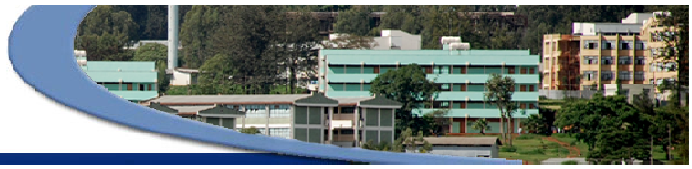
ጅማ ዩኒቨርሲቲ
JIMMA UNIVERSITY
The National Pioneer in Community Based Education of Higher Learning



SCHOOL OF GRADUATE STUDIES, DEPARTMENT OF STATISTICS

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

By:

Getachew Tekle

A Thesis Submitted to the Department of Statistics, College of Natural Science,
Jimma University as a Partial Fulfillment for the Requirements of
Masters of Science (MSc) Degree in Biostatistics

December, 2014
Jimma, Ethiopia

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

MSc thesis

Getachew Tekle
E-mail: getch.55tekle@gmail.com

Main advisor: Wondwosen Kassahun (Assi. Prof, PhD)

Co-advisor: Abdisa Gurmessa (MSc)

December, 2014
Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we herby certify that we have read the thesis prepared by GETACHEW TEKLE under our guidance, which is entitled "**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**", in its final format it is consistent and acceptable; fulfills university and department style requirements; the final manuscript is satisfactory to the graduate committee.

Wondwosen Kassahun (Assi. Prof, PhD)	_____	_____
Main Advisor	Signature	Date
Abdisa Gurmessa (MSc)	_____	_____
Co-advisor	Signature	Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Main Advisor	Signature	Date
_____	_____	_____
Name of Co-advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

STATEMENT OF AUTHOR

I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Master of Science in Biostatistics, as the result of my striving and all sources of materials used, for writing it, have been appropriately acknowledged. The thesis can be deposited in the university library to be made available to borrowers for reference. I somberly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Getachew Tekle

Date: _____

Signature: _____

Jimma University, Jimma

DEDICATION

This thesis is dedicated to my family my mother, Ayelech Gaenamo, my father, Tekle Mekiso, and my sister, Tseganesh Tekle who have tried all to me throughout my study!!!!

ABSTRACT

Background: Human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) have caused the world most shocking tragedy and risk. Mortality among patients on HAART is associated with high baseline levels of HIV RNA, WHO stage III or IV at the beginning of treatment, low body mass index, severe anemia, low CD4+ cell count, type of ART treatment, gender, resource-poor settings, and poor adherence to HAART.

Objective: The main objective of this study was to make use of appropriate modeling approach to CD4+ cell progression and identify the potential risk factors affecting the CD4+ cell progression of ART patients in Hossana District Queen Elleni Mohamad Memorial Hospital.

Methods: In this longitudinal retrospective based study secondary data was used from Hossana District Queen Elleni Mohamad Memorial Hospital. The study population consists of 222 HIV-1-positive patients, measured repeatedly at least one time on each patient who are 15 years old or older those treated with ART drugs from September 2011 to May 2014. The data was analyzed using SAS 9.2 version procedure NLMIXED. Poisson, Poisson-gamma, Poisson-normal, and Poisson-normal-gamma models were applied to study over-dispersion and correlation in the data.

Results: A total of 222 adult ART HIV-1-positive patients were included in this study. Out of these ART patients, 131(59%) were female patients and 91(41%) were male patients; 65(29.30%) were followed the drug combinations properly; the mean and standard deviation of baseline CD4+ cell counts were 355.9 and 321.4 cells per milliliter of blood, respectively; the mean and standard deviation of age of patients (p=0.0001) were 31.06 and 8.50 years, respectively; patients were followed for a mean of 24 months (p=0.0001). The analysis showed that the covariates significant for the progression of CD4+ cell counts were age of the patient, time since seroconversion, and sex at 5% level of significance.

Conclusion: On average CD4+ cell count increases after patients initiated to the HAART program (the disease rate declines). The progression of end outcome depends on patient's baseline socio-demographic characteristics. For the presence of over-dispersion, and clustering, the Poisson-normal-gamma model results in improvement in model fit.

Key: CD4+ cell count, Poisson-normal-gamma model, Overdispersion, Correlation.

ACKNOWLEDGEMENT

Above all, I would like to thank my almighty God for he allowed me to live and availed his mercy to move and breath, as the result of this, to achieve every of my activities on this land.

My special gratitude goes to Wondewosen Kassahun (Assistant Professor, PhD), my major advisor and instructor, for his kind support, advice, guidance and constructive comments beginning from the proposal development up to the final work of this thesis.

Likewise, I am sincerely grateful to my co-advisor Mr. Abdisa Gurmessa, head department of statistics, for his valuable suggestions and comments beginning from the proposal development up to the final work of this thesis.

I would like to thank Hossana District Queen Elleni Mohamad Memorial Hospital staffs especially Dr. Adane, Medical director of the hospital who allowed and facilitated all the inputs for data collection. I am grateful to my families for their guidance and encouragement during thesis preparation.

I would like to thank Jimma University for providing me the financial support for this study.

I would also like to extend my gratitude to my senior colleagues especially Mr. Yasin Negash and Mr. Amsalu Arega for their friendly advice and provision of important materials. In this line, I am grateful to all authors, not listed here, whose materials I used.

Table of Contents

ACRONYMS

AIDS: Acquired immunodeficiency syndrome

ART: Anti retroviral treatment

ARV: Antiretroviral

CD4+/ CD4+ T: Cluster of differentiation in HIV-positive patients

GLM: Generalized linear models

GLMM: Generalized linear mixed models

HAART: Highly active antiretroviral therapy

HIV: Human immunodeficiency virus

NLMIXED: Non-linear mixed models

IDU: Injection Drug use

RIS: Random intercept slope

RNA: Ribonucleic acid

SAS: Statistical Analysis System

SNNPR: Southern nations and nationalities people's region

NRTI: Number of nucleotide reverse transcriptase inhibitors

PLWHA: People living with HIV/AIDS

PNG: Poisson-normal-gamma models

P-G: Poisson-gamma models

PN-: Poisson-normal models

P--: Poisson models

STI's: Sexually transmitted infections

VCT: Voluntary testing and counseling

UN: united Nations

WHO: World health organization

List of Tables

List of Figures

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the Study

Human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) have caused the world most distressing tragedy and danger. More than 25 million people worldwide have died of AIDS since 1981, as reported by AVERT. ORG (2009) HIV/AIDS. More than 66% of the 40 million people living with HIV/AIDS are in sub Saharan Africa, where AIDS is the leading cause of death. AIDS is a disease of the human immune system caused by the human immune deficiency virus (HIV). HIV infects primarily vital cells in the human immune system such as helper T cells (to be specific, CD4+ T cells), macrophages, and dendritic cells that are necessary to activate B-lymphocytes and induce the production of antibodies.

As found out by Marlink *et al*., (1994), there are two types of HIV: HIV-1 and HIV-2. Both types are transmitted by sexual contact, through blood, and from mother to child, and they appear to cause clinically indistinguishable AIDS. However, it seems that HIV-2 is less easily transmitted, and the period between initial infection and illness is longer in the case of HIV-2.Worldwide, the predominant virus is HIV-1, and generally when people refer to HIV without specifying the type of virus they will be referring to HIV-1. The relatively uncommon HIV-2 type is concentrated in West Africa and is rarely found elsewhere.

Many factors can affect how quickly HIV infection progresses to AIDS. Factors such as age, co-infections (infections other than HIV), ethnicity, poverty, illiteracy, gender inequality, geographic location, genetics, infection route (how the disease was transmitted), nutrition, pregnancy, stress, health care provider's experience in treating HIV patients and whether or not the patient smokes or uses recreational drugs can affect the rate at which an HIV patient develops AIDS.

Elly *et al*., (2008) put that the goals of treatment with antiretroviral drugs are to inhibit viral replication while minimizing toxicities and side effects associated with the available drugs. The inhibition of virus replication permits restoration of the immune system (suppression of HIV replication, as reflected in plasma HIV concentration, to as low as possible and for as long as

possible, the preservation or enhancement of the immune function (CD4 restoration), thereby preventing or delaying the clinical progression of HIV disease. Viral eradication from the host genome is not achievable, thus a cure for HIV is not yet possible. By using HAART, it is possible to promote growth in children and prolong the survival of all HIV infected patients, reduce their morbidity and improve their quality of life.

Yibeltal *et al*., (2010) has put that in Ethiopia more than 1.3 million people living with HIV and an estimated 277,800 people requiring treatment. In 2003 the government of Ethiopia introduced ART program with the goal of reducing HIV related morbidity and mortality, improving quality of life of people living with HIV and mitigating some of the impact of epidemic. In 2005 Ethiopia launched free ART and over 71, 000 were initiated by the end of November 2006 and 241 hospitals and health centers are now providing HIV care and treatment services in regions of the country.

Elke *et al*., (2011) has found that the introduction of HAART has greatly improved the survival of HIV/AIDS infected people. HAART reduces morbidity and mortality by suppression of viral replication, restoration and preservation of immune function, and prevention of drug resistance. Mortality among patients on HAART is associated with high baseline levels of HIV RNA, WHO stage III or IV at the beginning of treatment, low body mass index, severe anemia, low CD4 cell count ,type of ART treatment, cotrimoxazole prophylaxis; gender, resource-poor settings and poor adherence to HAART. The benefits of highly active antiretroviral therapy (HAART) in the treatment of HIV infection have been well described including viral suppression, CD4 lymphocyte repletion, and durable reductions in AIDS related opportunistic diseases and death. However, the durability of the effectiveness of HAART remains to be delineated. Factors that limit the success of HAART include poor therapy adherence, regimen complexity, viral resistance, pharmacodynamic interactions, drug tolerability and toxicity, therapy costs, and presence of comorbid conditions such as substance abuse and addiction.

As was discussed by Françoise *et al*., (2006) a fundamental component of working towards the goal of providing, by 2010, universal access to antiretroviral treatment for patients with acquired immunodeficiency syndrome (AIDS) is an increased and secured production of antiretroviral drugs (ARTs) in order to meet the increased demand from lower- and middle-income countries. The vast majority of adults (96%) were reported to be receiving first-line regimens. Reporting

compliance was very high for this group, with information on the specific regimens used available for 97% of this set of patients. The programs reported that 95% of all adults receiving first line regimens were using regimens consistent with the preferred first-line approach including: Stavudine (d4T) lamivudine (3TC) Nevirapine (NVP) (61%), Zidovudine (ZDV) +3TC+NVP (16%), ZDV+3TC+ efavirenz (EFV) (9%), and d4T+3TC+EFV (8%). Less than 1% of these groups were reported to be taking either alternative first-line regimens, including the triple nucleoside combinations of ZDV+3TC+abacavir (ABC) and d4T+3TC+ABC, or taking regimens not considered or not recommended by WHO.

Bayeh *et al*., (2010) has conducted a longitudinal survey of HIV-positive patients treated with ART at Felege-Hiwot Hospital and the result showed that the ART-naive HIV patients were from low levels of education and with minimum monthly income. Moreover, he has tried to recommend the implementation of appropriate interventions in order to promote and enable HIV positive individuals to enter into ART programs as early as possible.

Since HIV was first claimed to be the cause of AIDS in 1984, the CD4 count has been widely used to make treatment and diagnostic decisions, but the use of CD4 count has been controversial and recommended actions on how to use them have changed several times over the years (Stohr et al. 2007). There are two major arms in the immune system, one which works through antibody produced by B cells and plasma cells, and the other that works through cells including CD4+T lymphocyte cells. The first is called antibody mediated or humoral immunity and the second is called cell mediated immunity. It is this cell mediated immunity that is profoundly suppressed in people diagnosed with AIDS.

There are two main approaches regarding when to start antiretroviral treatment as out lined by the guidelines. The more aggressive approach recommends starting when the CD4 counts fall below 500, and the second approach is more used in the United States which recommends starting antiretroviral medications immediately in all patients regardless of the patients CD4 counts (WHO protocols for CIS countries 2004).

Models for count data, when conditionally specified, will naturally have a subject-specific interpretation. However, taking on their purposefully modified marginalized versions leads to a

direct marginal or population-averaged interpretation for parameter estimates of covariate effects, which is the primary interest in many applications.

Following part of this thesis is organized as follows: The statement of the problem and objectives of the study are presented next in this Chapter. Chapter 2 describes some literatures related to the associated factors for the progression of CD4+ cell counts and different modeling approaches. In Chapter 3, the data and the detail methods of data analyses are explained. Then, basic results of the study are presented in Chapter 4 and discussed in Sub-Sections. Finally, some concluding remarks, recommendations and some limitations are provided in Chapter 5.

1.2. Statement of the Problem

Some motivations for identification of problem of this study were tried to put as below:
Stohr *et al.*, (2007) found out that apart from the management of large numbers of HIV/AIDS clients and the costs involved, another challenge exists on the ART programs. This is monitoring the progress of clients/patient's immune response to HAART in HIV/AIDS patients'.

Currently, there is no enough evidence showing that all the ART centers in Ethiopia have implemented research tools to monitor patients' immune (CD4) response to HAART within a specified time frame and identification of factors that might be associated with the poor CD4-Lymphocyte response to HAART. Although trend analysis studies have been carried out by international research institutions, there has been less focus on the local institutions to carry out and strengthen CD4/viral load trends analysis studies to give a clear indication on the response of HIV/AIDS patients to ART. It should still remain the responsibility of the active ART centers to employ a tool that will constantly monitor the clients' progress in CD4 count since it still remains the major indicator of an individual's immunity.

This identifies serious consideration of managing and monitoring the high number of clients on HAART with well defined strategies to measure their immune response to treatment and identify factors that have the ability to influence their CD4 count recovery.

Count data are collected repeatedly over time in many applications, such as biology, epidemiology, and public health. Such data are often characterized by the following features. First, correlation due to the repeated measures is usually accounted for using subject-specific

random effects, which are assumed to be normally distributed. Second, the sample variance may exceed the mean, and hence, the theoretical mean–variance relationship is violated, leading to over-dispersion. This is usually allowed for based on a hierarchical approach, combining a Poisson model with gamma distributed random effects (Kassahun et al., 2014).

In practice, however, all these features can appear simultaneously. Hence, appropriate modeling approaches which can overcome these issues and which lighten data analysis are needed.

However, statistical modeling of such data pretenses several challenges. This is because repeatedly measured CD+ cell counts often exhibit the following features. First, there are correlated observations per subject, which result from clustering of measurements within subjects. Second, the variance exceeds the mean, leading to so-called over-dispersion.

Study Questions

This study seeks to answer the following questions:

1. Does HAART have a positive effect on the HIV/AIDS patient's immune system based on an indication of their gained CD4+cell counts trend analysis at Hossana District Queen Elleni Mohamad Memorial Hospital, SNNPR, Ethiopia?
2. What is/are the appropriate longitudinal model/s to handle over-dispersed and correlated individual subjects in this data?
3. What are important potential determining factors in HIV/AIDS patients' response to HAART at Hossana District Queen Elleni Mohamad Memorial Hospital, SNNPR, Ethiopia?

1.3. Objectives of the study

1.3.1. General objective:

General objective of this study was to make use of appropriate modeling approach to CD4+ cell counts progression and identify the potential risk factors affecting the CD4+ cell progression of ART patients in Hossana District Queen Elleni Mohamad Memorial Hospital.

1.3.2. Specific objectives of this study

Specific objectives of this study were to:

- To explore how CD4+ cell counts of HIV-1-positive patients under ART in Hossana District Queen Elleni Mohamad Memorial Hospital change over time;
- To fit an appropriate statistical model for the average progression of CD4+ cell counts of HIV-1- positive patients;
- To identify the potential factors affecting the progression of CD4+ cell counts among HIV-1- positive patients under ART in Hossana District Queen Elleni Mohamad Memorial Hospital.

1.4. Significance of the Study

The results of this study will be useful in the development of an effective care and patient monitoring system on use of antiretroviral agents in HIV-1-infected adults and adolescents.

Specifically:

- It helps to identify the potential risk factors influencing the absolute CD4 count measurements in HIV infected patients.
- It helps the respective policy makers of the health sector monitoring frequency of CD4 Count, monitoring therapeutic response, and judge urgency of ART initiation using CD4+ cell Counts of the patients.
- It can be used as a reference for those who want to apply the techniques of handling correlation and over-dispersion in counts data longitudinally collected; so that it serves as a base for further study for the question what brings this variation and others.

CHAPTER TWO

2. LITERETURES REVIEW

2.1. Impact of HIV/AIDS in developing countries of Africa and ART in Ethiopia

According Johannessen *et al*., (2008), 39.4 million peoples were living with HIV/AIDS in developing countries of Africa. Adults contribute 37.2 million. About 5 million peoples were newly infected of which 4.3 million were adults from these more than 95% of new infections were in developing countries. Over 6 million infected need ART but 350,000-400,000 was treated in developing countries. By December 2006 two million people in low and middle income countries were receiving ART but this was still only 28% of those estimated to be in urgent need of it.

Yibeltal *et al*., (2010) and Hoyos *et al*., (2007) reported that in Ethiopia, the highest prevalence occurs in the age group 15-24, and 90% of HIV infection occurs in adults between 15-49 years. In Ethiopia 443,964 PLWHA were enrolled, 246,347 PLWHA started treatment and 179,183 were currently on ART.

A study was conducted by Johnson and Dorrington (2006) applying survival analysis to describe the impact of HIV/AIDS and the effects of HIV/AIDS prevention and treatment programmes. From this study they found out that HAART was expected to have a significant impact on HIV prevalence, due to the improved survival prospects of infected individuals. As well the study supported the reduction of HIV prevalence in 2005 roughly by 2% due to prevention programmes. In the same manner, Gange *et al*., (2002) conducted a study on 1691 HIV seropositive women in USA and revealed that HAART improved immunological function, suppressed HIV disease activity, and reduced morbidity and mortality.

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

According to Stohr *et al.,* (2007) knowledge and the picture HIV on an individual's immune system were changing rapidly. There is a need for close supervision and improve on the existing techniques that would ensure monitoring CD4 trends and variations in trends, in HIV/AIDS patients on HAART.

Florence *et al*., (2003) conducted a prospective observational study in Eurosida on more than 8500 HIV/AIDS patients who were followed in 63 hospitals of the 20 European countries using absolute CD4 counts coupled with viral load monitoring. The patients for this study were all those who started HAART with moderate immunologic suppression, with base line CD4 count below 350cells/µl(measured within previous 6 months at most). Allison (2008) reported that, as untreated HIV progresses, CD4 count decreases by about approximately 4% every year. With successful ART the CD4 count might increase by greater than 50 cells per micro litre within weeks after viral suppression. Additionally, it may increase by 50-100 cells per micro litre per year thereafter, until a threshold is reached (Allison, 2006).

However, it was discussed by Johannessen *et al*., (2008) as, it is important to note that in some patients, CD4 count may not increase this quickly or steadily even with durable viral suppression. Association between risk factors potentially associated with a low CD4 count response in the Eurosida study were assessed by using a logistic regression model and expressed as odds ratio, with 95% confidence intervals. The risk factors taken into account for low CD4 count response were demographic factors (age, sex ratio, ethnic origin, HIV transmission group, region), antiretroviral treatment factors (number of nucleotide reverse transcriptase inhibitors(NRTI) use before HAART, start date of HAART, type of HAART, HAART duration), immunological factors(previous episodes of opportunistic infections). Using the SAS statistical software version 6.12, the findings of the study were that although there is an increased CD4 count in patients commenced on HAART there was a poor immune reconstitution despite a good virological control among patients with a baseline CD4 count of <350 cells/µl. The underlying mechanism leading to this condition seemed mainly driven by age and baseline immunological and virological status of patient.

Patients starting treatment at CD4 50-199 and <50 cells/µl have net health benefits of 7.6 and 7.3 life years. Without treatment, HIV patients with CD4 counts 200-350; 50- 199 and < 50 cells/µl can expect to live 4.8; 2.0 and 0.7 life years, respectively. Individuals with incomplete CD4+ T

cell recovery to <500 cells/microL had more advanced HIV- 1 infection at baseline. CD4+ T cell changes during the first 3-6 months of ART already reflect the capacity of the immune system to replenish depleted CD4+ T lymphocytes as discussed by Kaufmann *et al*., (2005) and Zachariah *et al*., (2006).

As reported by Dzimnenani (2007), WHO stage IV diseases, however, was found to be a strong predictor of mortality in all studies reporting on this. In three studies comparing patients with WHO stage IV disease at baseline with those with WHO stages I–III, WHO stage IV was associated with more than a doubling in the hazard of death (summary hazard ratio, 2.2; 95% CI: 1.5–3).

Yanis *et al*., (2005) reported that female patient had significantly lower age, higher prevalence of heterosexual contact and lower prevalence of intravenous drug use as risk factors for HIV infection than male subjects. They were also reported to have higher previous exposure to antiretroviral therapy, higher CD4 cell counts and lower viral loads than male individuals.

Moore and Keruly (2007) found that patients tends to reach a plateau CD4 increase after four years of HAART with good or highest CD4 peaks among those who started treatment with CD4 counts above 350cell/μL. The study showed that people who commence treatment with CD4 counts much lower than 350 cell/μL are less likely to attain normal CD4-cell counts of (<750 cells/mm^3). On the other hand, good immunological and virological responses to HAART can be achieved regardless of the CD4 count at initiation. For the purpose of this study, it was assumed that the plateau CD4 rise will not be reached within the first two years of and that good CD4 response to HAART was a gain of 50-100 cells/μl per year (Allison, 2006).

A study was conducted with an objective to determine the timing of initiation of antiretroviral therapy (ART) in routine clinical practices, reflected treatment guidelines that evolved toward recommending starting therapy at lower CD4 cell counts. The study analyzed the longitudinal data on 10820 patients. The study analyzed the effects of non-clinical factors such as (age, sex, ethnicity, and exposure category) by logistic regression. Kaplan-meir analysis was used to estimate the proportion of patient who had initiated ART by particular CD4 count among early presenters (initial CD4 count>500cells/μl). The results showed that there was a tendency to initiate ART at lower CD4 counts over the years 1997-2000, especially in the range 200-500cells/μL with little change thereafter.

The conclusion of the study was that initiation of ART in the clinics included in this study reflected evolved treatment guidelines. A variety of consensus guidelines had been written and all recommend starting HAART well before the CD4 count fall below 200cells/μL, this being the threshold below which the risk of opportunistic infections(OI's) is reported to increase significantly (Stohr *et al*, 2007). The above studies strongly recommend trends analysis research in HIV patients on ART to determine if there are similar findings in other geographic areas.

According to Katubulushi and Chanda (2009), a population-based cohort study of unselected adults in Misisi, a shanty compound in Lusaka, was under follow-up since 1999, and CD4 cell counts have been followed in participants since the initial survey. No antiretroviral drugs were used by any of the participants over the period of the study. Approval was given by the research ethics committees of the University of Zambia and the London School of Hygiene and Tropical Medicine. The initial cohort of 261 adults included 65 HIV-seropositive participants, of whom 12 died, and the researchers were able to obtain repeated measurements (in 1999, 2000, 2001 and 2003) in 24.

Among the survivors, the mean age of men (n = 7) was 35 years (SD 7.0) and of women (n = 17) it was 28.6 years (SD 5.7), and the median initial CD4 cell count was 389 cells/μl (interquartile range 255-537). The median initial CD4 cell count was 122 cells/μl among the 12 who died. In the survivors, the mean decline over 4 years was 29 cells/μl per year. The mean percentage decline from baseline was 30%. In six of those who died and who had had at least two measurements separated by at least one year, the mean decline was 15 cells/μl per year. This data indicated that the decline in CD4 cell counts over the period from 1999 to 2003 in adults not treated with highly active antiretroviral therapy was slow, and the estimate of the rate of loss of CD4 cells is in very close agreement with the estimate of 21.5 cells/μl per year from Tanzania. This data support the idea that HIV progression to AIDS and death is slower than at first thought, even in very under-resourced populations living in crowded conditions. The data were unlikely to be skewed by depressed initial CD4 cell counts as a result of inter current infection because close analysis of the data depicted reveals only three individuals whose CD4 cell counts rose from an initially depressed level.

It was well established that infection occurs in young women at an earlier age than in young men, and AIDS-related deaths start to occur in women in their third decade of life. It could not

have taken 19 years for these young women to progress from infection to AIDS if their infections were acquired sexually after puberty, so we must consider alternative explanations. This leads to the understanding that other factors are at play in influencing the rate of decline of the CD4 counts over a period of time. Some studies have demonstrated a number of factors that affect the CD4 count response to HAART as discussed below (UNICEF et al., 2002).

2.2. Risk Factors for CD4 cell counts progression

2.2.1. Social/demographic factors

Age

Most doctors confirm that CD4 recovery is slow and less perfect in older people. As in many other diseases, age is an important prognostic factor in HIV infection.

Similar to others as reported next to this, Sophie *et al*., (2006) reported that age at sero-conversion and age at a given CD4 cell count were shown to be important determinants of progression and survival before the wide spread introduction of HAART, starting in 1996. From this date time to next of it, several studies, such as the ART Cohort Collaboration (ART-CC)(Egger *et al*.,2002), which included 13 cohort studies conducted in Europe and North America, had shown that age remains an independent predictor of clinical progression on HAART. The impact of age in the ART-CC study seemed to be less marked than in the pre-HAART era, but a threshold effect was noted at 50 years. Because older patients are usually excluded from clinical trials, controlled data were lacking on this age group. Studies of the response to HAART in elderly patients have mostly involved small populations and relatively short follow-up, as reported by Perez and Moore (2003).

Gender

As reported by Hoyos *et al*., (2007), a journal of women's health reported findings that there are no differences in HIV progression and response to HAART attributable to gender among patients accessing the Spanish hospital network. It was a multicenter, hospital-based cohort of HIV-infected patients attending 10 hospitals in Spain from January 1997 to December 2003. Kaplan-Meier and Cox regression were used to assess the effect of sex on time to AIDS, survival from AIDS, onset of a new AIDS event or death, and viral suppression from HAART. The study concluded no differences in HIV progression and response to HAART attributable to gender among patients accessing the Spanish hospital network.

2.2.2. ART factors

Adherence to any of the drugs

According to published research from the United States concluded by the authors Kitahata *et al*., (2004), HAART adherence predicts treatment response and progression to AIDS and death. Although adherence to HAART at a level above 95% has been associated with optimal viral suppression, the impact of different levels of adherence on long-term clinical outcomes has not been determined. Researchers used an objective pharmacy based measure to examine the association between three levels of adherence to HAART and disease progression among a population-based cohort of HIV infected patients attending an urban HIV specialty clinic. Findings were that higher levels of adherence to HAART were significantly associated with longer time to virologic failure ($p<0.001$), greater increase in CD4 cell count ($p=0.04$), and lower risk of progression to clinical AIDS or death ($p<0.007$). After controlling for other factors, patients with low adherence had over five times the risk of disease progression than patients with moderate adherence ($p=0.007$) or patients with high adherence ($p=0.001$). There was no significant difference in the risk of progression between patients with moderate and high levels of adherence ($p>0.2$). Patients who progressed to AIDS or death had significantly higher viral loads ($p=0.01$) and lower CD4 cell counts ($p=0.03$) than patients who experienced virological failure, but did not progress.

Evan *et al.*, (1996) similarly conducted a study in that looked at adherence to antiretroviral therapy and CD4+ T-cell count responses among HIV-infected injection drug users and reported that research findings clearly stated that overall, the CD4 cell count response rate was slower among injection drug users in Kaplan-Meier analyses (log-rank: $p<0.05$). Injection drug users were poor in adherence to HAART. However, no differences existed when the analyses were restricted to adherent patients (log-rank: $p=0.349$). Similarly, the differences in the time to CD4 cell count response observed in univariate Cox regression analyses for patients with a history of injection drug use [relative hazard: 0.85 (95% CI: 0.75-0.97)] diminished after adjustment for adherence [adjusted relative hazard: 1.02(95% CI: 0.89-1.16)]. In Conclusion, these data demonstrated the importance of adherence on CD4 cell count responses and highlighted the need for interventions to improve antiretroviral adherence among injection drug user.

Furthermore, in a crude form, the other potential factors were marital status, level of education, baseline CD4 counts, religion, employment status of the patient, residence of the patient and WHO Stage of the disease as discussed by (Gezahegn, 2011; Smith *et al*., 2004; Ketema, 2011; Marlon, 2010).

2.3. Modeling Aspects

Some of the modeling approaches in longitudinal data analysis particularly count data like CD4+ cell counts proposed by different authors were reviewed as follows:

Gezahegn (2011) applied Cox proportional-hazard regression to calculate the bivariate and adjusted hazard rate and then determine independent predictors of time to death in CD4 cell counts data in Durame and Hosanna hospital. The estimated mortality was 7%, 8%, 11.3 %, 15.7% and 21% at 6, 12, 24, 36 and 48 months respectively. After adjustment, the independent significant predictors of death in patients living with HIV/AIDS after initiation of ART remain poor ART adherence(AHR=5.09[95% CI: 5.51-49.48]), Advanced WHO staging (AHR=1.5[95% CI: 1.18-2.16]), positive TB test (AHR=3.9[95% CI: 1.89-8.07]), not married or single (AHR=10.27[95% CI: 1.35-78.3), male gender (AHR=1.704[95% CI: 1.23-2.24])and older age(AHR=1.45[95% CI: 1.1-1.96). This study demonstrated that simple laboratory and clinical data, available to health care providers prior to ART initiation, can predict which patients are at increased risk of death when they start therapy.

Smith *et al*., (2004) used linear mixed models to investigate factors associated with increases in CD4 cell counts from 3 months onward patients at the Ian Charleson Centre at the Royal Free Hospital, London. The analysis of this study revealed that After 6, 12, and 24 months of HAART, the median increases in CD4 cell counts were 114, 181, and 248 cells/µl, respectively; 84%, 84%, and 80% of subjects had a virus load of <400 copies/mL during the same periods. White ethnicity, higher pre-HAART virus load, and lower pre-HAART CD4 and CD8 cell counts were associated with greater increases in CD4 cell counts during the first 3 months of HAART. From 3 months onward, a greater cumulative proportion of time spent with virus load <400 copies/mL was associated with a more favorable change in CD4 cell count (an average increase of 5.2 cells/µl/year [95% confidence interval [CI]: 3.8–6.7 cells/µl/year] for each extra 10% cumulative time spent with a virus load <400 copies/mL) ($P < .0001$).

Ketema (2011) conducted a retrospective cohort study in Armed Forces General Teaching Hospital (AFGTH) located in Addis Ababa, Ethiopia and applied Kaplan-Meier survival curves and Log-Rank test to compare the survival experience of different category of ART patients, and employed proportional hazards Cox model to identify independent predictors of mortality. 734 patients on ART were followed for a median of 38.5 months (IQR 10.75, 53). The independent predictors of mortality were low CD4 cell count at baseline, (HR = 0.995, 95% CI: 0.991 - 0.999), ambulatory and bedridden functional status, (HR=2.011, 95% CI: 1.018 - 3.973) and (HR=3.358, 95% CI: 1.734 - 6.500), respectively, WHO clinical stages III and IV (HR=7.052, 95% CI: 1.677- 29.658) and (HR=12.64, 95% CI: 3.003 - 53.199), respectively, TB co-infection, (HR=1.734, 95% CI: 1.039 - 2.893) and OIs (HR=8.985, 95% CI: 1.240 - 65.085).

Marlon (2010) conducted a retrospective cohort design on a sample size of 340 files of clients at Chreso Ministries VCT and ART centres to identify factors that affect CD4+T Lymphocyte count response in patients commenced on HAART within 24 months of treatment and used the Chi-Square test at 5% with cross tabulation tables to determine associations between identified variables and CD4-Lymphocyte count response to HAART and used Logistic regression analysis to predict the probability of CD4 count response to HAART using the variables of this study. In this study it was found that gender, alcohol consumption, nadire and regimen affects CD4 count response to HAART. It was found that men, non-alcohol consumers and those that start HAART with baseline CD4 count above 350 cell/μL experienced a good CD4 response to HAART. Additionally, those who commenced treatment on Truvada and devoted themselves to 95% adherence also experienced a good CD4 count response to HAART. On the other hand, age, smoking and employment status did not affect CD4 count response to HAART.

Vernon *et al*., (2013) performed a longitudinal cohort study of HIV-infected adults who had started highly active antiretroviral therapy (HAART) selected from three outpatients HIV medical clinics in Cleveland, Ohio. They used Linear mixed effects models to assess the associations between immune function and periodontal disease (PD). The analysis of this study shown that forty (40) subjects with median 2.7 months on HAART and median nadir CD4+ T-cell count of 212 cells/μl completed a median 3 visits. Over 24 months, CD4+ T-cell count increased by a mean of 173 cells/μl ($p<0.001$) and HIV RNA decreased by 0.5 log10 copies/ml ($p<0.001$); concurrently, periodontal probing depth, clinical attachment level and bleeding on

probing decreased by a mean 11.7%, 12.1%, and 14.7% respectively (all $p<0.001$). Lower nadir CD4+ T-cell count was associated with worse baseline regression (REC) (-6.72%; $p=0.04$) and clinical attachment level (9.06%; $p<0.001$). Further, lower nadir CD4+ T-cell count was associated with a greater relative longitudinal improvement in periodontal probing depth (PPD) in subjects with higher baseline levels of Porphyromonas gingivalis ($p=0.027$), and bleeding on probing (BOP) in subjects with higher baseline levels of Porphyromonas gingivalis or Treponema denticola ($p=0.001$ and $p=0.006$ respectively). Longitudinal changes from baseline in CD4+ T-cell count and level of HIV RNA were not independently associated with longitudinal changes in any clinical markers of PD.

From the analysis the following conclusion was made as the degree of immune suppression was associated with baseline gingival recession. After HAART initiation, measures of active PD improved most in those with lower nadir CD4+ T-cell counts and higher baseline levels of specific periodontopathogens. Nadir CD4+ T-cell count differentially influences periodontal disease both before and after HAART in HIV-infected adults.

Mustapha and Albert (2013) conducted a retrospective study at the Builsa District hospital in the Upper East Region of Ghana. They applied Linear mixed effects model to model the change in CD4+ cell count over time. They found from the analysis of this data that the correlation between CD4+ cell counts at different times had a first order autoregressive moving average variance-covariance structure. The Initial CD4+ cell count of a patient, the duration of treatment and the drug type used in the treatment, were the factors that significantly determined a patient's current CD4+ cell count. From the study they concluded that this study is useful to guide education to the Public, particularly Patients, and also guide policy and management of treatment, and they tried to recommend that further studies are recommended to expand the scope of study as well as to include more covariates.

Reda *et al.*, (2013) conducted a retrospective cohort study among HIV/AIDS patients taking ART from 2005 to 2010 in Hiwot Fana, Jugal and Dil Chora hospitals located in eastern Ethiopia. They applied mixed models regression to examine changes in CD4 cell count and weight after the baseline measurement. The analysis of the data of this study found that both the median CD4 lymphocyte counts and weight showed improvements in the follow up periods. The multivariate analysis showed that the duration of ART was an important predictor of

improvements in CD4 lymphocyte count (beta 7.91; 95% CI: 7.48–8.34; p =0.000) and weight (beta 0.15; 95% CI: 0.13–0.18; p= 0.000). Advanced WHO clinical stage, lower baseline CD4 cell count and baseline hemoglobin levels were factors associated with decline in weight. Actively working patients had higher CD4 lymphocyte count and weight compared to those that were ambulatory (p= 0.05).

Correlation and over-dispersion can crop up simultaneously, in practice. So that some authors have been proposed some techniques and they have formulated a flexible and unified modeling framework to overcome this, which afterwards termed as the so-called combined model, to handle a wide range of hierarchical data, including count, binary, and time-to-event data (Molenberghs *et al*., 2010). They also come up with two sets of random effects. The normally distributed subject-specific random effects which take into custody correlation and a certain amount of over-dispersion, while a conjugate measurement-specific random effect on the natural parameter scale is used to accommodate the remaining over-dispersion.

For counts, the Poisson model with normal and gamma random effects can be specified as $Y_{ij} \sim$ Poisson ($\lambda_{ij} = \theta_{ij}\kappa_{ij}$), with conditional mean $\lambda_{ij} = \theta_{ij}\kappa_{ij}$ and κ_{ij} =exp ($\mathbf{x}'_{ij}\,\beta + \mathbf{z}'_{ij}\mathbf{b}$), where θij ~ Gamma (α, β), and the rest of the notation is as before, corresponding to (Molenberghs *et al*., 2010). When over-dispersion random effects θ_{ij} are gone astray, with then $\lambda_{ij} = \kappa_{ij,}$ the Poisson-normal GLMM results as special case of the Poisson-normal-gamma (combined) model.

As Wedderburn (1974) proposed, one possible itinerary to contend with over-dispersion is to introduce an over-dispersion parameter and only specify a relationship between the mean and the variance, and then apply quasi-likelihood, whereby the extra variability in the data will be captured by the dispersion parameter. According to the following authors: Breslow (1984), Hinde and Demétrio (1998a), and Hinde and Demétrio (1998b), for count data like CD4+ cell counts, it is familiar to combine Poisson distribution with a gamma distributed random effect, so that the unconditional distribution of the outcome turns out to be a negative binomial distribution.

Conversely, as Engel and Keen (1992), Molenberghs and Verbeke, (2005), and Pinheiro and Bates (2000) focus on hierarchical data, the GLM is usually extended to generalized linear mixed

models (GLMMs), with a subject-specific random effect, habitually a Gaussian form, added in the linear predictor to capture a hierarchy-induced association or to account for over-dispersion.

Different statistical methods were used in the above literatures. However, none of them did not contribute to the issue of over-dispersion and correlation simultaneously. Thus, though linear mixed models were used to handle correlation, it could not handle over-dispersion, nor did Cox proportional hazards. Hence, this current study contributes to the problem of over-dispersion and correlation by applying combined models (P-N-G) which takes in to consideration both over-dispersion and correlation simultaneously.

CHAPTER THREE

3. METHODS OF DATA ANALYSIS

3.1. Study area and period

The study was conducted in Hossana Queen Elleni Mohamad Memorial Hospital, SNNPR, Ethiopia from September, 2011 to May, 2014. Hadiya zone is one of 13 zones in SNNPR. There are 10 woredas and one town administration in Hadiya zone. Hosanna town is administrative center for Hadiya zone and it is 235 km away from Addis Ababa. In Hosanna town there is one hospital and health center which gives preventive, curative and rehabilitative service for the catchment area population. ART clinics in Hosanna hospital and health center give an ART service for HIV patients, a total of 1,745 have been enrolled, 910 ever started ART and 623 patients are currently on ART.

3.1.1. A short history of Queen Elleni Mohamed Memorial Hospital

The hospital was built in 1976 E.C. (1984) and its first name was Colonel Mengistu Haile Mariam hospital.

The hospital is one of the first of its kind in the region and has been serving millions since its establishment. Following the fall of the Derg regime, the name of the hospital was changed to Hosanna Hospital in 1985 E.C. (1993). It has got its current name, Queen Elleni Mohamad Memorial Hospital, in 1997 E.C. (2005). The change was made to recognize the roles played by Queen Elleni at the national level.

[*Source: Glan Clayd Hossana link, Central North Wales, UK, and staff at Hossana hospital in Southern Ethiopia*].

3.2. Study design

A retrospective longitudinal study design was conducted to apply appropriate modeling approach to CD4+ cell progression and identify the potential risk factors affecting the CD4+ cell progression.

3.3. Study Population

3.3.1. Source population and sample size

Person living with HIV/AIDS, age ≥15 years and started ART treatment in Hossana Queen Elleni Mohamad Memorial Hospital. The sample size for this study was 222 HIV positive patients in the HAART with a minimum of one and maximum of nineteen measures. The sample was selected by using systematic sampling technique by assigning a number to each of the eligible patients and having determined the first one randomly. Then, every of the tenth patient in the registration book up to proposed sample size based on the interest of the researcher was selected.

3.4. Eligibility Criteria

Inclusion criteria

- HIV positive adults aged 15 years or older who started ART
- HIV positive patients with complete intake form, registers and follow up form

Exclusion criteria

- Diagnosis made outside of the hospital (suspicious for their test result)
- Loss to follow up (withdraw, referred to, and died)
- Women who were pregnant at the time of ART initiation and lactating mother (due to the reason that most of the time tablets are not ordered by doctors based on stage of pregnancy)

3.5.Variables and Data description

As described in section 3.3 the data was obtained from Hossana Queen Elleni Mohamad Memorial Hospital. Patents' follow up time was six months gab according to the order of the doctor, though does not work for all patients and the data were recorded on patients' medical follow up card by assigning an identification number per individual by health workers in the ART, which helps to find the patients profile easily during his/her next visit time, as described below in section 3.7.

The outcome variable

The outcome variable was Enumeration of CD4+ cell counts (CD4+ T cells) per milliliter cube of blood of ART patients, which is count (since each patient's CD4 cell is counted every six months). The outcome for each individual was measured in every six months interval.

Covariates

Eleven covariates were used to meet the goal of the study. Three of these are continuous and eight of them are categorical covariates.

Table-1: Covariates which were used for the analysis of data in this study

No.	Name	Values/ Codes
1.	Test time	Month
2.	Age of Patient at ART initiation	Year
3.	Sex of patient	Male =1, Female =0
4.	Level of Education	No education=1, Primary=2, Secondary=3, Tertiary=4
5.	WHO Stage of the disease	I=1, II=2, III=3, IV=4
6.	Area/residence of the patient	Rural=0, Urban=1
7.	Adherence to any of the drugs	No=0, Yes=1
8.	Employment Status	Unemployed=1, Employed=2, Working fulltime=3, Other=4
9.	Religion	Muslim=1, Orthodox=2, Protestant=3, Other=4
10.	Marital Status	Never married=1,Married=2,Widowed=3, Divorced=4, Separated=5
11.	Baseline CD4+ cell counts	Cells per ML cube of blood

The covariates in the above table described as baseline age, time since a month of seroconversion and

baseline CD4+ cell counts are continuous covariates. Sex of the patient is the first categorical covariate with two levels: male and female. The second one is level of education with four levels: no education (not attended any of schools), primary (who learnt at most eighth grade), secondary (who learnt any of grades nine, ten, eleven, or twelve), and tertiary (who attended any of higher institutions). The third categorical covariate is (WHO) clinical Stage which is classified into four stages; I, II, III and IV; where stage I indicates asymptomatic disease, stage II indicates mild disease, stage III indicates advanced disease and stage IV indicates severe disease. The fourth one is area/residence of the patient classified as rural which is far from any of clinical centers in the zone and not near to any of the towns; urban which is for towns including Hossana town. The fifth categorical one is adherence to any of the drugs classified as Yes for patients who followed any of the drug combinations properly as recommended; No for patients who did not follow any of the drug combinations properly as recommended. The sixth one is employment status classified as unemployed who has no work in any of organizations or dependent, employed who works in any of organizations, working fulltime who is employed at any of organizations, but has no extra work as source of income, Other for any cases other than listed here. The seventh one is religion classified as Muslim, Orthodox, Protestant, and other for any of the religion not listed here. The last covariate is marital status having the levels: never married, married, widowed, divorced and separated.

Time: Observation time of CD4+ T cells (Every six months).

CD4+ T cells: number of CD4 cells per cubic milliliter of blood measured for each individual in every six months.

3.6.Data collection and quality control

The data was collected by reviewing pre-ART register, laboratory request, monthly cohort form, and follow up form, ART intake form, patients' card and death certificate complemented by registration by home visitors. The most recent laboratory results before starting ART were used as a base line value. A total of two days training was given for all supervisors and data collectors. The overall activity was controlled by the researcher. Data quality was controlled by designing the proper data collection/reviewing materials and through continuous supervision. All completed data collection form was examined for completeness and consistency during data management, storage and analysis.

3.7.Check list for data collection

The checklist consists of the following data

- Socio demographic data
- Base line clinical, laboratory and ART data
- ART treatment
- Follow up data

3.8.Ethical Considerations

Ethical clearance was obtained from Jimma University, College of Public Health and Medicine. And, the official ethical clearance also was obtained from Hossana Queen Elleni Mohamad Memorial Hospital medical director. To keep the confidentiality, the data collectors extracted the necessary data from the patient baseline and follow up card. Moreover, no personal identifier was used on data collection form. The recorded data was not accessed by a third person except the researcher, and was kept confidentially. Thus, the data obtained by checklist was organized by the researcher.

3.9. Statistical Methods for Data Analysis

3.9.1. Exploratory data analysis

It is a technique to visualize the patterns of data relative to research interests. Since exploratory data analysis can serve to discover as much of the information regarding raw data as possible, plotting individual curves to carefully examined the data should be performed first before any formal model fitting is carried out. Thus, this study assessed the nature of the data by exploring individual profiles, and the average evolution.

a) **Exploring the individual profile:** To explore the individual profile, plot of the response with time was used to show whether there is a noticeable pattern common to most subjects. These individual profiles can also provide some information on within and between subject variability.
b) **Exploring the Mean Structure:** The major purpose of exploring the mean structure is to choose the fixed effects for the model.

To explore the overall mean, we plot the response variable against time including individual and overall mean profiles. In line with the overall mean, the possible differences between the groups will be studied by plotting the mean of each group separately with the same figure.

Variance Function

The variance function is used to model non-systematic variability (Kachman, 1998). Typically, with a generalized linear model, residual variability arises from three sources: (i), variability arises from the sampling distribution. For example, a Poisson random variable with mean μ has a variance of μ. (ii), additional variability, or over-dispersion, is often observed. Modeling variability due to the sampling distribution is straight forward. For Poisson sampling distribution, the variance function is given as: link function=log, inverse $link = e^{\eta}$, $\text{var}(\mu) = \mu$.Variability due to over-dispersion can be modeled in a number of ways. One approach is to scale the residual variability as $\text{var}(y_i \mid u) = \phi v(\mu_i)$, where ϕ is an over-dispersion parameter. A second approach is to add an additional random effect, $e_i \sim N(0;\phi)$, to the linear predictor for each observation. (iii) Select another distribution. For example, instead of using a one parameter (μ) Poisson distribution for count data, a two parameter (μ; ϕ) negative binomial distribution could be used. The three approaches all involve the estimation of an additional parameter, ϕ. Scaling the residual variability is the simplest approach, but can yield poor results (Kachman, 1998).

The discussion was made on how estimates of the variance components can be obtained. Theoretically the variance component problem can be busted into two gears: (i), the estimation of the variance components associated with the random effects. (ii), the estimation of the variance components associated with the error distribution.

3.9.2. Statistical Models

Generalized linear models are usually in use for modeling univariate non-Gaussian data. GLMs take in a wide range of statistical models for counts, binary data, rates and ratios, time-to-event data, and others (Molenberghs and Verbeke, 2005 and Agresti, 2002).

The first distinguishing feature of the Poisson model is that the variance is equal to the mean. However, in many applications with count data, the observed variance is greater than the mean, which is termed as over-dispersion and it can also be less than what the model has in which case it is said under-dispersion to occur (Agresti, 2002). We will generically use the term over-dispersion to actually encompass both, even though under-dispersion comes with its specific issues Kassahun *et al.*, (2012). For conventional GLM and over-dispersion models, full maximum likelihood is a noticeable estimation itinerary. As made use of by these authors, certain over-dispersion models come down to a modification of the score equation, for which so-called quasi-likelihood can be used.

One of the goals of this study will be to determine the appropriate model for CD4+ cell counts data and make available an implementation of the proposed general approaches with the software package, SAS procedure NLMIXED. Thus, as Kassahun et al., (2012) made use of, although modeling of count data like CD4+ cell counts to handle over-dispersion and correlation seems at first scene relatively complex, it can be implemented easily in the available standard software packages like SAS.

3.9.2.1. Generalized linear models

According to the authors, McCullagh and Nelder (1989), Agresti (2002), and Molenberghs and Verbeke (2005), although it is a very useful framework, there are some situations where general linear models are not appropriate: (i) the range of Y is restricted (e.g. count data like CD 4+ cell counts) (ii) the variance of Y (outcome) depends on the mean. Generalized linear models extend the general linear model framework to address both of these issues, to a linear combination of predictor variables. Generalized linear model is specified by three components: (i) the random component identifies a vector of observations of Y and its probability distribution; (ii) the systematic component is a specification for the vector μ in terms of a vector of p fixed unknown parameters β; (iii) and the link function specifies the function of E(Y) that the model equates to the systematic component.

A big family of probability density functions is called an exponential family distribution η if it can be expressed as:

$f(y) \equiv f(y/\eta,\phi) = e^{\{\phi^{-1}[y\eta-\psi(\eta)]+C(y,\phi)\}}$, where η and ϕ are unknown parameters, and $\psi(.)$ and C (.,.) are known functions, η and ϕ are termed 'natural parameter' (or 'canonical parameter') and 'scale parameter', respectively.

3.9.2.1.1. Poisson model

The Poisson distribution belongs to the exponential family and is commonly used for the analysis of count data, like CD4+ cell counts. In this case, the distribution of the response is Y ~ Poisson (θ). Anyone may be interested to explain variability between outcome(Y) values based on covariate values with density function.

$$f(y_i,\theta_i) = \frac{\theta_i^{y_i} e^{-\theta}}{Y_i}. \qquad (3.1)$$

The linear predictor $g(E(Y_i)) = \log(\mu_i) = \eta$, $\mu_i = e^{\eta} = e^{\beta_0+\beta_i X_i}$

The mean is given by $E(Y) = \mu = \theta$ and the variance by $\text{var}(Y) = \theta$. Here, the scale parameter equals $\phi = 1$. Suppose $Y_1,\ldots, Y_N$ is a sample of independent counts and let $x_1,\ldots, x_N$ represent the corresponding p-dimensional vectors of covariate values. The Poisson regression model, with β a vector of p fixed, unknown regression coefficients is given by $\log(\lambda) = x_i'\beta$.

3.9.2.2. Over-dispersion Models

According to Wedderburn (1974), one possible way to deal with over-dispersion for counts based on binary data is to allow for the over-dispersion parameter $\phi \neq 1$ and only specify a relation between the mean and the variance, and then apply quasi-likelihood estimation. A simple quasi-likelihood approach uses the variance function, $\text{var}(\pi_i) = \theta\pi_i(1-\pi_i)/n_i$. In this context, if $\theta > 1$, over-dispersion is said to occur. A key assumption of the GLM Poisson model is that the variance is equal to the mean, $\text{var}(\mu) = \mu = \theta$. However, as Agresti (2002) suggested, in many applications with count data, the observed variance is higher than the mean, leading to over-dispersion. In the CD4+ cell counts data which was described in section 3.5, suppose that Y_{ij} denote the number of CD4+ cell counts collected from patient i at time j. These counts may vary from patient to patient based on factors Age of Patient, Sex of patient, Level of Education, Adherence to any of the drugs, Test time, Drug Regimen, Base line CD4 cell counts, Marital Status, Religion, Location/ residence of the patient, and Employment Status at treatment initiation, which in turn will induce heterogeneity, leading to more variation in the data than predicted by the Poisson model (over-dispersed).

Like the clustered binary and binomial data, we can apply quasi-likelihood estimation here (Wedderburn, 1974). Furthermore, if $\theta > 1$, over-dispersion is said to occur. As some authors implied, an alternative approach to modeling over-dispersion in count data is combining a Poisson distribution with a random effect θ_i to handle the unobserved heterogeneity. Then, $Y_i \mid \theta_i \sim Poisson(\theta_i\mu_i)$. As the authors Breslow (1984), Hinde and Demétrio (1998a), and Hinde and Demétrio (1998b), brought about, since θ_i is unobserved, it is common to assume a gamma distribution, so that the uncondition distribution of the outcome turns out to be a negative binomial distribution. The negative binomial distribution has mean $E(Y) = \mu$ and variance $\text{var}(Y) = \mu(1+\sigma^2)$, where σ^2 is the variance of the unobserved term. If $\sigma^2 > 0$, the variance is greater than the mean, implying that the negative binomial allows for over-dispersion. When $\sigma^2 = 0$, the Poisson model results as a special case.

3.9.2.3. Generalized linear mixed models

A generalized linear mixed model is a model which gives us extra flexibility in developing an appropriate model for the data (Breslow, and Clayton, 1993). Linear mixed models provide a powerful means of predicting changes in CD4+ cell counts.

However, for many circumstances the assumptions of linear responses, constant variance, and normality are questionable. Generalized linear mixed models provide a means of modeling these deviations from the usual linear mixed model. This study will examine what constitutes a generalized linear mixed model, issues involved in constructing a generalized linear mixed model (Kachman, 1998).

In this section the linear mixed model and when the implied assumptions are not appropriate as was discussed. A linear mixed model is $Y \mid u \sim N(X\beta + Zu; R)$ where $u \sim N(0; G)$, X and Z are known design matrices, and the covariance matrices R and G may depend on a set of unknown variance components. The linear mixed model assumes that the relationship between the mean of the dependent variable Y and the fixed and random effects can be modeled as a linear function, the variance is not a function of the mean, and that the random effects follow a normal distribution. Any or all these assumptions may be violated for certain circumstances.

A case where the assumption of linear relationships is questionable is number of CD4+ cell counts. Thus, due to some factors it may continuously decreases and with the modification of the treatments it may somehow improve.

A number of approaches have been taken to address the deficiencies of a linear mixed model. Transformations have been used to stabilize the variance, to obtain a linear relationship, and to normalize the distribution. However, the transformation needed to stabilize the variance may not be the same transformation needed to obtain a linear relationship. For example, a log transformation to stabilize the variance has the side effect that the model on the original scale is multiplicative. Linear and multiplicative adjustments are used to adjust to a common base and to account for heterogeneous variances. We have a set of estimation procedures which are based on a linear mixed model and manipulate the data to make it fit a linear mixed model.

It seems more reasonable to start with an appropriate model for the data and use an estimation procedure derived from that model. A generalized linear mixed model is a model which gives us extra flexibility in developing an appropriate model for the data (Breslow, and Clayton, 1993).

According to Engel and Keen (1992), Molenberghs and Verbeke (2005), and Pinheiro and Bates (2000),

when non-Gaussian data are repeatedly measured like CD4+ cell counts, the GLM is usually extended to generalized linear mixed models (GLMMs), with a subject-specific random effect, usually a Gaussian type, added in the linear predictor to capture the correlation. According to these authors, GLMMs combine the properties of linear mixed models and generalized linear models.

Suppose that Y_{ij} is an outcome, number of CD4+ cell counts for the i^{th} subject measured at the j^{th} time point(in month), and b_i are assumed to be normally distributed with mean 0 and variance-covariance matrix D, that is $b_i \sim N(0, D)$, with $E(b_i) = 0$ and $\text{var}(b_i) = D$. Then, it is assumed that the conditional distribution of the response, $Y_{ij} | b_i$ is independent and belongs to the following exponential family density

$$f_i(y_{ij} \mid b_i, \theta) = \exp\{\theta - 1[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y, \theta)\} \quad (3.2)$$

The expectation is $E(Y_{ij} \mid b_i) = \mu_{ij} = \eta^{-1}(x_{ij}^{'}\beta + z_{ij}^{'}b_i)$, where η (.) is a known link function, x_{ij} is a p-dimensional design matrix of the fixed effect parameters β, and z_{ij} is a q-dimensional design matrix of the random effects b_i (Kassahun *et al*., 2014).

The likelihood contribution of subject i is

$$f_i(y_{ij} \mid \beta, b_i, \varphi) = \int \prod_{j=1}^{ni} f_{ij}(y_{ij} \mid \beta, b_i, \varphi) f(b_i \mid D) db_i. \quad (3.3)$$

From this, the likelihood for β, D and φ is given as

$$L(\beta, D, \varphi) = \prod_{i=1}^{N} \int \prod_{j=1}^{ni} f_{ij}(y_{ij} \mid \beta, b_i, \varphi) f(b_i \mid D) db_i, \quad (3.4)$$

Thus, as Molenberghs and Verbeke (2005) and Skrondal and Rabe-Hesketh (2004) show, expression (3.4) does not have an analytical solution, and hence numerical approximations are needed.

For count data (CD4+ cell counts, in this study), let Y_{ij} be the value of the number of CD4+ cell counts for patient i and time point j. It is customary to assume that

$$Y_{ij} \sim Pois(\theta_{ij}) \quad (3.5)$$

With the conditional mean θ_{ij} modeled as $\theta_{ij} = \exp(x_{ij}^{'}\beta + z_{ij}^{'}b_i)$.

The variance components can be estimated using the approximate REML quasi-likelihood (Breslow and Clayton 1993) which after some algebra is

$$ql(\beta; \sigma) = -1/2ln|V| - 1/2ln|X'H'\text{v} - 1HX| - 1/2(\text{y} * -HX\beta)'\text{v} - 1(\text{y} * -HX\beta) \quad (3.6)$$

, where σ is the vector of variance component and $V = R + HZGZ'H'$. For the variance components

associated with the random effects in G the estimating equations remain the same.

Estimation of the variance components associated with the error distribution is more problematic. The quadratic form becomes

$$\left(y-\hat{\mu}\right)' R-1\frac{\partial_R}{\partial_{\emptyset}} R1\left(y-\hat{\mu}\right). \tag{3.7}$$

However, the corresponding functions for the linear mixed model assumes that R=Iσ^2_0. The functions of the left hand sides for ϕ are

$$f_{00}(C) = [tr(\Phi) - 2\,tr(\Omega\,\Phi) + tr(\Omega\,\Phi\,\Omega\,\Phi)]$$

$$f_{i0}(C) = \left[tr\left(C^i\begin{pmatrix} X'H'\Phi HX & X'H'\Phi HZ \\ Z'H'\Phi HX & Z'H'\Phi HZ \end{pmatrix}C^{i'}\right)\right]$$

$$where\ C^i = (C^{i0}\quad C^{i1}\quad \ldots\quad C^{ir}),\ \Phi = R-1\frac{\partial_R}{\partial_{\emptyset}}R-1$$

$$\Omega = (\mathrm{HX}\quad \mathrm{HZ})\mathrm{C}\begin{pmatrix}\mathrm{X'H'} \\ \mathrm{Z'H'}\end{pmatrix}. \tag{3.8}$$

3.9.2.4. Poisson-normal-gamma models for over-dispersion and correlation

Correlation and over-dispersion can crop up simultaneously, in practice. So that some authors have been proposed some techniques and they have formulated a flexible and unified modeling framework to overcome this, which afterwards termed as the so-called combined model, to handle a wide range of hierarchical data, including count, binary, and time-to-event data.

Molenberghs *et al*., (2010) together with other authors brought two sets of random effects. The normally distributed subject-specific random effects capture the correlation, while the conjugate measurement specific random effects on the natural parameter which leads to the beta-binomial model for binary data and the negative-binomial model for count data, and are used to put up over-dispersion. The Poisson-normal-gamma model having both the over-dispersion and normal random effects takes the form:

$f_i(y_{ij} \mid b_i, \phi) = \exp\{\phi - 1[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y,\phi)\}$. The expectation is,

$E(Y_{ij} \mid b_i) = \mu_{ij} = \theta_{ij}\kappa_{ij}, where\,\theta_{ij} \sim g_{ij}(\vartheta_{ij}, \sigma^2_{ij}), \vartheta_{ij} and \sigma^2_{ij}$ are mean and variances of θ_{ij}, respectively. The likelihood contribution of subjection for the model is given as:

$f_i(y_i \mid \vartheta, D, \vartheta_i, \Sigma_i) = \int \prod_{j=1}^{n} f_{ij}(y_{ij} \mid \vartheta, b_i, \theta_i) f_i(b_i \mid D) f(\theta_i \mid \vartheta_i, \Sigma_i) db_i d\theta_i$. From the above, the likelihood contribution of all subjects in common is given as below:

$$L(\vartheta, D, \vartheta, \Sigma) = \prod_{j=1}^{N} f_i(y_i \mid \vartheta, D, \vartheta_i, \Sigma_i)$$

$$= \prod_{j=1}^{N} \int \prod_{j=1}^{n} f_{ij}(y_{ij} \mid \vartheta, b_i, \theta_i) f(b_i \mid D) f(\theta_i \mid \vartheta_i, \Sigma_i) db_i d\theta_i \quad (3.9)$$

The Poisson-normal-gamma model assuming that counts are generated from a Poisson-normal-gamma process has mean $\lambda c_{ij} = \theta_{ij} \kappa_{ij} with \theta_{ij} \sim Gamma(\alpha, 1/\alpha)$, where the subscript 'c' refers to the conditional, as Kassahun *et al*., (2014).

For this study, CD4+ cell count data, let Y_{ij} be the j^{th} outcome repeatedly measured for patient for i=1,…, N, time point j=1, …, n_i. The Poisson model with normal and gamma random effects can be specified as

$$Y_{ij} \sim Pois(\lambda_{ij} = \theta_{ij} \kappa_{ij}) \quad (3.10)$$

, with the conditional mean λ_{ij} modeled as $\theta_{ij}\kappa_{ij}$ and $\kappa_{ij} = \exp(x_{ij}^{'}\beta + z_{ij}^{'} b_i)$

$b_i \sim N(0, D), and \theta_{ij} \sim Gamma(\alpha, \beta), x_{ij} and \ z_{ij}$ P-dimensional and q-dimensional vectors of known covariate values, and β a p-dimensional vector of unknown fixed regression coefficients.

Molenberghs *et al.,* (2007) and Molenberghs *et al.*, (2010) marginalized the Poisson-normal-gamma model analytically over the gamma random effect, whereby this partially marginalized model takes the form:

$$f(y_{ij} \mid b_{1i}, \beta) = \int f(y_{ij} \mid b_{1i}, \beta, \theta_{ij}) f(\theta_{ij} \mid \alpha_j, \beta_j) d\theta_{ij}$$

$$= \binom{a_j + y_{ij-1}}{a_j - 1} \left(\frac{\beta_j}{1 + \kappa_{ij}\beta_j} \right)^{y_{ij}} \left(\frac{1}{1 + \kappa_{ij}\beta_j} \right)^{a_j} \kappa_{ij}^{\ y_{ij}}$$

Herein after, further numerical integration over the normal random effects can be made to obtain the maximum likelihood estimates, marginal expressions for the mean vector $E(Y_{ij})$ and variance-covariance of Y_i.

Here one has to mind that the Poisson-normal GLMM results as an exceptional case of the Poisson-normal-gamma model, when over-dispersion random effects θ_{ij} are gone astray, with a conditional mean given by:

$$Y_{ij} \sim Pois(\kappa_{ij}) \quad (3.11)$$

As Kassahun *et al*., (2014) used, one can apply the following notations conventionally. The Poisson-normal-gamma model which brings both features together is denoted as (PNG), where 'P' refers to basic Poisson model, 'N' refers to normal random effects and 'G' stands for gamma random effects. The extraordinary case, which follows by excluding the gamma random-effects structures, i.e., the Poisson-normal GLMM is denoted as (PN-), and leaving out only the normal-random effects by (P-G), but the simplest case is when both random-effects are dropped, leading to Poisson GLM model (P--).

Note that there are other examples of models including non-normal random effects. For example, Lee and Nelder (2001, 1996), in a series of influential papers, proposed hierarchical generalized linear models, where random effects can be non-normal. In addition, these authors proposed a computationally feasible estimating method, both in numerical as well as in statistical terms.

Variable selection technique: In all models, to select significant variables, first the main effect and main effect by time interaction was incorporated to the initial candidate model. After that, avoiding non-significant variables one by one starting from the most non-significant terms or simply backward selection technique.

The NLMIXED procedure for model comparison

The PROC NLMIXED procedure fits nonlinear mixed models—that is, models in which both fixed and random effects enter nonlinearly.

It fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects. Different integral approximations are available, the principal ones being adaptive Gaussian quadrature and a first-order Taylor series approximation. A variety of alternative optimization techniques are available to carry out the maximization; the default is a dual quasi-Newton algorithm.
Successful convergence of the optimization problem results in parameter estimates along with their approximate standard errors based on the second derivative matrix of the likelihood function. PROC NLMIXED uses adaptive Guassian quadrature method by default. This procedure enables you to analyze data that are normal, binomial, or Poisson or that has any likelihood programmable with SAS statements.
Poisson and negative binomial regression models are designed to analyze count data.

However, Poisson and negative binomial regression models differ in regards to their assumptions of the

conditional mean and variance of the dependent variable. Poisson models assume that the conditional mean and variance of the distribution are equal. Negative binomial regression models do not assume an equal mean and variance and particularly correct for over-dispersion in the data, which is when the variance is greater than the conditional mean (Osgood, 2000; Paternoster *et al*., 1997).

Although the AIC can be used in association with mixed models, it is not common to be used with the models discussed above to select either the optimal set of explanatory variables or other structures. Hence, finally the four models were compared using −2log-likelihood comparison technique.

Nonlinear Mixed-Effects Models

We suppose there are N independent clusters, with n_i members in the i^{th} cluster. Let Y_{ik} denote the outcome for the k^{th} member of the i^{th} cluster, i = 1, . . . , N; k = 1, . . . , n_i. Let x_{ik} be the covariate vector for the k^{th} member of the i^{th} cluster, and b_i a random effect associated with the i^{th} cluster. The random effects have density function $f(b_i|\theta)$, with θ the vector of parameters for the distribution of b_i. The distribution of Y_{ik} given x_{ik} and b_i has density $f(y_{ij}|x_{ij}, b_i, \beta)$, indexed by the parameter vector β. We allow both $f(y_{ij}|x_{ij}, b_i, \beta)$ and $f(b_i|\theta)$ to have any general parametric density. Given the random effects b_i, we assume the Y_{ik}'s within a cluster are mutually independent. The MLE of (β, θ) is obtained by maximizing the marginal likelihood. The i^{th} cluster's contribution to the marginal likelihood is

$$L_i(\beta,\theta) = f(y_{i1},...,y_{ini} \mid x_{i1},...,x_{ini},\beta,\theta) = \int_{b_i}[\prod_{k=1}^{n_i} f(y_{ik} \mid x_{ik},b_i,\beta) f(b_i \mid \theta)]db_i, \quad (3.12)$$

And thus, the likelihood to be maximized is

$$L_i(\beta,\theta) = \prod_{i=1}^{N} \int_{b_i}[\prod_{k=1}^{n_i} f(y_{ik} \mid x_{ik},b_i,\beta) f(b_i \mid \theta)]db_i, \quad (3.13)$$

Clearly, (3.13) involves integration over the distribution of b_i and, in general, does not have a closed form. A number of methods are available for maximizing this likelihood directly (Geyer and Thompson, 1992) or via approximations (Pinheiro and Bates, 1995; Lindstrom and Bates, 1990; Breslow and Clayton, 1993; McGilchrist, 1994; Liu and Pierce, 1994; Wolfinger and Lin, 1997), respectively). Many of these methods assume $f(b_i \mid \theta)$ is normal. Numerical integration techniques, such as Gaussian quadrature (Davidian and Gallant, 1992; Liu and Pierce, 1994), are used increasingly for fitting NLME's. These methods are available, for example, in PROC NLMIXED in SAS, but only for the case of one or more normally distributed random effects. As we shown, non-normal random effects can be accommodated within the

numerical integration techniques available in PROC NLMIXED, via the use of the probability integral transform.

Suppose that the random effects (assumed continuous) have a non-normal distribution $f(b_i \mid \theta)$, but available software programs are restricted to normal random effects. Let a_i be a random effect from a standard normal distribution, that is, a_i~normal (0, 1). Then, using the probability integral transform (Hoel *et al*., 1971), $u_i = \Phi(a\_i)$ has a uniform (0, 1) distribution, where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF).

Applying the probability integral transform once more, $F_\theta(b_i)$ also has a uniform (0, 1) distribution, where $F\theta(\cdot)$ is the CDF of bi, with parameter θ. It then follows that $b_i = F_\theta^{-1}(u_i)$ has density $f(b_i|\theta)$, where$F_\theta^{-1}(\cdot)$ is the inverse CDF of b_i. Thus, $b_i = F_\theta^{-1}(\Phi(a\_i))$ has the non-normal distribution of interest. Almost all major statistical software packages have the $\Phi(\cdot)$ function built-in, as well as most common inverse CDF's (e.g., beta, gamma, t, and chi-square). Therefore, in terms of the normal random effect a_i, using probability theory for transformations, the i[th] cluster's contribution to the marginal likelihood can be re-written as

$$L_i(\beta,\theta) = \int_{b_i} [\prod_{k=1}^{n_i} f(y_{ik} \mid x_{ik}, b_i, \beta) f(b_i \mid \theta)] db_i,$$

$$db_i = \int_{a_i} [\prod_{k=1}^{n_i} f(y_{ik} \mid x_{ik}, F\theta^{-1}(\Phi(a_i)), \beta) \phi(a_i)] da_i, \qquad (3.14)$$

Where $\phi(.)$ is the standard normal probability density function. The marginal likelihood in (3.14) can be approximated by quadrature and standard maximization methods (e.g., Newton-Raphson, quasi-Newton) can be used to solve for $(\hat{\beta}, \hat{\theta})$.

The idea of using a probability integral transform for modeling a single non-normally distributed random effect can be extended in a straightforward manner to two or more independent non-normal random effects (b_{i1}, b_{i2}, . . .), with densities $f_i(b_{ij})$ and CDFs $F_j(b_{ij})$.

In the case of two independent non-normal random effects (b_{i1}, b_{i2}), the probability integral transform can be applied as follows. Let (a_{i1}, a_{i2}) be two independent normally distributed random effects, with zero means and each with a variance of one, such that$(a_{i1}, a_{i2}) \sim Normal(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix})$.

Let $u_{i1} = \Phi(a_{i1}) and u_{i2} = \Phi(a_{i2})$, both with a uniform (0, 1) distribution. Then applying the probability

integral transform to each variable, F_1 (b_{i1}) and F_2 (b_{i2}) also have independent uniform distributions, so that $b_{i1} = F_1^{-1}(u_{i1})$ has density f_1 (b_{i1}) and $b_{i2} = F_2^{-1}(u_{i2})$ has density f_2 (b_{i2}). The extension to more than two independent non-normal random effects follows in a similar manner. When correlation is present between two or more non-normal random effects, the probability integral transform method becomes more complicated and requires the use of a multivariate probability integral transform approach (Genest and Rivest, 2001).

However, if one defines $g_i = \theta_1 g_{i2}$, then g_i has the gamma distribution. Letting $F_{\theta 1}^{-1}(\cdot)$ denote the inverse CDF (often referred to as the "quantile function") of the gamma distribution, and $\Phi(\cdot)$ be the standard normal CDF, then, when using PROC NLMIXED in SAS, the gamma random effect g_i is obtained via the following set of transformations:

1. $a_i \sim N(0, 1)$
2. $p_i = \Phi(a_i)$
3. $gi2 = F_{\theta 1}^{-1}(p_i)$
4. $g_i = \theta_1 g_{i2}$.

As θ_1 approaches 0, observations within a cluster are independent, while large values of θ_1 induce high within-cluster correlation.

3.9.2.5. Estimation Techniques

It is straightforward to obtain the fully marginalized probability by numerically integrating over the normal random effects, and using a tool such as the SAS procedure NLMIXED that allows for normal random effects in arbitrary, user-specified models as Kassahun *et al*., (2014).

Estimation Methods for over-dispersion models

Maximum Quasi-Likelihood

For constant over-dispersion models with a variance of the form $\text{var}(Y_i) = \mu v_i(\mu_i)$ to estimate the regression parameters maximize the quasi-likelihood

$Q = 1/2 \sum_{i=1}^{n} \left\{ \frac{D(yi,\mu i)}{\phi} \right\}$, where D the variance function

$D(y, \mu) = -2 \int_y^{\mu} \frac{y-t}{v}(t)dt$, $\hat{\beta}$ is the same as the maximum likelihood estimation for the exponential family with variance function V (μ), e.g. for V (μ) = μ – Poisson (John, 2007).

For over-dispersed Poisson model, $\hat{\phi} = 1/(n-p)\sum_{i=1}^{n}(yi - \hat{\mu}i)^2 / \hat{\mu}i$

The standard errors of the $\hat{\beta}$ are used as for the non dispersed model inflated by $\sqrt{\tilde{\phi}}$.

Model selection

Fitzmaurice (1997) uses the Akaike information criterion (AIC) and -2likehood to compare deviances from different models. But, as the models are not nested, -2likehood was used to select the best model.

Testing Over-dispersion

Obtaining a goodness of fit test for over-dispersed models is not as simple as in say fitting a binomial or Poisson model where the residual deviance or Pearson Ch-square can often be used, at least as an approximate test.

Model Checking Technique

In GLMM, it is assumed that the random effects are normally distributed and uncorrelated with the error term. Residual plots can be used visually to check normality of these effects and to identify any outlying effect categories. Examining the plot of the standardized residuals versus fitted values by any covariates of interest can give a better feeling. The assumption of normality for the within-group error was assessed with the normal probability plot of the residuals by covariates. Similarly, Normality of the random effects is assessed using Normal Plot of each random effect. Normal plot of estimated random effects helps for checking marginal normality and to identify outliers.

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Baseline Information and Descriptive Statistics of CD4+ cell counts

A total of 222 adult ART HIV-1-positive patients (1047observations) were used with a minimum of one and maximum of nineteen measures of CD4+ cell counts per individual patients. The baseline demographic characteristics and descriptive statistics of patients are displayed in table-2 below. Out of these ART patients, 131(59%) were female patients and 91(41%) were male patients, 175(78.8%) were living around rural out of Hossana town and 47(21.2%) were living around urban centers including Hossana town, the mean and standard deviation of baseline CD4+ cell counts are 355.9 and 321.4 cells per milliliter of blood, respectively, the mean and standard deviation of age of patients are 31.06 and 8.5 years, respectively, 57(25.67%) are not educated, 98(44.1%) learnt primary education, and 52(23.4%) learnt secondary and above, 35(15.80%) are never married, 128(57.70%) are married, 24(10.80%) are widowed, 19(8.60%) are divorced, and 16(7.20%) separated, 26(11.70%) are Muslim, 63(28.4%) are Orthodox, 119(53.60%) are protestant, 5(2.30%) are Catholic, and 9(4.1%) are Others, 57(25.70%) are unemployed, 60(27.00%) are employed, 39(17.60%) are working full time, 65(29.30%) are following the drug combinations properly and 157(70.70%) are not taking the drug combinations properly, most of the patients 68(30.60) are in WHO stage III, the minimum and maximum CD4+ cell counts are 7 and 2000 cells per milliliter of blood, respectively, the mean and standard deviation of CD4+ cell counts were 438.511 and 307.75 cells per milliliter of blood, respectively.

Table-2: Baseline average value and standard deviation of covariates with their Percentages and frequencies for CD4+ cell counts of HIV-1-positive patients' data.

		CD4+ cell counts [mean (St.d) & # patients (%)]	
Variables	Levels	Mean(St.d)	# Patients (%)
Baseline CD4+ cell counts	-	355.891(321.455)	222(100)
Age of patients	-	31.067(8.521)	222(100)
Educational level	No education	288.000(66.30)	57(25.67)

of patients	Primary	489.491(323.71)	98(44.10)
	Secondary	409.077(279.44)	52(23.4)
	Tertiary	444.469(328.37)	15(6.76)
Sex of patients	Female	338.061 (330.467)	131(59.00)
	Male	381.560(308.00)	91(41.00)
Marital Status of patients	Never married	432.714(416.537)	35(15.80)
	Married	331.335(276.971)	128(57.70)
	Widowed	369.500(352.421)	24(10.80)
	Divorced	323.526(347.029)	19(8.60)
	Separated	402.312(349.103)	16(7.20)
Religion of patients	Muslim	567.730(355.185)	26(11.70)
	Orthodox	376.333(371.689)	63(28.40)
	Protestant	337.969 (286.336)	119(53.60)
	Catholic	317.400(174.321)	5(2.30)
	Other	437.222(385.016)	9(4.10)
Employment Status of patients	Unemployed	289.666(245.476)	57(25.70)
	Employed	1072.000(453.234)	60(27.00)
	Working full time	331.641(261.385)	39(17.60)
	Other	383.687(335.383)	66 (29.73)
Adherence to any of the drugs	No	349.758 (306.857)	157(70.70)
	Yes	370.707(356.357)	65(29.30)
WHO Stage of the disease	I	359.416(332.894)	60(27.00)
	II	413.873(323.824)	71(32.00)
	III	333.647(317.347)	68(30.60)
	IV	233.478(269.075)	23(10.40)
Area of the patients	Rural	364.595(358.506)	175(78.80)
	Urban	353.554(311.835)	47(21.20)

4.1.2. Exploratory analysis for CD4+ cell counts data

The first step in any model-building process is exploratory data analysis as done below.

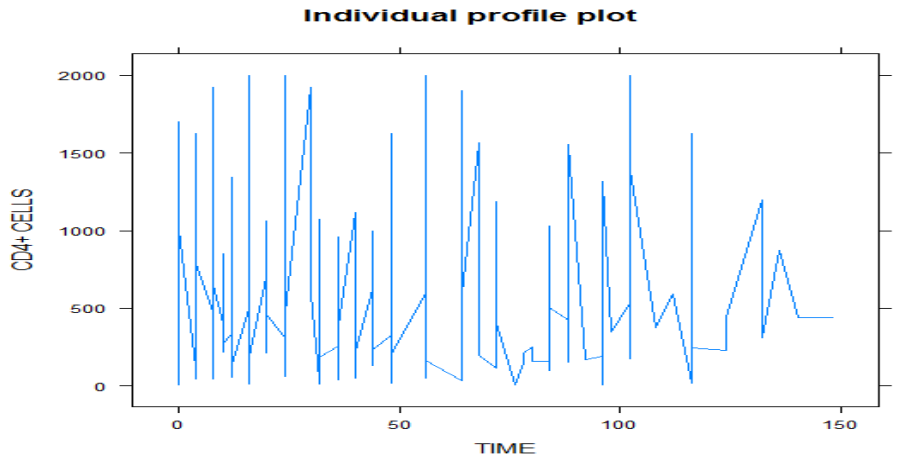


Figure-1: Individual Profile plot for CD4+ cell counts of HIV-1-positive patients

As depicted in figure-1, the individual profile plot indicates that most of the CD4+ cell counts are concentrated around below 500 and there is high variation in CD4+ cell counts at the baseline than at the end and the CD4 cell counts appear to be increasing and decreasing over time, the degree is very high.

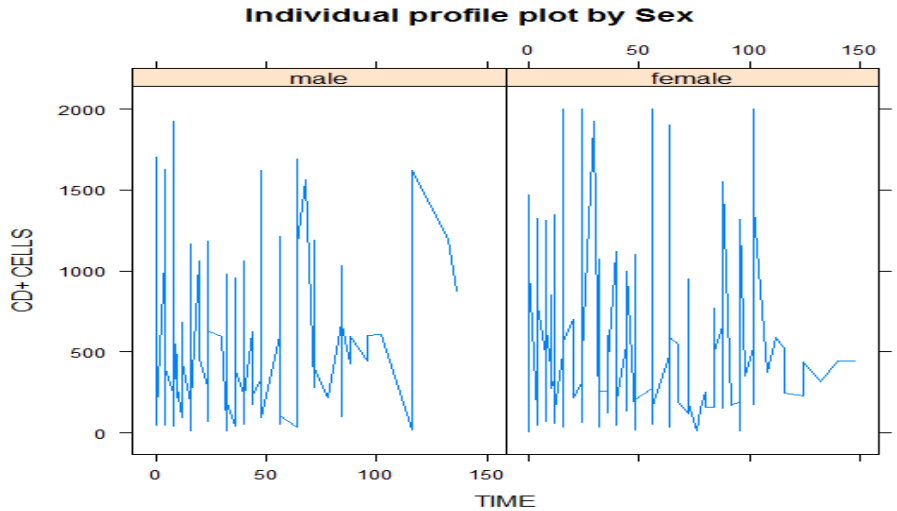


Figure-2: Individual Profile plot for CD4+ cell counts of HIV-1-positive patients by Sex

Furthermore, besides plotting the response over time, it is also useful to include different subgroups on the same graph to illustrate the relationship between the response and an explanatory variable over time.

Figure-2 shows that there is high variation in CD4+ cell counts at the baseline than at the end in both sex, but the variation of CD4+ cell counts in female is more pronounced than in male and the same feature can be seen as for figure-1 above. Thus, between and within-subject specific difference in variation can't be ignored.

Loess Curve

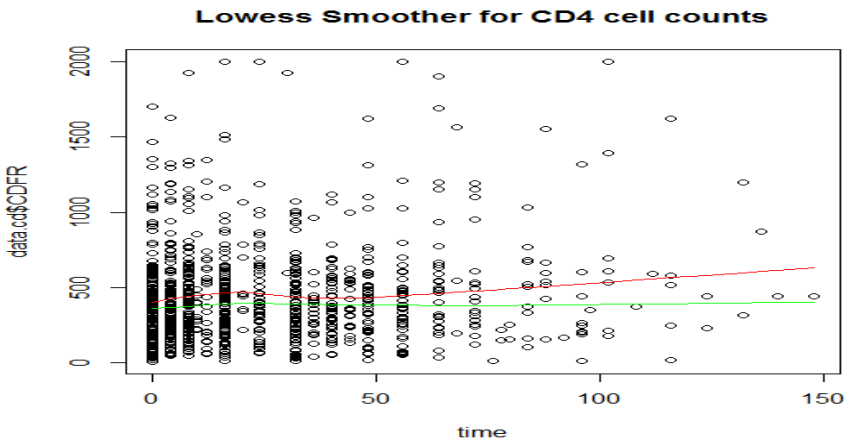


Figure-3: Loess smoother for CD4+ cell counts of HIV-1-positive patients

The loess smooth curve, as shown in figure-3, suggests that the average profile of CD4+ cell counts has a linear relationship over time and almost fairly constant around below 500 cells per milliliter of blood. It indicates that CD4+ cell counts show a slight increasing pattern over time, but the rate of increasing is very low. And also it indicates the linear time effects to be included as fixed-effects in the model.

NB:- Loess smoother and individual profile for most categories of the covariates with two/more levels like adherence to any of the drug combinations, area/residence of patients, and marital status, not displayed here for the sake of redundancy, seem very similar to those in figures 2 and 3, (see appendix 1).

4.2. Statistical Models of Data Analysis

4.2.1. Comparison of Poisson, Poisson-gamma, Poisson-normal, and Poisson-normal-gamma Models

Since the response variable in this study is counts and the data is over dispersed as the variance of CD4+ cell counts (94710.06) is greater than the mean(438.511), the Poisson model is fitted including different random effects(gamma and normal) to handle both over dispersion and correlation, respectively. Table-3 displays the comparison among four models, GLM (Poisson model without random effect to handle over

dispersion), GLMM (Poisson model with normal random effects to oversee the correlation), Poisson model with the gamma random effects (negative-binomial model) to grip extra variation in the data, and Poisson-normal-gamma model, model with both normal and gamma random effects to discuss about both correlation and over dispersion simultaneously

Table-3: Comparison of Poisson, Poisson-gamma, Poisson-normal, and Poisson-normal-gamma Models.

		Models			
Effects and parameters		Poisson	Poisson-normal	Poisson-gamma	Poisson-normal-gamma
Effects	Parameters	Estimate (Se.)	Estimate (Se.)	Estimate (Se.)	Estimate (Se.)
Intercept	β_0	6.055(0.038)	6.050(0.067)	6.131(0.081)	6.136(0.087)
Time	β_1	0.002(0.001)	0.002(0.001)	-0.005 (0.002)	-0.006(0.003)
Age	β_2	-0.006(0.001)	-0.006(0.001)	0.004(0.002)	0.004(0.002)
Sex	β_3	0.052(0.004)	0.052(0.004)	0.073(0.058)*	0.071(0.061)*
Sex*Time	β_4	-0.004(0.001)	0.093(0.031)	-0.004 (0.002)	-0.003(0.002)*
Sigma	σ	0.550(0.030)	1.568E-8(0.082)	-	0.164(0.036)
Negative-binomial parameter	α	-	-	2.170(0.089)	2.287(0.101)
1/ alpha	β	-	-	0.460(0.019)	0.437(0.019)
Variance	d	-	2.46E-16(-)	-	0.027(0.012)
RIS					
−2log-likelihood	-	151870	151975	14536	14527

* Estimates which are not significant at 5% level of significance.

Let's consider the above models where the random intercept b_i is assumed to be zero-mean normally distributed with variance d.

We consider special cases (1) the ordinary Poisson model, (2) the negative-binomial model, (3) the Poisson-normal model, together with (4) the Poisson-normal-gamma model. Estimates (standard errors, p-values) are presented in table-3. Clearly, both the negative-binomial model and the Poisson-normal-gamma model are important improvements, in terms of the likelihood, relative to the Poisson-normal model and ordinary Poisson model. This should come as no surprise since the latter (Poisson model) unrealistically assumes there is neither over-dispersion nor correlation within the outcomes, while clearly both are present. In addition, when considering the Poisson-normal-gamma model, there is a very strong improvement in fit when gamma and normal random effects are simultaneously allowed for as also the over dispersion parameter, Variance RIS (d) is significant (P= 0.0208) implying the presence of considerable extra variability due to the grouped nature of the data, which is beyond what can be accommodated by the combined model. Furthermore, the AIC comparison if one may interested on, also gives the same decision that the combined model (P-N-G) has the least AIC value.

This strongly affects the point and precision estimates of the coefficients of covariates, where all covariates except, time sex interaction in Poisson-normal-gamma model and sex in Poisson-gamma and Poisson-normal-gamma model, are significant in all models for the progression of CD4+ cell counts of ART patients.

Of course, one must not forget that, while the negative-binomial model accommodates over-dispersion, the θ_{ij} random effects are assumed independent, implying independence between repeated measures. Again, this is not realistic and therefore the Poisson-normal-gamma model is a more viable candidate, substantiated further by the abovementioned likelihood comparison.

Table-4: Comparison of Poisson-normal model among random effects for CD4+ cell counts data

Effects and parameters		**Models**	
		Poisson-normal(random intercept only)	Poisson-normal(with Random intercept& slope)
Effects	Parameters	Estimate (Se.)	Estimate (Se.)
Intercept	β_0	6.026(0.038)	6.072(0.007)
Time	β_1	0.003(0.001)	0.004(0.000)
Age	β_2	-0.006(0.001)	-0.004(0.001)
Sex	β_3	0.102(0.005)	0.042(0.004)
Sex*Time	β_4	-0.002(0.002)	-0.003(0.001)
normal random effect	b	-1.188(0.099)	-
var(b_1)	d_{11}	-	0.145(0.002)
Var(b_2)	d_{22}	-	0.001(1.552E-6)
Cov(b_1,b_2)	d_{12}	-	-0.004(0.001)
−2log-likelihood	-	151689	137475

The random slope model strongly improves the fit of the model based on the likelihood comparison. The estimates and standard errors of the covariates are similarly significant in both models for the response variable. All the normal random effects in the slope model are significant implying that the correlation among subjects is evident.

Table-5: Comparison of Poisson-normal-gamma model without and with random effects for CD4+ cell counts data

Effects and parameters		**Models**	
		Combined model(with no random effect)	Combined model(random effect)
Effects	Parameters	Estimate (Se.)	Estimate (Se.)
Intercept	β_0	6.136(0.087)	5.876(0.006)
Time	β_1	-0.006(0.003)	0.003(0.001)
Age	β_2	0.004(0.002)	-0.006(0.001)
Sex	β_3	0.071(0.061)*	0.106(0.005)
Sex*Time	β_4	-0.003(0.002)*	-0.002(0.001)
Negative-binomial parameter	α	2.287(0.102)	2.170(0.089)
1/ alpha	β	0.438(0.019)	0.461(0.019)
Variance RIS	d	-	0.027(0.012)
theta_1	θ	-	0.089(0.001)
−2log-likelihood	-	14527	152476

** Estimates which are not significant at 5% level of significance. Combined model=Poisson-normal-gamma model.*

As the above table-5 shows, allowing for the extension of the Poisson-normal-gamma model to include both random intercept and random slope does not improve the fit based on the likelihood comparison. As the respective quantities are significant, it is evident that there are both over-dispersion and correlation as the individual models (table-4 and table-5), respectively shows, and also supported by data exploration and descriptive statistics that the observed sample variance is greater than the sample mean which leads to over-dispersion.

The estimates of the covariates in both models are more or less similar, except an improvement in the slope model for the time sex interaction and sex, where time has a linear relationship with CD4+ cell count, which is what we observed in the graph showing the average trend. Thus, in the slope model, all the covariates are significant including the time sex interaction and sex for the change in the CD4+ cell counts of ART patients following HAART. However, the random intercept model supports the same thing except for the time sex interaction and sex which may be of interest to estimate. In plural term, the random intercept Poisson-normal-gamma model is chosen based on likelihood comparison and further discussion can be made on if any.

Model diagnostics

Residual versus fitted value plot for final Poisson-normal-gamma model is presented in Appendix-2, it does not show any systematic pattern this point out the model fits the data well and the Q-Q plot also verifies the residuals are normally distributed and symmetric around zero. Thus, it meets the assumption of error term. Besides to the above, the non linearity of the Q-Q plot confirms the model is not linear, which is somewhat contradictory to analytical model results above, i.e., further overlook at it should be made. Furthermore, the residual versus predicted recommended that there is a uniformity of residuals specifies that homogeneity of error variances. Plots of observed versus fitted value of CD4 count is given in Appendix-2, it verifies that there is a close agreement between observed and fitted values suggested that this model is good in predicting CD4 count. Q-Q plots for normality of random effects are also given in the same Appendix; which illustrates the random effects are normally distributed with mean zero and variance covariance matrix G. Thus, the fitted Poisson-normal-gamma model is fine for ART data.

4.3. DISCUSSION

A retrospective longitudinal study was conducted at Hossana Queen Elleni Mohamad Memorial Hospital, SNNPR, Ethiopia to determine the appropriate model for CD4+ cell counts and to characterize the time course of CD4+ cell progress with the software package, SAS procedure.

The data was unbalanced because some subjects were not keeping the regular time schedule and they were measured at different time points and the number of measurements was different across a subject which is similar to Vernon *et al.*, (2013). The time scale was used in a monthly

format though a six months interval schedule was not worked for some patients which might be reluctance of subjects to follow up. The data was analyzed by version SAS 9.2 using PROC NLMIXED procedure that fits nonlinear mixed models—that is, models in which both fixed and random effects enter nonlinearly. PROC NLMIXED fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects. This makes the approach to analyze data of this study unlike to other studies done on CD4+ cell counts.

As the result of this study reveals that the mean and standard deviation of CD4+ cell counts per milliliter of blood are not consistent with the result of the study conducted by Gezahegn (2011) at Durame and Hosanna hospital found. This disagreement in result may be due to the sample size the study included, due to it was made at two different hospitals and due to differences in educational and socio-economic levels. Exploratory data analysis reveals that there is a time trend in the data and the average profile of CD4+ cell counts has a little linear relationship over time which is unlike the SAS **sasuser.aids** data analyzed using linear mixed model by Michael (2002). This difference may be due to smoothing technique, time scale, number of observations, and progressive pattern of baseline CD4+ cell counts. However, in this current study, this could not affect the random effect structure to be included in the model. Thus, no improvement in the analysis when non-linearity was assumed and hence, as shown from the relationship a linear time effect was studied on CD4+ cell counts.

Most of the patients were females and they had lower mean CD4+ cell counts than males before ART was initiated which is similar to Moges *et al.*, (2013). This is because females are biologically and socially more vulnerable to HIV infection in the developing countries. However, this is inconsistent with Kumarasamy *et al.*, (2008) reported from India. This difference could be due to several reasons as described in that study; HIV associated TB could be the contributing factor for the low CD4+ count in males as the proportion of patients having TB was significantly higher in male HIV positive patients than females. In addition, it may be due to a sex-related difference in the overall CD4+ counts among males and females. HIV sero-negative Ethiopian females had relatively higher CD4+ cell counts than HIV seronegative males like reported by Yanis et al., (2005) and Tsegay *et al.*, (1999).

Most of the HIV infected patients enrolled in this study were living around rural out of Hossana town as found out by Nuredin (2007) in another study at Adama hospital and most of the patients

were young mean age of 31 years old who were sexually more active and thus have a higher risk of infection which is comparable to the study conducted by Moges *et al*., (2013) at Zewuditu Hospital Addis Ababa, Ethiopia. These findings as found by these authors could conform as previous reports from elsewhere in Ethiopia which reported that HIV prevalence decreases significantly to increasing level of education as well as their socio economic status.

The data in this study indicates that the majority of HIV patients started antiretroviral treatment with more advanced immunodeficiency status. Since the majority of HIV patients had AIDS as defined by their CD4 cell counts < 200 cells/μl, indicating advanced immune suppression at initiation of ART. This was significantly higher when compared to the studies conducted in Nigeria, south eastern United States and Thailand which reported a lower rate of AIDS at the initiation of ART (Chasombat *et al*., 2009; Nwokedi *et al*., 2007). Therefore, in the hospital of this current study, delayed enrollment in ART program could be attributed by several factors such as due to fear of stigma. In Ethiopia, as described by the above authors, only one third of HIV infected persons disclosed their HIV status to their partner further compromising the utilization of the counseling and testing and ART services. A similar observation was made among South Africans where patients stared ART program with advanced immunodeficiency status. These findings indicate urgent need to promote early and enhanced HIV testing to enable HIV/AIDS patients to benefit from the expanding ART services (Stohr *et al.,* 2007).

Using stepwise selection technique, enter and remove, and backward technique, the most non-significant covariates are removed and the rest in the model are refitted and so on. At the last step the procedure ends with (the most likely selected covariates): time since month of seroconversion, age of the patients, and sex of the patient. Except sex time interaction term in the model, the other covariates, time since month of seroconversion and age of patient are significant for the change in the CD4+ cell counts of ART patients at HAART which was supported by (Nuredin, 2007; Moges *et al*., 2013).

Thus, as in many other diseases, age is an important prognostic factor in HIV infection. Age at sero-conversion and age at a given CD4 cell count were shown to be important determinants of progression and survival before the widespread introduction of HAART, starting in 1996 (Sophie *et al*., 2006). This supports the current study. One study is reviewed which supports the significance of gender like this study, but it showed no difference among male and female. Thus,

no differences in HIV progression and response to HAART attributable to gender among patients accessing the Spanish hospital network (Hoyos *et al*., 2007). This difference may attribute to method of data analysis used in that study, Kaplan-Meier and Cox regression were used to assess the effect of sex on time to AIDS, survival from AIDS and attribute to other factors.

The comparison among four models were made, GLM (Poisson model without random effect to handle over dispersion), GLMM (Poisson model with normal random effects to administer the correlation), Poisson model with the gamma random effects to grip extra variation in the data, and Poisson-normal-gamma model, model with both normal and gamma random effects to capture both correlation and over-dispersion simultaneously. Estimation was done by maximum likelihood using numerical integration over the normal random effects, if present as was done by Molenberghs *et al*., (2007).

As Kassahun *et al*., (2014) summarized, based on Molenberghs *et al*., (2010), it is argued that the normal and non-normal, a gamma random effect, can usefully be integrated together into a single model to induce association between repeated Poisson data and to correct for the over-dispersion.

One possible route to deal with over-dispersion is to introduce an over-dispersion parameter and only specify a relationship between the mean and the variance, and then apply quasi-likelihood, whereby the extra variability in the data could captured by the dispersion parameter which is as Wedderburn (1974) did.

For the count data, it is common to combine Poisson distribution with a gamma distributed random effect, so that the unconditional distribution of the outcome turns out to be a negative binomial distribution, SAS procedure NLMIXED displayed this as shown on table-3 (fifth column) (Breslow, 1984; Hinde and Demétrio, 1998a, and Hinde and Demétrio, 1998b). On the other hand, focusing on hierarchical data, the GLM is usually extended to generalized linear mixed models (GLMMs), with a subject-specific random effect, typically a Gaussian form, added in the linear predictor to take into custody a hierarchy-induced association or to account for over-dispersion (Engel and Keen, 1992; Molenberghs and Verbeke, (2005); Pinheiro and Bates, 2000), displayed on table-3 fourth column. As used by Kassahun *et al.,* (2014) and Molenberghs *et al*., (2010) proposed a flexible and unified modeling structure, termed the Poisson-normal-gamma model, to simultaneously capture over-dispersion and correlation for a

wide range of clustered data, including count, binary and time-to-event. Thus, two sets of random effects were brought together. The normally distributed subject specific-random effects take into custody the correlation, while a conjugate measurement-specific random effect on the natural parameter, is used to accommodate over-dispersion, as shown on table-3 last column.

Molenberghs and Verbeke (2005), considered a Poisson-normal model with random intercepts as well as random slopes in time. It is interesting to note that, when allowing for such an extension in our models, the random slopes improve the fit of the Poisson-normal model, but not of the Poisson-normal-gamma model (details are shown in table-4&table-5). Recall that the same procedure was applied, too, by Booth *et al*., (2003). While in this study it is considered four different models; but those authors focused on the Poisson-normal and Poisson-normal-gamma implementations. There are further differences in actual fixed-effects and random effects models considered.

Therefore, as tried out to lay the ground to talk, the four models were compared based on their likelihoods since they are not nested to use AIC though some authors used. Accordingly, as was discussed by Kassahun *et al*., (2014), the Poisson-normal-gamma model which combines both normal and gamma random effects to capture together both over-dispersion and correlation was selected to improve the fit to the model based on −2log-likelihood which is 14527.

NB:- CD4+ cell counts is to mean the CD4 cell counts after a patient is tested and notified that he/she is HIV positive, i e., CD4 cell counts after the first visit.

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

In conclusion, in this study, although good CD4+ cells recovery in response to ART was recognized, HIV-positive patients were enrolled in ART program at decreased CD4 cells levels. There is a poor recovery of CD4+ cell when they start at <200 CD4+cell counts than when they start ART at >200 CD4+ cell counts.

The covariates in the analysis were identified as risk factors affecting the progress in CD4+ cell counts of the patients. Time since month of seroconversion, sex of the patient and age of patient are significant for the change in the CD4+ cell counts of ART patients at HAART. Thus, time since month of seroconversion, sex of the patient and age of patient are potential risk factors for the change in the CD4+ cell counts of ART patients at HAART, in this study by applying models handling over-dispersion and correlation.

The analysis showed that in the presence of over-dispersion, and clustering, the Poisson-normal-gamma model results in improvement in model fit. Thus, Poisson-normal-gamma model is the best chosen appropriate model for CD4+ cell counts data in this study.

5.2. Recommendations

As the numerical figure in table-1 reveals high number of patients were not following their drug combinations this may affect their CD4+ cell counts progress and the response to HAART may not be as expected and high number of patients were married and were living around rural areas out of Hossana town which are far from the hospital. Hence, the HAART and any health related concerning bodies should have to support in giving advice to the patients to take care in making relation with others except their partners as they are married and they have to encourage the patients to follow the ART though the service center may be far.

Moreover, interventions need to be designed to promote early HIV testing and early enrollment of HIV infected individuals into ART services. As socio-demographic factors and lack of awareness about ART services, fear of stigma and discrimination compromise the utilization of

ART program, improving public awareness by advocacy and social mobilization should be included in the ART service. ART considerably improves the immune recovery. It is strongly recommended that underline the need of anti-retroviral therapy in HIV infected patients for immune reconstitution.

The natural feature of CD4+ cell counts is non-linear as supported by some authors, but in this study there was a slight linearity in the data. So, this difference needs some overlook at it.

Further Longitudinal studies with better number of repeated measurements per subject should be conducted on CD4+ cell counts to get better insight on the trends and to account for both over-dispersion and correlation. And also further overlook should be made on the disagreement among the models in selecting potential risk factors for the outcome variable. Thus, some important covariates which are selected as potential risk factors for the change in CD4+ cell counts are not so in this study; hence it needs further investigation.

Limitations of the study

The following were the limitations of the study;

1. Limited number of variables was captured during patient enrolment: In order to determine probabilities of predicting CD4 response to HAART, it was needed to identify some variables that were found in the records of the respondents, captured during commencement of ARV therapy. The problem was that variables considered for this study was not recorded in all the files of clients on HAART.
2. Limited number of variables to measure social economic status. Only income and employment status were used as proxy measures of social economic status.
3. One limitation of all observational studies is that of unmeasured confounding, none of the co-infectious diseases like TB were included.
4. Some of important techniques like Monte Carlo simulation models were not used to track HIV disease progression and to indirectly estimate the outcomes and costs of treatment when initiated at various CD4 cell counts. Using this approach, initiation of HAART at a CD4 cell count more than 350 cells/μl can be seen to result in longer quality-adjusted survival compared to starting HAART at lower CD4 cell counts.

Reference List

Allison, A. (2006). HIV Laboratory Monitoring, www.aidsalliance.org/publications. *Agency for International Development (USAID).* Knowledge Hub Newswire - Vol.1 No.3 U.S.

Allison, (2008). The evidence base for programming for children affected by HIV/AIDS in low prevalence and concentrated epidemic countries. UNICEF, 3 UN Plazas, NY, NY 10017.

Agresti, A. (2002). Categorical Data Analysis, 2nd edn. John Wiley and Sons: New York.

AVERT. ORG (2009). HIV and AIDS in ZAMBIA: The epidemic and its impact. Republic of Zambia.

Bayeh, A., Fisseha, w., Tsehaye, T., Atnaf, A., Mohammed, Yessin. (2010). ART-naive HIV patients at Feleg-Hiwot Referral Hospital Northwest, Ethiopia. *Ethiop. J. Health Dev.;* **24**(1).

Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*; **33**:38–44.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc*., **88**:9–25, 1993.

Breslow, N. and Clayton, D. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*; **88**, 9–25.

Booth, J., Casella, G., Friedl, H., Hobert, J. (2003). Negative binomial log-linear mixed models. *Stat Model;* **3**:179–181.

Chasombat, S., McConnell, M., Siangphoe, U., Yuktanont, P., Jirawattanapisal, T., Fox *et al*., (2009). National expansion of antiretroviral treatment in Thailand, 2000-2007: program scale-up and patient outcomes. *J Acquir Immune DeficSyndr*; **50**(5):506-12.

Davidian, M. and Gallant, A. (1992). Smooth Non-parametric Maximum Likelihood Estimation for Population Pharmacokinetics, With an Application to Quinidine. *Journal of Pharmacokinetics and Biopharmaceutics*; **80**, 475–488.

Dzimnenani, M. (2007). HIV/AIDS is epidemic outcome of poverty in sub Saharan Africa? *Croate MedJ*; **48**(1):605-617.

Egger, M., May, M., Chene, G., Christiank, R. (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lance*t; **360**: 119– 29. [CrossRef][ISI][Medline].

Elke, K., Yirga, A., Katherine, D., Peter, G, Tesfaye, A. and Bud, C.(2011). Implications of adopting new WHO guidelines for antiretroviral therapy initiation in Ethiopia. *Bulletin of the World Health Organization*; **90**:659-663. doi: 10.2471/BLT.11.089599.

Elly, T., Katabira, FRCP., Moses, R., Kamya, M., Med, M., MPH, I., Med, A., (2008). Ugandan Antiretroviral Treatment and Care Guidelines for Adults, Adolescents, and Children. Kampala, Uganda.

Engel, B. and Keen, A. (1992). A Simple Approach for the Analysis of Generalized Linear Mixed Models. LWA-92-6, Agricultural Mathematics Group (GLW-DLO). Wageningen, the Netherlands.

Engel, B. and Keen, A. (1994). A simple approach for the analysis of generalized linear mixed models. *Statist. Neerlandica*; **48**(1):1–22.

Evan, W., Richard, H., Morgan, B. (1996). Adherence to antiretroviral and CD4 T-cell count response among HIV- infected drug users. Volume 9. International media press. p 36-40. London.

Fitzmaurice, G. (1997). Model selection with over-dispersed data. *The Statistician*; **46**, 81-91.

Florence *et al*., (2003). HIV medicine journal (factors associated with a reduced CD4 lymphocyte count response to HAART despite full viral suppression in the EUROSIDA study. Black well publishing.

Françoise, R., Chris, D., Stephen, K., Sigrid, T., Joseph, P. (2006). Use of antiretroviral therapy in resource-limited countries: distribution and uptake of First and second line regimens. *Department of HIV/AIDS, World Health Organization, CH-1211 Geneva 27, Switzerland*; Email: theryf@who.int.

Gange, S., Barrón, Y. Greenblatt, R., Anastos, K., Minkoff, H., Young, M., Kovacs, A., Cohen, M., Meyer III, W. and Muñoz, A. (2002). Effectiveness of highly active antiretroviral therapy among HIV-1 infected women. *J Epidemiol Community Health; **56**:153–159.*

Genest, C. and Rivest, L. (2001). On the Multivariate Probability Integral Transformation. *Statistics and Probability Letters*; **53**, 391–399.

Geyer, C. and Thompson, E. (1992). Constrained Maximum Likelihood for Dependent Data (with discussion). *Journal of the Royal Statistical Society, Series B*; **54**, 657–699.

Gezahegn, A. (2011). Survival Status among patient living with HIV AIDS who are on ART treatment in Durame and Hossana Hospitals. DOI: 10.1371/journal.pone.0087392 .doi:10.1371/journal.pone.0087392.t002.

Hinde, J. and Demétrio, C. (1998) a. Over-dispersion: Models and Estimation, São Paulo: Associação Brasileira de Estatística.

Hinde, J. and Demétrio, C. (1998) b. Categorical Data Analysis, São Paulo: Associação Brasileira de Estatística.

Hoel, P., Port, S. and Stone, C. (1971). Introduction to Probability Theory, Boston, MA: Houghton Mifflin Company.

Hoyos, P., Paul, K., Henry, T. (2007). Progression to AIDS and death and response to HAART in men and women from multicenter Hospital based cohort. Journal of women health. Vol 16. Mary Ann liebet Inc.

Johannessen *et al.*, (2008). Predictors of mortality in HIV-infected patients starting antiretroviral Therapy in a rural hospital in Tanzania; **8:**52doi:10.1186/1471-2334-8-52.

John, H. (2007). Over-dispersion: Models and Estimation. A Short Course for SINAPE. MSOR Department, Laver Building, University of Exeter, North Park Road, Exeter, EX4 4QE, UK.

Johnson, L. and Dorrington, R. (2006). Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demographic research*; **14**(22):541-574.

Kachman, S. (1998). An introduction to Generalized Linear Mixed Models. Department of Biometry, University of Nebraska–Lincoln.

Kassahun, W., Neyens, T., Molenberghs, G., Faes, C., and Verbeke, G. (2014). Modeling Hierarchical Data, Allowing for Over-dispersion and Zero Inflation, in Particular Excess Zeros. *Stat Med*; doi: 10.1002/sim.6237. [Epub ahead of print] PMID: 24957791[Pub Med- as supplied by publisher].

Kassahun, W., Neyens, T., Molenberghs, G., Faes, C., and Verbeke, G. (2012). Modeling over-dispersed longitudinal binary data using a combined beta and normal random-effects model. *Arch Public Health*; **70**(1):7. doi: 10.1186/0778-7367-70-7. PMID: 22958735 [Pub Med].

Katubulushi, M. and Chanda, W. (2009). Slow decline in CD4 cell counts in a cohort of HIV-infected adults living in Lusaka, Zambia. Department of Medicine, University of Zambia, School of Medicine, University Teaching Hospital. Volume 19 - Issue 1 - p 102-103.

Kaufmann, G., Furrer, H., Ledergerber, B. (2005). Characteristics, determinants, and clinical relevance Of CD4 T cell recovery to <500 cells/microL in HIV type 1-infected individuals receiving Potent antiretroviral therapy. *Clin Infect Dis; **41** (3): 361-372. doi: 10.1086/431484*

Ketema, K. (2011). Determining factors that affect the survival rate of HIV-infected patients on ART: The case of Armed Forces General Teaching Hospital, Addis Ababa, Ethiopia. *Ethiop. J. Health Dev*; **22**(1).

Kitahata, M., Reed, S., Dillingham, P., Gleen, H., Fecher, M. (2004). Pharmacy-based assessment of adherence to HAART predicts virologic and immunologic treatment response and clinical progression to AIDS and death. *Int J STD AIDS. Dec*; **15**(12):803-10.

Kumarasamy, N., Venkatesh, K., Cecelia, A., Devaleenol, B., Saghayam, S., Yepthomi *et al*., (2008). Gender-based differences in treatment and outcome among HIV patients in South India. *J Women Health (Larchmt)*; **17**(9):1471-5.

Lee, Y. and Nelder, J. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*; **58**, 619-678.

Lee, Y. and Nelder, J. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika*; **88**, 987-1006.

Lindstrom, M. and Bates, D. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*; **46**, 673–687.

Liu, Q. and Pierce, D. (1994). A Note on Gauss-Hermite Quadrature. *Biometrika*; **81**, 624 -629.

Marlink *et al*., (1994). Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science* **265**, 1587–1590.

Marlon, B. (2010). Factors affecting CD4+T-Lymphocyte count response to HAART in HIV/AIDS patients within 24 months of treatment at Chreso Ministries ART Centers. *BMC Infectious Diseases*; http://www.biomedcentral.com/.

McCullagh, P. and Nelder, J. [1989]. Generalized Linear Models, Second Edition. Chapman and Hall, New York, NY.

McGilchrist, C. (1994). Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society, Series B*; **56**, 61–69.

Michael, P. (2002). Longitudinal data analysis with discrete and continuous responses course notes. *SAS Institute Inc*., Cary, NC 27513; 58710.

Moges, D., Monga, D. and Deresse, D. (2013). Immunological response among HIV/AIDS patients before and after ART therapy at Zewuditu Hospital Addis Ababa, Ethiopia. *American Journal of Research Communication;* **1**(1): 103-115. www.usa-journals.com, ISSN: 2325-4076.

Molenberghs, G. and Verbeke, G. (2005). Models for Discrete Longitudinal Data, Springer-Verlag, Berlin.

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

Molenberghs, G. and Verbeke, G. (2007). Models for Discrete Longitudinal Data. Springer: New York.

Molenberghs, G., Verbeke, G., Demétrio, C., Vieira, A., (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*; **25**:325–347.

Moore, R. and Keruly, J., (2007). CD4+ cell count 6 years after commencement of highly active antiretroviral therapy in persons with sustained virologic suppression. *Clin Infect Dis*; **44**:441-6.

Mustapha, A. and Albert, L. (2013). Longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on Antiretroviral Therapy (Art) in the Builsa District Hospital. *European Scientific Journal vol.9;* **No.33 ISSN**: 1857 – 7881 (Print) e - ISSN 1857-7431.

Nuredin, I. (2007). Evaluation of factors affecting chance of survival/death status among HIV-positive people under the Anti- Retroviral Treatment Program: The Case of Adama Hospital.

Nwokedi, E., Ochicha, O., Mohammed, A., Saddiq, N. (2007). Baseline CD4 lymphocyte count among HIV patients in Kano, Northern Nigeria. *Afr J Health Sci*; **14**:4.

Osgood, D. 2000. Statistical Models of Life Events and Criminal Behavior. In the Handbook of Quantitative Criminology, Alex R. Piquero and David Weisburd Eds, p. 375-396. New York: Springer.

Paternoster, R., Brame, R., Mazerolle, P. (1997). Using the correct statistical test for the equality of regression coefficients. *A Piquero Criminology*; **36** (4), 859-866.

Perez, J. and Moore, R. (2003). Greater effect of highly active antiretroviral therapy on survival in people aged > or = 50 years compared with younger people in an urban observational cohort. *Clin Infect Dis*; **36**: 212–18. [CrossRef][ISI][Medline].

Pinheiro, J. and Bates, D. (1995). Approximations to the Log-likelihood Function in the Nonlinear Mixed Effects Model. *Journal of Computational and Graphical Statistics*; **4**, 12–35.

Pinheiro, J., and Bates, D. (2000). Mixed-Effects Models in S and S-PLUS, Springer.

Recsky, M., Zabrina, L., Keith *et al*., (2004). Antiretroviral resistance among HIV-infected persons who have died in British Columbia, in the era of modern antiretroviral therapy. *The Journal of Infectious Diseases; **190**:285-92.*

Reda, A., Biadgilign, S., Deribew, A., Gebre, B., Deribe, K. (2013). Predictors of Change in CD4 Lymphocyte Count and Weight among HIV Infected Patients on Anti-Retroviral Treatment in Ethiopia: A Retrospective Longitudinal Study. *PLoS ONE;* **8**(4):e58595. doi:10.1371/journal.pone.0058595.

Skrondal, A. and Rabe-Hesketh, S. (2004). Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Boca Raton: Chapman & Hall/CRC.

Smith *et al*., (2004). Factors influencing increases in CD4 cell counts of HIV-positive persons receiving long-term highly Active Antiretroviral Therapy (ART). 1860 • JID: 190.

Sophie *et al*., (2006). HIV infection in older patients in the HAART era. Journal of Antimicrobial chemotherapy Volume 57, Oxford university press.

Stephen, L. (2008). A constellation of change and commitment' in response to AIDS. Zambia. p 2, 6-7.

Stohr *et al*., (2007). Factors affecting CD4+T-lymphocyte count response to HAART in HIV/AIDS patients. HIV medicine journal. Volume 8, Blackwell publishing, uk.pp135 **141**(7).

Tsegaye, A., Messele, T., Tilahun, T., Hailu, E., Sahlu, T., Doorly *et al*., (1999). Immunohematological reference ranges for adult Ethiopians. *Clin Diagn Lab Immunol*; **6**(3):410-4.

Vernon, L., Demko, C., Babineau, D., Wang, X., Toossi, Z. *et al.,* (2013). Effect of Nadir CD4+ T cell Count on Clinical Measures of Periodontal Disease in HIV+ Adults before and during Immune Reconstitution on HAART. *PLoS ONE;* **8**(10): e76986.doi:10.1371/journal. pone.0076986.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models and the gauss-newton method. *Biometrika;* **61**, 439-47.

Wolfinger, R. and Lin, X. (1997). Two Taylor-Series Approximation Methods for Nonlinear Mixed Models. *Computational Statistics and Data Analysis*; **25**, 465–490.

WHO protocols for CIS countries (2004). HIV/AIDS treatment and care clinical protocols for the WHO European Region. *DK-2100 Copenhagen, Denmark.* http://www.euro.who.int/pubrequest.

Yanis, L., Alycia, K., Jian, G., Thomas, W., Lawrence, A., Patrick, S., and Yuan, C. (2005). Gender difference in clinical progression of HIV-1 infected individuals during long term highly active antiretroviral therapy. AIDS volume19, **No6**:577-583.

Yibeltal, A., Eleni, S., Achamyeleh, A., Abinet, A., Belen, K., Hana, T., Dule, T., Yezihalem, A., and Netsanet, H. (2010). HIV/AIDS prevention and control office (HAPCO) monthly HIV and ART update http.www//arc.org.et.

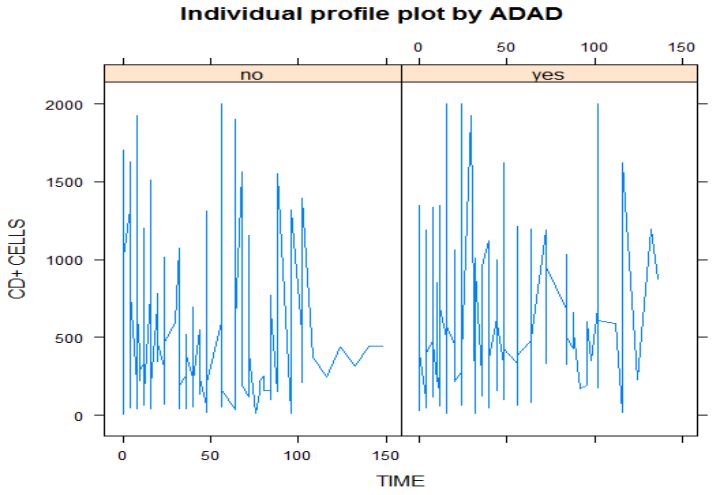
Zachariah, R., Fitzgerald, M., Mass quoi *et al*., (2006). Risk factors for high early mortality in Patients on antiretroviral treatment in a rural district of Malawi. *Nati. Institut. Health*; **20**(18):2355-60.

APPENDICES

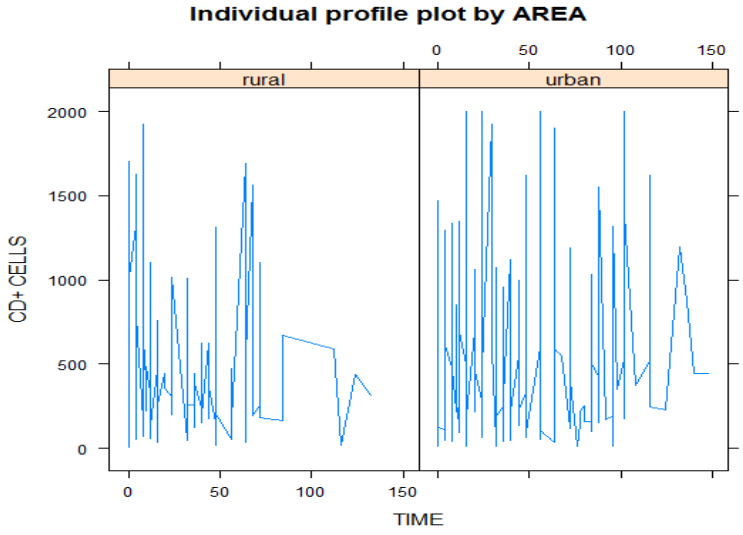
APPENDIX-1: EXPLORING DATA AND GRAPHS FOR MODEL DIAGNOSIS CHECK

APPENDIX-1 A: Individual Plots of CD4+ cell counts data by some categorical covariates

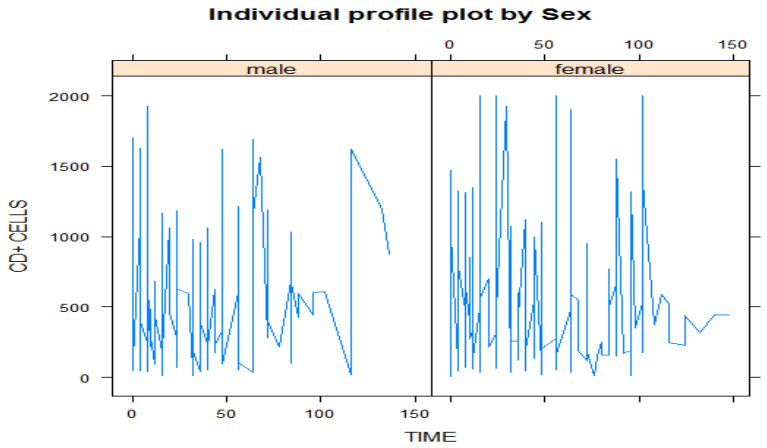
A-1: Individual Plot for CD4+ cell counts data by Adherence to any of the drug combinations



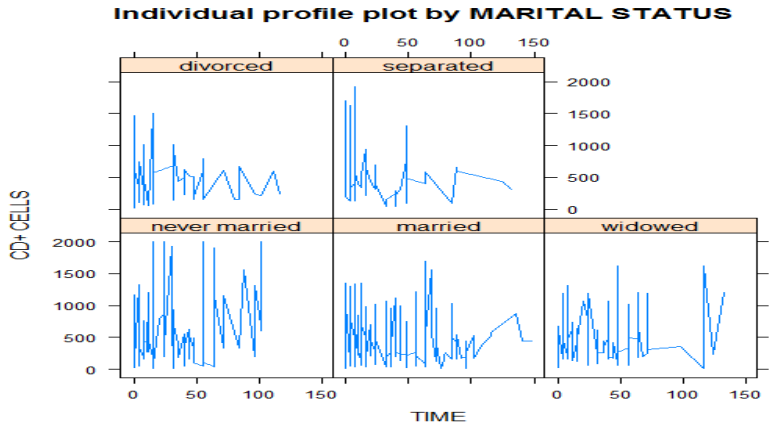
A-2: Individual Plot for CD4+ cell counts data by Residence or Area of the patient



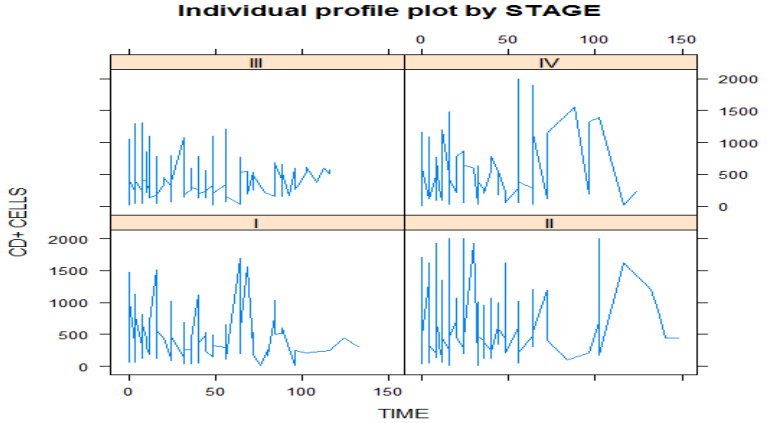
A-3: Individual Plot for CD4+ cell counts data by Sex of the patient



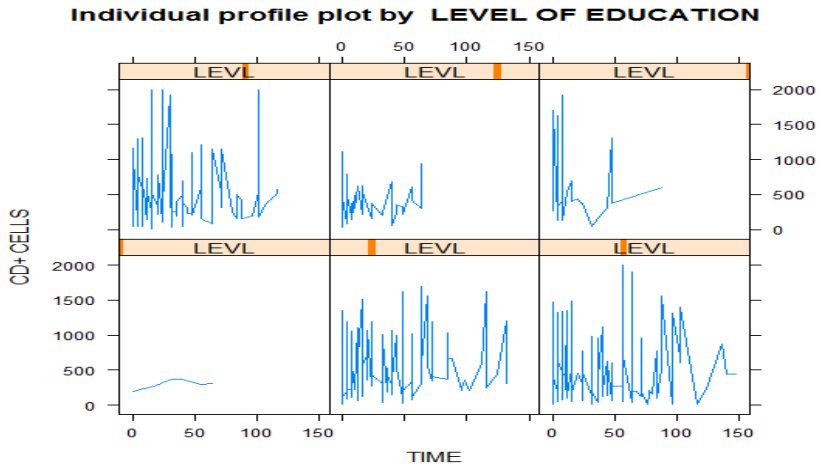
A-4: Individual Plot for CD4+ cell counts data by marital status of the patient



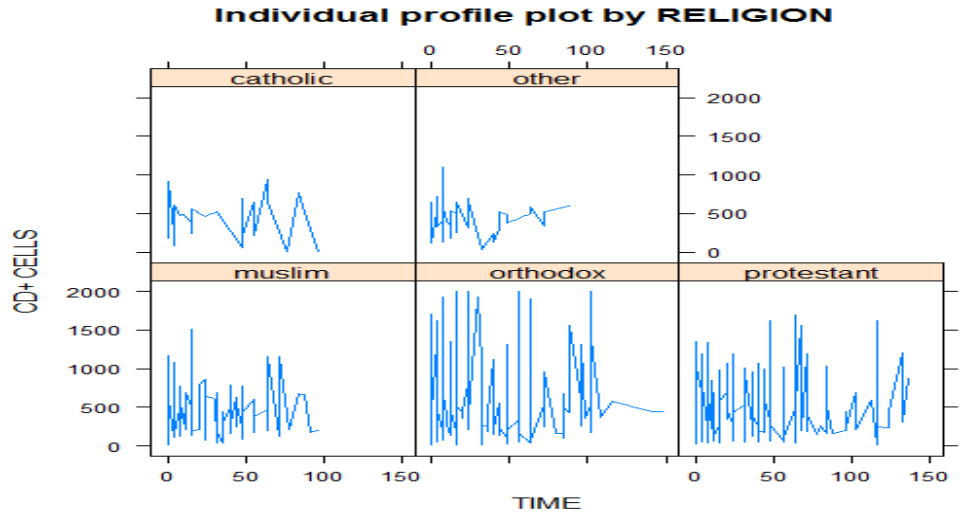
A-5: Individual Plot for CD4+ cell counts data by Stage of disease of the patient



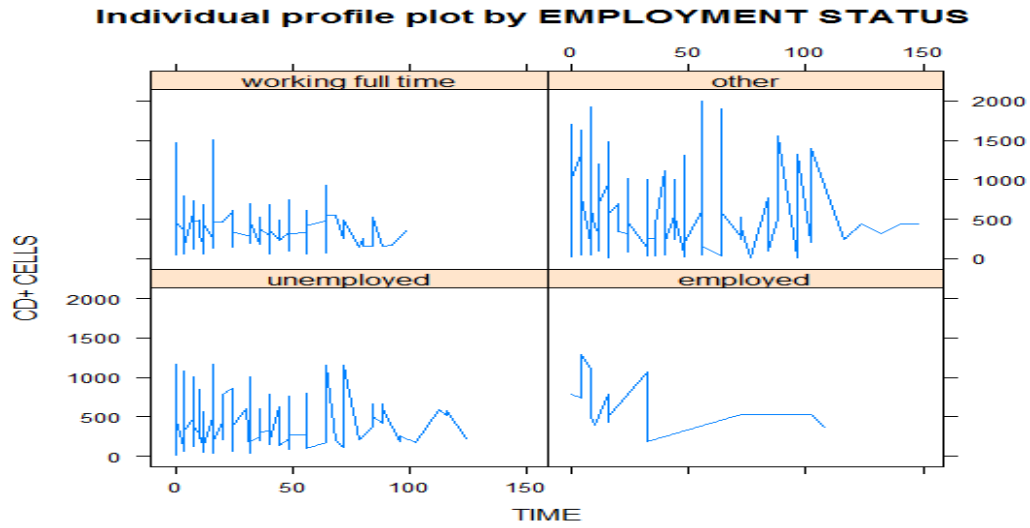
A-6: Individual Plot for CD4+ cell counts data by Level of education of the patient



A-7: Individual Plot for CD4+ cell counts data by Religion of the patient

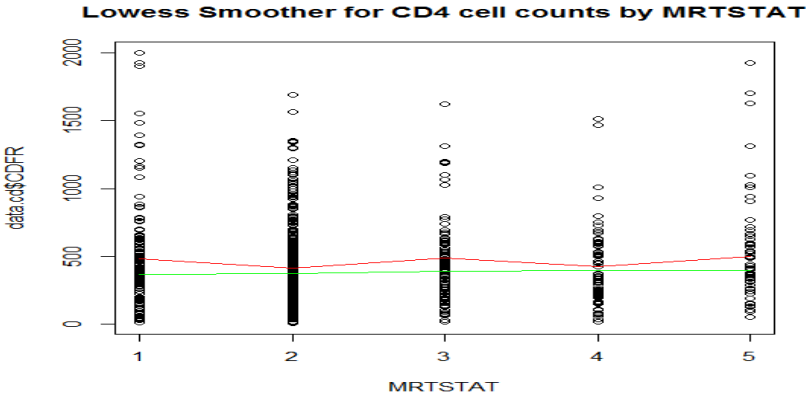


A-8: Individual Plot for CD4+ cell counts data by employment status of the patient

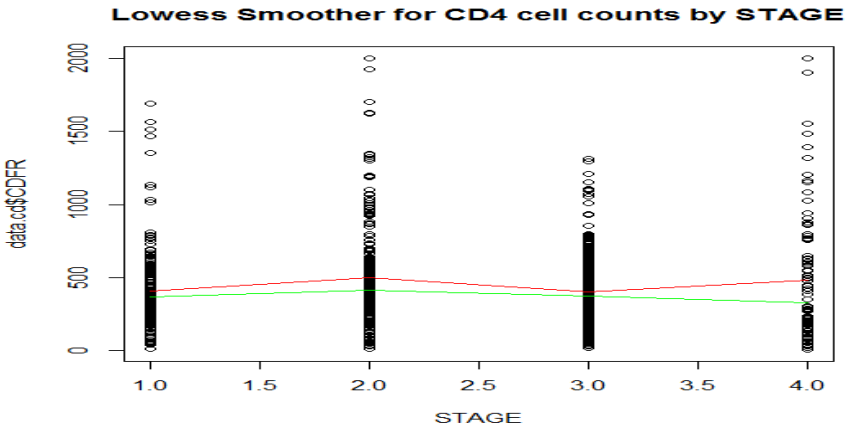


APPENDIX-1 B: Loess smoother of CD4+ cell counts data by some categorical covariates

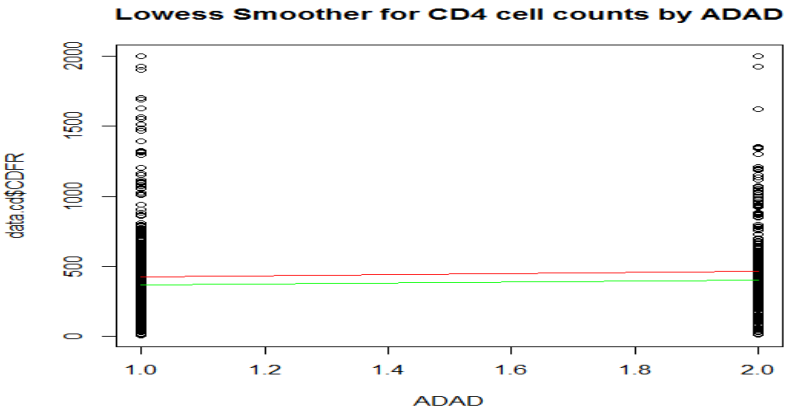
B-1: Loess smoother for CD4+ cell counts data by marital status of the patient



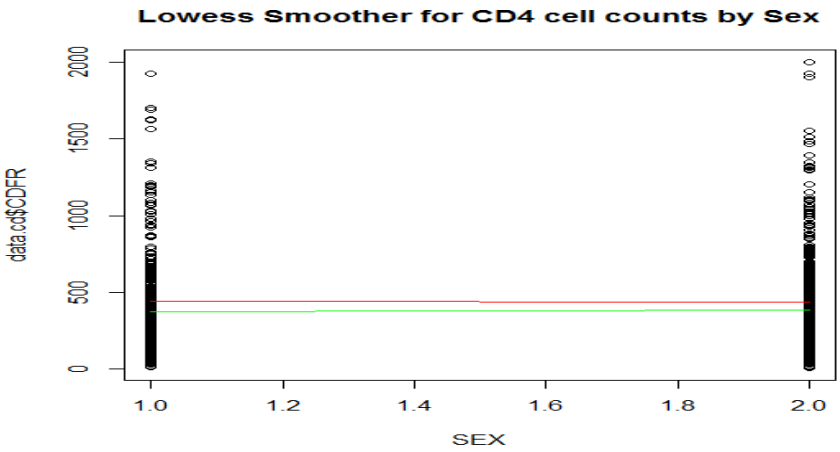
B-2: Loess smoother for CD4+ cell counts data by WHO Stage of disease of the patient



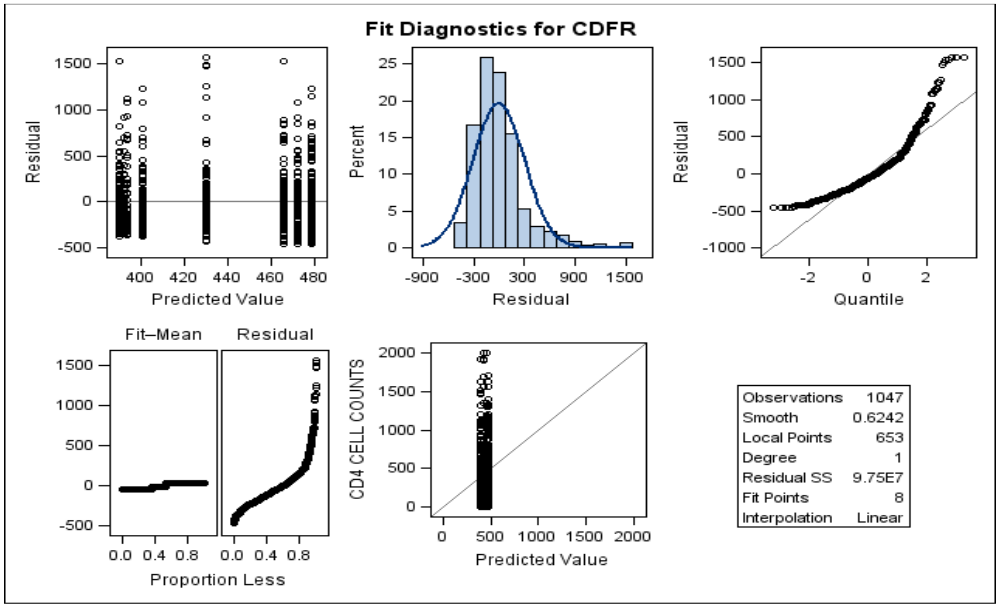
B-3: Loess smoother for CD4+ cell counts data by Adherence to any of drug combinations of the patient

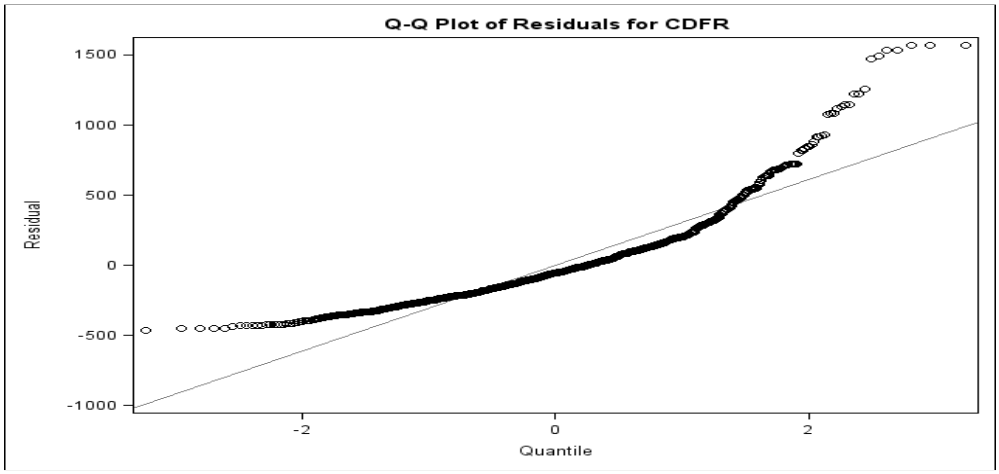


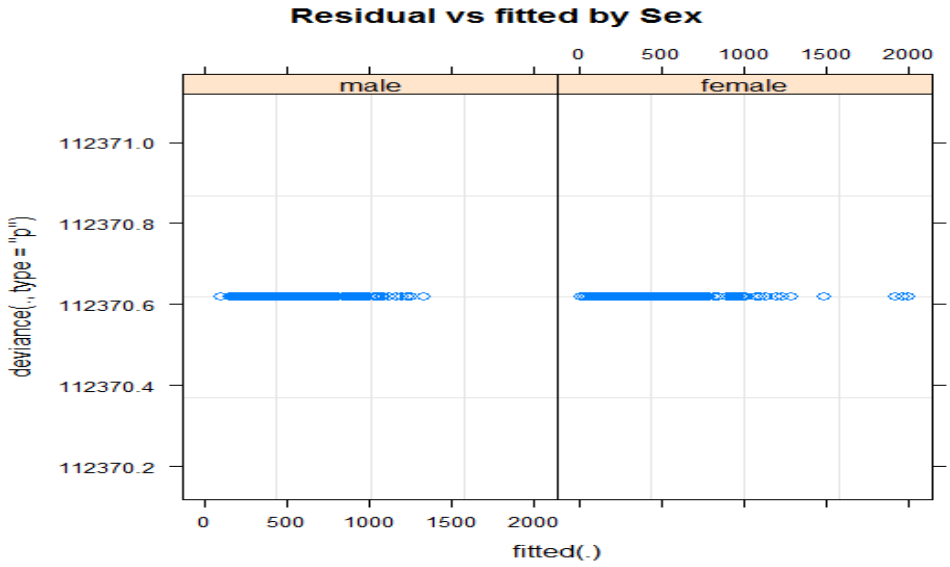
B-4: Loess smoother for CD4+ cell counts data by Sex of the patient



APPENDIX-2: GRAPHS FOR MODEL DIAGNOSIS CHECK




Q-Q Plot of Residuals for CDFR
Residual
Quantile
1500
1000
500
0
-500
-1000
-2
0
2


Residual vs fitted by Sex
male
female
deviance(., type = "p")
fitted(.)
112371.0
112370.8
112370.6
112370.4
112370.2
0
500
1000
1500
2000

APPENDIX 3: WHO Staging for HIV Infection and Disease in Adults & adolescents

Clinical Stage I:

1. Asymptomatic
2. Persistent generalized lymphadenopathy

Performance Scale 1: Asymptomatic, normal activity

Clinical Stage II:

1. Moderate weight loss (less than 10% of presumed or measured body weight)
2. Minor mucocutaneous manifestations (seborrhoeic dermatitis, prurigo, fungal nail infections, recurrent oral ulcerations, angular stomatitis)
3. Herpes zoster within the last 5 years
4. Recurrent upper respiratory tract infections, e.g., bacterial sinusitis, tonsillitis, otitis media and pharyngitis

And/or Performance Scale 2: Symptomatic but normal activity

Clinical Stage III:

1. Severe weight loss (more than 10% of presumed or measured body weight)
2. Unexplained chronic diarrhoea for more than 1 month
3. Unexplained prolonged fever, intermittent or constant, for more than 1 month
4. Oral candidiasis
5. Oral hairy leukoplakia
6. Pulmonary tuberculosis (current)
7. Severe bacterial infections such as pneumonias, pyomyositis, empyema, bacteremia or meningitis
8. Acute necrotizing ulcerstive stomatitis, gingivitis or periodontitis
9. Unexplained anemia (<8gm/dl), neutropenia (<0.5× 10_9 per litre), or chronic thrombocytopenia (<50× 10_9 per litre)

And/or Performance Scale 3: Bed-ridden for less than 50% of the day during the last month

Clinical Stage IV:

1. HIV wasting syndrome – weight loss of more than 10%, and either unexplained chronic diarrhoea for more than 1 month, or chronic weakness or unexplained prolonged fever for more than 1 month
2. *Pneumocystis pneumonia* (PCP)
3. Recurrent severe bacterial pneumonia
4. *Toxoplasmosis of the brain*
5. Cryptosporidiosis with diarrhoea for more than 1 month
6. Chronic isosporiasis
7. Extrapulmonary cryptococcosis including meningitis
8. Cytomegalovirus infection (retinitis or infection of other organs)
9. Herpes simplex virus (HSV) infection, mucocutaneous for more than 1 month, or visceral at any site
10. Progressive multifocal leukoencephalopathy (PML)
11. Any disseminated endemic mycosis such as histoplasmosis, coccidioidomycosis
12. Candidiasis of the oesophagus, trachea, bronchi or lungs
13. Atypical mycobacteriosis, disseminated
14. Recurrent non-typhoid salmonella septicaemia
15. Extrapulmonary tuberculosis
16. Lymphoma
17. Invasive cancer of the cervix
18. Kaposi's sarcoma
19. HIV encephalopathy – disabling cognitive and/or motor dysfunction interfering with activities of daily living, progressing slowly over weeks or months, in the absence of concurrent illness or condition other than HIV infection that could account for the findings
20. Atypical disseminated leishmaniasis
21. Symptomatic HIV-associated nephropathy or symptomatic HIV associated cardiomyopathy

And/or Performance Scale 4: Bed-ridden for more than 50% of the day during the last month