.09	282.00	354.62	415.07	368.70	411.21
	Std	153.24	173.74	168.69	169.15	239.44	178.80	197.22
Functional Status								
Working	n	346	291	256	214	210	220	211
	Mean	168.32	304.06	330.68	371.34	398.90	386.94	401.08
	Std	106.02	167.79	168.80	205.36	198.21	182.26	186.72
Ambulatory	n	99	81	64	69	61	59	50
	Mean	114.19	258.06	295.20	309.97	339.80	338.90	385.06
	Std	96.02	148.55	158.15	164.62	222.88	181.56	190.01

ambulatory (Table 4.3).

Summary of patients' CD4 cell count progress for the other categorical covariates is available in the Appendix (Tables A2, A3 and A4). About 82.70% of the 445 HIV patients were TB negative and 27.30% were positive at baseline. Patients who were TB negative at baseline have higher mean CD4 cell count at all time points than patients with Positive TB status (Table A2).

4.2 Exploratory Data Analysis

4.2.1 Exploring Individual Profiles

Figure 4.1 depicts the individual profile plot of CD4 count of HIV infected patients included in this study. It is the plot of CD4 count of each patient over time.The plot provides some information on between patients CD4+ count variability and illustrate that there is change in patients' CD4+ count over time.

Some individuals have erratic CD4 count and others have a slowly increasing CD4 count over time. As one could easily see from the graph, there is considerably large difference in the intercepts of individual trajectories.

Similarly, some trajectories were steeper while others were almost horizontal, indicating the possible variability in the slope of CD4 counts. Therefore, because of the variability in the intercept and slope of trajectories, using a mixed model could fit the data very well. The individual

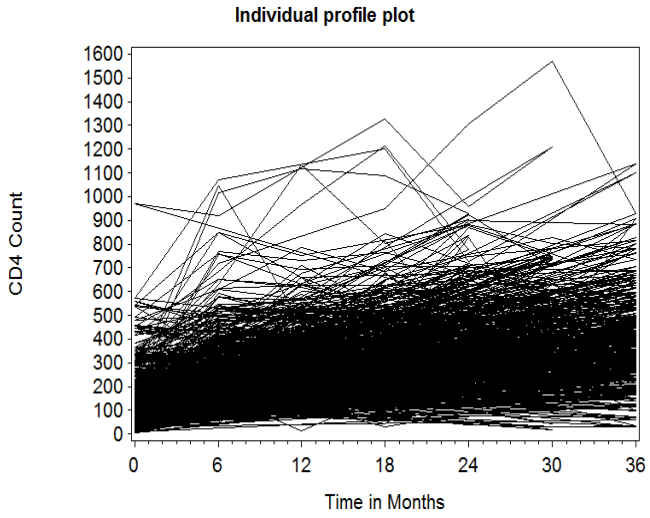

Figure 4.1: Individual Profile plot for CD4 count of patients

profile plot of CD4 count for both male and female groups is displayed in Figure 4.2. As could be seen from the Figure, it seems that there is high within variation in female patients as compared with the male patients. The between patients variation is high at the end as compared with at baseline for both groups. Some female patients have erratic evolution of CD4 count over time

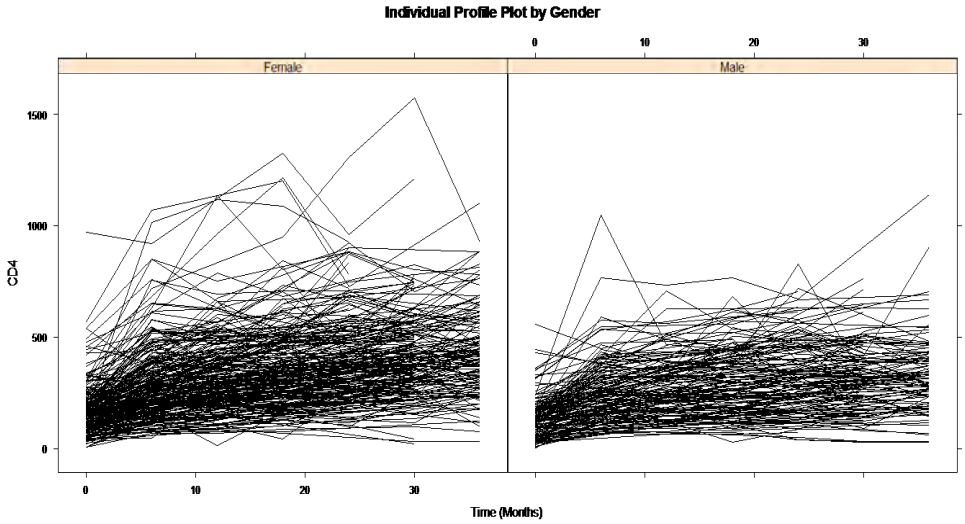

Figure 4.2: Individual Profile plot by Sex

but it seems almost regular for male patients. For the other categorical covariates, the individual profile plot of CD4 cell count is presented in the Appendix (Figures A1, A2, A3, A4 and A5).

4.2.2 Exploring Mean Profiles

The overall mean profile plot of CD4 cell counts shows somehow a linear increasing pattern of CD4 cell count over time (Figure 4.3) which suggests that a linear time effect is reasonable. The mean CD4 cell count increases in a high rate from baseline till the 6^{th} month and then it starts to increase slowly from six-24^{th} month and decreases at month 30. Then it starts to increase after the 30^{th} month.

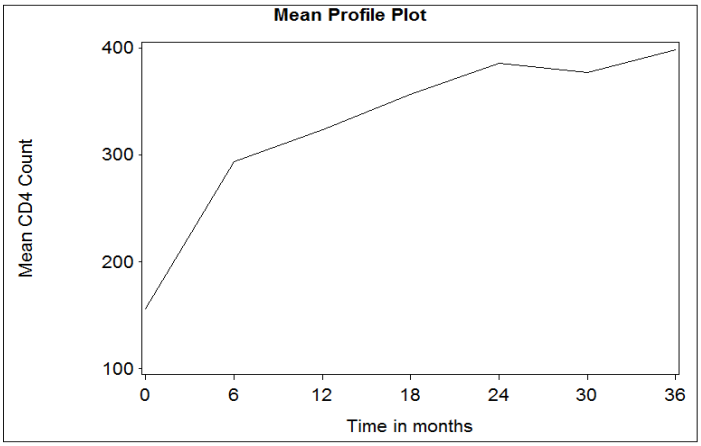

Figure 4.3: Mean Profile Plot of CD4 Count

As can be seen from Figure 4.4, the mean CD4 count profile of females is higher than that of males. Although the plot shows an increasing pattern over time for both groups, it can be observed that the mean CD4 count of females increases at some time points and show a mild decreases at another time points. It can also observed that the difference in mean CD4 count of the two groups increases after the 6^{th} month.

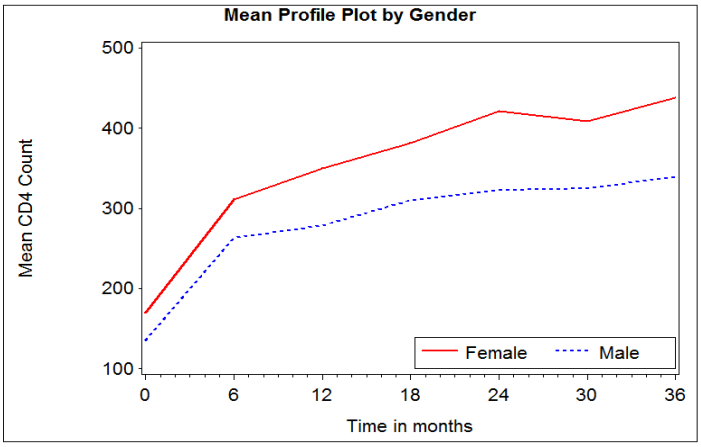

Figure 4.4: Mean Profile Plot of CD4 Count by Sex

The mean profile plots of the other three categorical covariates (WHO stage, Marital status and education level) suggests time interaction effect of covariates as the plots crossed each other (Figures A6, A8 and A9). This means we need to include two way interaction of time with these three covariates.

Patients with negative TB status at baseline have higher mean profile than patients with positive TB status at all time points implying the joint association of HIV and TB (Figure A10). This implies that baseline TB status of patients has an impact on CD4 cell count evolution.

4.2.3 Exploring Variance Function

The variance structure for CD4 count shows an irregular pattern over time (Figure 4.5). It increases at some point and decreases at another point suggesting a non-constant variance.

High variation among female patients is observed in Figure 4.6 at all times as compered with male patients. The variance of both groups increases at some point and decreases at another point which suggests as there is no constant variance.

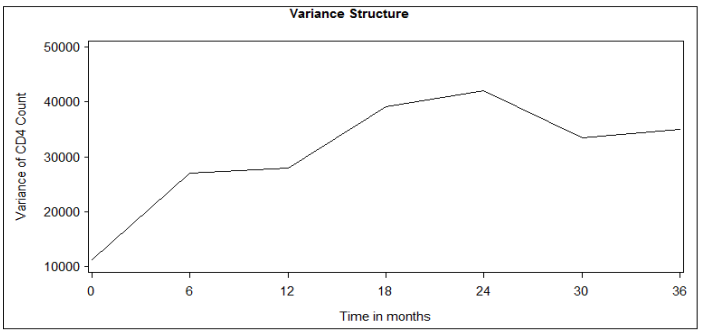

Figure 4.5: Variance Function of CD4 Count

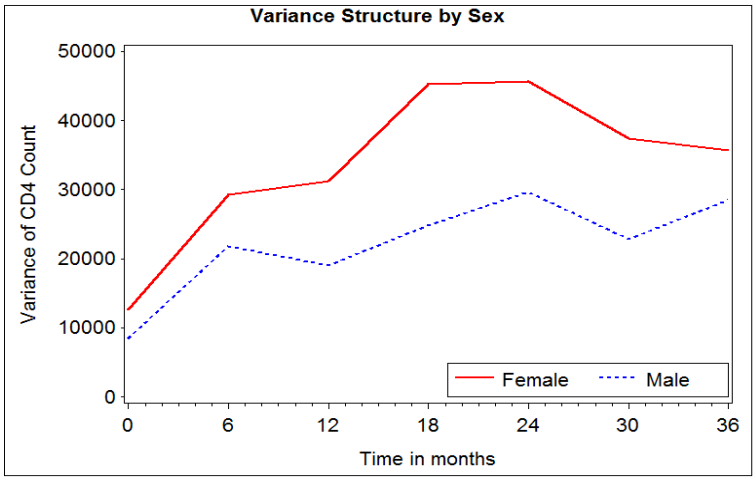

Figure 4.6: Variance Function of CD4 Count by Sex

The variance function plot of CD4 cell count for the other covariates can be seen in the Appendix (Figures A11, A12, A13, A14 and A15).

4.3 Statistical Analysis

Our data includes 445 HIV positive patients who were under the treatment antiretroviral therapy in Debre Markos Referral Hospital. The response variable of the study was CD4 cells/mm^3 per cubic millimeter of blood which is a count variable. Some researchers used linear mixed model, which requires a normality (multivariate) assumption. However, for our data we have checked

that multivariate normality assumption failed suggesting that using linear mixed model is not appropriate without data transformation. Both univariate normality tests at each time point (Table A1 in the Appendix) and multivariate normality tests (Table 4.4) revealed that the data is not normally distributed.

Table 4.4: Multivariate Normality Test

Test	Value	P value	Result
Henze-Zirkler's Test	1.099	$8.10e^{-06}$	Data are not multivariate normal
Mardia's Test			Data are not multivariate normal
Skewness(22.51)	138.842	0.0001572687	
Kurtosis(71.24)	2.231	0.0256385	
Royston's Test	25.294	0.0002289688	Data are not multivariate normal

Although the minimum observed CD4 count in our study was three, CD4 count of HIV patients can take a value of zero (Saravo1atz *et al*, 1996, L Al-Harthi et al, 2004 and Littell *et al*, 2006) which makes the use of Poisson models reasonable to analyze CD4 count data. A zero CD4 cout reveals that the immune system has been severely damaged and the patients have a very high risk from many infections. This is a sign that the immune system is being weakened and the more likely the person will get sick.

Hence, some of the candidate count data models were applied to fit CD4 count of HIV positive patients and the final model was selected using different selection criteria. The candidate models employed to fit the CD4 count of HIV patients were Poisson, Negative Binomial, Poisson-Normal and Poisson-Gamma-Normal (Negative Binomial log-linear mixed) models.

4.3.1 Poisson and Poisson-Gamma Models

Generalized linear models are used for analysing univariate non-Gaussian data. Poisson model is one of these models which is commonly used for the analysis of count data. Table 4.5 summarizes the parameter estimates of fixed effects Poisson and Negative Binomial regression models employed on CD4 cell counts. It shows that all parameters included in the Poisson regression model are significant at 5% level of significance.

The variable of interest in this study was CD4 cell count and the data is overdispersed as the sample variance of CD4 cell counts at all time points is greater than its corresponding sample means (Table 4.1). A likelihood ratio (LR) test was used to test the null hypothesis that the

Table 4.5: Poisson and Negative Binomial Models

	Poisson		Negative binomial	
Effect	Estimate (s.e)	P-value	Estimate (s.e)	P-value
Intercept	4.6438 (0.0118)*	$2e^{-16}$	4.5387 (0.0954)*	$2e^{-16}$
Time	0.0201 (0.0001)*	$2e^{-16}$	0.0248 (0.0008)*	$2e^{-16}$
Sex				
Female	0.1418 (0.0032)*	$2e^{-16}$	0.1069 (0.0260)*	$3.98e^{-05}$
MaritalStatus				
Married	0.0632 (0.0031)*	$2e^{-16}$	0.0110 (0.0255)	0.664672
Never	0.1290 (0.0047)*	$2e^{-16}$	0.0853 (0.0382)*	0.025572
Window	0.0508 (0.0039)*	$2e^{-16}$	0.0065 (0.0319)	0.839708
Level of Education				
No	0.0264 (0.0033)*	$4.6e^{-16}$	0.0157 (0.0267)	0.557679
Primary	-0.0066 (0.0034)*	0.0544	-0.0005 (0.0273)	0.985658
Tertiary	0.0565 (0.0046)*	$2e^{-16}$	0.0245 (0.0386)	0.525688
Functional Status				
Working	0.0067 (0.0033)*	0.0457	-0.0123 (0.0263)	0.639653
WHOStage				
Stage I	0.0281 (0.0046)*	$7.17e^{-10}$	0.0204 (0.0377)	0.588228
Stage III	0.1125 (0.0035)*	$2e^{-16}$	0.0989 (0.0279)*	0.000387
Stage IV	0.0257 (0.0058)*	$1.05e^{-05}$	0.0061 (0.0473)	0.896811
TB Status				
Negative	0.0477 (0.0034)*	$2e^{-16}$	0.0570 (0.0272)*	0.036378
Age	-0.0038 (0.0001)*	$2e^{-16}$	-0.0033 (0.0011)*	0.003999
Weight	0.0036 (0.0002)*	$2e^{-16}$	0.0021 (0.0013)	0.106629
BaseCD4	0.0024 (0.0000)*	$2e^{-16}$	0.0034 (0.0001)*	$2e^{-16}$
Dispersion Parameter($1/\alpha$)			4.666 (0.139)	
AIC	145,787		27,848	

*indicates parameter estimates with p-value<5%

restriction in the Poisson model is true. The test revealed that the null hypothesis is rejected with a very large Chi-Square Test Statistic = 117,940.9213 (p-value = $2.2e^{-16}$). This implies that there is overdispersion in our data and Negative Binomial model should be used to fit the data.

Table 4.5 displays the parameter estimates of the Negative Binomial model along with standard errors. It is clearly seen that there is an improvement in Negative Binomial model with Akaike information criterion (AIC) value of 27,848 as compared with the Poisson model of AIC 145,787.

4.3.2 Intraclass Correlation

The value of the intraclass correlation which is the ratio of the between-patients variance to the total variance is presented in Table 4.6. It tells us the proportion of the total variance in CD4 cell count that is accounted for by the clustering(the variance among patients). 52.65% of the the total variance in CD4 cell count is accounted by the clustering(the variance among patients). It can also be interpreted as the correlation among observations within the same patient. That

Table 4.6: Intraclass Correlation with 95% confidence interval

	95% CI					
ICC	lower	upper	N	K	varw	vara
0.5265	0.4836	0.5699	445	5.0129	17179.64	19104.06

is the correlation among observations within the same patient is 0.5265.

The intraclass correlation helps us to determine whether or not a mixed model is even necessary. The value 0.5265 of the intraclass correlation of our data tells that mixed model is necessary to fit the data. Consequently, Poisson-Normal and Poisson-Gamma-Normal models are fitted to our data in order to account for the correlation in repeated measures of CD4 count of HIV patients.

4.3.3 Poisson-Normal and Poisson-Gamma-Normal Models

The Poisson-Normal model with only random intercept, and both random intercept and slope were fitted using Laplace approximation method. Poisson-Normal model with both random intercept and slope was found to be the best fit as it has small information criteria values as compared with the one with random intercept only. The parameter estimates of this model are displayed in Table 4.7. Depending on this model time, WHO stage and initial CD4 cell count are found to be significant factors of patients' CD4 cell progression.

An improvement in both Poisson-Gamma and Poisson-Normal models as compared with the Poisson model in fitting the data is an indication of the occurrence of both correlation and overdispersion in the data. Therefore, we need a model that can handle both overdispersion and correlation simultaneously. The negative binomial log-linear mixed (Poisson-Normal-Gamma) model proposed by Booth *et al* (2003) and Molenberghs *et al* (2007 and 2010) is the one that

Table 4.7: Poisson-Normal Model

	Random Intercept only		Random intercept and Slope	
Effects	Estimate (s.e.)	P-value	Estimate (s.e.)	P-value
Intercept	4.5063 (0.1637)*	$2e^{-16}$	4.2967 (0.1514)*	$2e^{-16}$
Time	0.0211 (0.0001)*	$2e^{-16}$	0.0243 (0.0008)*	$2e^{-16}$
Sex				
Female	0.1242 (0.0451)*	0.00595	0.0767 (0.0422)	0.0692
Marital Status				
Married	0.0175 (0.0443)	0.69206	0.0204 (0.0408)	0.6165
Never	0.1004 (0.0656)	0.12599	0.0881 (0.0607)	0.1464
Window)	0.0340 (0.0554)	0.53931	0.0543 (0.0510)	0.2876
Level of Education				
No	-0.02109 (0.0463)	0.64872	0.0178 (0.0427)	0.6768
Primary	-0.0255 (0.0474)	0.59010	0.0164 (0.0437)	0.7068
Tertiary	-0.0282 (0.0663)	0.67022	0.0099 (0.0611)	0.8710
Functional Status				
Working	-0.0036 (0.0449)	0.93574	0.0456 (0.0415)	0.2721
WHOStage				
Stage I	0.0361 (0.0651)	0.57910	-0.0092(0.0600)	0.8779
Stage III	0.1194 (0.0479)*	0.01262	0.0964 (0.0442)*	0.0292
Stage IV	-0.2138 (0.0631)*	0.00070	-0.0666 (0.0622)	0.2844
TBStatus				
Negative	0.0730 (0.0470)	0.12053	0.0671 (0.0434)	0.1217
Age	-0.0027 (0.0020)	0.17255	-0.0028 (0.0018)	0.1224
Weight	0.0022 (0.0022)	0.32375	0.0031 (0.0020)	0.1354
BaseCD4	0.0033 (0.0002)*	$2e^{-16}$	0.0039 (0.0002)*	$2e^{-16}$
AIC	71,331.7		59,073.9	
BIC	71,434.5		59,188.1	
logLik	-35,647.8		-29,516.9	
Random intercept variance	0.1289		0.1203	
Random Slope Variance			0.0003	
Cov(Random Effects)			-0.32	

*indicates parameter estimates with p-value<5%

can be used to overcome this problem in correlated and overdispersed count data analysis.

As clearly shown in Table 4.8, the random intercept Poisson-Gamma-Normal Model is a much better fit because it has a much lower AIC (27,379.9), BIC (27,488.4) and -2loglikelihood (27,342) as compared with the Poisson-normal models. Therefore, the final model to fit our data is random intercept Poisson-Gamma-Normal model. This model simultaneously captures overdispersion and correlation in overdispersed longitudinal count data.

Table 4.8: Poisson-Gamma-Normal Model

	Random Intercept only		Random intercept and Slope	
Effects	Estimate (s.e.)	P-value	Estimate (s.e.)	P-value
Intercept	4.4105 (0.1555)*	$2e^{-16}$	4.3591 (0.1078)*	$2e^{-16}$
Time	0.0243 (0.0007)*	$2e^{-16}$	0.0239 (0.0010)*	$2e^{-16}$
Sex				
Female	0.1147 (0.0427)*	0.00725	0.0743 (0.0295)*	0.01171
MaritalStatus				
Married	0.0173 (0.0419)	0.67898	0.0183 (0.0289)	0.52769
Never	0.1024 (0.0623)	0.10014	0.0903 (0.0430)*	0.03583
Window	0.0406 (0.0524)	0.43900	0.0424 (0.0362)	0.24241
Level of Education				
No	-0.0012 (0.0438)	0.97866	0.0256 (0.0302)	0.39547
Primary	-0.0052 (0.0449)	0.90695	0.0188 (0.0312)	0.54747
Tertiary	0.0039 (0.0630)	0.95074	0.0004 (0.0432)	0.99277
Functional Status				
Working	0.0139 (0.0427)	0.74563	0.0282 (0.0298)	0.34484
WHO Stage				
Stage I	0.0260 (0.0616)	0.67255	-0.0055 (0.0424)	0.89762
Stage III	0.0989 (0.0455)*	0.02986	0.0870 (0.0314)*	0.00563
Stage IV	-0.0136 (0.0773)	0.86068	-0.0257 (0.0544)	0.63637
TB Status				
Negative	0.0661 (0.0446)	0.13852	0.0525 (0.0310)	0.08975
Age	-0.0027 (0.0019)	0.14401	-0.0028 (0.0013)*	0.02860
Weight	0.0023 (0.0021)	0.27205	0.0027 (0.0014)	0.06115
BaseCD4	0.0034 (0.0002)*	$2e^{-16}$	0.0039 (0.0001)*	$2e^{-16}$
Dispersion parameter($1/\alpha$)	7.7009		7.7009	
AIC	27,379.9		27,487.7	
BIC	27,488.4		27,607.6	
logLik	-13,671.0		-13,722.8	
Random intercept variance	0.08837		0.0000	
Random Slope Variance			0.00025	
Cov(Random Effects)			0	

*indicates parameter estimates with p-value<5%

4.3.4 Model Comparison

Model selection was done based on likelihood tests (Table A5 in the Appendix) and different information criteria (AIC,BIC) and -2log-likelihood. Poisson, Poisson-Gamma (Negative Binomial), Random intercept and slope Poisson-Normal, random intercept Poisson-Gamma-Normal Models were compared to select a best model that fits our data. The parameter estimates of the models are given in sections 4.3.1 and 4.3.3.

The summary of different information criteria and -2log-likelihood of these models is presented in Table 4.9.

Table 4.9: Summary of Information Criteria of Different Models

	Models			
Criteria	Poisson	Poisson-Normal	Poisson-Gamma	Poisson-Gamma-Normal
AIC	145,787	59,073.9	27,848	27,379.9
BIC	145,884	59,188.1	27,951	27,488.4
-2logLik	145,754	59,033.8	27,812	27,342

Depending on the results displayed in Table 4.9, the Poisson-Gamma-Normal (Negative Binomial log-linear mixed) model is found to be the best model to fit our data. As can be seen from the table both Poisson-Normal and Poisson-Gamma models showed good improvement in fitting the data as compared with the Poisson model.

Poisson-Gamma model has lower AIC, BIC and -2loglikelihood values as compared with both Poisson and Poisson-Normal models which is because the Poisson-Gamma model handles overdispersion in the data and fits the data well. Because good improvements were observed in both Poisson-Normal (which handles correlation in the data) and Poisson-Gamma (handles overdispersion in the data) models, we preferred to fit a model that can handle both correlation and overdispersion in the data.

In line with Booth *et al* (2003) and Molenberghs *et al* (2007 and 2010), we applied the Negative Binomial log linear mixed (Poisson-Gamma-Normal) model to our data and we found that it fits our data well with small information criteria values. Accordingly, the Poisson-Gamma-Normal model which combines both normal and gamma random effects to capture together both overdispersion and correlation was selected to improve the model fit (Table 4.10).

For the final model, to select the important variables, first the main effect and main effect by time interactions were incorporated to the initial candidate model and, then the non-significant interaction effects were removed and the model was refitted again. We checked all possible interactions and found that all were insignificant. Hence, we included only the main effects in our final model.

Based on the results obtained from the Poisson-Gamma-Normal model, time in months, sex,

Table 4.10: Poisson-Gamma-Normal Model

Random Intercept only				
		95% CI		
Effects	Estimate (s.e.)	Lower	Upper	p-value
Intercept	4.4105 (0.1555)	4.1057	4.7153	$2e^{-16}$
Time	0.0243 (0.0007)	0.0229	0.0257	$2e^{-16}$
Sex				
Female	0.1147 (0.0427)	0.0310	0.1984	0.00725
MaritalStatus				
Married	0.0173 (0.0419)	-0.0648	0.0994	0.67898
Never	0.1024 (0.0623)	-0.0197	0.2245	0.10014
Window	0.0406 (0.0524)	-0.0621	0.1433	0.43900
Level of Education				
No	-0.0012 (0.0438)	-0.0870	0.0846	0.97866
Primary	-0.0052 (0.0449)	-0.0932	0.0828	0.90695
Tertiary	0.0039 (0.0630)	-0.1196	0.1274	0.95074
Functional Status				
Working	0.0139 (0.0427)	-0.0698	0.0976	0.74563
WHO Stage				
Stage I	0.0260 (0.0616)	-0.0947	0.1467	0.67255
Stage III	0.0989 (0.0455)	0.0097	0.1881	0.02986
Stage IV	-0.0136 (0.0773)	-0.1651	0.1379	0.86068
TB Status				
Negative	0.0661 (0.0446)	-0.0213	0.1535	0.13852
Age	-0.0027 (0.0019)	-0.0064	0.0010	0.14401
Weight	0.0023 (0.0021)	-0.0018	0.0064	0.27205
BaseCD4	0.0034 (0.0002)	0.0030	0.0038	$2e^{-16}$
Dispersion parameter($1/\alpha$)	7.7009			
Random intercept variance	0.08837			

CI stands for Confidence Interval

WHO Stage and baseline CD4 cell count were found to be significant factors of CD4 cell count of a patient. For a given patient, keeping the random intercept and other covariates constant, a one month change in time increases CD4 cell count by a multiplicative factor of $e^{0.0243}$=1.0246.

A female patient has CD4 count of $e^{0.1147}$=1.1215 times that of a male patient, adjusting for other covariates and random intercept. Holding other covariates constant, a patient with WHO's stage I has higher CD4 cell count with a multiplicative factor of $e^{0.0260}$=1.0263 as compared with a patient at stage II.

Keeping other covariates constant, a patient at WHO's stage III has higher CD4 cell count

with a multiplicative factor of $e^{0.0989}$=1.1040 as compared with a patient at stage II. Keeping other covariates constant, a patient with WHO stage IV has lower CD4 cell count with a multiplicative factor of $e^{-0.0136}$=0.9865 as compared with a patient at stage II.

A unit change in baseline CD4 cell count increases the CD4 cell count of a patient by a factor of $e^{0.0034}$=1.0034, fixing the values of the other covariates and the random intercept constant. As depicted in Table 4.10, the variance of the random intercept is found to be 0.08837 which implies that conditioning on the covariates, the remaining among-patients variability is 0.08837.

4.3.5 Discussion of Results

The effects of demographic and clinical factors on the recovery of CD4 counts over time of HIV patients taking ART treatment in Debre Markos Referral Hospital were assessed using Poisson longitudinal models as the response variable of interest CD4 count is a count variable.

The results of the summary statistics displayed in Table 4.1 revealed that the variances at all time points are maximum which indicates that there was high variation among the patients' CD4 count at baseline as well as at different time points after the initiation of ART treatment. This variation might caused by the year at which the patients started ART treatment as there were different WHO's CD4 cell count cut-off points to initiate ART treatment at different times.

As explained in the model selection part, different longitudinal count data models were applied to the data to determine the appropriate model to CD4 cell count and Poisson-Gamma-Normal (Negative Binomial log-linear mixed) model with a random intercept was found to be the best fit of the data (Table 4.10). The analysis was done using R software (version 3.4.0) under lme4 package and functions glmer for the Poisson model and glmer.nb for the Poisson-Gamma-Normal (Negative Binomial log linear mixed model). Results from Table 4.10 revealed that time in months, sex, WHO Stage and baseline CD4 cell count are found to be significant factors of CD4 cell count of a patient.

The sign of the parameter estimate of WHO stage 3 is positive which implies that a patient with WHO stage III has higher CD4 cell counts as compared with a patient of WHO stage II. It might be because the number of patients with WHO stage 3 are much higher (non-comparable)

than patients with WHO stage II (Table 4.4).

The finding of our study is in agreement with the studies of Lubyayi *et al*, (2015) and that of Adams M., Luguterah A.,(2013), in revealing the benefit of early treatment. A higher baseline CD4 cell count would result in a better CD4 cell count progression of patients on ART which is inline with the recent WHO 'treat all' recommendation. Duration on treatment also have a positive effect on CD4 count progression of HIV positive patients. This means patients with longer time on ART treatment have good recovery of CD4 cell count than that of patients with short duration on the treatment.

Getting to ART treatment at early WHO clinical stages is positively related with progression of CD4 cell count of HIV patients over time. In our study we revealed that a patient starting ART treatment with WHO stage I is beneficial in recovering his or her CD4 count as compared with the one with WHO stage II.

Keeping the other covariates constant, a patient who started ART treatment with WHO stage IV has a disadvantage of 0.0077 times CD4 count of a patient with WHO stage II, or the CD4 count of a patient who started ART treatment with WHO stage IV decreases by 0.77% as compared with patients started ART treatment at WHO stage II.

The result in the final model also showed that starting ART treatment at WHO stage III has an advantage over WHO stage II for CD4 cell count recovery which seems unrealistic. This result is not consistent with Awoke *et al* (2016) which showed that being at later WHO stages has disadvantage than being at lower WHO stages. It might be happened because the number of patients of WHO stage III (282) is not comparable with the number of patients of WHO stage II (77) included in our study.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Analysis of CD4 cell count data using conventional models like linear mixed models is inadequate as the data is highly skewed and may not satisfy normality (multivariate) assumption as demonstrated in our data.

In this study, CD4 cell count data of 445 HIV patients under ART in Debre Markos Referral Hospital was analysed using different longitudinal count data models and finally, Poisson-Gamma-Normal model was selected as a final model to fit the data based on different selection criteria. Poisson-Gamma-Normal model is the most appropriate model for CD4 cell counts data to handle overdispersion and correlation simultaneously.

Duration on ART treatment (time in months), sex of patients, WHO stage at baseline and baseline CD4 cell count were identified as potential risk factors of CD4 cell count progression. Having good CD4 cell count at baseline has a positive impact on CD4 cell count evolution over time.

Although good CD4 cell count progress in response to ART was observed, most of the patients (77.98%) were at decreased CD4 cell counts (< 200 cells/mm^3) when enroled for ART treatment which might contribute to low CD4 count recovery in some patients.

5.2 Recommendations

From the results of our final model, we observed that being male, being at later WHO stage (e.g. stage IV) and having a lower CD4 cell count at baseline have negative impact on progression of CD4 cell count. Therefore, patients with such characteristics need special guidance and due attention to improve their response to the treatment.

This study was restricted only on nine covariates and a linear trend time effect. So, it is

recommended that further studies of this nature include other important independent variables that were not included in this study such as, adherence to drugs, viral load results, opportunistic infections, etc. We considered the values of all independent variables included in our study at baseline. Therefore, it is also recommended to consider time dependent covariates and use statistical methods that can handle missingness in the data.

For HIV patients, they are advised to start antiretroviral therapy treatment as early as possible, with higher CD4 cell counts and while at lower WHO stage to improve their CD4 cell count progression. Patients are also recommended not to be discouraged and stay on the treatment as duration on treatment (time in months) affects CD4 cell count progression positively. Although we couldn't consider adherence to antiretroviral therapy as a covarite in our models, it is obvious that adherence to antiretroviral therapies is possitively ralated with CD4 cell count evolution and patients are recommended to seriously adhere to their drugs.

REFERENCES

Abbastabar H.,Abbas Rezainzadeh,Abdolreza Rajaeefard,Haleh Ghaem,Mohammad Motamodifar and Parvin Afsar Kazeroon (2016). Determining factors of CD4 cell count in HIV patients: In a historical cohort study.*International Journal of Life Science and Pharma research*.SP 2016-Issue 1 p.No 93 to 101.

Adams M., Albert Luguterah, (2013). Longitudinal analysis of change in cd4+ cell counts of hiv-1 patients on antiretroviral therapy (art) in the builsa district hospital, Department of Statistics University for Development Studies, Ghana, *European Scientific Journal* November 2013 edition vol.9, No.33 ISSN: 1857-7881 (Print) e-ISSN 1857-7431.

Allison D.,(2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: SAS Institute Inc.

Avert (2016) what are HIV and AIDS? http://www.avert.org/about-hiv-aids/what-hiv-aids; accessed on December 10, 2016.

Awoke Seyoum, Principal Ndlovu and Temesgen Zewotir, (2016). Quasi-Poisson versus negative binomial regression models in identifying factors affecting initial CD4 cell count change due to antiretroviral therapy administered to HIV-positive adults in NorthWest Ethiopia (Amhara region). *AIDS Research and Therapy*. DOI 10.1186/s12981-016-0119-6.

Bhatta L., Elise Klouman, Keshab Deuba, Rachana Shrestha, Deepak Kumar Karki, Anna Mia Ekstrom and Luai Awad Ahmed, (2013). Survival on antiretroviral treatment among adult HIV-infected patients in Nepal: a retrospective cohort study in far-western Region, 2006-2011.*BMC Infectious Diseases* 13:604.

Booth J., George Casella, Herwig Friedl and James P Hobert, (2003). Negative binomial loglinear mixed models.*Statistical Modelling* 2003; 3:179191.

Chakravarty J., Narendra K. Tiwary, Shashi Ranjan Prasad, Saurabh Shukla, Anurag Tiwari, Rabindra Nath Mishra and Shyam Sundar, 2014. Determinants of survival in adult HIV patients on antiretroviral therapy in Eastern Uttar Pradesh: A prospective study. *Indian Journal of Medical Research*, 2014 Oct; 140(4): 491500.

Daniele de Brito Trindade, Raydonal Ospina and Leila D. Amorim, (2015). Choosing the right strategy to model longitudinal count data in Epidemiology: An application with CD4 cell counts. *Epidemiology Biostatistics and Public Health*, Volume 12, Number 4.

Getachew Tekle, Wondwosen Kassahun and Abdisa Gurmessa, 2016. Statistical Analysis of CD4+ Cell Counts progression of HIV-1-positive Patients enrolled in Antiretroviral Therapy at Hossana District Queen Elleni Mohamad Memorial Hospital, South Ethiopia. *Biometrics and Biostatistics International Journal*. Volume 3 issue 1.

Grover G., Ravi Vajala and Prafulla Kumar Swain(2015). On the assessment of various factors affecting the improvement in CD4 count of AIDS patients undergoing antiretroviral therapy using generalized Poisson regression.*Journal of Applied Statistics*,42:6,1291-1305.

Hedeker, D. and Gibbons, R.D. (2006). *Longitudinal Data Analysis*. John Wiley and Sons, Inc., Hoboken, New Jersey.

Hilbe M. (2011). *Negative Binomial Regression*. Second edition. Cambridge University Press, New York.

Jemal Ayalew, Helen Moges, Omprakash sahu, Anteneh Worku, (2014). Identifying Factors Related to the Survival of AIDS Patients under the Follow-up of Antiretroviral Therapy (ART): The Case of South Wollo. *International Journal of Data Envelopment Analysis and Operations Research*, 2014, Vol. 1, No. 2, 21-27.

Jennifer Hoffman, Johan van Griensven, Robert Colebunders and Mehri McKellar, (2010); Role of the CD4 count in HIV management.

Joint United Nations program on HIV/AIDS (UNAIDS) global AIDS ubdate (2016). www.unaids.org/en/resources/documents/2016/Global-AIDS-update-2016; (accessed on December 10, 2016).

Joint United Nations program on HIV/AIDS (UNAIDS)Prevention gap report, 2016. http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf;(accessed on March 06, 2017)

Ketema Kebebew and Eshetu Wencheko, (2012). Survival analysis of HIV-infected patients under antiretroviral treatment at the Armed Forces General Teaching Hospital, Addis Ababa, Ethiopia. *Ethiopian Journal of Health Development*. 26(3).

L Al-Harthi, J Voris, BK Patterson, S Becker, J Eron, KY Smith, R DAmico, D Mildvan, J Snidow, B Pobiner, L Yau and A Landay. Evaluation of the impact of highly active antiretroviral therapy on immune recovery in antiretroviral naive patients. *HIV Medicine* (2004), 5, 5565.

Lemma Derseh Gezie, 2016. Predictors of CD4 count over time among HIV patients initiated ART in Felege Hiwot Referral Hospital, northwest Ethiopia: multilevel analysis. *BMC Research Notes* (2016) 9:377 DOI 10.1186/s13104-016-2182-4.

Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. 2006. *SAS for Mixed Models*, Second Edition. Cary, NC: SAS Institute Inc.

Lubyayi L., Ziv Shkedy, Dan Kajungu1, Daniel Mwehire1, Christine Kirungi, Esther Kawuma, Mary Odiit, Barbara Mukasa and Paul Janssen, 2015. Evolution of CD4 Cell Counts over Time for HIV/AIDS Patients on Antiretroviral Therapy (ART) In Mildmay Uganda. *Biometrics and Biostatistics International Journal*. Volume 2 Issue 2-2015.

Molenberghs G. and Verbeke G., (2000). *Linear mixed models for longitudinal data*. Springer-verlag New York, Inc.

Molenberghs G. and Verbeke G., (2005).*Models for Discrete Longitudinal Data*. Springer Science+Business Media, Inc.

Molenberghs G., Verbeke G. and Clarice G. B. Demtrio (2007). An extended random-effects approach to modeling repeated, over dispersed count data. *Lifetime Data Anal* 13:513-531.

Molenberghs G., Verbeke G., Clarice G. B. Demetrio and Afranio M. C. Vieira (2010). A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects. *Statistical Science* Vol. 25, No. 3, 325-347.

Montarroyos UR, Miranda-Filho DB, Cesar CC, Souza WV, Lacerda HR, et al. (2014) Factors Related to Changes in CD4+ T-Cell Counts over Time in Patients Living with HIV/AIDS:A Multilevel Analysis. *PLoS* ONE 9(2): e84276. doi:10.1371/journal.pone.0084276.

Mulugeta Asmare, Mekonnen Aychiluhem, Mulatu Ayana and Dube Jara, (2014). Level of ART Adherence and Associated Factors among HIV Sero- Positive Adult on Highly Active Antiretroviral Therapy in Debre Markos Referral Hospital, Northwest Ethiopia. *J Antivir Antiretrovir*, 6:3.

Nurilign Abebe, Kassahun Alemu, Tadesse Asfaw and Amanuel Alemu Abajobir , (2014). Predictors of mortality among HIV positive adults on antiretroviral therapy in Debremarkos Referral Hospital, Northwest Ethiopia. *Journal of AIDS and HIV Research.* Vol.6(1),pp.19-27.

Saravolatz L., James D. Neaton, Leonard Sacks, Lawrence Deyton, Frank Rhame, and Renslow Sherer. CD4+ T Lymphocyte Counts and Patterns of Mortality Among Patients Infected with Human Immunodeficiency Virus Who Were Enrolled in Community Programs for Clinical Research on AIDS. *Clinical Infectious Diseases* 1996;22:513-20.

WebMD Medical Reference Reviewed by Melinda Ratini, DO, MS on November (labelt162016), How CD4 Counts Help Treat HIV and AIDS
http://www.webmd.com/hiv-aids/cd4-count-what-does-it-mean#1; accessed on December 11, 2016.

World health Organization (WHO), Regional Office for South-East Asia New Delhi, (2007) Laboratory Guidelines for enumerating CD4 T Lymphocytes in the context of HIV/AIDS, (www.who.int/hiv/amds/LaboratoryGuideEnumeratingCD4TLymphocytes.pdf; accessed on December 11, 2016).

Zhou J., Thira Sirisanthana, Sasisopin Kiertiburanakul, Yi-Ming A Chen, Ning Han, Poh-Lian Lim, Nagalingeswaran Kumarasamy, Jun Yong Choi, Tuti Parwati Merati, Evy Yunihastuti1, Shinichi Oka, Adeeba Kamarulzaman, Praphan Phanuphak, Christopher KC Lee, Patrick CK Li, Sanjay Pujari, Vanthanak Saphonn, Matthew G Law, 2010. Trends in CD4 counts in HIV-infected patients with HIV viral load monitoring while on combination antiretroviral treatment: results from The TREAT Asia HIV Observational Database. *BMC Infectious Diseases* 2010, 10:361.

Appendix

Table A1: Univariate Normality Test at Each Time Point

	Shapiro-Wilk Test		Kolmogorov-Smirnov Test	
Time	W	P-value	D	P-value
0	0.86602	$2.2e^{-16}$	0.12623	$1.388e^{-06}$
6	0.90101	$7.72e^{-15}$	0.091729	0.003822
12	0.9325	$7.189e^{-11}$	0.070935	0.07988
18	0.9094	$4.946e{-}12$	0.097392	0.009321
24	0.94355	$1.076e^{-08}$	0.10231	0.006875
30	0.91962	$4.179e^{-11}$	0.067528	0.157
36	0.95793	$6.961e^{-07}$	0.086115	0.04167

Table A2: Summary of CD4 cell Progress by TB Status

TB Status	Time	0	6	12	18	24	30	36
Negative	n	368	311	268	232	231	225	213
	Mean	158.74	296.93	331.31	367.03	390.10	380.91	401.70
	Std	109.64	163.51	169.91	205.73	206.01	188.57	190.17
Positive	n	77	61	52	51	40	54	48
	Mean	144.48	279.34	283.73	307.92	359.55	359.57	381.63
	Std	87.55	171.26	146.81	148.25	200.23	157.14	173.69

Table A3: Summary of CD4 cell Progress by Marital Status

Marital Status	Time	0	6	12	18	24	30	36
Never	n	54	44	37	35	29	31	30
	Mean	142.39	309.11	307.51	334.27	421.93	380.10	418.5
	Std	72.36	183.51	160.36	155.15	205.31	148.82	214.79
Married	n	205	167	152	130	120	123	121
	Mean	158.15	290.28	328.72	363.62	388.27	371.14	403.11
	Std	103.84	171.02	163.73	207.19	206.54	196.49	199.21
Divorced	n	109	92	80	72	69	74	65
	Mean	158.28	296.82	326.96	350.75	366.64	375.29	375.38
	Std	137.73	159.68	176.47	209.83	215.21	197.40	170.29
Window	n	77	69	51	46	53	51	45
	Mean	158.19	289.86	314.63	361.57	384.34	390.51	403.31
	Std	78.59	144.80	170.49	183.35	190.49	145.67	157.72

Table A4: Summary of CD4 cell Progress by Education Level

Education Level	Time	0	6	12	18	24	30	36
No	n	160	135	105	106	101	100	86
	Mean	174.74	317.21	368.73	382.78	412.15	407.47	424.85
	Std	124.47	172.38	182.53	219.04	223.01	222.17	188.81
Primary	n	131	105	95	79	87	80	78
	Mean	138.39	278.50	306.26	318.84	348.60	343.55	361.95
	Std	84.10	138.16	138.01	146.52	186.58	163.05	167.62
Secondary	n	111	95	86	72	62	76	69
	Mean	149.51	279.92	288.71	347.39	381.19	366.82	380.35
	Std	96.67	163.31	167.80	185.71	175.02	134.62	170.75
Tertiary	n	43	37	34	26	21	23	28
	Mean	159.51	289.89	320.74	387.65	424.10	391.83	459.54
	Std	108.51	202.26	166.02	256.53	254.79	184.96	246.00

Table A5: Likelihood Ratio Tests

	Df	AIC	BIC	logLik	deviance	Chisq	ChiDf	P-value
pnml	18	71332	71434	-35648	71296			
pnm2	20	59074	59188	-29517	59034	12262	2	$2.2e^{-16}$
pnml	18	71332	71434	-35648	71296			
pgnm1	19	27380	27488	-13671	27342	43954	1	$2.2e^{-16}$
pnm2	20	59074	59188	-29517	59034			
pgnm2	21	27488	27608	-13723	27446	31588	1	$2.2e^{-16}$
pgnm1	19	27380	27488	-13671	27342			
pgnm2	21	27488	27608	-13723	27446	0	2	1

pnml= Poisson-Normal model with random intercept only
pnm2= Poisson-Normal model with both random intercept and slope
pgnm1= Poisson-Gamma-Normal with random intercept only
pgnm2= Poisson-Gamma-Normal with both random intercept and slope

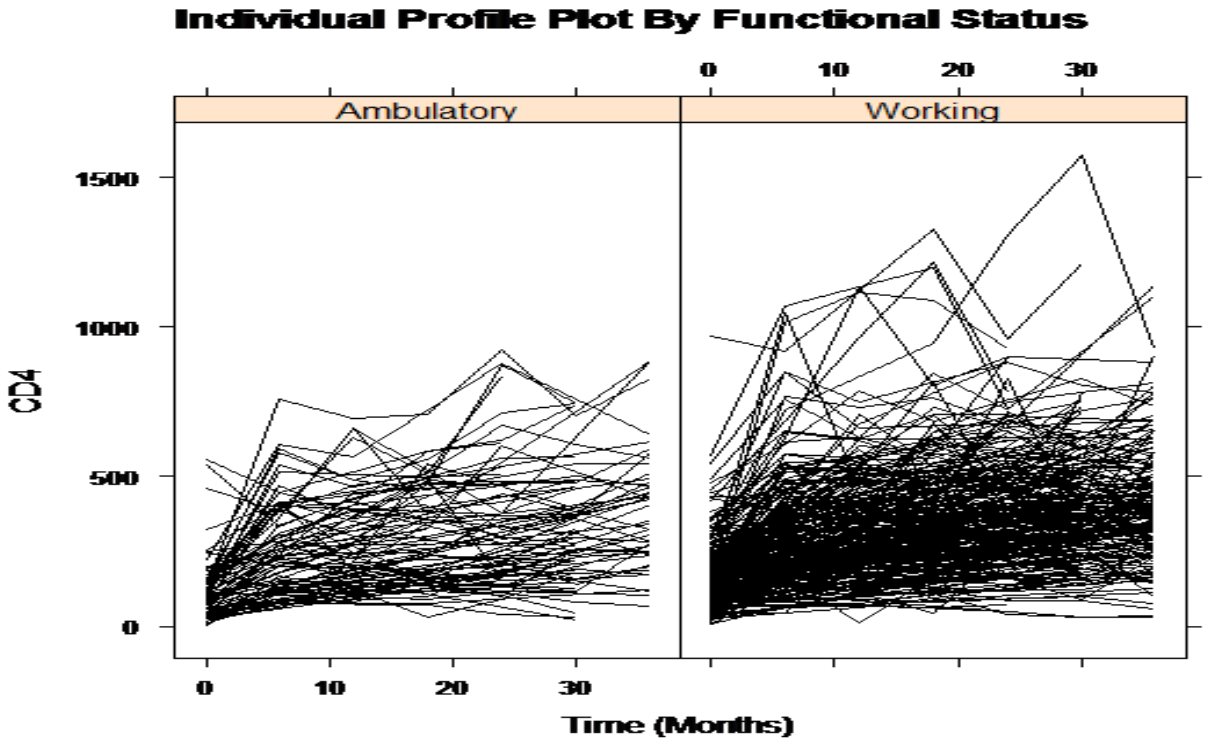

Figure A1: Individual Profile plot by Functional Status

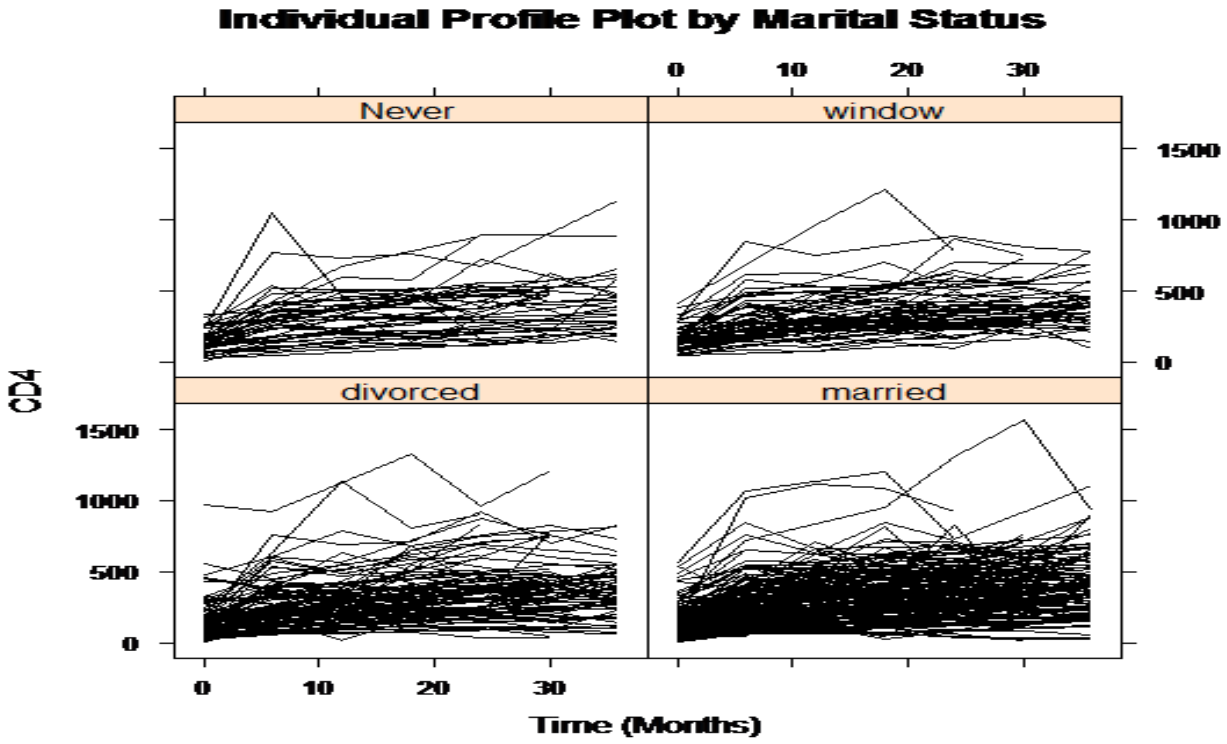

Figure A2: Individual Profile plot by Marital Status

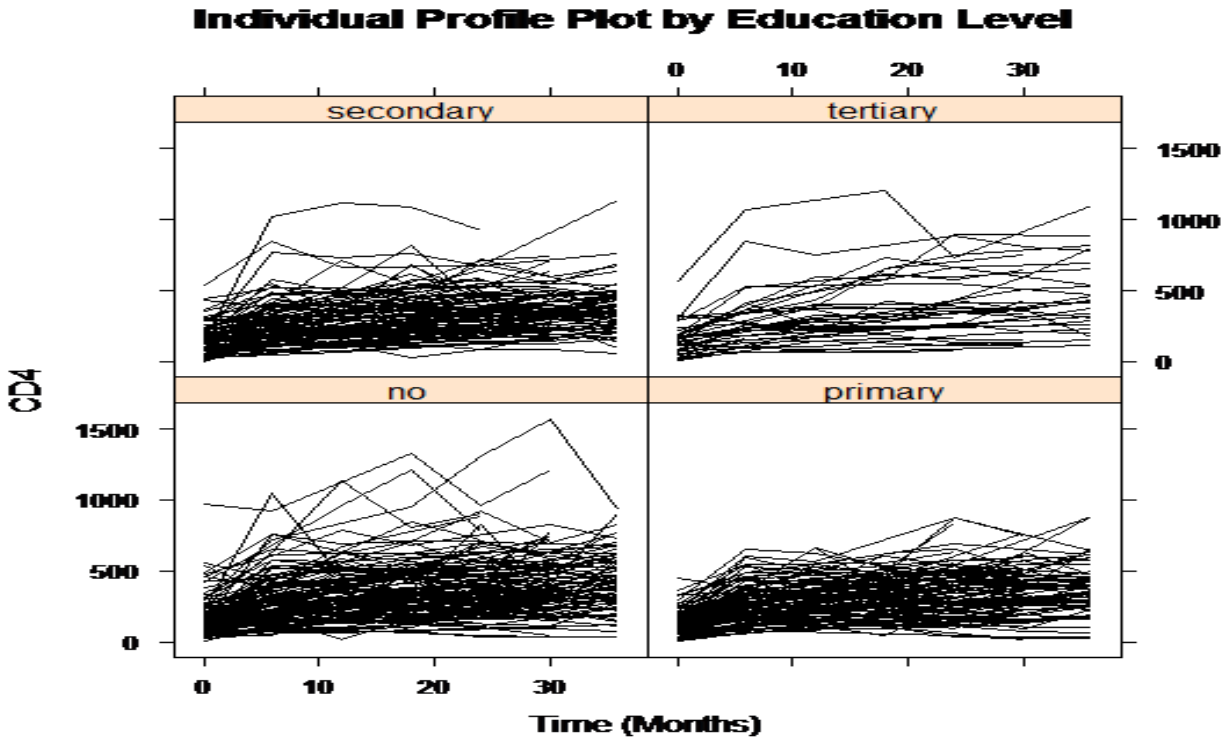

Figure A3: Individual Profile plot Education Level

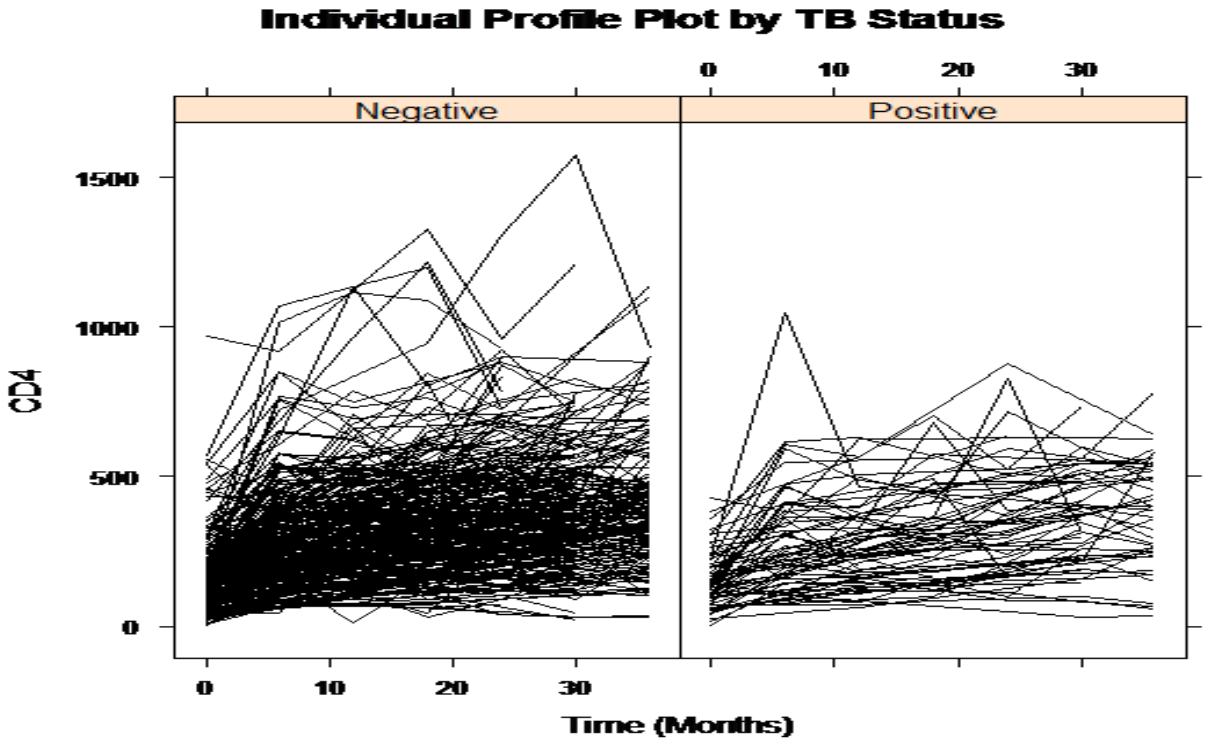

Figure A4: Individual Profile plot by TB Status

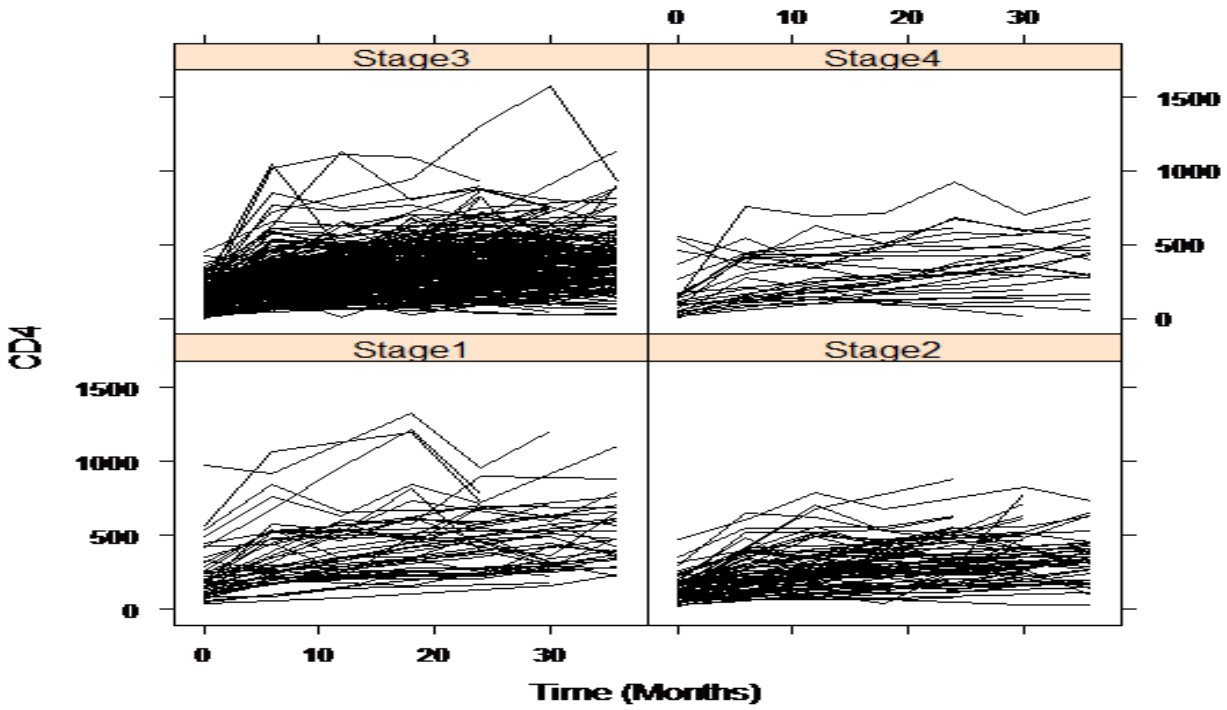

Figure A5: Individual Profile plot by WHO Stage

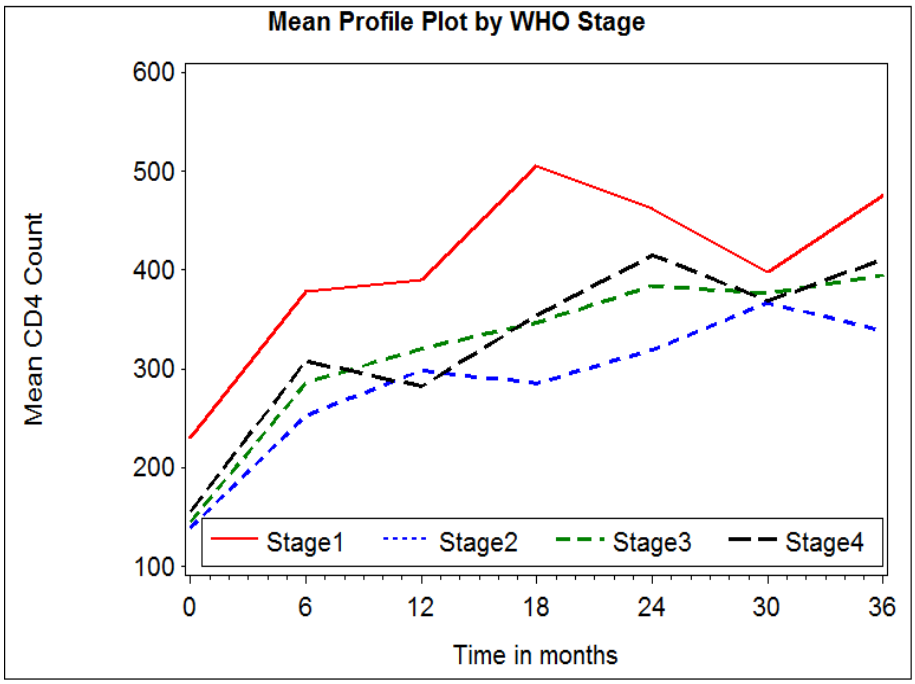

Figure A6: Mean Profile plot by WHO Stage

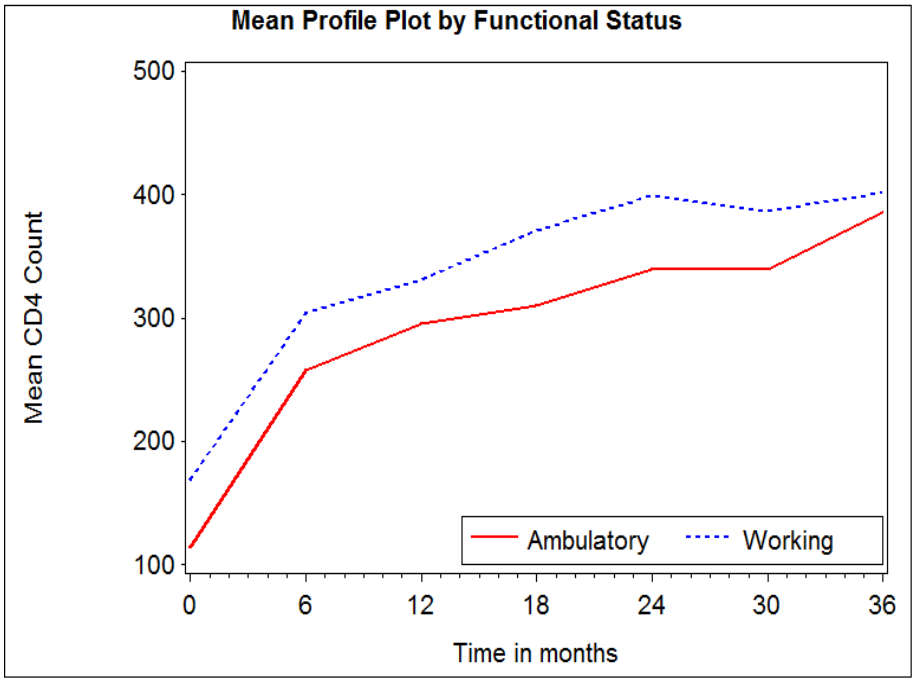

Figure A7: Mean Profile Plot of CD4 Count by Functional Status

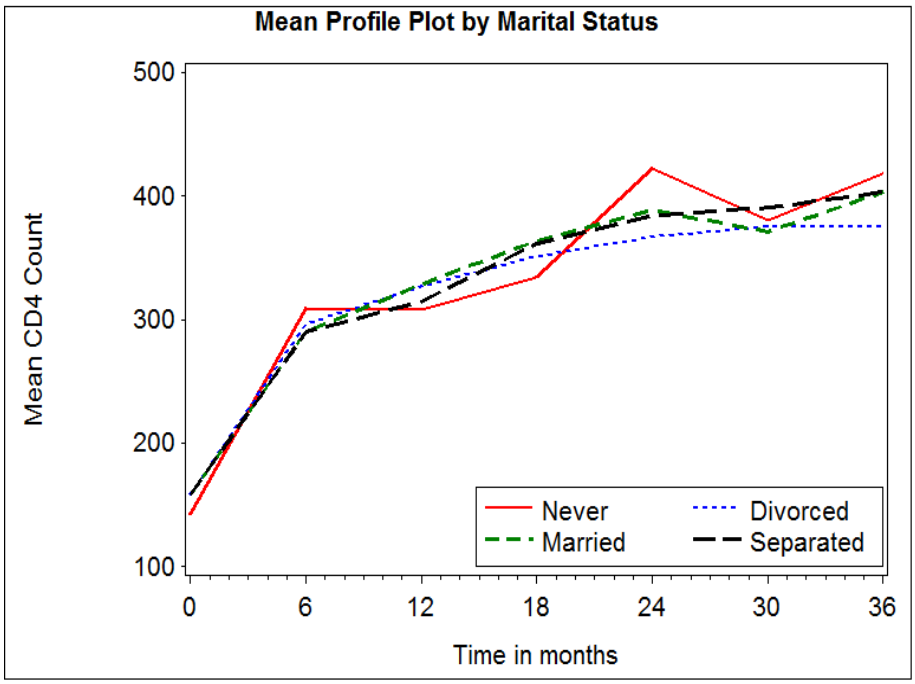

Figure A8: Mean Profile Plot of CD4 Count by Marital Status

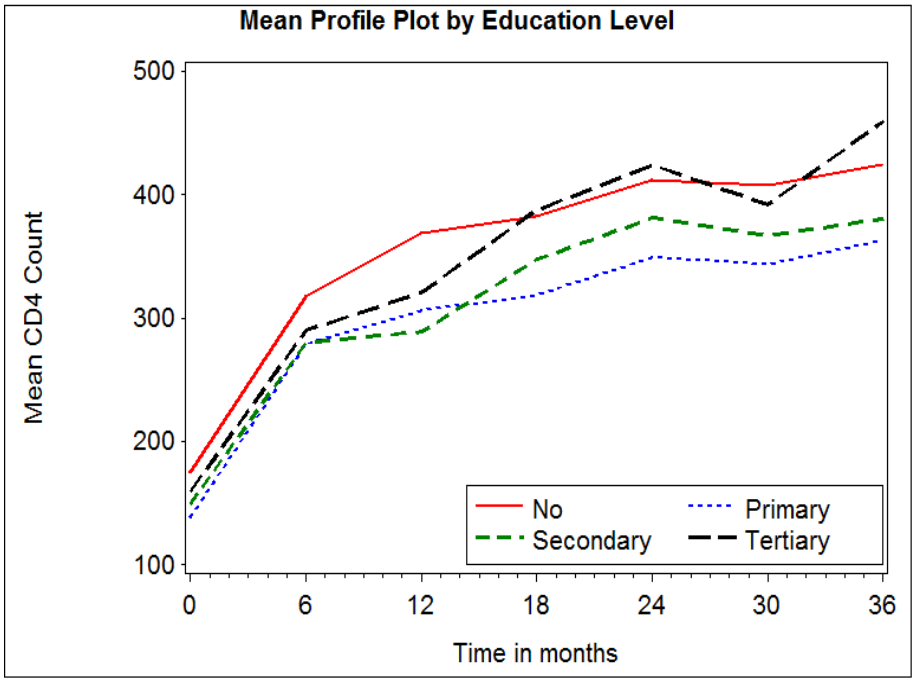

Figure A9: Mean Profile Plot of CD4 Count by Education Level

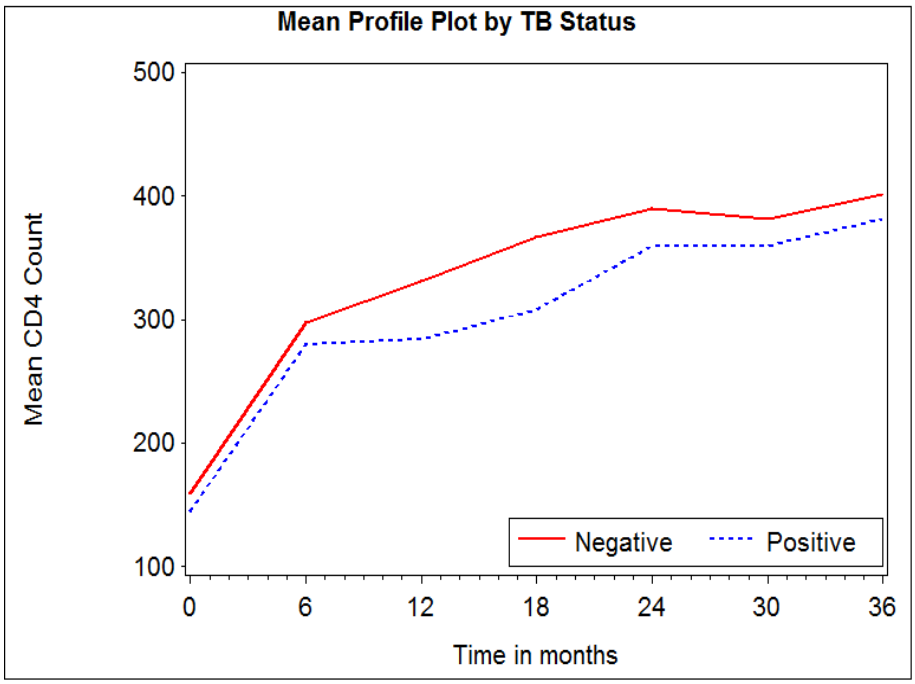

Figure A10: Mean Profile Plot of CD4 Count by TB Status

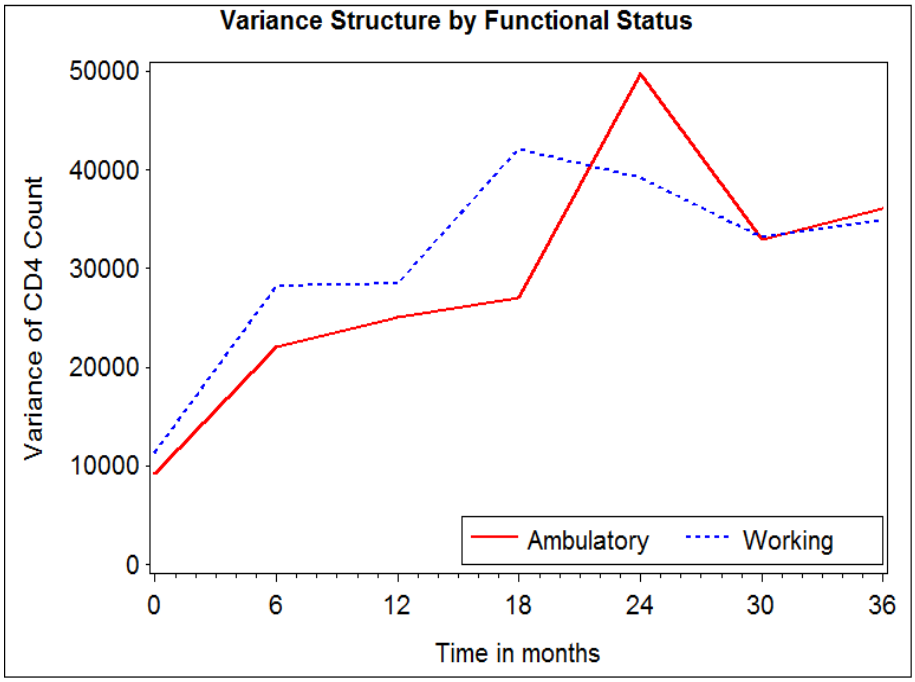

Figure A11: Variance Function of CD4 Count by Functional Status

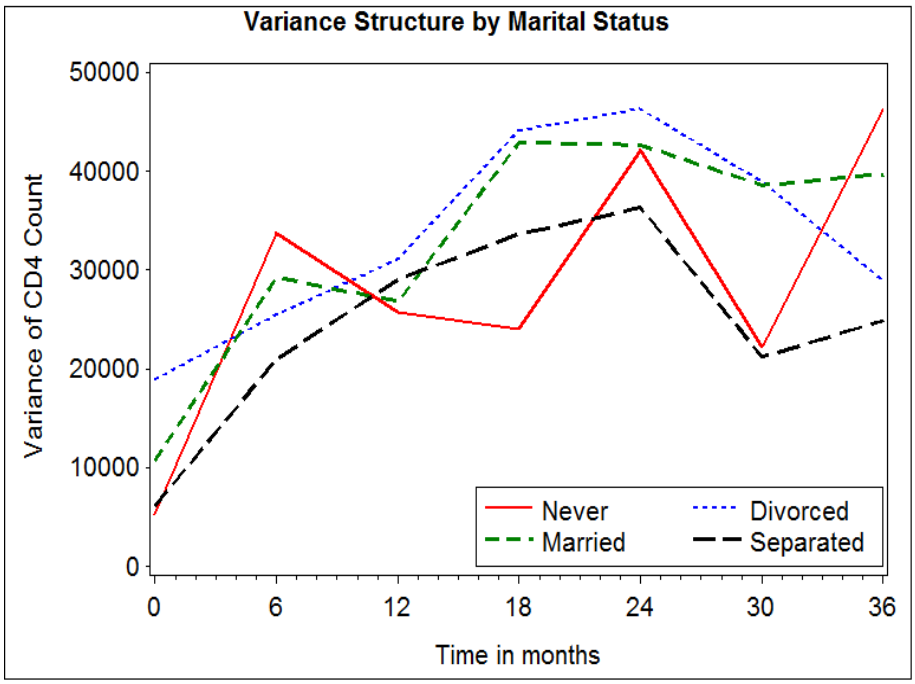

Figure A12: Variance Function of CD4 Count by Marital Status

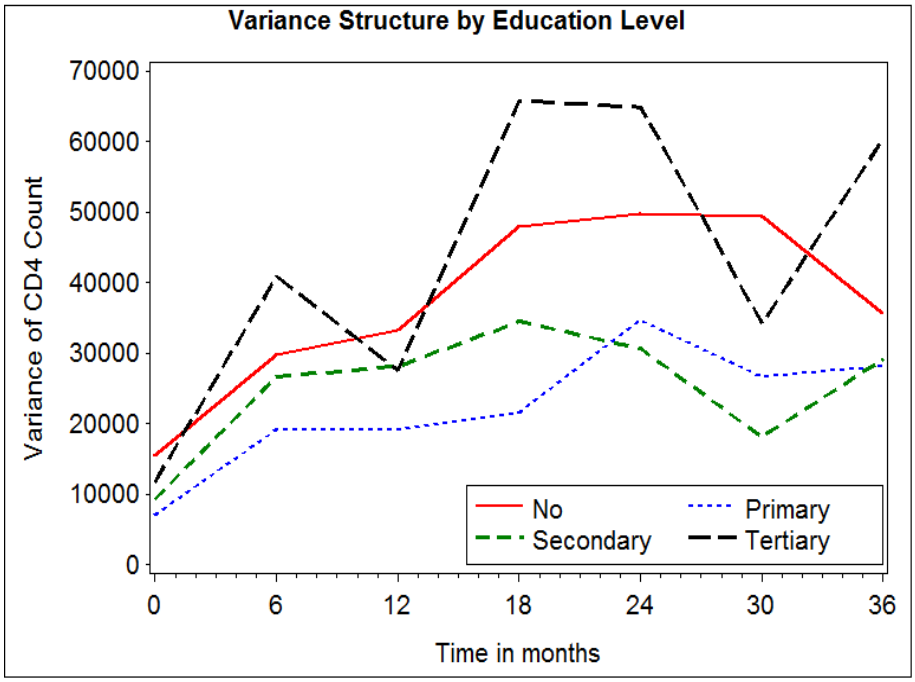

Figure A13: Variance Function of CD4 Count by Education Level

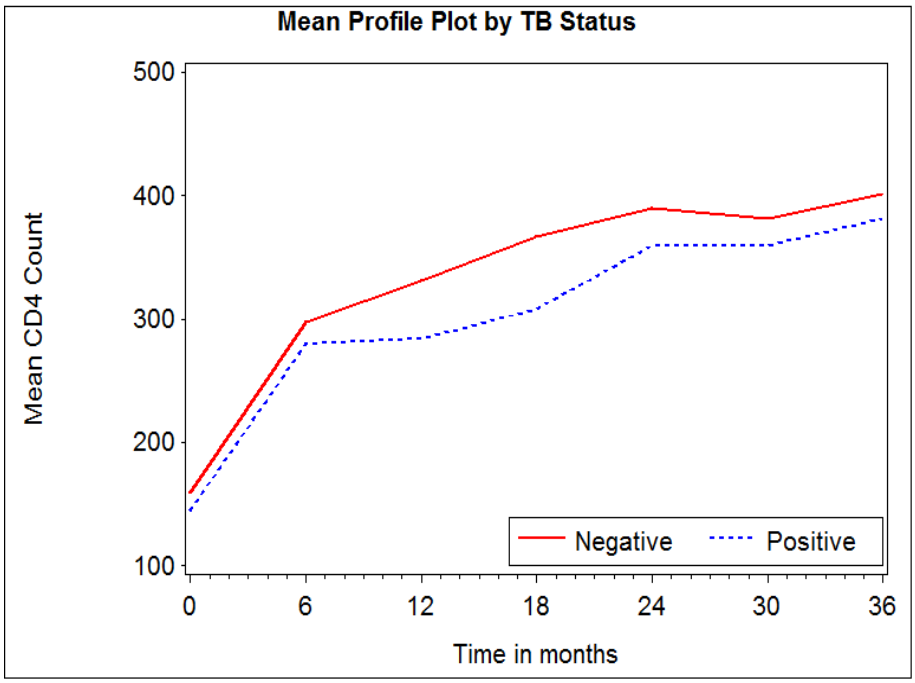

Figure A14: Variance Function of CD4 Count by TB Status

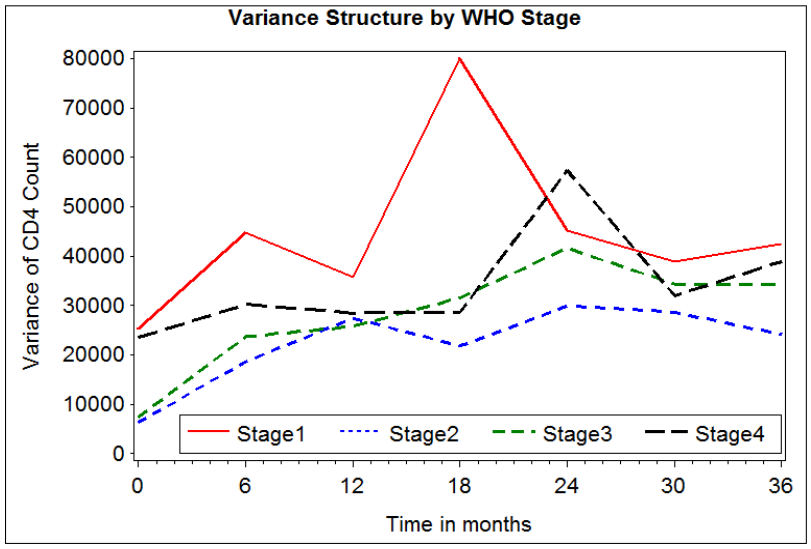

Figure A15: Variance Function of CD4 Count by WHO Stage