JOINT MODELING OF LONGITUDINAL CD4 CELL COUNT AND TIME-TO-VIRAL REBOUND OF HIV INFECTED PATIENTS INITIATING ANTIRETROVIRAL THERAPY AT JIMMA UNIVERSITY MEDICAL CENTER

MSc. THESIS

BY

DELELEGN LAMBAMO ANITO

ADVISOR

ABDISA GURMESSA (PhD candidate)

CO-ADVISOR

Mr. KIBREALEM SISAY (MSc.)

MARCH, 2022

JIMMA, ETHIOPIA

JOINT MODELING OF LONGITUDINAL CD4 CELL COUNT AND TIME TO VIRAL REBOUND OF HIV/AIDS INFECTED PATIENTS INITIATING ANTIRETROVIRAL THERAPY

BY

DELELEGN LAMBAMO ANITO

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL SCIENCES, JIMMA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTERS OF SCIENCE IN BIOSTATISTICS

MARCH, 2022

JIMMA, ETHIOPIA

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

MSc THESIS APPROVAL SHEET

We, the undersigned, member of the Board of Examiners of the final open defense by **DELELEGN LAMBAMO** have read and evaluated his/her thesis entitled JOINT MODELING OF LONGITUDINAL CD4 CELL COUNT AND TIME-TO-VIRAL REBOUND OF HIV INFECTED PATIENTS INITIATING ANTIRETROVIRAL THERAPY AT JIMMA UNIVERSITY MEDICAL CENTER " and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree Master of Science in **Statistics (Biostatistics).**

Jaleta Abdisa Name of the Chairperson	 Signature	05/04/2022 Date
Kibreab Sisay Name of Major Advisor	 Signature	 Date
Kenenisa Tadesse Name of the Internal Examiner	 Signature	 Date
Girma Taye, PhD Name of the External Examiner	 Signature	02/04/2022 Date

DECLARATION

I hereby declare that this thesis is a result of my genuine work, and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in partial fulfillment for the degree of Master of Science in Biostatistics. The thesis can be deposited in the University Library to be made available to borrowers for reference. I solemnly 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 acknowledgment of the source is made. In all other instances, however, permission must be obtained from the author.

Name: Delelegn Lambamo Anito

Date: **…….**

Signature: **…….**

March, 2022

Jimma, Ethiopia

ACKNOWLEDGEMENT

First and foremost, I would like to thank the almighty **God** for being there with me in every step of my life. Next, I would like to express my grateful and sincere gratitude to my advisor Abdisa Gurmessa (PhD candidate), and co-advisor Mr. Kibrealem Sisay (MSc.), for their time, support, guidance, encouragement, and precious comments from the beginning until the completion of this work.

My sincere thank goes to staffs in Jimma University Medical center who are working in ART unit, for their kind cooperation in providing me all the data. Then, I sincerely thank my entire family for their love, encouragement, and support throughout this phase of my academic journey.

Finally, I would like to gratefully thank the department of Statistics, Jimma University, and Duna Worde Waters, Mines and Energy Office for the chance they gave me to pursue my master's study.

ABSTRACT

Back ground: HIV infection leads to severe depletion of CD4 cell with subsequently reduced levels of circulating CD4+lymphocytes in the peripheral blood. CD4 cell counts are the primary laboratory markers used to track the progression of HIV to AIDS. Time-to-viral rebound and CD4 count measures are the outcome variables of HIV patients after starting ART in this study. The time-to-viral rebound from ART is determined by month time interval among dates of ART commencement to rebound, as documented by the health information data administrator. In such follow-up trials, joint models are used because both longitudinal and survival data are generated.

Objective: The objective of the study is to compare separate and joint models of longitudinal CD4 cells measurements and to identify factors affecting change in CD4 cell count over time.

Methods: A retrospective cohort study design was conducted among 309 HIV/AIDS patients who were 18 years old and who are under ART follow-up from February 1; 2016 to May 30; 2021 at Jimma University Medical Center, West Ethiopia. First, the data were analyzed using longitudinal and survival models separately. Then, based on the separate model's several joint models with different random effects and shared parameters have been explored and compared using AIC score.

Results: Among 309 HIV patients considered in this study; 235 (76.1%) of them were rebound while the remaining 74 (23.9%) were censored. The result from the joint model of the estimated association parameter α is -0.102, this indicates both outcomes are negatively associated and higher values of the CD4 cell count are associated with better survival. The two outcomes were associated. The joint model was used to handle the associations between them to obtain a valid and efficient estimate. The result of the longitudinal model reviled that age; adherence; functional status; WHO clinical stage; interaction effect of adherence; functional status and WHO clinical stage with linear time had significantly associated with mean change in the square root of CD4 count. Furthermore, from the survival model we found the survival probability of HIV infected patient were determined by age; viral load; adherence; WHO stage and peripheral neuropathy.

Conclusion: The joint model reveals an association between time to viral rebound and repeated CD4 cell measurement. When evaluating the overall performance of both the separate and joint models in terms of model parsimony, the goodness of fit, smaller total AIC, and the statistical significance of both the association parameters, the joint model performs better. Thus, we concluded that the joint model is preferred for simultaneous analyses of repeated measurement and survival data.

Recommendation: In the future, the study recommends the application of joint model of bivariate longitudinal and time to viral rebound of survival analysis of HIV progression.

Key words: HIV/AIDS; Joint Modeling; Linear mixed model; Longitudinal Analysis, Survival Analysis; Viral rebound.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
ARV	Anti-Retroviral
AZT	Azidothymidine/ Zidovudine
BIC	Bayesian Information Criteria
CD4	Cluster of Differentiation 4
DHS	Demographic and Health Survey
EFV	Efavirenz
EPHI	Ethiopian Public Health Institute
HIV	Human Immune Deficiency Virus
HR	Hazard ratio
LMM	Linear Mixed Model
LRT	Likelihood Ratio Test
ML	Maximum Likelihood
NVP	Nevirapine
PH	Proportional Hazard
PLWHA	People Living with HIV/AIDS
UNAIDS	United Nations Programs on HIV/AIDS
WHO	World Health Organization
3TC	Lamivudine

CHAPTER ONE

1. INTRODUCTION

1.1. BACK GROUND OF THE STUDY

The Human Immunodeficiency Virus causes Acquired Immune-Deficiency Syndrome (AIDS), a condition in which the immune system begins to decline, exposing infected individuals to life-threatening opportunistic infections; the first acquired immunodeficiency syndrome case emerged in the early 1980s and since then, the AIDS prevalence has been increasing Silva et al. (2015). It is an epidemic that affects every part of the globe. According to (UNAIDS, 2021) an estimated 37.7 million people globally were living with HIV of these 27.5 million people were accessing antiretroviral therapy (ART); among total infected patients, 36.0 million were adults 74% had to access ART. Among 37.7 million infected patients 25.6 million people were living in Africa and 16,384000(64%) had to access ART. There were 1.3 million newly infected aged +15years globally with an estimated 765000(66%) those occurring in Africa and 680000 people were died due to AIDS and related illnesses globally followed by 467900(69%) people were in Africa.

In Ethiopia, an estimate of 669,236 people was living with HIV including 625,007 adults +15years. There are 14842 new infections of which 11,613(78.2%) were adults aged +15years. Furthermore, the number of deaths due to AIDS-related illnesses for the same period was estimated to be 11546 in the country and 9491(82%) were adults aged greater than 15years (EPHI, 2020).

According to Loures et al., (2020) the UNAIDS set 90–90–90 target goals to end AIDS epidemic by 2030 as part of global sustainable development goals. By 2020, 90% of all the people living with HIV will know their status, 90% of all the people with a diagnosed HIV infection will receive ART and 90%of all people receiving ART will have viral suppression. When this goal is achieved, it is believed that at least 73% of all people living with HIV worldwide will be virally suppressed.

WHO guideline recommended that, both zidovudine (AZT) and tenofovir (TDF)-based regimens containing efavirenz (EFV) or nevirapine (NVP) was first-line options Rights et al., (2019). However, the WHO consolidated guideline recommends that TDF/3TC/EFV regimen type was preferred first-line (Keele & Li, 2016). The first and second-line antiretroviral consolidated WHO guideline on ART for prevention, care, and treatment was developed (Brief & Treatment, 2019).

Non-adherence results in antiretroviral agents not being able to maintain sufficient concentration to suppress HIV replication in infected cells and to lower the plasma viral load (Mosha, 2018). Poor adherence also accelerates drug-resistant Chaiyachati et al. (2014) & Adeniyi et al. (2018).

The development of drug-resistant variants that can develop in HIV/AIDS patients under ART makes it not feasible to completely eradicate the virus (Shoko & Chikobvu, 2018). But, with proper adherence to treatment, ART has the potential to suppress viral replication, often below the level of detection by commercially available tests Olariu et al., (2017).

Antiretroviral therapy for HIV- infection can very effectively control the infection and hold the amount of a circulating virus below the level detectable by a clinical assay, improving both the quality and length of life Id et al., (2019), and consequently, the standard of care for people living with (PLWH) is to maintain life-long ART. However, there is significant heterogeneity in rebound times. In a pooled analysis of participants from six AIDS Clinical Trials Group analytic treatment interruption studies to identify predictors of viral rebound, the widely varying times to viral rebound, with a significant number of participants maintaining viral suppression to undetectable levels for up to 2 or more months in the absence of ART (Suppression & States, 2020).

The total CD4+ T-cell count $<$500cells/mm^{3}at 2–12years after ART initiation, with undetectable plasma VL; CD4+ T-cell count $>$500cells/mm^{3}at 2–12 years after ART initiation, with an undetectable plasma VL in both immunological non-responder and responder respectively Yang et al. (2020). For asymptomatic patients with higher CD4 counts (e.g.,$> 350\,\text{cells}/\text{mm}^3$), the question of when to initiate ART remains an area of research and debate. In persons with CD4 counts of $> 200\,\text{cells}/\text{mm}^3$, effective ART dramatically decreases morbidity and mortality. A variety of data from cohort studies show that a reduction in death as well as in AIDS and non-AIDS related complications among persons who have initiated ART with CD4 counts of $> 350\,\text{cells}/\text{mm}^3$ rather than $< 350\,\text{cells}/\text{mm}^3$ (Oguntibeju, 2012).

In recent years, the longitudinal analysis interest has grown rapidly through the development of new methods and the increase in computational power to aid and further develop this field of research. These processes are typically correlated, where both types of data are associated through unobserved random effects. When these processes are correlated, the use of independent models can cause biased estimates (Little, 2002; Ratclie, 2004; Yi-Kuan Tseng, 2015); with joint models resulting in a reduction in the standard error of estimates. Thus, with a more accurate estimate of the parameter, valid inferences on the longitudinal and survival processes can be obtained.

Joint modeling of longitudinal and time-to-event data is an area of increasing research (Tsiatis & Davidian, 2004), which allows the simultaneous modeling of a longitudinal outcome such as weekly biomarker measurements, and a time-to-event (survival) outcome such as time to death. Joint modeling enables the simultaneous study of a longitudinal marker and a correlated time-to-event. Among them, the shared random-effect models that are defined as a mixed model for the longitudinal marker and a survival model for the time-to-event including characteristics of the mixed effect model as covariates received the main interest. Indeed, they extend naturally the survival model within a time-dependent covariate and offer a flexible framework to explore the link between a longitudinal biomarker and a risk of event Wang et al. (2014).

The approach that this study used to build a joint model is simultaneously modeling the longitudinal CD4 cell count measurements and time-to-death processes by linking those using the shared random effects parameter model. In the proposed model, to characterize the longitudinal CD4 cell count measurements a linear mixed-effects model that incorporates patient-specific CD4

cell count intercept and slopes is used for the longitudinal sub-model while the Cox PH model is used to describe the time to-death survival data for the survival sub-model. Then, the two sub-models are linked with shared parameters McHunu et al., (2020), with different forms, since these random effects characterize the subject-specific longitudinal process.

The thesis is organized as follows: The statement of the problem; objectives of the study and significance of the study are presented next in this section. Section 2 describes some literature related to HIV infection and different joint modeling approaches. In Section 3, the data and the methods of data analyses employed are detailed explained. Then, in Section 4 basic results of the study are presented, and in Section 5 are discussed based on the result. Finally, some concluding remarks and recommendations are provided in Section 6.

1.2. STATEMENT OF THE PROBLEM

In many clinical studies, longitudinal data and survival data are frequently observed together in practice. CD4 counts measurement as a biomarker of disease progression are regularly measured repeatedly at different time points and time-to-event of a patient (e.g., viral rebound) is recorded under follow-up. In clinical trials longitudinal and survival data have been usually analyzed considering time-to-event data or repeated measurements separately Wu et al., (2012).

In HIV infection many well-established methods exist for analyzing longitudinal and survival data separately; including linear mixed-effects models for longitudinal modeling part, and semi-parametric or parametric models for survival part. But their separate use was inappropriate since the longitudinal measured CD4 cell count process is correlated with patient health status, Guo et al., (2004), and survival outcome (either with the subject's status as well as the possibility of dropout) may not be adequate and can lead to inefficient estimation or biased results because they fail to take into account the association between the two components of the data (Lim, 2017).

In a situation, where both outcomes are observed for each subject, separate modeling does not take into account the dependence between the two types of responses. The interrelationships of the two responses can be investigated by joint modeling. It generally behaves that when the association between the two processes exists, incorporate all information simultaneously, less biased and more efficient inferences will be obtained by using a joint model Guo et al., (2004).

The main aim of this thesis was studying the joint model of longitudinal estimated CD4 cell count measurement and time-to-viral rebound of HIV/AIDS-infected patients treated under antiretroviral therapy (ART) at Jimma University Medical Center.

In general, the study addresses the following major research questions:

1. What are the determinant factors for longitudinal CD4 cell count and survival time to a viral rebound of HIV-infected patients initiating ART?

2. How the average progressions of CD4 count of HIV patients initiating ART change over time?

3. How the association of the evolution between longitudinal CD4 count and the risk for viral rebound look like over time?

1.3. OBJECTIVES OF THE STUDY

1.3.1. General Objective

The main aim of this study was joint modeling of both the longitudinal CD4 cell count and time-to-viral rebound of HIV-infected patients using shared parameters.

1.3.2. Specific Objectives

Specifically, the study addresses the following specific objectives:

- To identify determinant factors for the change in CD4 cell count and viral rebound in HIV-infected patients in the ART.
- To estimate the average progression of CD4 cell counts of HIV-infected patients over time.
- To determine the association between longitudinal CD4 count and time-to-viral rebound of HIV-infected patients.

1.4. SIGNIFICANCE OF THE STUDY

The outcome of this study would provide information about the risk factors or the most influential covariate that would be a significant impact on the survival of HIV patients during treatment. Laboratory measurements, such as numbers of CD4 cells and levels of plasma HIV RNA (level of viral load), help determine the stage of infection and may serve as prognostic markers.

The results of this study will give a deeper insight into how the concept of a standardized measure of variability and Akaike information criterion can be applied in joint model analysis.

It also helps the health sectors as inputs to create awareness for the community on the risks for the survival of HIV infection.

It also used input for researchers who want to investigate HIV-infection-related areas by pointing directions to be addressed in the future. It also helps the clinicians to give consultancy and awareness for their infected patients depending on the identified risk factors.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. GENERAL REVIEWS ON HIV- INFECTIONS

The CD4 cell count is a critical measure of the immune system and is used as an important marker in describing the progression to AIDS. Some studies have demonstrated that subjects who have initiated ART at higher baseline CD4 cell count levels have better chances of immune recovery compared to subjects who are initiated at lower levels and an adequate CD4 response for most patients on therapy is defined as an increase in the range of 50–150 cells/mm^3 per year with an accelerated response in the first 3 months of treatment Manosuthi et al. (2015). Each CD4 cell count was considered to hold predictive value for no more than the subsequent 6month periods, with individual patients contributing multiple 6month periods of follow-up (Chaiyasin & Sungkanuparph, 2016). Lower CD4 counts are associated with a greater risk of disease progression. CD4 counts from 350-500$cell/mm^3$are associated with risks of $\leq 5\%$ across all age and HIV- RNA strata, while the risk of progression to AIDS increases substantially at CD4 count measure < 350 $cell/mm^3$, the greatest risk increase occurring as CD4 cell counts fell below 200 $cell/mm^3$. The risk of disease progression at 200 $cell/mm^3$, the threshold for ART initiation in resource-limited settings, is generally double the risk of 350 $cell/mm^3$, the treatment threshold in resource-rich countries (Chaiyasin & Sungkanuparph, 2016). Monitoring the levels of CD4 cell count is the standard used in decision making concerning the initiation of antiretroviral therapy and response to ART over time Andrew et al., (2008).

In a follow-up study, Li and his team identified post-treatment controllers from ATI studies, who maintained viral loads of $\leq$400 HIV RNA copies/mL for $\geq$ 24 weeks (Namazi & Sharaf, 2018). In previous reports of these rare PTCs include the VISCONTI cohort study, 14 PLWH who have initiated ART within three months of their estimated date of infection who were able to control HIV infection for a prolonged period after stopping ART Prazuck et al., (2013). Results from the VISCONTI study and others suggest that Post-treatments control may control HIV by mechanism distinct from that of spontaneous HIV controllers Lambotte et al., (2005). However, the factors that mediate delayed timing of HIV rebound are not well understood Woldemeskel et al., (2020).

A study conducted on the viral rebound among persons with diagnosed HIV who achieved viral suppression; united states with the objective of therapies to achieve sustained antiretroviral therapy-free HIV remission will be required to the validation in analytic treatment interruption (ATI) trials. Identifying biomarkers that predict time to viral rebound could accelerate the development of such therapeutics showed that of the participants who initiated ART among those with $\geq$ 2 viral load tests who achieved viral suppression, 7.5% demonstrated viral rebound. In multivariable analyses, viral rebound was higher among non-Hispanic blacks, person ages 18–39, persons with public insurance, persons recently experiencing homelessness, persons with higher

numbers of viral load tests, persons who missed HIV care appointments, and persons with suboptimal adherence to antiretroviral therapy (Suppression & States, 2020).

Study conducted by Ioannidis et al., (2000) Dynamics of HIV-1 viral load rebound among patients with previous suppression of viral replication, with the main objective to model the dynamics of HIV-1 rebound in patients receiving suboptimal therapy after the suppression of the plasma viremia to $< 200\,\text{cell}/\text{mm}^3$ by triple combination therapy. A rebound of plasma HIV RNA followed a sigmoid curve with an initial exponential phase. There was significant heterogeneity in the slopes of rebound for individual patients. In the indinavir maintenance rebounds, the average initial slope was estimated to be 0.587/day (doubling time 1.2 days). The slopes of the rebound in patients on zidovudine/lamivudine maintenance tended to be less steep on average. Among patients taking indinavir maintenance, the average Ro for the initial rebound of viremia was 4.3; in multivariate regressions, the slope of rebound was steeper during early rebound and in patients with higher viral load at the start of triple therapy or a higher CD4 cell count when indinavir monotherapy was initiated. The slope was less steep in patients with a greater increase in the number of CD4 cells during triple therapy.

According to Id et al., (2019) predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers conducted in the Pennsylvania State University Institutional Review Board, the Los Alamos National Laboratory Institutional Review Board, and the Partners Institutional Review Board with the main objective is to construct a model that predicts the viral rebound time, i.e., the time between suspension of therapy and detectable viremia the results represent first steps towards a model that can make predictions on a person living with HIV (PLWH)'s rebound time distribution based on biomarkers, and help identify PLWH with long viral rebound delays.

According to Mendoza et al. (2014) on the different viral rebound following discontinuation of ART in case of infection with viruses carrying L74V or thymidine-associated mutations conducted in service of infectious diseases, hospital Carlos III, Calle Nueva zelanda on a total of 76 patients discontinued treatment with didanosine plus hydroxyurea after 1 year of maintenance therapy. The greatest human immunodeficiency virus (HIV)-RNA rebounds were seen in 10 patients harboring an L74V mutation, and the presence of viruses with this mutation rapidly waned. In contrast, viral rebounds were significantly less pronounced in 12 subjects harboring thymidine-associated mutations; these mutations persisted in all instances. Thus, the selection of an L74V mutation during didanosine therapy may compromise HIV replication in vivo.

A cohort study was conducted by Phillips et al. (2001) in Europe between 1996 and 2000 with the main objective to characterize the relationship of viral load response to ART with baseline CD4 cell count and baseline viral load; using Cox proportional hazards models the results have shown that of 3226 patients during the median follow-up of 119 weeks, 2741 (85%) experienced viral suppression to less than 500 copies/mL by 32 weeks. Relative hazards (RHS) of achieving this

were 1.08 (95% CI: 0.98 - 1.21) and 0.94 (95% CI: 0.84 - 1.04) for baseline CD4 counts between 200 and $349 \times 10^6/\text{L}$ and baseline CD4 cell counts were lower than $200 \times 10^6/\text{L}$, respectively, compared with baseline CD4 cell counts of $350 \times 10^6/\text{L}$ or higher, after adjustment for several factors including baseline viral load. For baseline viral load, the RHS were 0.95 (95% CI: 0.84 - 1.07), and 0.65 (95% CI: 0.58 - 0.74), for 10 000 to 99 999 and 100000 copies/mL or greater, respectively, compared with less than 10000 copies/mL, but the probability of viral load lower than 500 copies/mL at week 32 was similar in all three groups. Subsequent rebound above 500 copies/mL was no more likely with a lower baseline CD4 cell count or higher viral load.

According to Min et al. (2020) on evaluating HIV viral rebound among persons on suppressive ARV treatment in the era of "Undetectable Equals Untransmittable (U = U)" in Southeastern New England between January 1, 2016, through December 31, 2019, using logistic regression models the results have shown that a total of 1242 patients with viral suppression were included in the baseline cohort and retention in care was significantly associated with viral suppression, while younger age, black race, high school or equivalent education, non–men who have sex with men, and history of incarceration were significantly associated with a viral rebound.

According to Smith et al. (2009) on factors associated with viral rebound among highly treatment-experienced HIV-positive patients who have achieved viral suppression British HIV Association with main objective more and more highly treatment-experienced patients are achieving viral suppression. However, the durability of suppression remains unclear of two hundred and forty-seven patients who contributed 723 person-years and 114 viral rebounds. More recent calendar years of viral suppression and a greater number of ARVs in the regimen not previously failed were associated with lower viral rebound rates.

According to Kayode et al. (2020) on assessment of the effect of antiretroviral therapy on hematological parameters in HIV positive individuals in Zaria a total of 230 patients receiving HAART for the first time and followed regularly were retained and their information has gotten using a questionnaire. Of this were Stavudine + Lamivudine + Nevirapine (Regimen 1) and they were zidovudine + Lamivudine + Nevirapine (Regimen 5). The data were analyzed using Graph pad In-Stat version 3. All patients had an appreciable increase in CD4 levels, patients on regimen 1 had a significant increase in Hb, PCV, and Lymphocyte count. Patients on regimen 5 on the other hand had a significant decrease in HB, PCV, and Lymphocyte count. In this study, a hematological response is better in regimen 1 than regimen 5.

A prospective cohort study conducted by Dessie et al., (2020) among HIV-infected women from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) on Modeling Viral Suppression, Viral Rebound and State-Specific Duration of HIV Patients with CD4 Cell Count Adjustment: Parametric Multistate Frailty Model Approach from August 2004 to December 2017 in which a total of 8760 follow-up visits were recorded for 219 HIV-infected women. The viral

rebound was found to be significantly associated with many sex partners, higher eosinophils count, younger age, lower educational level, higher monocyte counts, having abnormal neutrophils count, and higher liver enzyme abnormality. Furthermore, viral suppression was also found to be significantly associated with higher quality of life (QoL) scores, and having a stable sex partner. The analysis result also showed that patients with a stable sex partner, higher educational levels, higher quality of life (QoL) scores, lower eosinophils count, lower monocyte counts, and higher RBC indices were more likely to spend more time in an undetectable viral load state.

A cohort study conducted by (Shoko & Chikobvu, 2018) on HIV-infected patients receiving antiretroviral therapy in Bela, South Africa, from the year 2005 to the year 2009 with the main objective of the determinants of viral load rebound on HIV/AIDS infected patients receiving antiretroviral therapy of the total of 320 HIV-1 infected patients on antiretroviral therapy (ART) used continuous-time Markov model showed that of the results show no gender differences on transition intensities. The effects of the covariates including combination give a significantly better fit to the observed data. From almost all states, rates of viral suppression were higher than rates of viral rebound except for patients in state 2 where rates of viral rebound to state 3 were higher than rates of viral suppression to undetectable levels. For this transition, confidence intervals were very small. This was quite notable for patients who were administered with AZT-3TC-LPV/r & FTC-TDF-EFV. Although patients on d4T-3TC-EFV also had higher rates of viral rebound from state 2 than suppression, the difference was not significant.

The study cohort comprised 399 patients undergoing treatment follow up at a wellness clinic in Bela, South Africa; the Continuous-time non-homogeneous Markov model is used to the progression of HIV/AIDS in patients on combination antiretroviral therapy (ART). Results show that when the viral load of a patient is below 100000 copies/mL, rate of viral suppression are higher than rates of viral rebound. The undetectable viral load (state 1) is accessible from all the states. The rate of attainment of an undetectable viral load depends on the condition of a patient. Patients with the highest viral copies/mL (state 5) have the lowest risk of attaining an undetectable viral load whereas patients with the lowest viral copies/mL (state 2) have the highest chance of attaining an undetectable viral load (Claris, 2019).

A cross-sectional study conducted by(Tony & Emmanuel, 2020) on viral suppression and predictor among adolescents receiving care for HIV/AIDS in a tertiary health center in Uyo, South-South, Nigeria with the objective of the level of viral suppression and its predictors among adolescents living with HIV (ALHIV), who knew their status, at the pediatric infectious diseases unit in Nigeria; The socio-demographic data and responses to the possible factors that influencing viral suppression was obtained and recorded in a proforma. Viral load ranged from <40 to 522,244 HIV RNA copies/ml of blood. Parents being alive; caregivers being on routine medications; missing medications; a number of missed doses of ARV medications, and the current regimen of antiretroviral therapy were factors significantly associated with viral suppression.

A retrospective study by Maina et al. (2020) on incidences and factors associated with viral suppression or rebound among HIV infected patients on combination antiretroviral therapy from three countries in Kenya with the main objective to investigate the incidence rates of viral rebound following viral suppression, factors associated with a viral rebound, and the durability of viral suppression among HIV-infected individuals on ART from Kilifi, Meru, and Nakuru counties in Kenya. The covariates significant for the progression of CD4+ cell counts were good ART adherence; widow marital status, and WHO clinical stage I were associated with viral suppression, while poor ART adherence; WHO clinical stage II, and duration on ART of 36 months were associated with a viral rebound.

A study conducted by Asfaw et al., (2015) on CD4 cell count trends after commencement of ART among HIV-infected patients by using logistic regression analysis in Tigray, Northern Ethiopia. The median change from baseline to the most recent CD4 cell count was 292cell/mm^3. By 5years old, the overall median (inter-quartile range) CD4 cell count was 444(263 - 557)cell/mm^3 while the median (inter-quartile range, IQR) CD4 cell count was 342(246 - 580)cell/mm^3 among the HIV/ADS patients with a baseline CD4 cell count $\leq$200cell/mm^3, 500(241-557) cell/mm^3 among HIV-infected patients with baseline CD4 cell counts of 210-350 cell/mm^3, and 652(537-767)cell/mm^3 among those with baseline CD4 cell counts $>$350 cell/mm^3. Higher baseline CD4 counts and being male were independently associated with the risk of immunological non-response at 12months. Furthermore, it was also studied that these factors were significant predictors of subsequent CD4 cell count recovery.

A retrospective cohort study was conducted on East Shewa zone, Oromiya, Ethiopia, between October 3, 2011, and March 1, 2013, with the main objective of the time to viral load suppression and its associated factors in a cohort of patients taking antiretroviral treatment, showed that of the Plasma viral load was suppressed below the detection level in 72% of individuals taking a different regimen of ART. The median of Human Immunodeficiency Virus (HIV)-1 plasma viral load in the cohort study was estimated to be log 5.3111 copies/ml. The study observed Survival curve differences in the category of the marital status group (p-value 0.023) and baseline cluster of differentiation 4 (CD4) count value (p-value 0.023). The estimated median time to plasma viral load suppression of patient was 181 days within the age group between 30–39 years of having minimum time to achieve suppression of patient with 92 days and the maximum time required to reach the level was found among the age group between 50 and 59 years old (Ali, 2019).

According to Desta et al., (2020), a retrospective cross-sectional study was conducted in the Tigray region, Ethiopia, between April 2015 to March 2019; the main objective of this study was HIV virological non-suppression and factors associated with non-suppression among adolescents and adults on antiretroviral therapy. The study showed the covariates being male; age; WHO stage II; poor and fair ART adherence, and AZT-3TC-NVP and TDF-3TC-ATV/R regimen types were significantly associated with viral non-suppression.

2.3. REVIEWS ON JOINT MODELING APPROACHES

According to (Abebe, 2019) the result from the joint model showed longitudinal CD4 cell count is significantly associated with survival time. The estimated association parameter α is -0.10, this indicates both outcomes are negatively associated and higher values of the CD4 cell count are associated with better survival. The result from the longitudinal sub-model reviled that observation time, age, WHO clinical stage, history of TB, and functional status had significantly associated with mean change in the square root of CD4 cell count. Furthermore, from the survival sub-model, we found the survival probability of HIV-infected children was determined by WHO clinical stage, functional status, history of TB, and BMI.

A study on statistical joint modeling on longitudinal body weight and CD4 cell progression with survival time-to-death predictors on HIV/AIDS patients in Mekelle General Hospital, Ethiopia. The relationship between the two biomarkers CD4 cell and body weight with risk for survival time-to-death were statistical insignificant. In event process the sub-model, baseline CD4, fair and good adherence, HIV/TB (yes) and sex (male)were significant factors of risk to short survival time-to-death on HIV/AIDS patients. In the 1^{st}longitudinal process sub-model, baseline CD4, ambulatory functional status, HIV/TB (yes), time*ambulatory functional status, time*working functional status and time*baseline CD4 were the significant factors of square root CD4 count progression. Moreover, in 2^{nd} longitudinal process sub-model, visit Time of follow-up, age, sex (male), baseline weight, time*ambulatory and time*working functional status were the significant factors of $\log_{10}$(body weight) progression (Gebrerufael et al., 2020).

According to (Tiruneh et al., 2021) on joint modeling for predicting the association of CD4 count measurement and time to death of people living with HIV who enrolled in ART; a total of 358 HIV-positive patients. Males constitute the larger proportion, 51.68%. The square root of CD4 count has declined on average over time. The study further unveiled the factor where age significantly determined HIV patients' CD4 cell count, while sex determined the survival time of the patients.

A study by Temesgen & Kebede (2016) on the joint model of longitudinal CD4 cell count and weight measurements of HIV/TB co-infected patients found that sex, educational level, and functional status were the factors contributing to the prediction of HIV/TB co-infected patients weight at baseline among other variables.

According to (Tegegne et al., 2018) conduct joint longitudinal data analysis to identify the determinants of CD4 cell count change and adherence to highly active antiretroviral therapy at Felege Hiwot Teaching and Specialized Hospital identified age, baseline CD4 cell count, ownership of cell phone, visiting times, adherence to HAART, marital status, WHO stage, residence area and level of disclosure of the disease to family members had significantly affected

both outcomes. The joint model with linear predictor indicates that CD4 cell count change was positively correlated with adherence to HAART.

According to Basit et al. (2018) types of regimens contributed better to patients' survival time and CD4 cell count growth. The study further unveiled those factors such as preCD4 cell count, gender, and duration of treatment significantly determined HIV patient's CD4 cell count, while drug regimen, functional status, age, and preCD4 cell count determined the survival time of the patients.

A study conducted on an illustration using CD4 count and mortality in a cohort study of patients initiated on ART between June 2004 and August 2013 in South Africa showed that of using joint modeling, we found that lower CD4 cell count over time associated with a 1.3-fold increase in the risk of death. Whereas, results from the time-varying Cox model showed that lower CD4 count over time was associated with a 1.2-fold increase in the risk of death McHunu et al. (2020).

Some subjects drop out of the study before the occurrence of the terminal event of interest. One may then wish to evaluate the relationship between time to dropout & the internal covariate. The Cox model is a standard framework for that purpose (Ameraoui & Boukhetala, 2016) addressed this problem in situations where the value of the covariate at dropout is unobserved. They suggested a joint model which combines a first-order Markov model for the longitudinally measured covariate with a time-dependent Cox model for the dropout process by likelihood estimation of their model and shows how estimation can be carried out via the EM-algorithm. They state that the indicated joint model may have an application in the context of longitudinal data with non-ignorable dropout.

The accelerated failure time (AFT) model is an attractive alternative to the Cox model when the proportionality assumption fails to capture the relation between the survival time and longitudinal covariates (Saikia & Barman, 2017). They are several complications that arise when the covariates are measured intermittently at different time points for different subjects, possibly with measurement errors, or measurements are not available after the failure time.

Used a penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data and they proposed to use an estimation procedure based on a penalized joint likelihood generated by Laplace approximation of a joint likelihood and by using a partial likelihood instead of the full likelihood for the event time data. The results of their simulation study showed that this penalized likelihood approach performs as well as the corresponding EM algorithm under a variety of scenarios, but only requires a fraction of the computational time. They also identified an additional advantage of this approach which does not require estimation of the baseline hazard function and they applied the proposed procedure to a data set for evaluating the effect of the longitudinal biomarker PSA on the recurrence of prostate cancer Ye et al., (2008).

Proposed methods for joint modeling of survival time and longitudinal data assuming a mixed-effects model with subject-specific change points for the longitudinal covariates and the proportional hazards model for the survival times and they develop the conditional score and

corrected score estimators, which do not require the distributional assumption on the random effects or the change points and also they showed that the two functional methods are equivalent asymptotically and the demonstrated use of joint modeling in the analysis of an HIV dataset with CD4 cell count measurements and survival time. In their joint modeling, they combined a linear Gaussian random effects sub-model for the repeated CD4 cell count measurements and Cox or Weibull survival sub-model, linked through their shared dependence on the latent variable and they showed that the hazard rate of rebound depended on the longitudinal progression of CD4 cell counts, i.e., a patient's baseline CD4 cell count and the rate of change in CD4 cell counts significantly impact on his or her survival time (Lim, 2017).

Propose to estimate all the parameters using the nonparametric maximum likelihood estimators (NPMLE) on their Joint Models of Longitudinal Data and Recurrent events with the informative terminal event and they provide that the easy and efficient EM algorithms implement the proposed inference procedure. Asymptotic properties of the estimators are shown to be asymptotically normal and semi-parametrically efficient. Finally, they evaluate the performance of the method through extensive simulation investigation and a real-data application (Manuscript & Event, 2012).

Investigated the known association between hemoglobin fluctuations and the survival of dialysis patients and their joint model agrees that those patients with higher hemoglobin levels have a greater survival rate. They identified the significance of the shared parameter that links the two processes, and the reduction in the standard error of the parameter estimates when compared to independent model estimates, indicates the need for a joint analysis of for data compared to the use of independent models Rezende et al., (2020).

On their early review of the shared random-effect model and details of its implementation and evaluation through a real data example from the study of prostate cancer progression after radiation therapy. In particular, the different specifications of the dependency between the longitudinal biomarker, the prostate-specific antigen (PSA), and the risk of clinical recurrence are investigated to better understand the link between the PSA dynamics and the risk of clinical recurrence. They built different joint models that are compared in terms of goodness-of-fit and adequacy to the joint model assumptions but also in terms of predictive accuracy using the expected prognostic cross-entropy. Indeed, in addition, to better understand the link between the PSA dynamics and the risk of clinical recurrence, they used perspective in prostate cancer studies is to provide dynamic prognostic tools of clinical recurrence based on the biomarker history Sène et al., (2013).

According to Henderson, (2000) a joint model consists of two sub-models, which are referred to as the measurement model for the longitudinal process, and the intensity model for the survival process, and a latent association function of the random effects in which the two sub-models are linked. And these two processes are assumed to be conditionally independent given unobserved random effects (Wulfsohn & Tsisatis, 1997).

CHAPTER THREE

3. MATERIALS AND METHODS

3.1. DATA SOURCE

3.1.1. Study Area and Period

This study was conducted at Jimma University Medical Center, Oromia Region, Ethiopia on HIV-infected patients who initiated ART follow-up from February 1, 2016, until May 30, 2021.

3.1.2. Study Design

Data obtained through a retrospective cohort study design where basically joint longitudinal and survival modeling has considered to determine potential predictors. In this study, patients were followed on ART retrospectively for 54 months from February 1, 2016, to May 30, 2021, at Jimma University Medical Center, and age greater than or equal to 18 years were eligible. Joint modeling between CD4 cell count variation and the time-to-viral rebound was employed.

3.1.3. Study Population

All HIV/AIDS positive patients whose ages were 18 years and above were treated on ART follow up from February 1, 2016 to May 30, 2021 in Jimma University Medical Center.

3.1.4. Data Collection Procedure

The study used secondary data and a data extraction checklist was prepared to collect the data by reviewing their records. The relevant data were extracted from HIV/AIDS patients under ART follow-up charts which contain epidemiological, laboratory, and clinical information of all HIV infected patients under ART follow-up including detailed antiretroviral therapy history and have been collected by two professional data collectors and also one supervisor.

3.1.5. Quality of Data

The quality of data was controlled by data controllers from an antiretroviral therapy (ART) section of the Jimma University Medical Center. The necessary amendments were made on the final data collection sheet and the filled formats were checked daily by the supervisor and authors. The data extraction mechanism and variables included in this investigation checked its reliability of understanding and the completeness of data.

3.1.6. Inclusion and Exclusion Criteria

Patients who were 18 years old or older and who were attending a minimum of three visits of ART treatment between February 1, 2016, to May 30, 2021 GC in Jimma University Medical Center are included. In addition, patients out of the study period are excluded.

3.2. VARIABLES OF THE STUDY

The response and predictor variables considered in this study are defined as follows.

3.2.1. The Response Variable

The survival and longitudinal response variables in this study were CD4 cell count and time-to-viral rebound after ART treatment started.

The survival outcome variable was the survival time-to-viral rebound of the infected patients. Time-to-viral rebound of the patients was the time from date of started ART treatment to the viral rebound of the patients during the time period which was measured in month. Patients who lost the follow; transferred to another hospital before experience the event and did not rebound at thirty May 2021 (at end of the study) is considered as the censoring. This was measured from the starting of treatment till the patient's time-to-viral rebound or censored of the last visit (*i.e* in our study right censoring was faced).

The longitudinal continuous outcome which is a biomarker variable was the number of CD4 cell counts per cubic milliliter (mm3) of blood which was measured within six months interval.

3.2.2. The Covariates

The covariates (predictor) variables in this study were considered to potentially affect the CD4 cell count progression and then, aggravate the viral rebound of HIV/AIDS patients.

Notice that WHO Standardized Clinical Stage which is classified into four; 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. Hence disease severity increases from Stage I to Stage IV. The functional status of the patients is also a categorical covariate with three categories: Working, Ambulatory and Bedridden. Working patients are those patients who can able to work day to day while ambulatory patients are those patients who can able to work sometimes but bedridden patients cannot able to work due to infectious disease.

In general, the independent covariates considered for the separate longitudinal and survival modeling as well for the joint modeling are listed in the following Table 1.

Table 1: List of covariates used to an analysis by separately and joint modeling

No.	Variable name	Values of the Variable	Type
1.	Age	Years(baseline)	Continuous
2.	Viral load baseline	Copies/ml	Continuous
3.	CD4baseline (CD4)	Cells/mm^3	Continuous
4.	Treatment change	No, Yes	Categorical
5.	WHO Clinical stage	Stage-I, Stage-II, Stage-III, Stage-IV	Categorical
6.	Regimen type	ART regimens combinations	Categorical
7.	Functional status	Working, Ambulatory, Bed ridden	Categorical
8.	Gender	Female, Male	Categorical
9.	Marital status	Single, Married, Divorced, Widowed	Categorical
10.	Residence	Rural, Urban	Categorical
11.	Adherence	Poor, Fair, Good	Categorical
12.	Peripheral neuropathy	No, Yes	Categorical
13.	Education Level	Not Educated, Primary, Secondary, Tertiary	Categorical

3.3. MODEL SPECIFICATION

To extract information from the given data, the collected data was analyzed using different methods depending on the objective of the study to give a certain conclusion about the collected data. Both descriptive and inferential data analyses were considered.

In this study the author was used the following four types of different statistical data models:

- A linear mixed-effects model was used for continuous response variables for the longitudinal data like CD4 cell count.
- Survival model for the continuous time-to-viral rebound from antiretroviral therapy (ART) response variable like Cox proportional hazard model.
- Survival model for the continuous time-to-viral rebound from ART response variable like accelerated failure time (AFT) model when the proportionality assumption was fails.
- Joint model of longitudinal (CD4 cell count) analysis for longitudinal measurements with survival time-to-viral rebound.

3.3.1. Longitudinal Data Modeling

Longitudinal responses may arise in two common situations; one is when the measurements are taken from the same subject at different times and the other is when the measurements are taken

on related subjects (clusters). In both of these cases, the measurements are likely to be correlated. Therefore; the longitudinal model considers two sources of variations which are known to be; within-subject variation which is the variation in the measurements within each subject and between-subject variations; which is the variation in the data between different subjects. Modeling within-subject variation allows studying changes over time while modeling between-subject variations allows understanding differences between subjects.

3.3.1.1. Exploratory Data Analysis

The first step in any model-building process is exploratory data analysis. Data exploration is a very important tool to fit appropriate models and to look at the pattern of data over time. It shows as much of the raw data as possible rather than summarized values, highlight aggregate patterns of scientific interest. Some of the data explorations used for the study include individual profiles; for identification of within and between variability of CD4 count measurement of the patients at different time points, the average evolution; for identification the mean structure of the CD4 count measurements over time and the variance evolution; for identification of the variance structure of CD4 count measurement taken at different time points. In all exploration graphical inspection can be used by connecting each value computed at each time point separately. Since the data do not balance loess smoothing is used instead that gives us visualization of data to choose fixed-effects and random effects for the linear mixed model.

3.3.1.2. Linear Mixed Modeling (LMM)

This model arises when multiple observations are made on the same subject over time. Measurements made on the same variable for the same subject are likely to be correlated and another important outcome that is commonly measured in a longitudinal study is the time until a key clinical event of interest occurs such as disease recurrence or rebound (time-to-event data).

LMM is a parametric linear model for repeatedly measured data that quantifies the relationships between a continuous dependent variable and various predictor variables when the response variable has been followed a normal distribution and also included accounting for the correlation and this might include both fixed-effect parameters associated with one or more continuous or categorical covariates and random effects associated with one or more random factors.

The linear mixed model (LMM) is used to model longitudinal outcomes by accounting for and between-subject sources of variations. This is due to the measurement taken from the same subject at different time points or the measurements taken from the same clusters are likely to be correlated.

Linear mixed models (LMM) is statistical models for longitudinal or repeated-measures studies, in which subjects are measured repeatedly over time or under different conditions and measurements in which the residuals are normally distributed but may not be independent (have correlations) this LMM is proposed by Laird & Ware (1982) on which their work was based

on Harville (1977), included a unified approach using growth models and repeated-measures models for the sequence of the longitudinal measurements $y_{i1}, y_{i2}, \dots, y_{ini}$ for the i^{th} subject at times $t_{i1}, t_{i2}, \dots, t_{ini}$ is model as:

$$\boldsymbol{y_i} = \boldsymbol{X}^T(\boldsymbol{t})\boldsymbol{\beta} + \boldsymbol{Z}_{\boldsymbol{i}}^T(\boldsymbol{t})\boldsymbol{b}_i + \boldsymbol{\varepsilon_i}$$

$$\boldsymbol{y_i} = \boldsymbol{\mu_i}(\boldsymbol{t}) + \boldsymbol{U_{1i}}(\boldsymbol{t}) + \boldsymbol{\varepsilon_i} \quad \dots\dots\dots (1)$$

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \boldsymbol{\delta}_{\varepsilon}^{2}\boldsymbol{I})$$

Where $\boldsymbol{y}_i$is the $n_i \; x \; 1$ vector of observed response values, $\boldsymbol{\beta}$ is the $p \; x \; 1$ vector of fixed-effects parameters, $\boldsymbol{X}(\boldsymbol{t})$ is the $n_i \; x \; p$ observed design matrix corresponding to the fixed-effects, $\boldsymbol{b_i}$ is the $q \; x \; 1$ vector of random-effects parameters, $\boldsymbol{Z_i}$ is the $n_i \; x \; q$ observed design matrix corresponding to the random-effects, and $\boldsymbol{\varepsilon_i}$ is the $n_j \; x \; 1$ vector of within-group errors which is normally distributed.

In this model, $\boldsymbol{\mu_i}(\boldsymbol{t}) = \boldsymbol{X}^T(\boldsymbol{t})\boldsymbol{\beta}$ represent is the mean square root of CD4 cell measurement and $\boldsymbol{U_{1i}}(\boldsymbol{t}) = \boldsymbol{Z}_{\boldsymbol{i}}^T(\boldsymbol{t})\boldsymbol{b_i}$ incorporates the part of the random effects which is the true individual level CD4 cell count measurement trajectories after they have been adjusted for the overall mean. Here, in mixed-effects models, random effects $\mathbf{b_i}$ is introduced for each subject to incorporate the correlation between the repeated measurements within a subject. Since each subject shares the same random effects, the measurements within-subject are correlated. Moreover, the random effects facilitate subject-specific inference.

In general, the above model (1) specifically incorporates both sources of variations: it uses random effects or subject effects to represent deviations of subject longitudinal trajectories from the population average. Thus, a mixed-effects model allows subject-specific inference, in addition to standard population average inference and the model was fitted in two stages in which the first stage involves the fitting of the appropriate fixed-effect model which is developed using a linear model and the second stage involves the selection of appropriate random effects parts for the selected fixed effects.

3.3.1.3. Covariance Structures

A model for the covariance must be chosen on the basis of some assumed model for the mean response. To reduce the number of parameters in the variance-covariance structure Σ, we can fit models with more parsimonious structures. The following are commonly used variance-covariance structures (Σ) among others: Independent (IND); Compound symmetry (CS); First-order autoregressive (AR(1)), and Unstructured (UN). Lead to more efficient inferences for the mean parameters & particularly useful when many repeated measurements are taken per subject.

Independent (IND)

The simplest covariance structure is the independent structure, where the within-subject error correlation is zero.

$$\Sigma = \begin{bmatrix} \delta^2 & 0 & 0 & \cdots & 0 \\ 0 & \delta^2 & 0 & \cdots & 0 \\ 0 & 0 & \delta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \delta^2 \end{bmatrix} = \delta^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Compound Symmetry (CS)

The covariance structure with the simplest correlation model is the CS structure. It assumes that the correlation is constant regardless of the distance between the time points. The corresponding correlation is given by:

$$\Sigma = \begin{bmatrix} \delta^2 & \delta^2 p & \delta^2 p & \cdots & \delta^2 p \\ \delta^2 p & \delta^2 & \delta^2 p & \cdots & \delta^2 p \\ \delta^2 p & \delta^2 p & \delta^2 & \cdots & \delta^2 p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta^2 p & \delta^2 p & \delta^2 p & \cdots & \delta^2 \end{bmatrix} = \delta^2 \begin{bmatrix} 1 & p & p & \cdots & p \\ p & 1 & p & \cdots & p \\ p & p & 1 & \cdots & p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p & p & p & \cdots & 1 \end{bmatrix}$$

First Order Autoregressive (AR(1))

The AR(1) structure is often used to fit models to data sets with equally spaced longitudinal observations on the same units of analysis. This structure implies that observations closer to each other in time exhibit a higher correlation than observations further apart in time. The general form of the Σ matrix for this covariance structure is as follows:

$$\Sigma = \begin{bmatrix} \delta^2 & \delta^2 p & \delta^2 p^2 & \cdots & \delta^2 p^{n-1} \\ \delta^2 p & \delta^2 & \delta^2 p^2 & \cdots & \delta^2 p^{n-2} \\ \delta^2 p & \delta^2 p^2 & \delta^2 & \cdots & \delta^2 p^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta^2 p^{n-1} & \delta^2 p^{n-2} & \delta^2 p^{n-3} & \cdots & \delta^2 \end{bmatrix} = \delta^2 \begin{bmatrix} 1 & p & p^2 & \cdots & p^{n-1} \\ p & 1 & p^3 & \cdots & p^{n-2} \\ p & p^2 & 1 & \cdots & p^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p^{n-1} & p^{n-2} & p^{n-3} & \cdots & 1 \end{bmatrix}$$

The AR(1) is a special case of the Toeplitz covariance structure and is useful for modeling first-order temporal autocorrelation structure.

Unstructured (UN)

The most complex covariance structure is UN covariance because it is estimating unique correlations within-subject errors for each pair of time intervals. It adds a significant number of free parameters to the fitting process since a $p \times p$ covariance matrix has $\frac{p(p+1)}{2}$ non-redundant elements.

$$\Sigma = \begin{bmatrix} \delta_1^2 & \delta_{12} & \delta_{13} & . & . & . & \delta_{1p} \\ \delta_{21} & \delta_2^2 & \delta_{23} & . & . & . & \delta p \\ \delta_{31} & \delta_{32} & \delta_3^2 & . & . & . & \delta p \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ \delta_{p^1} & \delta_{p^2} & \delta_{p^3} & . & . & . & \delta p^2 \end{bmatrix} = \delta^2 \begin{bmatrix} 1 & p_{12} & p_{13} & . & . & . & p^{n-1} \\ p_{12} & 1 & p_{23} & . & . & . & p^{n-2} \\ p_{31} & p_{32} & 1 & . & . & . & p^{n-3} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ p_{p^1} & p_{p^2} & p_{p^3} & . & . & . & 1 \end{bmatrix}$$

Each of these can be described in a fairly intuitive manner. The correlation and/or the covariance structure will be obtained in order to determine the type of correlation structure of the random effects to be considered in the model. Therefore, in any given analysis, we try to determine the structure for the Σ_i the matrix that seems most appropriate and parsimonious, given the observed data and knowledge about the relationships between observations on an individual subject. Hence, a common recommendation is to only choose the covariance structures that make sense given data.

Random Intercept Model

The random-effects model or subject-specific model assumes that extra correlation arises among longitudinal response Diggle et al., (2002). The random intercepts model allows intercepts to vary across groups. In particular, a basic example of a random intercepts model which is included in order to illustrate the model fitting is formed by two distinct parts,

$$Y_i = \beta_0 + \beta_1 X_{ij} + u_{0i} + \varepsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

These are a fixed part (which is the intercept and the coefficient of the explanatory variable times the explanatory variable) and a random part. The random part is composed of two random terms $\varepsilon_i \sim N(0, \delta^2)$ and $u_i \sim N(0, \delta_u^2)$ the random effect u_i and within-subject error ε_i are independent for different subjects and independent of each other for the same subject. $i.e.$ $cov(\varepsilon_i, \varepsilon_j)$ if $i \neq j$, $cov(u_i, u_j) = 0$ if $i \neq j$ and $cov(u_i, \varepsilon_j) = 0$. In the mixed model formulation, the design matrices are replaced by:

$$X_i = \begin{bmatrix} 1_1 & x_{i1} \\ . & . \\ . & . \\ . & . \\ 1_{ni} & x_{ini} \end{bmatrix}, \ Z_i = \begin{bmatrix} 1_1 \\ . \\ . \\ . \\ x_{ini} \end{bmatrix}, \boldsymbol{\beta} = [\beta_0, \ \beta_1]^T$$

And the random effects model covariance structure, and $u_i \sim N(0, D_i)$, with $D_i = \delta_u^2$

Random Intercept and Random Slope Model

An intuitive extension that allows a random shift in the subject-specific slopes is known as random intercepts and random slopes model. Consider the simple random intercepts and slopes model,

$$Y_i = \beta_0 + \beta_1 X_{ij} + u_{0i} + u_{1i} t_{ij} + \varepsilon_i \text{ .... (3)}$$

In this model we have additional u_{1i} which represents the random slope effect of the coefficient X_{ij}, $j = 1,2, \ldots n_j$, of $j'th$ response on $i'th$ subject. As a result, actually, two extra parameters should be estimated: the variance in intercepts between groups δ_{u0}^2 and the variance in slopes between groups δ_{u12}^2. In this case random effect design matrix Z_i has the form,

$$Z_i = \begin{bmatrix} 1_1 & x_{i1} \\ . & . \\ . & . \\ . & . \\ 1_{ni} & x_{ini} \end{bmatrix}, \text{ and}$$

The random-effects model covariance structure,

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N(0, D_I) \text{ with } D_i = \begin{bmatrix} \delta_{u0}^2 & \delta_{u0u1}^2 \\ \delta_{u0u1}^2 & \delta_{u1}^2 \end{bmatrix}$$

Where δ_{u0ui}^2 denotes the covariance between the intercepts and slopes.

3.3.1.4. Estimation of Linear Mixed Model

Estimation is more difficult in the mixed model than in the general linear model. This is because in the mixed model estimation of random effects and covariance structure of the random error is necessary besides the fixed effect. The maximum likelihood (ML) will consider for the estimation of the parameters of the model. The maximum likelihood estimation method finds the parameter estimates that are most likely to occur given the data. The parameter estimates are derived by maximizing the likelihood function, which is a mathematical expression that describes the joint probability of obtaining the data expressed as a function of the parameter estimates.

Maximum likelihood estimation: the maximum likelihood (ML) method used to estimate $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$. Let $\boldsymbol{V}$ be the variance of the response the maximum likelihood provides unbiased estimators under normal errors. The log-likelihood function for observed responses is given by:

$$\boldsymbol{L}(\boldsymbol{D}, \boldsymbol{\Sigma}) = -\frac{1}{2} \boldsymbol{log}|\boldsymbol{Y}| - \frac{1}{2}(\boldsymbol{M})^T \boldsymbol{Y}^{-1} \boldsymbol{M} - \frac{n-p}{2} log(2\pi) \text{ .... (4)}$$

Where; $\boldsymbol{M} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{Y},$ and p is the rank of $\mathbf{X}$ estimating the fixed effect $(\boldsymbol{\beta})$ and random effect $(\boldsymbol{b})$ parameters in the Mixed Model. Once getting estimates values of $\mathbf{D}$ and $\boldsymbol{\varepsilon}$, which are denoted by $\widehat{\boldsymbol{D}}$ and $\hat{\boldsymbol{\varepsilon}}$ hat respectively the estimated values of random effect and fixed

were based on these two estimated values (Proust-lima & Liquet, 2016). The computation of values parameters is based on statistical software packages R version 4.0.5.

3.3.2. Survival Data Modeling

Survival models are seeking to explain how the risk, or hazard, of an event occurring at a given time, is affected by covariates of theoretical interest. In a single event analysis, the survival function is defined as the probability that the survival time is greater or equal to t which is given by:

$$S(t) = P(T \geq t) = \int_t^{\infty} f(t)dt \quad \text{for } t \geq 0 \text{ .....} \quad (5)$$

Where $f(t)$ is the probability density function of event time T for continuous cases and the integration value becomes summation when we have a discrete-time event. Whereas the hazard rate is the instantaneous risk of experiencing the event at a given time given that it has survived (i.e., not experienced the event) up to that time which is given by:

$$\lambda(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T < \triangle t / T \geq t)}{\triangle t}, t \geq 0$$

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d \log(S(t))}{dt} \quad \text{.....} \quad (6)$$

Or in other words hazard function is the probability that an individual will experience an event. In general, survival techniques can be applied to a wider range of different situations, subject to the three requirements as stated by (Cox and D.Oakes, 1999); firstly a well-defined time origin must be determined, then a scale for measuring the progress of time must be defined, and finally, the exact definition of failure must be clear.

3.3.2.1. Non-parametric Survival Methods

Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and gets an idea of whether or not the groups are proportional, i.e., if the estimated survival functions for two groups are approximately parallel (do not cross).

The Kaplan-Meier estimator is a nonparametric estimator of the survival function which is not based on the actual observed event and censoring times, but rather on the order in which events occur. This principle of nonparametric estimation of the survival function is to assign a probability to and only to uncensored failure times. Suppose there are n observations, $t_1, \dots, t_n$, with corresponding censoring indicators $\delta_1, \dots, \delta_n$. Let the number of distinct event times be r $(r \leq n)$, with the ordered event times given by $t_{(1)} <, \dots, < t_{(r)}$ and the corresponding number of events $d_{(1)}, \dots, d_{(r)}$. And also let $R(t_{(j)})$ denote the risk set at the event time $t_{(j)}$, $i.e.$, the set of subjects that did not yet experience the event and were not yet censored before time $t_{(j)}$ and thus still at risk

for the event at that time. Therefore, the Kaplan-Meier estimate of the survival function at time t is given by:

$\hat{S}(t) = \prod_{j=1}^{k}\left(\frac{R(t_{(j)})-d_{(j)}}{R(t_{(j)})}\right)$, for $t_{(j)} < t < t_{(j+1)}, k = 1,2,\dots,r$ ..... (7)

3.3.2.2. Log-rank Test

The estimated Kaplan- Meier survival curves show the pattern of one survivorship function lying above another, this means the group defined by the upper estimated curve lived longer or had a more favorable survival experience than the group defined by the lower estimated curve. But, the statistical question is whether the observed difference seen on the curve is significant. One way of which give an answer for such statistical question is a log-rank test which is the most widely used to test the significant difference between the estimated Kaplan Meier survival curves where its computed statistics is given by:

$\frac{\left(\sum_{i=1}^{m} d_{1i}-\hat{e}_{1i}\right)^2}{\sum_{i=1}^{m}\hat{V}_{1i}}$, $\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}$ and $\hat{V}_{1i} = \frac{n_{0i}n_{1i}d_i(n_i-d_i)}{n_i^2(n_i-1)}$

Where;
m is the number of rank-ordered failures (rebound) times.
n_{0i} is the number of individuals at risk at observed survival time $t_{(1)}$ in group 0
n_{1i} is the number of individuals at risk at observed survival time $t_{(1)}$in group 1
d_{0i} is the number of observed rebounds in group 0
d_{1i} is the number of observed rebounds in group 1
n_i is the total number of individuals or risks prior to the time $t_{(i)}$
d_i is the total number of rebounds at a time (i)
The computed test statistics have a chi-square distribution.

3.3.2.3. Semi Parametric Survival Model: Cox PH Model

In survival analysis, to determine if the variation in subjects' survival experience is partially explained by covariates or to find any possible relationship between survival times and important covariates, a popular approach is to model the hazard function rather than the mean of the survival times as in the classical regression models. That is, survival models are most often defined in terms of the hazard function. Since a hazard function may be complicated, a parametric assumption can be avoided and the hazard function allowed being nonparametric. The most commonly used semi-parametric survival model which does not require the distributional assumption of the survival time is the Cox proportional hazard model proposed by (Cox, 2007) which expresses the hazard of an event at time *t* as:

$\boldsymbol{\lambda}_i(t) = \boldsymbol{\lambda}_0(t)exp(\boldsymbol{W}^T\boldsymbol{\gamma})$ ..... (8)

Where; $\boldsymbol{W}$ is the matrix of baseline covariates which may or may not have the same element with linear mixed-effects covariates $\boldsymbol{X}(t)$, $\boldsymbol{\gamma}$ is the vector of parameters and the term and $\boldsymbol{\lambda}_0(t)$ is the baseline hazard where the effects of covariates are zero.

If all of the covariates are zero the model (8) above become $\boldsymbol{\lambda}_i(t) = \boldsymbol{\lambda}_0(t)$ because of this, we call the term $\boldsymbol{\lambda}_0(t)$ the baseline hazard function.

The only assumption of this model is that the hazards ratio $\psi = \frac{\boldsymbol{\lambda}_i(t)}{\boldsymbol{\lambda}_j(t)}$ does not change over time (i.e., proportional hazards) which is why this model is also known as a semi-parametric model.

The parameter of the Cox proportional hazard model refers to the hazard ratio of one group in comparison to the other groups for categorical covariates and change in hazard ratio with a unit change of the covariate for the continuous variables when other covariates are fixed. The elements of (covariates in survival model) $\boldsymbol{W}$ may or may not be the same as that of longitudinal matrix covariates or $\boldsymbol{X}$ and the change in the hazard ratio for the continuous covariate is given by:

$\frac{\lambda(t,\ w_k+1)}{\lambda(t,\ w_k)} = \frac{exp(\gamma_1 w_1 + \ldots + \gamma_k(w_k+1) + \ldots)}{exp(\gamma_1 w_1 + \ldots + \gamma_k(w_k) + \ldots)} = exp(\gamma_k)$ which represents change (equivalently, $exp(\gamma_k) * 100\%$ percentage change) hazard function with a unit change in covariate provided that other covariates remain fixed. For a categorical covariate $\boldsymbol{W}$with l levels, the model contains $(l-1)$ dummy variables defined as $Z_i = 1$ if $\boldsymbol{W} = i$, and 0 otherwise for $= 1, 2, \ldots., l-1$. Let $\gamma_1 \ldots \gamma_{i-1}$ denote the coefficients in front of the appropriate dummy variables. Then the ratio of the hazard of two subjects, one with $\boldsymbol{W}$ at level $\boldsymbol{j}$ and the other with $\boldsymbol{W}$ at level k $(j, k = 1,2, \ldots l-1)$, provided the values of all other covariates for these subjects are the same, the hazard ratio between these two categories is given by:

$\frac{\boldsymbol{\lambda(t,z_i)}}{\boldsymbol{\lambda(t,z_k)}} = \frac{\boldsymbol{exp(\gamma_j)}}{\boldsymbol{exp(\gamma_k)}} = \boldsymbol{exp(\gamma_j - \gamma_k)}$ Which represents hazard functions for subjects at level j and at level k of the covariate $(j, k = 1,2, \ldots l-1)$, provided the other covariates have equal values. There are also some assumptions of the Cox proportional hazards model to fulfill that is; The ratio of the hazard function for two individuals with different sets of covariates does not depend on time, time is measured on a continuous scale and censoring occurs randomly.

3.3.2.4. Parametric Survival Models

Parametric survival models are models requiring the specification of a probability distribution for the survival times and survival times need to follow a certain parametric distribution. Parametric models assume that the survival data follow some probability distribution. The effect of covariates on survival time is through the conditional hazard function. The PH model of the parametric survival model is the same as a model (8) but in the parametric PH model, the baseline hazard function $\boldsymbol{\lambda}_0(t)$ is modeled parametrically which have a certain parametric distribution which represents the baseline hazard function for parametric survival model when all covariates are zero

and the influence of covariates are multiplicative through $exp(\boldsymbol{W^T\gamma})$. The proportional hazard for the different individuals for the parametric model also is given by:

$\frac{\lambda(t/W_j)}{\lambda(t/W_k)} = \frac{\lambda_0(\boldsymbol{t})exp(W_j^T\boldsymbol{\gamma})}{\lambda_0(\boldsymbol{t})exp(W_k^T\boldsymbol{\gamma})} = exp(W_j - W_k)^T\boldsymbol{\gamma}$ this is constant. In addition to the Cox PH model different parametric survival, models are also considered for the study by assuming different parametric distributions for the time to rebound to have an appropriate survival model for the infected patients. However, the hazard model for the parametric model is also the same as the Cox PH but the only difference is that the baseline hazard for the parametric survival model is modeled parametrically which has a specified parametric time distribution. If the proportional hazard is no longer valid an alternative method is survival regression modeling is the accelerated failure time (AFT) model since the model does not require the proportional hazard assumption. In the AFT model, we consider the log scale of time which is given by:

$$\text{Log}(\boldsymbol{T}) = \boldsymbol{W^T\gamma} + \boldsymbol{\sigma\zeta_i} \quad \text{.....} \quad (9)$$

Where $\boldsymbol{\zeta_i} \sim F$ and F is parametric error distribution and σ is the scale parameter.

Different distributional choices for $\boldsymbol{\zeta_i}$ lead to different models and the most common choice for the distribution of $\boldsymbol{\xi_i}$ is the Gumbel distribution which is an extreme value distribution. If $\boldsymbol{\zeta_i}$follows the Gumbel distribution, the survival time Ti follows a Weibull distribution. If $\boldsymbol{\zeta_i}$follows the Gumbel distribution and $\boldsymbol{\sigma = 1}$, then it will be reduced to an Exponential model and another common choice for the distribution of $\boldsymbol{\zeta_i}$ is the standard normal distribution N(0,1). If $\boldsymbol{\zeta_i}$ follows N(0,1), the survival time Ti follows a log-normal distribution. The logistic distribution is also another possible choice if $\boldsymbol{\zeta_i}$ follows a logistic distribution, the survival time Ti follows a log-logistic distribution, and the hazard function of the AFT model is given by:

$$\lambda(\boldsymbol{t/w}) = \lambda_0\Big(\boldsymbol{t}\, exp(-\boldsymbol{\gamma^T W})\Big)exp(-\boldsymbol{\gamma^T W}) \quad \text{.....} \quad (10)$$

We deal with the effect of the covariates through $exp(-\boldsymbol{\gamma^T W})$ that is the time scale is changed by a factor of $exp(-\boldsymbol{\gamma^T W})$.

3.3.2.5. Estimation methods of survival models

Semi parametric model parameter estimation method: In the Cox proportional hazards model we can estimate the vector of parameters $\boldsymbol{\gamma}$ without having any assumptions about the baseline hazard $\lambda_0(\boldsymbol{t})$. As a consequence, this model is more flexible and an estimate of the parameters can be obtained easily. Consider n independent individuals, the data that we need for the Cox proportional hazard model is represented by (T_i, δ_i, W_i) $i = 1,2,\dots,n$, Where, $t_i =$ the survival time for the i^{th} individual $\boldsymbol{\delta_i} =$ an indicator of censoring for the $\boldsymbol{i^{th}}$ an individual is given by 0 for censored and 1 for event $\boldsymbol{W_i} =$ a vector of covariates for individual i $(w_1, w_2, \dots, W_p)$.

The full likelihood for right-censored data can be constructed as

$\boldsymbol{L(\gamma)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{h}(T_i, \boldsymbol{W}, \boldsymbol{\gamma})^{\delta_i}\, \boldsymbol{S}(T_i, \boldsymbol{W}, \boldsymbol{\gamma})$ ………………………………………………. (11)

Where;

$h(t_i \;\; W_i, \gamma\,) = h_0(t_i)exp(\boldsymbol{\gamma^T} W_i)$ is the hazard function for individual i

$h(t_i \;\; W, \gamma\,) = \;(S_0(t_i))^{exp(\boldsymbol{\gamma^T W})}$ is the survival function for individuals i.

It follows that $\boldsymbol{L(\gamma)} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} (h_0(t_i)exp(\boldsymbol{\gamma^T W}))^{\delta_i} (S_0(t_i))^{exp(\boldsymbol{\gamma^T W})}$

The full maximum likelihood estimator of $\boldsymbol{\gamma}$ can be obtained by differentiating $\boldsymbol{L(\gamma)}$ with respect to the components of $\boldsymbol{\gamma}$ and the baseline hazard. This implies that unless we explicitly specify the baseline hazard, as in the case of parametric PH, we cannot obtain the maximum likelihood estimators for the full likelihood. To avoid the specification of the baseline hazard, (Cooke, 2007) proposed a partial likelihood approach that treats the baseline hazard as a nuisance parameter and removes it from the estimating equation. Instead of constructing a full likelihood, we consider the probability that an individual experiences an event at a time t_i given that an event occurred at that time.

Partial likelihood: Let R_i denote the set of individuals at risk at a time just prior to $t_{(i)}$. Assume that for the present case there is only one failure at a time $t_{(i)}$, i.e, no ties. The probability that individual i with covariates w_i is the one who experience the event at a time $t_{(i)}$is given by:

$\frac{h(t,W)}{\sum_{j\epsilon R_{t(i)}} h(t,w_i)}$ and under the proportional hazards assumption on an equation, the ratio

$\frac{h_0(t)exp(\gamma^T W_i)}{\sum_{j\epsilon R_{t(i)}} h_0(t)exp(\gamma^T w_i)}$ shows the contribution to the partial likelihood at each event time $t_{(i)}$ by the individuals with covariate w_i in risk set $R_{t(i)}$.

where $R_{t(i)}$ are the overall subjects in the risk set at the time $t_{(i)}$ but by eliminating the baseline hazards function, the above equation becomes $\frac{exp(\gamma^T w_i)}{\sum_{j\epsilon R_{t(i)}} exp(\gamma^T w_i)}$

Thus, the partial likelihood is the product overall failure time $t_{(i)}$ for $i = 1,2,..,m$ of the conditional probability to give a partial likelihood

$\boldsymbol{L_P(\gamma)} = \prod_{i=1}^{m} \frac{exp(\gamma^T W_i)}{\sum_{j\epsilon R_{t(i)}} exp(\gamma^T W_i)}$ ……………..…………………………………. (12)

The product is over the m distinct ordered survival times and w_i denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$. The log partial likelihood function is

$l_P(\boldsymbol{\gamma}) = \sum_{i=1}^{m} \left[\boldsymbol{\gamma^T} W_i - \ln\left(\sum_{j\epsilon R_{t(i)}} exp(\boldsymbol{\gamma^T} W_i)\right)\right]$ …………………….……………. (13)

under this setting the maximum partial likelihood estimator was obtained by differentiating the function concerning $\boldsymbol{\gamma}$, setting the derivative equal to zero, and solving for the unknown parameters. But the partial likelihood derived above is valid when there are no ties in the data set. In most real situations tied survival times are more likely to occur. To handle this fact, partial likelihood algorithms have been adopted to handle ties.

There are different methods to estimate regression parameters when there are ties. The most popular and easy approaches are Breslow's and Efron approximation, in this study the Breslow approximation which is the default value of ties handling in statistical software packages R version 4.0.5 is used in case of ties.

Estimation of parameters for parametric survival model: In parametric modeling, the ML estimation method is commonly used estimation of parameters of the model. For parametric regression model with baseline hazard function and with a vector of regression coefficient $\boldsymbol{\gamma}$ including the intercept, parameter suppose that the random variable and suppose that $(t_i, \boldsymbol{\delta_i}, W_i)$ come from the parametric hazard rate regression with the parametric distribution. The likely hood function that maximizes the parameter $\boldsymbol{\gamma}$ is given by:

$$\boldsymbol{L(\theta)} = \prod_{i=1}^{n}\left(\lambda_0(t)exp(\boldsymbol{W^T\gamma})\right)^{\boldsymbol{\delta_i}} exp\left(-\int_0^t \lambda_0(u)exp(\boldsymbol{W^T\gamma})\right) du \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

The likely hood function is also constructed in terms of AFT perspective which is given by:

$$\boldsymbol{L(\theta)} = \prod_{i=1}^{n}\{\lambda_0(t\ exp(-\boldsymbol{W^T\gamma}))\exp\ (-\boldsymbol{W^T\gamma})\}^{\boldsymbol{\delta_i}} exp\left(-\Lambda_{\boldsymbol{0}}\left(t\ exp(-\boldsymbol{W^T\gamma})\right)\right) \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

The estimation of parameters for the model was based on the full likelihood function in both cases and the required parameters were obtained by maximizing the full log-likelihood function concerning the required parameter and statistical software packages R version 4.0.5 was used for all computations.

3.4. THE JOINT MODELING STRUCTURE

Recently, joint modeling research has expanded very rapidly in Biostatistics and medical research. This is due to the model enabling the simultaneous study of a longitudinal marker and a correlated time to event. Among them, the shared random-effect models that are defined as a mixed model for the longitudinal marker and a survival model for the time-to-event including characteristics of the mixed model as covariates received the main interest. They extend naturally the survival model with time-dependent covariates and offer a flexible framework to explore the link between a longitudinal biomarker and a risk of an event (Adele, 2011).

The main aim of this study was also to relate longitudinally measure CD4 biomarker with time to viral rebound for HIV-infected patients to understand the association between the two processes.

Therefore, after having appropriate separate models the longitudinal sub-model has the same specification as the separate linear mixed model (1). The survival sub-model includes a shared parameter association function to the specified Cox PH model (8). This shared association parameter associates the longitudinally measured CD4 cell measurement random effects with a time-to-viral rebound of infected patients which can be expressed as follows:

$$\boldsymbol{\lambda}_i(t/\boldsymbol{W}) = \boldsymbol{\lambda}_0(t)\exp\left(\boldsymbol{W}^T\boldsymbol{\gamma} + \boldsymbol{U}_{2i}(t)\right) \quad \text{(16)}$$

Where $\boldsymbol{\lambda}_0(t)$ is the baseline hazards rate, $\boldsymbol{U}_{2i}(t)$ defines the nature association structure of the shared parameters between the two processes which have a multivariate distribution function.

1. $U_{2i}(t) = \alpha^T m_i(t)$

2. $U_{2i}(t) = \alpha^T b_i \ and$

3. $U_{2i}(t) = \alpha^T(\beta_b + b_i)$. Here, the values of $m_i(t)$, denotes a current underlying value of the longitudinally measure CD4 cell measurement marker processes at the same time point; α measures the strength of association vectors between two processes; b_i is random effect parameters of the longitudinal part and β_b is fixed effect parameters corresponding to the random effects.

3.4.1. Joint Model Estimation Methods

The maximum likelihood estimation to jointly model the survival time and its longitudinal variables has been successful to model both processes in longitudinal data. Random effects in the longitudinal process are oftentimes used to model the survival times through a proportional hazards model, and this invokes an EM algorithm used for the maximum likelihood estimates (MLEs). Several intriguing issues are investigated here, including the robustness of the MLEs against departure from the normal random effects assumption, and difficulties with the profile likelihood approach to provide reliable estimates for the standard error of the MLEs Hsieh et al. (2006).

The main estimation method proposed for a joint model is maximum likelihood. The standard ML method involves maximizing the log-likelihood, corresponding to the joint distribution of the time-to-event and longitudinal data processes. Strictly, both processes share the same unobserved random effects and are conditionally independent given these random effects (Rizopoulos, 2012), thus

$$f\left(T_i, \delta_i, Y_i \,|\mathrm{U}_i; \Theta\right) = f\left(T_i, \delta_i \,|\mathrm{U}_i: \Theta\right) f\left(Y_i \,|\mathrm{U}_i;\ \Theta\right) \quad \text{(17)}$$

$$\text{With } f\left(Y_i \,|\mathrm{U}_i;\ \Theta\right) = \prod f\left\{Y_i(t_{ij}) \,|\mathrm{U}_i;\ \Theta\right\} \quad \text{(18)}$$

Because of the fact that the survival and longitudinal sub-models share the same random effects, joint models of this type are also known as shared random-effects models. Under these conditional

independence assumptions between longitudinal outcome and time-to-event has given the random effects U_i, the joint log-likelihood contribution of the $i'th$ subject is expressed as

$$logf(T_i, \delta_i, Y_i; \Theta) = log \int f(T_i, \delta_i, Y_i; \Theta) dU_i$$

$$= log \int f\left(T_i, \delta_i \,|U_i; \Theta_t, \boldsymbol{\beta}\right) \left[\prod f\left\{Y_i(t_{ij}) \,|U_i;\ \Theta_y\right\}\right] f(U_i;\ \Theta_b) dU_i \ \ldots\ldots \text{(19)}$$

Where Θ_t, Θ_y, and Θ_b represent the parameters for the survival process, the longitudinal process, and the random effects respectively, $f\left\{Y_i(t_{ij}) \,|U_i;\ \Theta_y\right\}$ is the density for the longitudinal process and $f(U_i;\ \Theta_b)$is the density for the random effects. The likelihood of the survival part $f\left(T_i, \delta_i \,|U_i; \Theta_t, \boldsymbol{\beta}\right)$ is written as

$$f\left(T_i, \delta_i \,|U_i; \Theta_t, \boldsymbol{\beta}\right) = \lambda_i \left[T_i \,|M_i(T_i); \Theta_t, \boldsymbol{\beta}\right]^{\delta_i} S\left(T_i \,|M_i(T_i); \Theta_t, \boldsymbol{\beta}\right) \ldots\ldots\ldots\ldots\ldots\ldots \text{(20)}$$

And, the survivor function for the $i'th$ subjects are given by,

$$S\left(t \,|M_i(t), \omega_i; \Theta_t, \boldsymbol{\beta}\right) = Pr\left(T_i > t(M_i(t), \omega_i; \Theta_t, \boldsymbol{\beta})\right)$$

$$= exp\left\{-\int_0^t \lambda_i \left(S \,|M_i(S); \Theta_t, \boldsymbol{\beta}\right) dus\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(21)}$$

The log-likelihood for the joint model is approximated using the Expectation-Maximization (EM) algorithm, because both the integral with respect to the random effects and survival function typically do not have an analytical solution, except in some special cases.

3.5. MODEL SELECTION TECHNIQUES

To select the model which is best fits the given data, it is important to compare different models by using different techniques and methods. Hence, the comparison between different models is an important issue in statistical inference. To have an appropriate separate longitudinal and survival model Akaki information criteria (AIC) proposed by Akaike (Akaike, 1974) and Bayesian information criteria (BIC) proposed by Spiegelhalter et al. (2002) of the model which can be expressed as follows are considered.

$$\text{AIC} = -2\text{logLikilehood} + 2\text{npar} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(22)}$$

$$\text{BIC} = -2\text{logLikilehood} + 2\text{npar} * \ln(\text{N}) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(23)}$$

Where logLik is the log likelihood function, npar is the number of parameters in the model and *N* is the total number of observations considered to estimate the model. The model with smaller values of AIC and BIC values is considered the preferred model.

3.5.1. Likelihood Ratio Tests (LRTs)

LRTs are a class of tests that are based on comparing the values of likelihood functions for two models (i.e., the nested (null hypothesis) and reference models) defined as

$$-2log\left[\frac{L_{nested}}{L_{refernce}}\right] = -2log[L_{nested}] - \left[-2log\left[L_{refernce}\right]\right] \sim \chi^2(df)$$

Where L_{nested} and $L_{refernce}$ denote the ML or REML estimates under the null and alternative hypothesis, respectively. Likelihood theory states that under mild regularity conditions the LRT statistic asymptotically follows a χ^2 distribution, in which the number of degrees of freedom, df, is obtained by subtracting the number of parameters in the nested model from the number of parameters in the reference model (Brazzale & Mameli, 2018).

When the number of variables is relatively large, it can be computationally expensive to fit all possible models. In this situation, automatic routines for variable selection that are available in many software packages might seem an attractive prospect. These routines are based on forward selection, backward elimination, or the combination of the two known as the stepwise procedure. The model selection strategy depends to some extent on the purpose of the study. In a situation where the aim is to identify variables upon which the hazard function depends, instead of using the automatic variable selection procedures, the following procedure is recommended.

1. The first step is fitting a univariable model for each of the explanatory variables and identifying the variables that are significant at some level from 20% to 25% is recommended (Zhang, 2016).
2. The variables that appear to be important in 1 are then fitted together in a multivariable model. In the presence of certain variables, others may cease to be important. Consequently, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables, in turn, should be examined.

3.6. MODELS DIAGNOSTICS

Model diagnostic checking is particularly important. A standard tool to perform model diagnostics are residual graphical methods, as many model checking procedures are based on quantities known as residuals plots, and formal statistical tests. Residuals are values that can be calculated for each observation and have the feature that their behavior is known, at least approximately, when the fitted model is satisfactory. The following residuals have been proposed for use Jones et al., (2012) in connection with the types of residuals for joint models.

3.6.1. Standardized Marginal and Standardized Subject-Specific Residuals

For the longitudinal part of the joint model, two frequently used types of residuals are the standardized marginal and standardized subject-specific residuals, which are defined as

$r_i^{(ym)} = \hat{V}^{-1/2}(y_i - X_i\boldsymbol{\beta})$, and

$$r_i^{(ys)} = \{y_i(t_{ij}) - x_i^T(t_{ij})\hat{\beta} - Z_i^T(t_{ij})u_i\}/\delta_i \quad \text{..... (24)}$$

Where $\hat{\beta}$, δ and $\widehat{D}$ denote the maximum likelihood estimates under model longitudinal model $\widehat{u_i}$ are the empirical Bayes estimates for the random effects, and $V_i = Z_i\widehat{D}Z_i^T + \hat{\delta}^2\boldsymbol{I}$ with $\boldsymbol{I}$ denoting the identity matrix of appropriate dimensions. The marginal residual $r_i^{(ym)}$ predict the marginal errors $y_i - X_i\boldsymbol{\beta} = Z_iu_i + \varepsilon_{yi}$, and can be used to investigate miss-specification of the mean structure $X_i\boldsymbol{\beta}$ as well as to validate the assumptions for the within-subject covariance structure V_i. The subject-specific residuals $r_i^{(ys)}(t_{ij})$ predict the conditional errors $\varepsilon_i(t)$, and can be used for checking the homoscedasticity and normality assumptions.

3.6.2. Martingale and Cox-Snell Residuals

For the survival part of the joint model, a standard type of residuals is the martingale residuals defined as:

$$r_i^{(tm)} = \delta_i - \int_0^T h_i\left(S\,|\widehat{M}(S);\,\hat{\theta}\right)ds \quad \text{..... (25)}$$

These are commonly used for a direct assessment of excess events (i.e., to reveal subjects that are poorly fit by the model), and for evaluating whether the appropriate functional form for a covariate is used in the model. Another type of residuals for survival models, related to the martingale residuals, is the cox-snell residuals.

These are calculated as the value of cumulative risk function evaluated at the observed event times T_i.

$$r_i^{(tcs)} = \delta_i - \int_0^T h_i\left(S\,|\widehat{M}(S);\,\hat{\theta}\right)ds \quad \text{..... (26)}$$

If the assumed model fits the data well, we expect $r_i^{(tcs)}$ to have a unit exponential distribution; however, when T_i is censored, $r_i^{(tcs)}$ will be censored as well. To take censoring into account in checking the fit of the model, we can compare graphically the Kaplan-Meier estimate of the survival function of $r_i^{(tcs)}$ with the survival function of the unit exponential distribution.

CHAPTER FOUR

4. ANALYSIS AND RESULTS

4.1 DESCRIPTIVE ANALYSIS

The data consists of 309 patients who were HIV/AIDS infected and who were treated under ART between February 1^{st}, 2016 to May 30, 2021, in Jimma University Medical Center. All HIV infected patients who were below 18 years and those patients who started ART before 1^{st} February 2016 or after thirty May 2021 are excluded from the analysis.

As mentioned in Section 3.2.1., the two response variables were considered for the study; longitudinal and survival responses. The survival end point is the viral rebound of HIV-infected patients during February 1^{st}, 2016 to May 30, 2021 and those patients who missed the follow up; transferred to another hospital between the specified time period and did not rebound at thirty May 2021 were considered as right censoring values.

Descriptive statistics of baseline covariate for HIV-infected patients was illustrated in Table 2. Thus, among 309 HIV/AIDS positive patients considered in this study, 235(76.1%) of them were viral rebound while the remaining 74(23.9%) were right censored observation. The mean baseline age of HIV-infected patients was 36 years with a standard deviation of 10 years.

Among the 309 HIV-infected patients eligible for the study, 124(40.1%) were males and the remaining 185(59.9%) were females. Among the total marital status category 154(49.8%) of the infected patients were married while smaller number 32(10.5%) of the infected patients belong windowed marital status of the total viral rebound occurred in these groups 120(51.1%) and 25(10.6%) of the viral rebound occurred in married and windowed marital status respectively which represents the larger and smaller percentages according to the marital status category. There were 221(71.5%) patients who were able to work; 54(17.5%) were ambulatory and 34(11%) were bedridden in the functional status categories; out of the total patient's viral rebound in these categories 155(65.9%) of the viral rebound were occurred in the patient group were working.

Of the WHO clinical stages, 84(27.2%) of the patients were at clinical stage 1; 47(15.2 %) were at clinical stage 2; 134(43.4%) were at clinical stage 3 and the rest 44(14.2%) were at clinical stage 4 at the time of starting the ART treatment. More of these infected patients 218(70.6%) came from the urban and large number 164(69.8%) of viral rebound also occurred in this group and the remaining came from the rural areas.

Of the total of 309 patients, 195(63.1%) were getting poor adherence and a larger number of viral rebounds occurred in this group. The percentage of HIV-infected patients who were 174(56.3%) having peripheral neuropathy and 160(68.1%) viral rebounds have occurred.

Table 2: Descriptive summary of baseline characteristics from HIV-infected patients under ART at Jimma University Medical Center.

N_o. Covariates	Categories	Total n(%)	Status of the observation	
			Censored Observed n(%)	Observed events n(%)
1. Gender	Female	185(59.9)	37(50)	148(63)
	Male	124(40.1)	37(50)	87(37)
2. Marital status	Single	43(13.9)	15(20.3)	28(11.9)
	Married	154(49.8)	34(45.9)	120(51.1)
	Divorced	80(25.9)	18(24.3)	62(26.4)
	Widowed	32(10.5)	7(9.5)	25(10.6)
3. Functional status	Working	221(71.5)	66(89.1)	155(65.9)
	Ambulatory	54(17.5)	1(1.4)	53(22.5)
	Bed ridden	34(11)	7(9.5)	27(11.4)
4. Regimen type	AZT+3TC+ATV/r	33(10.7)	32(43.2)	1(0.42)
	AZT+3TC+LPV/r	35(11.3)	34(45.9)	1(0.42)
	TDF+3TC+DTG	5(1.6)	3(4.0)	2(0.8)
	TDF+3TC+EFV	76(24.6)	3(4.0)	73(23.6)
	AZT+3TC+DTG+DRV/r	160(51.8)	2(2.7)	158(67.2)
5. WHO Clinical stage	Stage-I	84(27.2)	50(67.6)	34(14.5)
	Stage-II	47(15.2)	11(14.8)	36(15.3)
	Stage-III	134(43.4)	8(10.8)	126(53.6)
	Stage-IV	44(14.2)	5(6.8)	39(16.6)
6. Residence	Rural	91(29.4)	20(27)	71(30.2)
	Urban	218(70.6)	54(73)	164(69.8)
7. Adherence	Poor	195(63.1)	6(8.1)	189(83)
	Fair	28(9.1)	8(10.8)	20(8.5)
	Good	80(25.9)	60(81.1)	20(8.5)
8. Peripheral neuropathy	No	135(43.7)	60(81.1)	75(31.9)
	Yes	174 (56.3)	14(18.9)	160(68.1)
9. Education Level	Not educated	60(19.4)	9(12.2)	51(21.7)
	Primary	109(35.3)	23(31.1)	86(36.6)
	Secondary	93(30.1)	27(36.4)	66(28.1)
	Tertiary	47(15.2)	15(20.3)	32(13.6)
10. Treatment change	No	229(67)	2(2.7)	227(96.5)
	Yes	80(33)	72(97.3)	8(3.5)

When we look at the educational level of the infected patient's larger number 109(35.3%) were attended their primary education while only 47(15.2%) attended their tertiary educations. Among the total, viral rebound occurred in educational level 86(36.6%) viral rebound occurred in primary education level while smaller number of viral rebound 32(13.6%) occurred in the tertiary

educational level. Of the total infected patients only 80(33%) had treatment (treatment line) change and the rest 229(67%) have no.

The longitudinal response was the number of CD4 cells counts per mm^3 of blood which were measured approximately every 6 months in standard clinical practice. To handle the longitudinal outcome with linear mixed model the square root transformed value were checked for normality and the normality of the transformed value was checked by using the box plot and normal Q-Q plot of figure 6 and 8 on annexes.

Without considering the censoring status of the HIV-infected patients the average number of the square root of CD4 cell count measurement with their standard deviation at each time point was reported in table 3.

Table 3 : Mean square root of CD4 count measurement with its standard deviation at each time points with respective of the sample sizes.

Follow-up time in month	0	6	12	18	24	30	36	42	48	54
Sample size	309	309	309	300	288	251	192	122	94	51
Mean square root of CD4	13.97	17.03	19.07	20.73	21.97	22.72	21.46	20.62	19.05	20.97
Standard deviation	1.85	1.39	1.38	1.50	1.49	2.29	4.58	5.16	6.49	6.85

As it can be observed from table 3 the mean square root of CD4 cell count measurement is 13.97 at baseline; it seems an increasing value from baseline up to 30 months and it seems like decreasing after 30 months. When we look at the standard deviation of the mean value of the square root of CD4 cell count measurement between baseline times up to 24 months there was no much variation between square root of CD4 count measurement and after 24 months the standard deviation values it seems increasing value.

With considering the censoring status of HIV-infected patients the mean and standard deviation square root of CD4 cell count measurement and the viral load of censored and rebound HIV-infected patients at each time points were given as follows on table 4.

As reported on table 4 in all-time points, the mean square root of CD4 count measurement of censored HIV-infected patients from baseline time to 24 months is less than that of viral rebound of infected patients whereas there was no big difference in standard deviation of square root of CD4 count measurement in censored HIV-infected patients from baseline time up to 24 months in comparison with viral rebound of infected patients.

When we look at the mean of viral load in all-time points the mean viral load value with in standard deviation of both censored and viral rebound HIV-infected patients it seems like decreasing value.

Table 4: Mean of the square root of CD4 count measurement and Viral load with their standard deviation at each time points for rebound and censored infected patients.

For Censored observation					**For the observed events**			
Time (months)	Square root of CD4		Viral load		Square root of CD4		Viral load	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
0	9.94	7.29	2921	62.43	16.60	6.38	12441	44.37
6	14.01	8.36	2921	62.43	17.56	7.43	12441	44.37
12	16.94	8.43	3158	63.43	19.47	7.50	12471	44.46
18	19.49	9.30	3455	64.44	20.86	9.70	12464	44.45
24	20.50	10.76	2648	54.60	22.15	11.05	12259	40.95
30	22.14	9.22	2289	52.73	21.92	13.80	12110	39.04
36	21.06	13.58	1904	52.01	21.71	15.29	11755	36.19
42	23.18	11.12	1205	41.88	20.46	16.04	11401	31.78
48	24.64	7.85	660	22.64	14.57	13.04	11245	28.67
54	26.04	3.21	598	20.20	11.51	7.73	10816	26.09

4.2. RESULTS USING SEPARATE MODELS

We initially analyzed data separately using both longitudinal and survival models described in Section 3.3.1 and 3.3.2. This is important for the full specification of the mean response of the model and determines the random effects and fixed effects to be included in the longitudinal sub-model, and to identify the covariates that have a contribution for the hazard of an event in the survival sub-model to provide initial values for the joint analysis.

4.2.1. Separate Analysis of Longitudinal Data

Before any data analysis, the assumptions of the data must be checked. In order to check the normality of the longitudinal data: box plots of the CD4 cell measurements over time, Q-Q plots, histogram, and Shapiro Wilk test of the CD4 cell count are used as shown in Figure 6, 8, 9 and Table 12 respectively.

Figure 6, shows a high degree of skewness toward high CD4 cell measurements, suggesting some transformation. After a square root transformation, the data attained normality. From the Shapiro Wilk tests of normality in Table 12 on the annexes, we found that the actual CD4 cell counts were not normal at all-time points as the test showed significant deviation from normality. This coincides with Figure 8 and 9 which takes the q-q plot and histogram of the overall data suggesting square root transformation of the data to normality.

4.2.1.1. Exploring Individual Profiles and Mean Structure

Data exploration is a very important tool to fit appropriate models and to look at pattern of data over time. In addition, the individual profile plots and mean structure plots were obtained to gain some insights of the data (Verbeke, G. and Molenberghs, 2000).

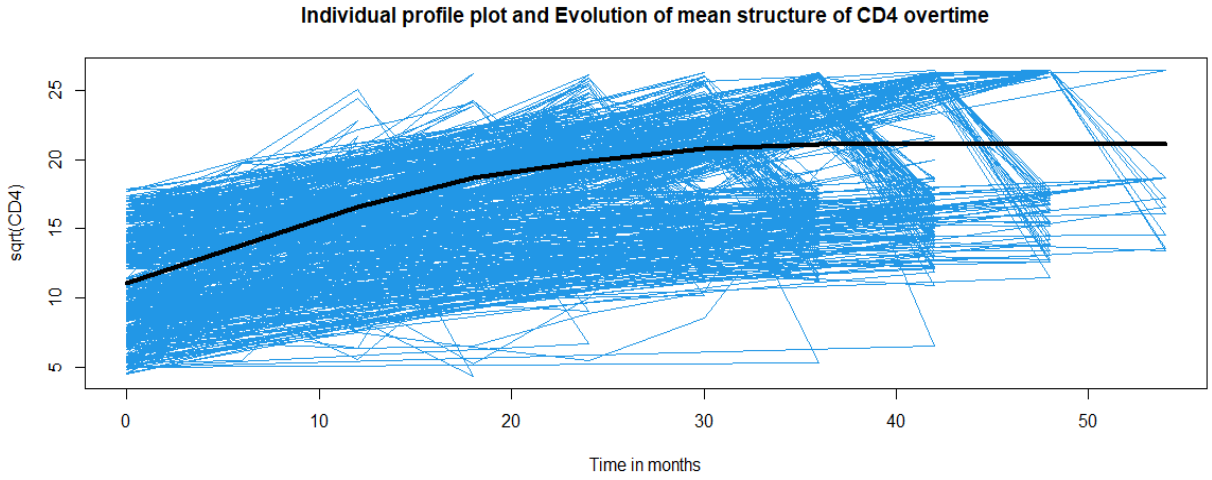


Figure 1: Individual profile plot with a loess smoothing technique

Figure 1 demonstrates the variability (within and between patients) in square root of CD4 cell count measurement of HIV/AIDS infected patients. Since the data is not balanced loess smoothing technique was used instead.

Exploring the Mean Profile plot:
To understand the possible relationships among the square root of CD4 cell count means overtime, a plot of a line connecting the average values computed at each time point is shown on Figure 1, the mean structure plot suggests that the mean of the square root of CD4 cell count profiles have a nearly linear (i.e., the relationship between square root of CD4 cell count and time seems to be approximately linear). The plotted profiles tend to generate a linearly increasing pattern which rationalizes the use of the linear mixed effects model to analyze the trajectory of CD4 cell count.

4.2.1.2. Linear Mixed-Effects Models Result

From the individual profile and mean structure exploratory analysis, linear time effects seem to be useful in modeling the random effects. After exploring, the data examine whether the assumption of heterogeneous within-subject variance for the square root of CD4 cell count is supported or not and identify the random effects (random intercepts, random slope, random intercept random slope, and random intercept and quadratic slope) to be included in the model.

The linear mixed model accounts both within and between subject sources of variations. The linear mixed model with minimum information criteria is considered as an appropriate one according to

the AIC and BIC model selection criteria. As reported in table 5, the selected fixed effects were fitted with different random effects starting with only random intercept up to random intercept; linear and quadratic slopes. Finally, we reach on an appropriate linear mixed model which have minimum AIC and BIC values that is 9237.49 and 9356.45 respectively which was obtained by fitting the selected fixed effects with adding subject specific intercept and slopes of time random effects. When the quadratic time slopes are included, both AIC and BIC values has increased indicating that the model becomes very complex as compared to those previous models. Therefore, both AIC and BIC criterion suggests not including the quadratic time slopes as random effects. As a result, the random quadratic time slopes are not included in the subsequent analyses.

Table 5: Selection of random effect models to be included in linear mixed effect model.

Random effect Model	AIC	BIC	Log-Likelihood
Random intercept	9272.29	9379.93	-4617.15
Random slope	9369.00	9506.63	-4670.50
Random intercept - Random slope	9237.49	9356.45	-4597.75
Random intercept and quadratic slope	9396.75	9633.13	-4682.37

Here from the following four (4) model examine the great reduction in the AIC of random intercept - random slope (RI RS) for the model incorporating subject-specific variances is an evident that subject-specific square root of CD4 cell count variances must be considered in the analysis. Also, the random effect of table tells us there is subject-specific variation. Hence, it supports the assumption of heterogeneous variance for the repeated square root of CD4 cell measurements. Table 13 on annexes showed that coefficient and standard error for the parameter in fixed effect and variance, standard deviation and 95% confidence interval for standard deviation in random effect of the four models.

4.2.1.3. Univariate and Multivariate Analysis of Longitudinal Data

We fitted univariate marginal models to explore the relationship between the square root of CD4 cell count and each covariates. In this analysis, we considered twelve variables. From these variables, those which are significant at 25% modest level of significance in the univariate analysis were used as a candidate for the multivariate analysis.

In statistical modeling, when the number of variables is relatively large, it can be computationally expensive to fit all possible models. Thus, one alternative is fitting a multivariate model that contains the variables which are significant at a modest level of significance in the univariate analysis case.

As the result in table 14 on annexes the variables age; viral load; adherence; functional status; peripheral neuropathy; regimen type; treatment change and WHO clinical stage were to be significant at 25% level of significance, and they are considered in multivariate analysis. However, educational level, gender, marital status and place of residence of the disease were not statistically

significant for square root of CD4 cell count at 25% revealing that these variables will not be included in the multivariate model. The covariates which were significant in the univariate analysis at 25% level of significance were all included in the multivariate analysis for the response variable of square root of CD4 cell count.

4.2.1.4. Covariance Structure

In a linear mixed effect model, covariance structures should be selected based on minimum AIC value to obtain valid inferences for the parameters of fixed effects in the model. Ignoring important correlations increase the probability of the type I error and underestimates the standard errors of an estimate Littell et al., (2000). In fitting the linear mixed effect model, a series of covariance structures of the longitudinal CD4 cell counts measure of HIV-infected patients were considered. From the possible covariance structures, the one with the smallest AIC and BIC with a convergence of the model in ML and REML was considered.

We considered four (4) different commonly used covariance structures such as; First Order Autoregressive (AR(1)); Compound symmetry (CS); Toeplitz structure and Unstructured (UN).

Table 6: Result of AIC; BIC and LogLik value for the selection of the best fitting covariance structure.

	Covariance structure			
Information criteria	**AR(1)**	**CS**	**Toeplitz**	**UN**
AIC	9152.36	9274.29	9154.97	9174.02
BIC	9265.65	9387.59	9276.59	9706.51
Log Likelihood	-4553.18	-4617.15	-4556.98	-4493.01

First Order Autoregressive (AR(1)) covariance structure, had the smallest AIC (9152.36) and BIC (9265.65) value, suggesting that the First Order Autoregressive covariance structure best fits our data compared to the other covariance structures. Therefore, First Order Autoregressive variance covariance was used in identifying the correlation structure.

4.2.1.5. Random Intercept-Linear Time Random Slope Model

In table 7, the best fitted linear mixed model for the longitudinal model was reported. The random intercept-linear time random slope version of the model was fitted. Fixed and random effect estimate from the separate longitudinal model for change in the square root of CD4 cell count were presented. The random effect (subject-specific level) model assumes that extra correlation arises among longitudinal response as the parameter estimates in the random-effects model vary across individuals Diggle et al., (2002). Individual profile plot Figure 1 showed that there is variation in the CD4 cell count of an individual over the follow-up period. This variation does not exist at baseline, but it exists over the follow-up period which implies that the random intercept linear time random slope should be included in the linear mixed model. The analysis of the longitudinal data was based on a linear mixed-effect model incorporating patient-specific CD4 cell count variability.

Table 7 showed the result of linear mixed effect model and found the variables age; adherence; functional status; WHO clinical stage; time*fair adherence; time*good adherence; time*bed ridden functional status; time*ambulatory functional status, and time* WHO clinical stage2; time* WHO clinical stage3 and time* WHO clinical stage4 interaction were statistically significant factors that affect change in the square root of CD4 cell count of HIV/AIDS patients at 5% level of significance.

Table 7: Parameter estimation of linear mixed effect model with random intercept-linear time random slope model.

Fixed effect				**95% CI**	
Parameters	$\hat{\beta}$	$Se(\hat{\beta})$	**p-value**	Lower	Upper
(Intercept)	21.40	0.39	<0.001*	20.63	22.17
Age	-0.42	0.08	<0.001*	-0.58	-0.26
Adherence; Poor **Ref**					
Fair	0.94	0.28	0.001*	0.39	1.49
Good	1.06	0.21	<0.001*	0.65	1.46
Functional status; Working **Ref**					
Bed ridden	-3.09	0.32	<0.001*	-3.71	-2.46
Ambulatory	-2.19	0.25	<0.001*	-2.67	-1.71
WHO Clinical stage; Stage1 **Ref**					
Stage2	-2.48	0.29	<0.001*	-3.05	-1.91
Stage3	-5.47	0.30	<0.001*	-6.06	-4.88
Stage4	-8.55	0.37	<0.001*	-9.28	-7.83
Time	-0.01	0.01	0.313	-0.04	0.01
Time: Adherence; Poor **Ref**					
Time: Fair	0.03	0.01	0.009*	0.01	0.05
Time: Good	0.11	0.01	<0.001*	0.08	0.13
Time: Functional status; Working **Ref**					
Time: Bed ridden	-0.04	0.01	0.009*	-0.06	-0.01
Time: Ambulatory	-0.21	0.10	0.030*	-0.41	-0.01
Time: WHO Clinical stage; Stage1 **Ref**					
Time: Stage2	0.04	0.01	<0.001*	0.01	0.07
Time: Stage3	0.03	0.01	0.031*	0.03	0.05
Time: Stage4	0.06	0.02	<0.001*	0.03	0.09

Random Effect	**Std.Dev**	**Corr**
Intercept	1.49	(Intr)
Time	0.04	-0.81
Residual	1.81	
	AIC	9151.22

** $p < 0.05$*

Time indicates observation time for CD4 cell count.

Under the random effect model, the estimated patient-specific variability was significant which supports the assumption of heterogeneous variances for the repeated CD4 count measurements.

From Table 7 the linear mixed effect model is given as;

$\boldsymbol{Y_{ij}} = 21.40 - 0.42 * \text{age} + 0.94 * \text{fair} + 1.06 * \text{good} - 3.09 * \text{bed ridden} - 2.19 * \text{ambulatory} - 2.48 * \text{stage2} - 5.47 * \text{stage3} - 8.55 * \text{stage4} + 0.03 * \text{time} * \text{fair} + 0.11 * \text{time} * \text{good} - 0.04 * \text{time} * \text{bedridden} - 0.02 * \text{time} * \text{ambulatory} + 0.04 * \text{time} * \text{stage2} + 0.03 * \text{time} * \text{stage3} + 0.06 * \text{time} * \text{stage4}.$

Where; $\boldsymbol{Y_{ij}}$ is the measured square root of CD4 count measurement measured for the HIV-infected patients at time t_{ij}, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, n_i$.

The estimated intercept value indicates that the mean square root of CD4 cell count measurement was 21.40 cells per mm^3 without the effect of other covariates.

The estimated coefficient age -0.42 shows that negative effect of age on square root of CD4 count measurement which indicates with unit change in age of HIV-infected patients decreases their mean square root of CD4 count measurement by 0.42 holding other covariates constant.

When compared with patients at adherence; change in the square root of CD4 cell count for HIV infected patient those are in fair adherence were 0.94 time higher compared to poor adherence, and good adherence were 1.06 time higher in the square root of CD4 cell count compared to poor adherence controlling for the other covariates.

Patients in bedridden and ambulatory functional status have lower square root of CD4 cell count compared to working. The change in the square root of CD4 cell count were 3.09 times lower for patients in bedridden functional status compared to working, and ambulatory were 2.19 time lower compared to working controlling for the other variable.

Regarding WHO clinical stage, change in the square root of CD4 cell count for patients those are in stage2 were 2.48 time lower compared to stage1, and stage3 and 4 were 5.47 and 8.55 time lower in the square root of CD4 cell count compared to stage1 respectively controlling for other covariates.

When we look at the time effect on the reported table linear time effect have both positive and negative effect on the square root of CD4 count measurement at 5% significance level.

The two within and between subject variations assumptions diagnosis of the linear mixed model were checked by using the graphical plots of figure 10 and 11 on annexes for between subject variation assumption and figure 12 and 13 on annexes for the within subject variation assumption and the plots showed with exception of some outlying values there was no problems of the assumptions and the fitted model is good fit the data.

4.2.2. Separate Analysis of Survival Data

4.2.2.1. Kaplan-Meier Survival Function Estimates and Log-Rank Tests

The Kaplan Meier curves for each study category provide an initial insight into the shape of the survival function for each category. We compared the survival time of patients by using Kaplan Meier and log-rank test.

The graph of Kaplan-Meier estimate of overall survival functions is showed in figure 2, which indicates decreasing pattern of survivorship function as we expected. In order to explore differences between survival time, separate Kaplan-Meier survival function curves are constructed for categorical covariates and results are given in the annexes. In figure 14, plot of the Kaplan-Meier estimates for adherence; functional status; marital status; peripheral neuropathy; treatment change and WHO clinical stage is displayed in the annexes.

The Kaplan-Meier survival plot illustrated in figure 14 showed the pattern of one survivorship function lying above another, indicating the group defined by the upper curve had a better survival probability than the group defined by the lower curve.

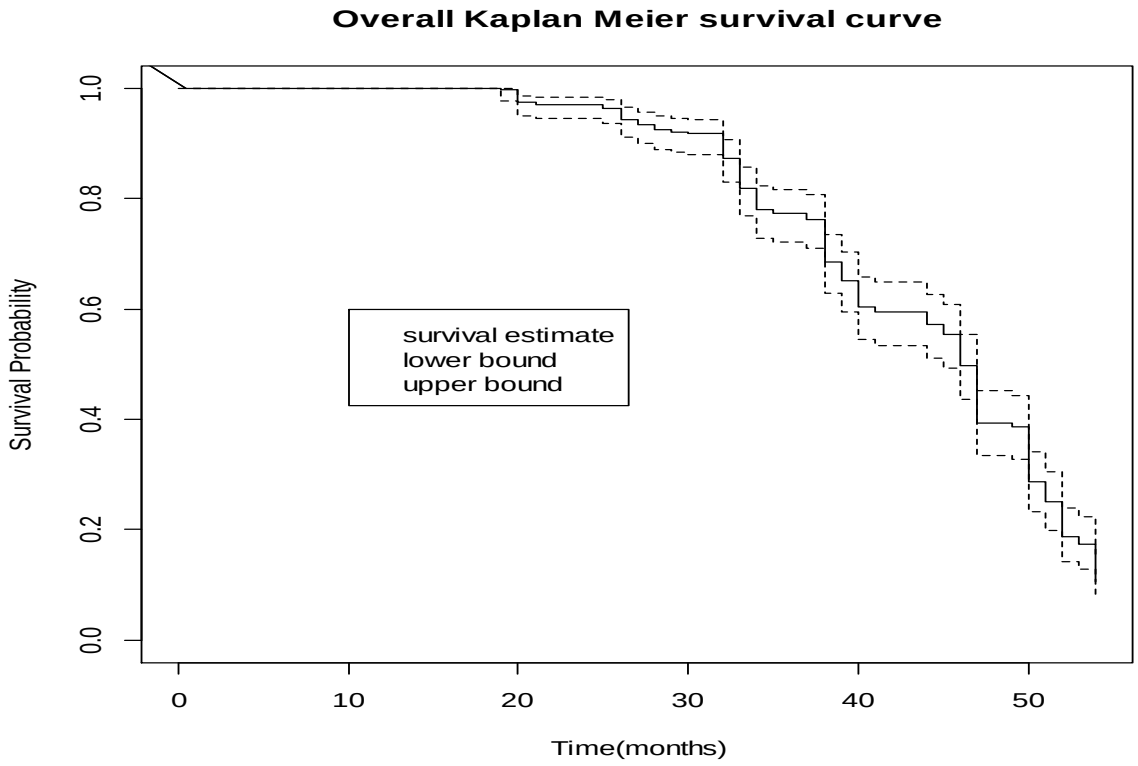


Figure 2: Estimated Kaplan-Meier curve of survival probability of HIV- infected patients for total survival functions.

To test the significant difference of the plotted curves by different covariates, the log-rank tested were employed (Table 8).

As per to the log-rank test, there is a significant difference in the survival probability of infected patients in different categories of adherence, marital status, functional status, WHO clinical stage, peripheral neuropathy and treatment change. However, there is no significant difference in the viral rebound rates between groups of education level, gender, place of residence and regimen type of infected patients.

Table 8: Log-rank test for categorical independent variables

Covariates	Test statistics	DF	P-Value
Adherence	80.3	2	<0.001*
Marital status	21.6	3	<0.001*
Functional status	96	2	<0.001*
Education level	10.2	3	0.400
Gender	0.6	1	0.410
Regimen type	10.3	4	0.510
WHO clinical stage	95.4	3	<0.001*
Place of residence	3.8	1	0.054
Peripheral neuropathy	39.2	1	<0.001*
Treatment changes	71	1	<0.001*

** Indicates the significance at* 5% *level of significance.*

4.2.2.2. Cox Proportional Hazards Model

We use univariate analysis to check all the covariates affecting the survival time before proceeding to higher models. Accordingly, the univariate Cox proportional hazards regression models are fitted for every covariate shown (table 15).

In the study, the variables that are significant in the univariate analysis in relation to time to the occurrence of an event (or viral rebound) due to HIV/AIDS were selected at 25% modest level of significance (Hosmer D. & Lemeshow S., 1999). The relationship between each single covariate and survival time of HIV/AIDS infected patients are presented in table 15.

4.2.2.3. Cox Proportionality Hazard Assumption

The proportional hazards assumption asserts that the hazard ratios are constant over time and it's important to use a fitted proportional hazard model. The risk of viral rebound must be the same no matter how long subjects have been followed. In order to test this assumption GLOBAL test and Schonfield residual was used.

Table 9: Cox proportional hazard model assumption

Covariates	Chisq	DF	P-value
Age	0.03	1	0.986
Adherence	2.65	2	0.266
WHO clinical stage	0.55	3	0.907
Peripheral neuropathy	9.76	1	0.006
Viral load	0.01	1	0.938
GLOBAL	14.11	8	0.088

From the table 9, it is clear to see that the p- value of the overall proportionality test (GLOBAL test) is not significant. This indicate that the PH assumption is not violated. Graphically schoenfield residual plots are presented in annexes (fig. 16) showed that the scaled Schoenfeld residuals are randomly distributed and a loess smoothed curve do not exhibit much departure from the horizontal line suggest that the proportional hazards assumption not violated. Age; viral load; adherence; WHO clinical stage and peripheral neuropathy were the variables to be included in the model.

Analysis of Multivariate Cox PH Model

After checking the assumption of proportional hazard, the survival data were analyzed based on Cox proportional hazard model. All potential variables that are supposed to have significant impact (p-value < 0.25) on the survival time of infected patients in univariate analysis were included in the multivariate cases. The results are presented in table 10.

The result of multivariable analysis of cox-PH model in table 10 indicate the covariates age; adherence; WHO clinical stage; peripheral neuropathy and viral load were significantly contributed to survival probability of HIV positive patients under ART at 5% level of significance.

Table 10: Multivariate analysis of Cox proportional hazards model for the selected variable

					95% CI	
Variables	$\hat{\beta}$	HR	$Se(\hat{\beta})$	Pr(|z|)	Lower	Upper
Age	0.56	1.78	0.21	<0.001*	1.18	2.67
Viral load	0.22	1.25	0.02	0.005*	1.19	1.30
Adherence; Poor **Ref**						
Fair	-0.01	0.99	0.26	0.097	0.59	1.64
Good	-0.60	0.55	0.30	0.004 *	0.30	0.99
WHO Clinical stage; Stage1 **Ref**						
Stage2	1.05	2.86	0.35	0.002*	1.45	5.62
Stage3	1.08	2.94	0.34	0.001*	1.49	5.78
Stage4	1.17	3.23	0.32	<0.001*	1.71	6.08
Peripheral neuropathy; Yes **Ref**						
No	-0.58	0.56	0.16	<0.001*	0.41	0.76

**Indicates the significance of the covariates at 5% level of significance.*

The results revealed there was an increase in hazard with an increase in age by the hazard of 1.78 significant at p<0.001, and an increase in hazard with an increase in viral load by the hazard of 1.25 significant at p = 0.005.

Regarding adherence of infected patients, good adherence was 0.55 times lower risk of viral rebound compared to poor adherence (HR = 0.55; 95% CI: 0.30 - 0.99).

Comparing hazard ratio (HR) by WHO clinical stage, showed being in stage2 were 2.86 times higher risk of viral rebound than in stage1 (HR = 2.86; 95% CI: 1.45 - 5.62); stage3 were 2.94 times higher risk of viral rebound than in stage1 (HR = 2.94; 95% CI: 1.49 - 5.78) and stage4 were 3.23 times higher risk of viral rebound compared to stage1 (HR = 3.23; 95% CI: 1.71 - 6.08, p-value < 0.001).

Comparing hazard ratio (HR) by peripheral neuropathy, HIV-infected patients with no peripheral neuropathy were 0.56 times lower risk of viral rebound compared to those have peripheral neuropathy (HR = 0.56; 95% CI: 0.41 - 0.76).

4.3. RESULTS USING JOINT MODELS

4.3.1. Joint Model of Survival and Longitudinal Analysis

After having appropriate separate models for the mean of the square root of CD4 cell count and time to the viral rebound of infected patients due to HIV/AIDS, the next step is to explore an appropriate joint model that associates the longitudinally measured CD4 cell count and time to viral rebound of infected patients from HIV/AIDS.

The result of the joint model could obtain by combining the selected random-intercept -linear time random slope model and cox-proportional hazard model.

In the joint modeling of longitudinal and survival data for HIV positive patients the variable age; adherence and WHO clinical stage were significant $(p - value < 0.05)$ both in longitudinal and survival sub-model as reported in table 11. However, age, adherence, WHO clinical stage, peripheral neuropathy and viral load were significant in survival sub-model and age, adherence, functional status, WHO clinical stage, and interaction effect of adherence, functional status and WHO clinical stage with linear time were found to be significant at 5% level of significance in longitudinal sub-model.

In longitudinal sub-model, the mean square root of CD4 cell count measurement was significantly lower for age and square root of CD4 cell counts decreased with the increment of age at diagnosis; this implies the estimated coefficient age -0.23 shows that negative effect of age on CD4 count measurement which indicates with unit change in age of HIV-infected patients decreases their

mean square root of CD4 cell count measurement by 0.23 with 95% CI (-0.25, -0.21) holding other covariates constant.

When compared with patients at a fair and good adherence on ART treatment, the mean change in the square root of CD4 cell count were 0.95 times greater for fair adherence compared to poor adherence and good adherence were 1.07 times greater square root of CD4 count measurements in comparison with poor adherence on ART controlling for the other variables.

Regarding functional status, the mean change in the square root of CD4 cell count were 3.09 times lower for bedridden functional status compared to working functional status and ambulatory functional status were 2.19 times lower square root of CD4 count measurements in comparison with working functional status controlling for the other variables.

Table 11: Results for the joint model of longitudinal and survival analysis

	Fixed effect	$\hat{\beta}$	$Se(\hat{\beta})$	**P - value**	**95% CI**	
					Lower	Upper
Longitudinal sub-model	Intercept	21.38	0.39	<0.001*	20.59	22.16
	Age	-0.23	0.01	<0.001*	-0.25	-0.21
	Adherence; Poor **Ref**					
	Fair	0.95	0.28	<0.001*	0.39	1.50
	Good	1.07	0.20	<0.001*	0.66	1.47
	Functional status; Working **Ref**					
	Bed ridden	-3.09	0.32	<0.001*	-3.72	-2.45
	Ambulatory	-2.19	0.24	<0.001*	-2.67	-1.71
	WHO Clinical stage; Stage1 **Ref**					
	Stage2	-2.46	0.29	<0.001*	-3.04	-1.88
	Stage3	-5.46	0.30	<0.001*	-6.06	-4.86
	Stage4	-8.54	0.37	<0.001*	-9.27	-7.80
	Time	-0.01	0.01	0.367	-0.03	0.01
	Time: Adherence; Poor **Ref**					
	Time: Fair	0.03	0.01	0.010*	0.01	0.05
	Time: Good	0.11	0.01	<0.001*	0.08	0.13
	Time: Functional status; Working **Ref**					
	Time: Bed ridden	-0.04	0.01	0.010*	-0.06	-0.01
	Time: Ambulatory	-0.19	0.01	0.030*	-0.21	-0.17
	Time: WHO Clinical stage; Stage1 **Ref**					
	Time: Stage2	0.04	0.01	<0.001*	0.01	0.06
	Time: Stage3	0.04	0.01	0.035*	0.03	0.06
	Time: Stage4	0.06	0.02	<0.001*	0.03	0.09
	Random effect	**StDev**		**Corr**		
	Intercept	1.48				
	Time	0.03				
	Residual	1.80				
	(Intercept, Time)			-0.80		

	Fixed effect	$\hat{\beta}$	$Se(\hat{\beta})$	P – value	HR	95% CI(HR) Lower	Upper
Survival sub-model	Age	0.59	0.20	<0.001*	1.80	1.21	2.67
	Adherence; Poor [Ref]						
	Fair	-0.74	0.25	0.008*	0.48	0.28	0.78
	Good	-0.65	0.29	0.002*	0.52	0.29	0.93
	WHO Clinical stage; Stage1 **Ref**						
	Stage2	0.95	0.33	0.004*	2.59	1.34	5.01
	Stage3	1.03	0.31	0.001*	2.80	1.51	5.21
	Stage4	1.00	0.33	0.002*	2.72	1.41	5.25
	Peripheral neuropathy; Yes [Ref]						
	No	-0.53	0.15	<0.001*	0.59	0.43	0.79
	Viral load	0.41	0.01	<0.001*	1.51	1.48	1.53
	Assoc	-0.102	0.02	<0.001*	0.90	0.86	0.94
	AIC	6439.26					

** Indicates the significance at* 5% *level of significance*.

Comparing mean change in the square root of CD4 cell count by WHO clinical stage, the mean change in the square root of CD4 cell count were 2.46 times lower for infected patient those are in stage2 compared to stage1, and 5.46 and 8.54 times lower for stage 3 and 4 compared to stage1 respectively controlling for other independent variables.

And in the survival sub-model all selected variables were associated with the hazard of viral rebound. The results revealed there was an increase in hazard with an increase in age by the hazard of 1.80 significant at p<0.001. Also, an increase in hazard with an increase in viral load by the hazard of 1.51 significant at p<0.001.

Good adherence on ART treatment effect lowers the hazard of viral rebound by 0.52 (95% CI: 0.29 - 0.93) than poor adherence on ART treatment and fair adherence on ART had 0.48 times lower risk (HR = 0.48; 95% CI: 0.28 - 0.78) of viral rebound compared to poor adherence on ART treatment. The other significantly important variable for the survival rate of the patient is WHO clinical stage. The hazard rate of patient whose WHO clinical stage2; stage3 and stage4 was increased by 2.59, 2.80 and 2.72 respectively, when we compared with those patients, whose WHO clinical stage was stage1.

Infected patient with no peripheral neuropathy had 0.53 times lower risk (HR = 0.53; 95% CI: 0.43 - 0.79) of viral rebound compared to having peripheral neuropathy.

In addition, the significant model association parameter revealed a negative association between the square root of CD4 cell counts and hazard of viral rebound, which means that viral rebound is less likely to occur in patients with higher square root of CD4 cell counts ((HR = 0.90) with 95% CI (0.86, 0.94), p-value = <0.001). The negative value of the association parameter (-0.102) indicated that the slope of the square root of CD4 cell counts was negatively associated with the

viral rebound, and with a unit increase in the square root of CD4 cell count the hazard of viral rebound was decreased.

4.3.2. Separate and Joint Model

The longitudinal sub-model in the fitted joint model was consistent with the results from the separate longitudinal analysis. In longitudinal sub-model of included in joint model age; adherence; functional status; WHO clinical stage; time*adherence; time* functional status and time* WHO clinical stage were statistically significant covariates. But follow- up time was not statistically significant.

In the survival sub-model of joint model age; adherence on ART; WHO clinical stage; peripheral neuropathy and viral load were significantly associated with the hazard of viral rebound. But fair adherence on ART was not statistically significant in the separate survival model.

In general, from the result of the estimated parameters of the two-model separate and joint models, the joint models had narrow confidence interval as compared with the separate models. This indicates that the joint model is more precise than the separate models. The estimates of the parameters of the separate and joint models are not identical. The estimates of the association parameters in the joint models are significantly different from zero, providing strong evidence of association between the two sub-models for the joint model. The estimate of the association $\hat{\alpha} = -0.102$ indicating that the higher square root of CD4 cell counts is associated with the lower hazard of viral rebound.

When evaluating the overall performance of both the separate and joint models in terms of model parsimony and goodness of fit, the joint model was performed better. Joint model has smaller AIC than the separate model. Also, the statistical significance of both the association parameters was evidenced that the joint model was better than the separate model Seid et al., (2014).

4.3.3. Joint Model Diagnostics

The joint models are fitted. The next step is to verify if all the necessary model assumptions are valid. Standard types of residual plots can be used to validate the assumptions behind mixed models and relative risk models.

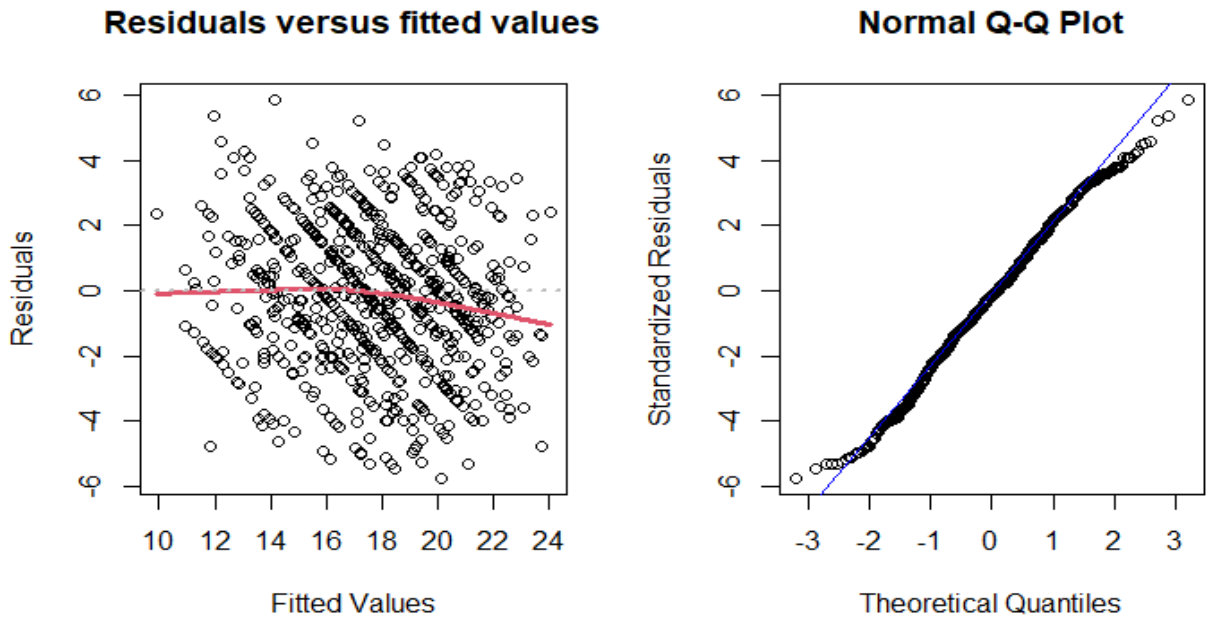


Figure 3: Residual plot against fitted values and Q-Q plot

The distributional assumption is checked by comparing theoretical quantiles. The Q-Q plot shows that the response variables of longitudinal sub-model (square root of CD4 cell counts) are normally distributed because the points are scattered on the line as well as the residual against fitted value plot didn't show any systematic pattern and no evidence of non-constant variance and the fitted LOWESS curve is close to zero. Hence, square root of CD4 cell counts measure is linear to the parameter and the error variance of the longitudinal sub-model are constant.

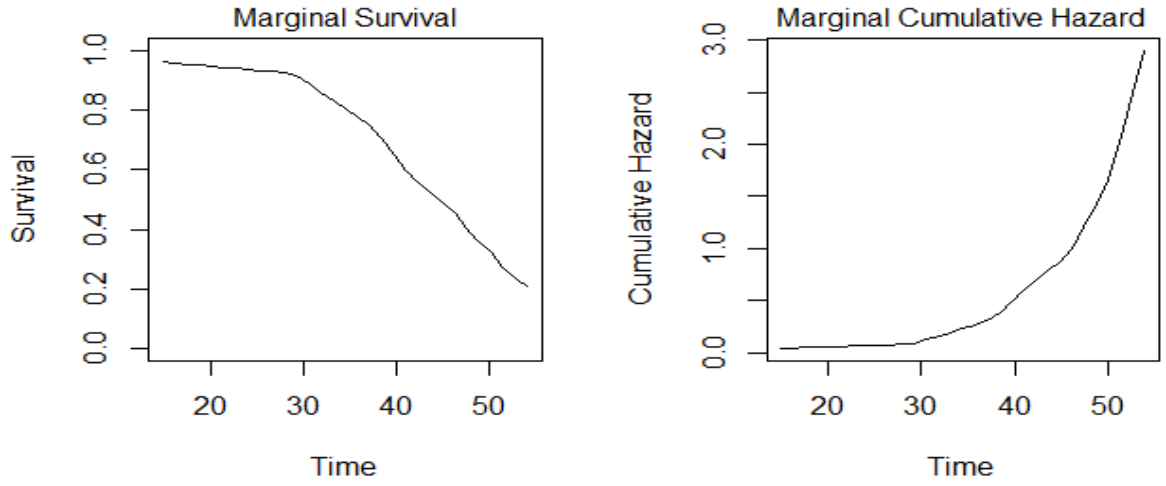


Figure 4: Marginal survival plot and marginal cumulative hazard plot

The marginal survival plot and marginal cumulative hazard plot in figure 5 showed that the survival probability (not developing viral rebound) come down and the probability of developing viral rebound come up to one respectively when the follow-up time increased.

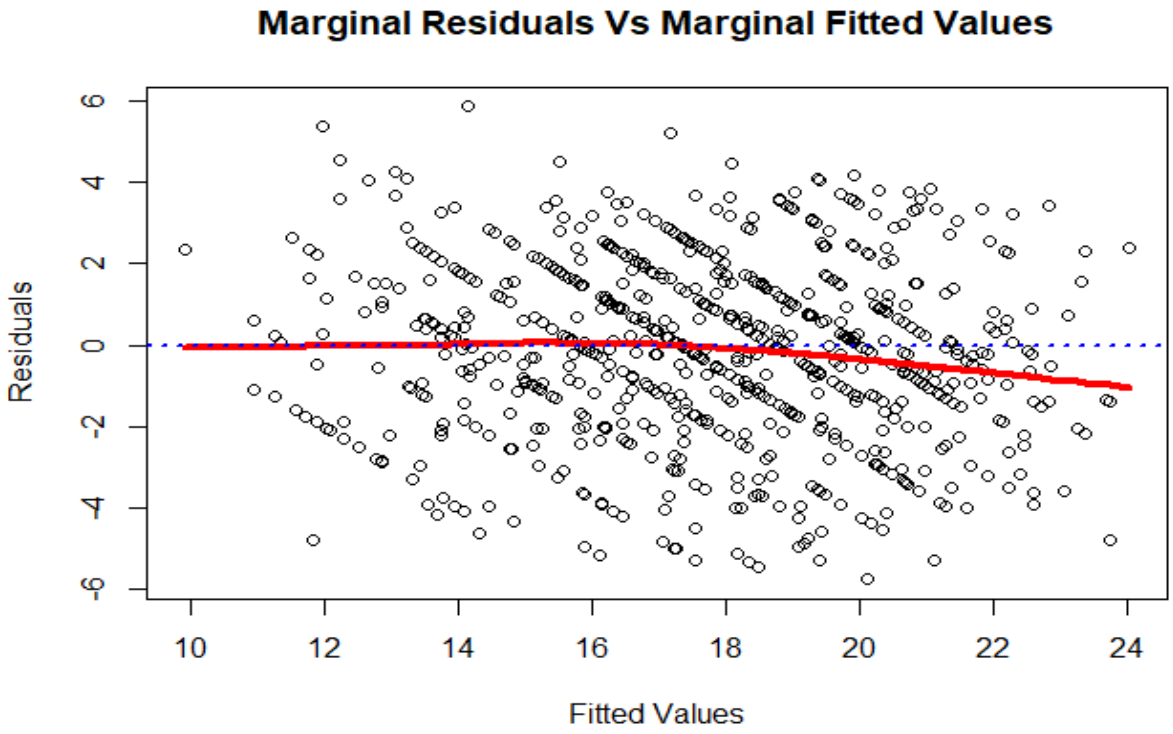


Figure 5: Marginal residual versus fitted values plot

A plot of marginal residual versus the fitted value of the standardized residuals for longitudinal process is nearly coincide with the reference passing through the origin. Hence, there is a validity of normality assumption of the error term in the longitudinal sub-model.

CHAPTER FIVE

5.1. DISCUSSION OF THE RESULTS

This study attempted to assess predictor that are associated with longitudinal CD4 count and survival time/time to viral rebound of patient due to HIV/AIDS at Jimma university medical center. So as to address mentioned objectives, three different models were explored; the linear mixed-effects model, Cox proportional hazards model for each outcome independently and joint modeling of the two outcomes together. Both approaches ended up with consistent results except that the joint model analysis added up information about the association between the two outcomes.

In the separate analysis of the longitudinal data, first the CD4 cell count measurements are checked for normality using box plot; histogram and Q-Q plots. The plots indicate that there is a deviation from normality and needs some transformation. After a square root transformation of the CD4 cell count, the mean response of the longitudinal square root of CD4 count is determined to be normal in time. Then, the data were analyzed using random intercept; random slope; random intercept-linear time random slope and random intercept and quadratic slope model. The parameter estimates of the four models are close to each other. But, the estimated patient-specific variability is significant which supports the assumption of heterogeneous variances. Also, the random intercept –linear time random slope model has a smaller AIC than the other models. The final selected model also showed age; adherence; functional status; WHO clinical stage, and the interaction of adherence; functional status and WHO clinical stage with linear time effects have significant effects on CD4 count measurements infected patients and the goodness of fit this model was also checked using the residual plot diagnosis tests the within and between variation assumption was satisfied and the selected random intercept-linear time random slope model for the longitudinal sub-model was good fit the data.

In the separate analysis of the survival data, the variables to be included in the survival model are determined using a variable selection method using statistical software packages R version 4.0.5. Next to variable selection proportional hazard assumption was checked. Then, of all candidate covariates consider, age; adherence; WHO clinical stage; peripheral neuropathy and viral load were extracted to be included in the survival models.

After the most suitable separate model have been decided for the data, the proposed joint model were applied to the data, with the aim of investigating the effects of repeated CD4 cell count measurements on time to viral rebound.

The study revealed that the variable age; adherence; WHO clinical stage were significant for both outcomes. The covariate such as age, adherence; WHO clinical stage; peripheral neuropathy, and viral load is a predictor that determines patient survival time, while age; adherence; functional status; WHO clinical stage, and interaction of adherence; functional status and WHO clinical stage with time was the predictor associated with CD4 cell count over time.

Baseline age had a significant negative association with CD4 cell count Tiruneh et al., (2021). Similarly, the mean square root of CD4 cell count decreases as the age patient increases. From the result of longitudinal sub-model of square root of CD4 cell count, as the age of patients increases by 1 year the mean square root of CD4 cell count decreases by 0.23 times (p-value < 0.001) by controlling other variables.

Adherence on ART has a significant effect on the progression change of square root of CD4 cell count of HIV infected patients. The mean square root of CD4 cell count were 0.95 times higher for fair adherence compared to poor adherence and good adherence were 1.07 times higher compared to poor adherence controlling for the other covariates. This result confirms the study conducted by Maina et al., & Desta, Kidane, et al., (2020).

On the functional status of patients on ART, mean change in the square root of CD4 cell counts for functional status patients the mean change in the square root of CD4 cell count were 3.09 times lower for bedridden compared to working, and ambulatory functional status were 2.19 times lower compared to working controlling for the other covariates. These results conformed to the studies conducted (Gebrerufael et al., 2020; Temesgen et al., 2018).

WHO Clinical stage has a significant effect on the progression change of square root of CD4 cell counts of HIV infected patients. The progression changes in square root of CD4 cell counts for stage2; stage 3 and stage 4 patients had decreased by 2.46; 5.46 and 8.54 as compared to stage 1 patients and this difference is statistically significant since the 95% confidence interval did not include zero by controlling the other variables constant. This indicates that the latter stage at diagnosis is associated with the lower mean change of square root of CD4 cell counts over time and similar to the study by (Abebe, 2019).

The interaction effect of time by functional status had statistically significant effect on the mean square root of CD4 cell counts, this suggesting that as the number of visit time increased the average square root of CD4 cell counts of HIV/AIDS patients who were bedridden and ambulatory functional status was lower than the average square root of CD4 cell counts of patients who were working functional status by 0.04 times (p-value= 0.010) and 0.19 times (p-value= 0.030) respectively when other variables constant. This study was agreed with a study by (Gebrerufael et al., 2020).

The interaction effect WHO clinical stage by time had a significant progression change on patients' square root of CD4 cell counts. That is patients with stage 2; stage 3 by visit time and stage 4 by visit time interaction increased the progression change of square root of CD4 cell counts by 0.04; 0.04 and 0.06 times respectively, as compared to patients with stage 1 by time interaction and this study was agreed with a study by (Abebe, 2019).

Baseline age had a significant positive association with the hazard of death (McHunu et al., 2020). Our finding also indicates that viral rebound of HIV/AIDS patient increases 1.80 times when age increased by 1 year by controlling other independent variables.

The risk of viral rebound HIV/AIDS patients whose adherence fair and good had lower than those whose adherence poor by controlling other predictor variables. The risk of viral rebound HIV-infected patient for adherence fair were 0.48 times lower (HR = 0.48; 95% CI: 0.28 - 0.78) compared to adherence poor and adherence good were 0.52 times lower (HR= 0.52; 95% CI: 0.29 - 0.93) compared to adherence poor. This shows that, HIV/AIDS patients whose adherence was fair and good have better survival time and understanding of the disease condition and comprehension of instructions given on drug usage than adherence poor patients and this result conforms to the study conducted by (Gebrerufael et al., 2020).

The other significantly important variable for the survival rate of the patient is WHO clinical stage. WHO clinical stage was significantly associated with patient survival time, and the risk of viral rebound of HIV- infected patient in stage2 were 2.59 times higher (HR = 2.59; 95% CI: 1.34 - 5.01) compared to stage1; stage3 were 2.80 times higher (HR = 2.80; 95% CI: 1.51 - 5.21) compared to stage1, and stage4 were 2.72 times higher (HR = 2.72; 95% CI: 1.41 - 5.25) compared to stage1 and this result was agreed with study by (Tegegne et al., 2018).

Study showed peripheral neuropathy was significantly associated with patient survival time and HIV infected patient in no peripheral neuropathy were 0.59 times (HR = 0.59; 95% CI: 0.43 - 0.79) lower risk of viral rebound compared to having peripheral neuropathy and this result was agreed with study by (Shoko & Chikobvu, 2018; Claris, 2019).

Baseline viral load had a significant positive association with the hazard of death (McHunu et al., 2020). Our finding also indicates that viral rebound of HIV/AIDS infected patient increases 1.51 times (HR = 1.51; 95% CI: 1.48 - 1.53) when viral load increased by one unit by controlling other independent variables.

On the other hand, we observed the estimated association parameter in the joint model was highly negative the significance of the association parameter $(\boldsymbol{\alpha}) =$ between two outcomes in joint modeling and the estimated association parameter is -0.102 indicating square root of CD4 cell count and time to viral rebound were negatively associated and similar with the study by Basit et al. (2018).

CHAPTER SIX

6. CONCLUSION AND RECOMMENDATION

6.1. CONCLUSION

From this specific study, the joint model was a better fit than the separate survival and longitudinal models. The separate and longitudinal sub-model showed that: Age; adherence; functional status; WHO clinical stage, and time interaction effect with adherence; functional status and WHO clinical stage of the patients were significant predictors of the progression change of square root of CD4 cell count. The covariates adherence and its interaction with time effect and time interaction with WHO clinical stage were positively associated with the progression change of square root of CD4 cell count, and time interaction with functional status were negatively associated with the progression change of square root of CD4 cell count.

Also, the separate and survival sub-model analysis showed that: Age; viral load; adherence; WHO clinical stage and peripheral neuropathy were statistically significant predictors of the time to viral rebound. The covariates adherence and no peripheral neuropathy were negatively associated with the time to viral rebound and age; viral load and WHO clinical stage have positive effects on hazard function of survival time that lowers the survival time of HIV infected patients at 5% level of significance in the study area.

The study showed the significance of the association parameter ($\boldsymbol{\alpha} = -0.102$) in joint models. The estimated association parameters indicate longitudinal CD4 cell count and the survival time was negatively associated, implies that the higher values of the CD4 cell count are associated with better survival. Therefore, due to the significance of association between the longitudinal and survival outcomes, the joint model analysis was suggested over separate models' analysis.

When evaluating the overall performance of both the separate and joint models in terms of model parsimony, the goodness of fit, smaller total AIC, and the statistical significance of both the association parameters, the joint model performs better. Thus, we concluded that the joint model was preferred for simultaneous analyses of repeated measurement and survival data.

6.2. RECOMMENDATIONS

Based on the finding of the study the following recommendations are forwarded:

- The risk of viral rebound is higher in older age; WHO clinical stage2, 3, 4 and higher viral load the infected patient should be cautious when in this category during HIV-infection period.
- WHO clinical stages; age and viral load are associated with higher risk of viral rebound and are indicators of the disease progression. Therefore, patients should need to diagnosis

and initiate ART early as per the recent WHO recommendation; HIV infected patients could better to initiate ART treatment early in respective of disease marker.

- Health professionals give attention for this epidemic disease to minimize the risk of viral rebound.
- HIV-positive patients at an older age have to start treatment early to upgrade their CD4 count progress over time.
- Health professionals give attention when a patient estimated CD4 cell count is increased through good and fair adherence on ART of patients.
- This thesis used a joint model longitudinal and survival data in analyzing the progression of HIV; that is, individual level. In the future, the study recommends the application of a joint model of bivariate longitudinal and time to viral rebound of survival analysis of HIV progression. This approach analyses the progression of HIV between different hierarchical models, for example, starting progression at the individual level then village level; zonal level; regional level, and finally national level.

6.3. LIMITATIONS OF THE STUDY

We were unable to include important various socio-demographic and socioeconomic variables like; lactic acidosis; consumption of alcohol; smoking; income level; liver abnormality and diet style that might have contributed to the survival times to viral rebound of the HIV/AIDS patients. Moreover, the present study was restricted to age group >=18, due to their different measures of CD4 cell count HIV/AIDS patients for children and adults.

Selection bias was possibly introduced due to the fact that patients with incomplete records of variables or charts which were lost for some patients were excluded. Therefore, those study subjects whose charts were not included in the study and with missing values may undermine the result if it is related to viral rebound.

7. References

Abebe, T. N. (2019). Longitudinal CD4 cell count and time to death among HIV infected children initating antiretrovial therapy. Arba Minch.

Adele, H. (2011). Joint Modelling of Longitudinal and Survival Data : A Comparison of Joint and Independent Models. 4971–4976.

Akaike, H. (1974). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Ali, J. H. (2019). Time to viral load suppression and its associated factors in cohort of patients taking antiretroviral treatment in East. 1–6.

Ameraoui, A., & Boukhetala, K. (2016). Modeling Algerian cars claims using LPC model under randomly right- censored data Modeling Algerian cars claims using LPC model under randomly right-censored data. December.

Andrew, B. P., Osler, M., Cohen, K., Channing, L., Hilderbrand, K., Mothibi, E., Zweigenthal, V., Slingers, N., Cloete, K., & Abdullah, F. (2008). Antiretroviral therapy and early mortality in South Africa. Bulletin of the World Health Organization, 86(9), 678–687. https://doi.org/10.2471/BLT.07.045294

Asfaw, A., Ali, D., Eticha, T., Alemayehu, A., Alemayehu, M., & Kindeya, F. (2015). CD4 cell count trends after commencement of antiretroviral therapy among HIV-infected patients in Tigray, Northern Ethiopia: A retrospective cross-sectional study. PLoS ONE, 10(3), 1–9. https://doi.org/10.1371/journal.pone.0122583

Basit, A., Rahaman, A., Luguterah, A., & Suleman, N. A. (2018). Joint Longitudinal and Survival Modeling of HIV in the Upper West Region of Ghana. 6(1), 56–63. https://doi.org/10.15640/ijhs.v6n1a7

Boutayeb, A. (2009). The impact of HIV/AIDS on human development in African countries. BMC Public Health, 9(SUPPL. 1), 1–10. https://doi.org/10.1186/1471-2458-9-S1-S3

Brazzale, A. R., & Mameli, V. (2018). Likelihood Asymptotics in Nonregular Settings A Review with Emphasis on the Likelihood Ratio. Working Paper Series, 4. http://www.stat.unipd.it

Brief, P., & Treatment, H. I. V. (2019). Update Of First- & Second-Line Antiretroviral Regimens.

Chaiyachati, K. H., Ogbuoji, O., & Price, M. (2014). Interventions to improve adherence to ART: a rapid systematic review. 187–204. https://doi.org/10.1097/QAD.0000000000000252

Chaiyasin, N., & Sungkanuparph, S. (2016). Rate of CD4 Decline and Factors Associated with Rapid CD4 Decline in Asymptomatic HIV-Infected Patients. Journal of the IAP of AIDS Care, 15(1), 3–6. https://doi.org/10.1177/2325957415616493

Claris, S. (2019). Continuous-time Markov Modelling of the Effects of Treatment Regimens on HIV / AIDS Immunology and Virology. 2015251768.

Cooke, R. M. (2007). The cox proportional hazard model. 34(2), 1–12.

Cox, D. R. (2007). Regression Models and Life-Tables. 34(2), 187–220.

D.Oakes, D. R. C. and. (1999). Analysis of Survival Data (D. B. R. and B. W. S. D.R. Cox, D.V. Hinkley, N.Reid (ed.)).

Dessie, Z. G., Zewotir, T., & Mwambi, H. (2020). Modeling Viral Suppression , Viral Rebound and State- Specific Duration of HIV Patients with CD4 Count Adjustment : Parametric Multistate Frailty Model Approach. Infectious Diseases and Therapy, 9(2), 367–388. https://doi.org/10.1007/s40121-020-00296-4

Desta, A. A., Kidane, K. M., Gebretsadik, A., Woldegebriel, Ajemu, K. F., Berhe, A. A., Zgita, D. N., Weldemhret, L., Teweldemedhn, Legesse, L., Woldegebriel, Bezabih, N. M., Wubayehu, T., & Woldearegay. (2020). Level of Adherence and Associated Factors Among HIV-Infected Patients on ART in Northern Ethiopia : Retrospective Analysis. 1585–1594.

Desta, A. A., Woldearegay, T. W., Futwi, N., Gebrehiwot, G. T., Gebru, G. G., Berhe, A. A., & Godefay, H. (2020). HIV virological non-suppression and factors associated with non-suppression among adolescents and adults on antiretroviral therapy in northern Ethiopia: A retrospective study. BMC Infectious Diseases, 20(1), 1–10. https://doi.org/10.1186/s12879-019-4732-6

Diggle, P. J., Heagerty, P. J., Liang, K., & Zeger, S. L. (2002). Analysis of Longitudinal Data, 2nd ed. (2nd ed.).

EPHI. (2020). HIV Related Estimats and Projections in Ethiopia for the Year-2019 February 2020 Addis Ababa Table of Contents. February.

Gebrerufael, G., Asfaw, Z., & Melese, D. (2020). Statistical Joint Modeling on Longitudinal Body Weight and CD4 Cell Progression with Survival Time-to-Death Predictors on HIV/AIDS Patients in Mekelle General Hospital, Ethiopia. 10. https://doi.org/10.21203/rs.3.rs-28299/v1

Guo, X., Carlin, B. P., Guo, X., & Carlin, B. P. (2004). Statistical Practice Data Using Standard Computer Packages. 58(1), 16–24.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72(358), 320–338. https://doi.org/10.1080/01621459.1977.10480998

Henderson, R. (2000). Joint modelling of longitudinal measurements and event time data. Biostatistics, 1(4), 465–480. https://doi.org/10.1093/biostatistics/1.4.465

Hosmer D. & Lemeshow S. (1999). Applied Survival Analysis. John Wiley and Sons, Inc., New York. In Statistical Methods in Medical Research (Vol. 8, Issue 4). https://doi.org/10.1177/096228029900800409

Hsieh, F., Tseng, Y. K., & Wang, J. L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. Biometrics, 62(4), 1037–1043. https://doi.org/10.1111/j.1541-0420.2006.00570.x

Id, J. M. C., Perelson, A. S., & Li, J. Z. (2019). Predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers. 2016, 1–26.

Ioannidis, J. P. A., Havlir, D. V., Tebas, P., Hirsch, M. S., Collier, A. C., & Richman, D. D. (2000). Dynamics of HIV-1 viral load rebound among patients with previous suppression of viral replication. Aids, 14(11), 1481–1488. https://doi.org/10.1097/00002030-200007280-00003

Jones, G., Johnson, W. O., Vink, W. D., & French, N. (2012). A Framework for the Joint Modeling of Longitudinal Diagnostic Outcome Data and Latent Infection Status: Application to Investigating the Temporal Relationship between Infection and Disease. Biometrics, 68(2), 371–379. https://doi.org/10.1111/j.1541-0420.2011.01687.x

Kayode, E. M., Usiegbodi, D. O., Ajiboye, M. E., Omonye, I. S., Febut, M. N., & Buru, A. S. (2020). Assessment of the effect of ART on haematological parameters in HIV positive

individuals in Zaria. 12(December), 17–23. https://doi.org/10.5897/JAHR2018.0467

Keele, B. F., & Li, Z. (2016). Origin of Rebound Plasma HIV Includes Cells with Identical Proviruses That Are Transcriptionally Active before Stopping of Antiretroviral Therapy. 90(3), 1369–1376. https://doi.org/10.1128/JVI.02139-15.Editor

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data Author (s): Nan M . Laird and James H . Ware Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529876 REFERENCES Linked references are available on JSTOR for this article. Biometrics, 38(4), 963–974.

Lambotte, O., Boufassa, F., Madec, Y., & Nguyen, A. (2005). HIV Controllers : A Homogeneous Group of HIV-1 — Infected Patients with Spontaneous Control of Viral Replication HIV / AIDS HIV Controllers : A Homogeneous Group of HIV-1 – Infected Patients with Spontaneous Control of Viral Replication. June 2016. https://doi.org/10.1086/433188

Lim, H. J. (2017). Joint modeling of longitudinal and event time data : application to HIV study. April. https://doi.org/10.7243/2053-7662-1-1

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). TUTORIAL IN BIOSTATISTICS Modelling covariance structure in the analysis of repeated measures data. July 1998, 1793–1819.

Little, Ratclie, Yi-Kuan Tseng, (2015). Routledge international handbook of advanced quantitative methods in nursing research. In Routledge International Handbook of Advanced Quantitative Methods in Nursing Research. https://doi.org/10.4324/9781315882307

Loures, L., Samb, B., & Sidibe, M. (2020). The UNAIDS 90 Á 90 Á 90 target : a clear choice for ending AIDS and for sustainable health and development. 1–2.

Maina, E. K., Mureithi, H., Adan, A. A., Muriuki, J., Lwembe, R. M., & Bukusi, E. A. (2020). International Journal of Infectious Diseases Incidences and factors associated with viral suppression or rebound among HIV patients on combination antiretroviral therapy from three counties in Kenya. International Journal of Infectious Diseases, 97, 151–158. https://doi.org/10.1016/j.ijid.2020.05.097

Manosuthi, W., Ongwandee, S., Bhakeecheep, S., Leechawengwongs, M., Ruxrungtham, K., Phanuphak, P., Hiransuthikul, N., Ratanasuwan, W., Chetchotisakd, P., Tantisiriwat, W., Kiertiburanakul, S., Avihingsanon, A., Sukkul, A., & Anekthananon, T. (2015). Guidelines for antiretroviral therapy in HIV-1 infected adults and adolescents 2014, Thailand. AIDS

Research and Therapy, 12(1), 1–9. https://doi.org/10.1186/s12981-015-0053-z

Manuscript, A., & Event, I. T. (2012). NIH Public Access. 4(2), 262–281. https://doi.org/10.1007/s12561-012-9061-x.Joint

McHunu, N. N., Mwambi, H. G., Reddy, T., Yende-Zuma, N., & Naidoo, K. (2020). Joint modelling of longitudinal and time-to-event data: An illustration using CD4 count and mortality in a cohort of patients initiated on antiretroviral therapy. BMC Infectious Diseases, 20(1), 1–9. https://doi.org/10.1186/s12879-020-04962-3

Mendoza, C. De, Paxinos, E., Barreiro, P., Camino, N., Nu, M., & Soriano, V. (2014). Different Viral Rebound following Discontinuation of Antiretroviral Therapy in Cases of Infection with Viruses Carrying L74V or Thymidine-Associated Different Viral Rebound following Discontinuation of Antiretroviral Therapy in Cases of Infection with Vir. March 2004. https://doi.org/10.1128/JCM.42.2.862

Min, S., Gillani, F. S., Aung, S., Garland, J. M., & Beckwith, C. G. (2020). Evaluating HIV Viral Rebound Among Persons on Suppressive Antiretroviral Treatment in the Era of " Undetectable Equals Untransmittable (U = U)." 1–9. https://doi.org/10.1093/ofid/ofaa529

Mosha, F. (2018). Predictors of non adherence to antiretroviral therapy at an urban HIV care and treatment center in Tanzania. 79–88.

Namazi, G., Fajnzylber, J. M., Aga, E., Bosch, R. J., Acosta, E. P., Sharaf, R., Hartogensis, W., Jacobson, J. M., Connick, E., Volberding, P., Skiest, D., Margolis, D., Sneller, M. C., Little, S. J., Gianella, S., Smith, D. M., Kuritzkes, D. R., Gulick, R. M., Mellors, J. W., … Henry, K. (2018). The Control of HIV After Antiretroviral Medication Pause (CHAMP) Study : Posttreatment Controllers Identified From 14 Clinical Studies. 218. https://doi.org/10.1093/infdis/jiy479

Oguntibeju, O. O. (2012). Quality of life of PLWH and AIDS and antiretroviral therapy. HIV/AIDS - Research and Palliative Care, 4, 117–124. https://doi.org/10.2147/HIV.S32321

Olariu, E., Cadwell, K. K., Hancock, E., Trueman, D., & Chevrou-severac, H. (2017). Current recommendations on the estimation of transition probabilities in Markov cohort models for use in health care decision-making : a targeted literature review.

Phillips, A. N., Weber, R., Kirk, O., Francioli, P., Miller, V., Vernazza, P., Lundgren, J. D., & Ledergerber, B. (2001). HIV Viral Load Response to Antiretroviral Therapy According to the

Baseline CD4 Cell Count and Viral Load. 286(20).

Prazuck, T., Descours, B., Guergnon, J., Viard, J., & Boufassa, F. (2013). Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy. 9(3). https://doi.org/10.1371/journal.ppat.1003211

Proust-lima, C., & Liquet, B. (2016). Estimation of Extended Mixed Models Using Latent Classes and Latent Processes : The R Package lcmm. 12.

Rezende, P. C., Hlatky, M. A., Hueb, W., Maria, R., Garcia, R., & Selistre, S. (2020). Association of Longitudinal Values of Glycated Hemoglobin With Cardiovascular Events in Patients With Type 2 Diabetes and Multivessel Coronary Artery Disease. 3(1), 1–9. https://doi.org/10.1001/jamanetworkopen.2019.19666

Rights, D., Barriers, B., People, R., & Services, W. H. I. V. (2019). COMMUNITIES AT THE.

Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data.

Saikia, R., & Barman, M. P. (2017). A Review on AFT Models. 12(2), 311–322.

Seid, A., Getie, M., Birlie, B., & Getachew, Y. (2014). Joint modeling of longitudinal CD4 cell counts and time-to-default from HAART treatment: A comparison of separate and joint models. Electronic Journal of Applied Statistical Analysis, 7(2), 292–314. https://doi.org/10.1285/i20705948v7n2p292

Sène, M., Bellera, C., Proust-lima, C., Sène, M., Bellera, C., & Shared, C. P. (2013). Shared random-effect models for the joint analysis of longitudinal and time-to-event data : application to the prediction of prostate cancer recurrence To cite this version : HAL Id : inserm-00870390 Shared random-effect models for the joint analysis of l.

Shoko, C., & Chikobvu, D. (2018). Determinants of viral load rebound on HIV/AIDS patients receiving antiretroviral therapy : results from South Africa. 1–13.

Silva, J. A. G., Dourado, I., de Brito, A. M., & da Silva, C. A. L. (2015). Factors associated with non-adherence to antiretroviral therapy in adults with AIDS in the first six months of treatment in Salvador, Bahia State, Brazil. Cadernos de Saude Publica, 31(6), 1188–11898. https://doi.org/10.1590/0102-311X00106914

Smith, C. J., Phillips, A. N., Dauer, B., Johnson, M. A., Lampe, F. C., Youle, M. S., Tyrer, M., & Staszewski, S. (2009). Factors associated with viral rebound among highly treatment-experienced HIV-positive patients who have achieved viral suppression. 19–27. https://doi.org/10.1111/j.1468-1293.2008.00650.x

Spiegelhalter, D. J., Best, N. G., & Carlin, B. P. (2002). Bayesian measures of model complexity and fit. 583–639.

Suppression, A. V., & States, U. (2020). HHS Public Access. 84(2), 133–140. https://doi.org/10.1097/QAI.0000000000002321.Viral

Tegegne, A. S., Ndlovu, P., & Zewotir, T. (2018). Determinants of CD4 cell count change and time-to default from HAART ; a comparison of separate and joint models. 1–11.

Temesgen, A., Gurmesa, A., & Getchew, Y. (2018). Joint Modeling of Longitudinal CD4 Count & Time-to-Death of HIV/TB Co-infected Patients at Jimma University Specialized Hospital. Annals of Data Science, 5(4), 659–678. https://doi.org/10.1007/s40745-018-0157-0

Temesgen, A., & Kebede, T. (2016). Joint Modeling of Longitudinal CD4 Count and Weight Measurements of HIV/TB Co-infected Patients at Jimma University Specialized Hospital. Annals of Data Science, 3(3), 321–338. https://doi.org/10.1007/s40745-016-0085-9

Tiruneh, F., Chewaka, L., & Abdissa, D. (2021). Statistical joint modeling for predicting the association of CD4 measurement and time to death of people living with HIV who enrolled in art, southwest Ethiopia. HIV/AIDS - Research and Palliative Care, 13, 73–79. https://doi.org/10.2147/HIV.S283059

Tony, D. O., & Emmanuel, I. E. (2020). Viral suppression and predictors among adolescents receiving care for HIV/AIDS in a tertiary health centre in. 12(December), 9–16. https://doi.org/10.5897/JAHR2020.0510

Tsiatis, A. A., & Davidian, M. (2004). Time-To-Event Data : An Overview. 14, 809–834.

UNAIDS. (2021). Fact Sheet 2021 Global HIV Statistics. Ending the AIDS Epidemic, June, 1–3.

Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. New York: Springer.

Wang, C. Y., De Dieu Tapsoba, J., Anderson, M. L., Vernon, S. W., Chubak, J., Fuller, S., & Green, B. B. (2014). Time to screening in the systems of support to increase colorectal cancer screening trial. Cancer Epidemiology Biomarkers and Prevention, 23(8), 1683–1688. https://doi.org/10.1158/1055-9965.EPI-14-0270

Woldemeskel, B. A., Kwaa, A. K., & Blankson, J. N. (2020). EBioMedicine Viral reservoirs in elite controllers of HIV-1 infection : Implications for HIV cure strategies. EBioMedicine, 62, 103118. https://doi.org/10.1016/j.ebiom.2020.103118

Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. Journal of Probability and Statistics, 2012. https://doi.org/10.1155/2012/640153

Wulfsohn, M. S., & Tsisatis, A. A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error Author (s): Michael S . Wulfsohn and Anastasios A . Tsiatis Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2533118 Accessed : 03-08-2016 21 : 3. Biometrics, 53(1), 330–339.

Yang, X., Su, B., Zhang, X., Liu, Y., Wu, H., & Zhang, T. (2020). Incomplete immune reconstitution in HIV/AIDS patients on antiretroviral therapy: Challenges of immunological non-responders. Journal of Leukocyte Biology, 107(4), 597–612. https://doi.org/10.1002/JLB.4MR1019-189R

Ye, W., Lin, X., & Taylor, J. M. G. (2008). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. 1, 33–45.

Zhang, Z. (2016). Model building strategy for logistic regression: Purposeful selection. Annals of Translational Medicine, 4(6), 4–10. https://doi.org/10.21037/atm.2016.02.15

8. ANNEXES

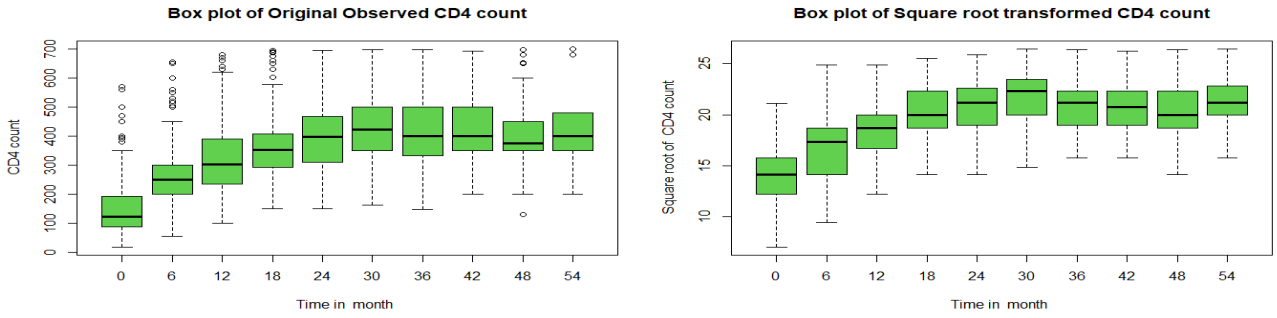
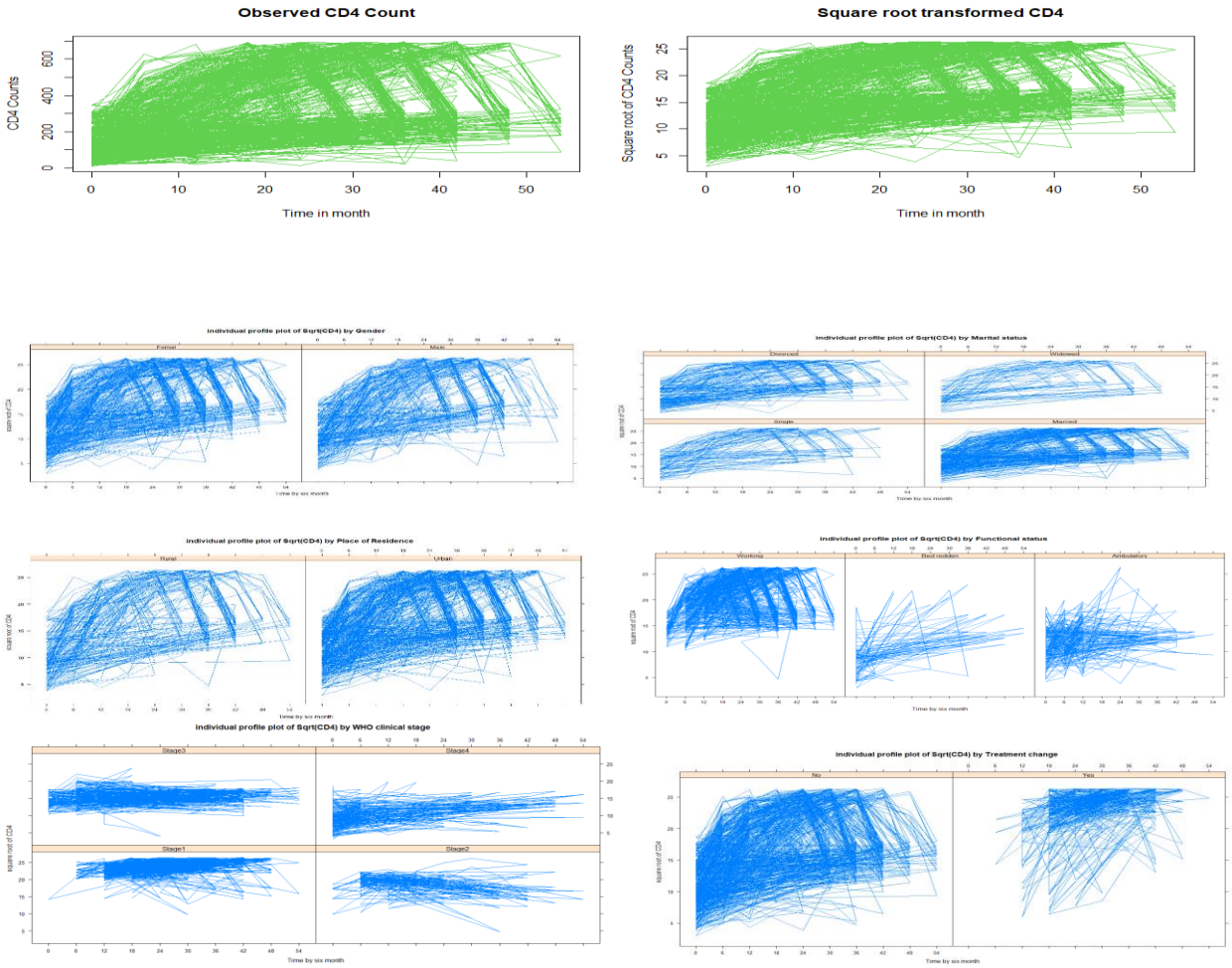


Figure 6: Box Plots of CD4 cell count by time for the original and square root transformed data for the normality check.



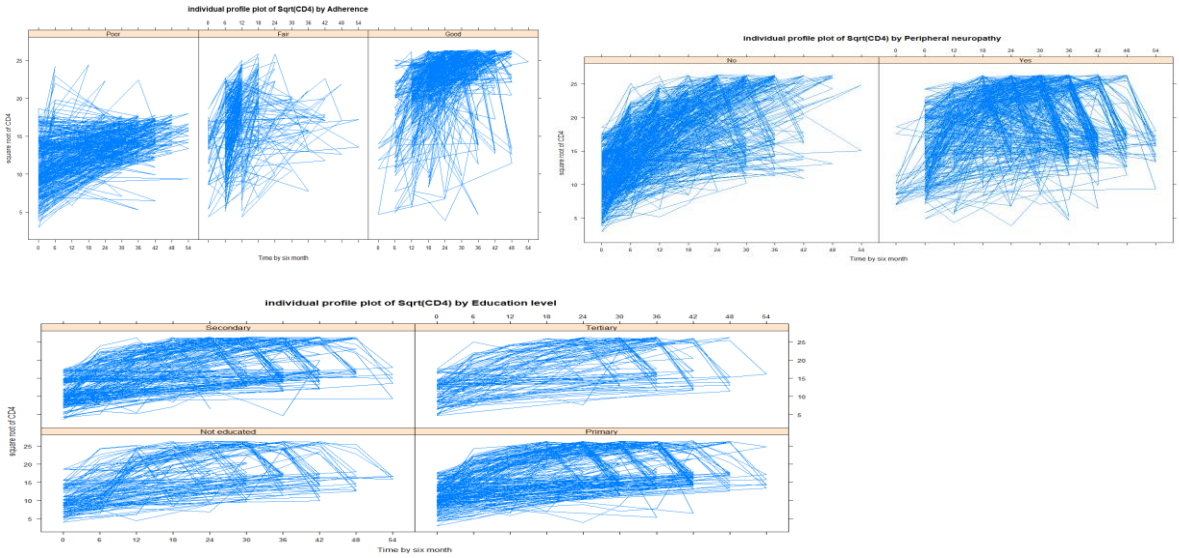


Figure 7: Individual profile plot for all covariates by the square root of CD4 cell count.

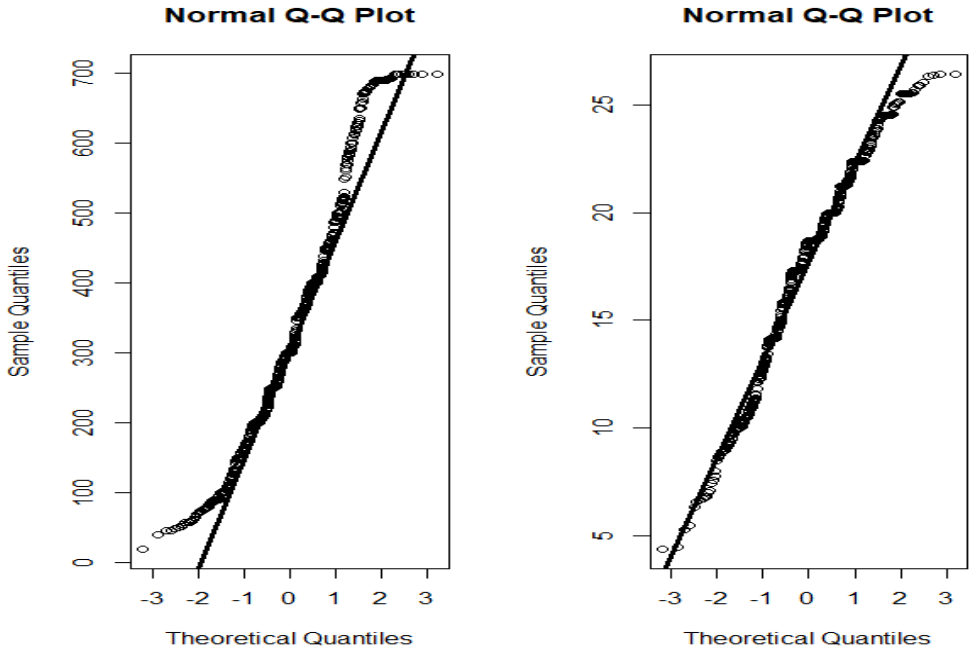


Figure 8: Q-Q plot for original CD4 cell count and square root of CD4 cell count progression to the normality check.

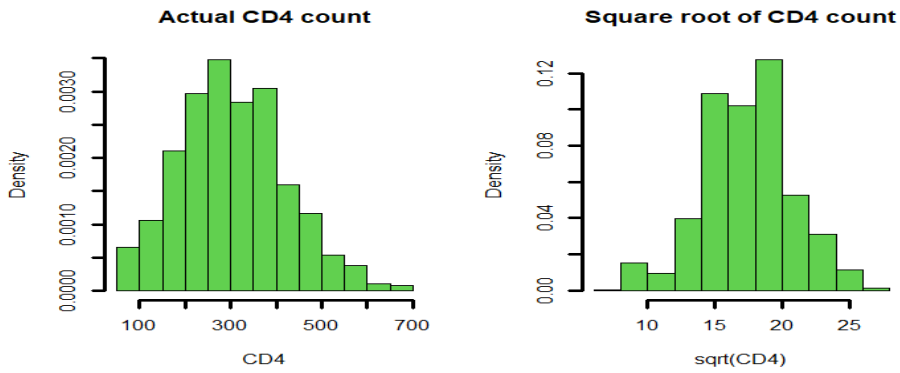


Figure 9: Histogram plot of the actual and square root of CD4 cell count for the normality check.

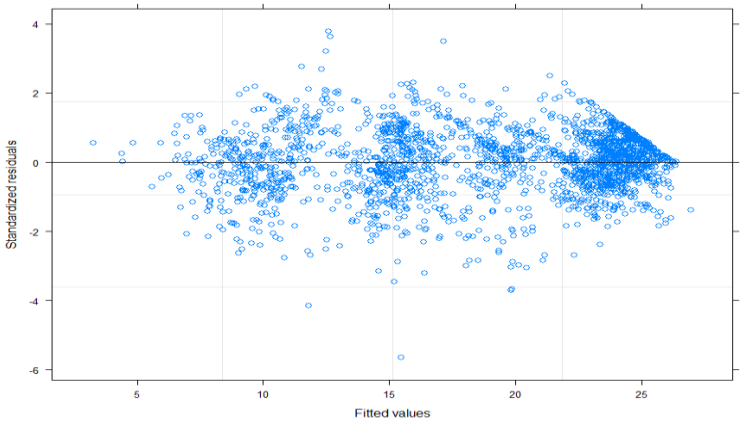


Figure 10: Residual versus fitted values of within-group error term of linear mixed model.

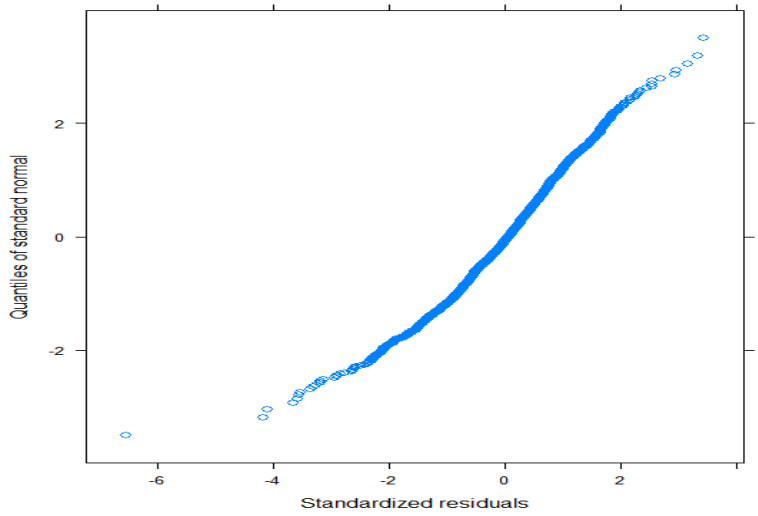


Figure 11: Quantile-quantile plots for the normality within-group error of the fitted linear mixed model.

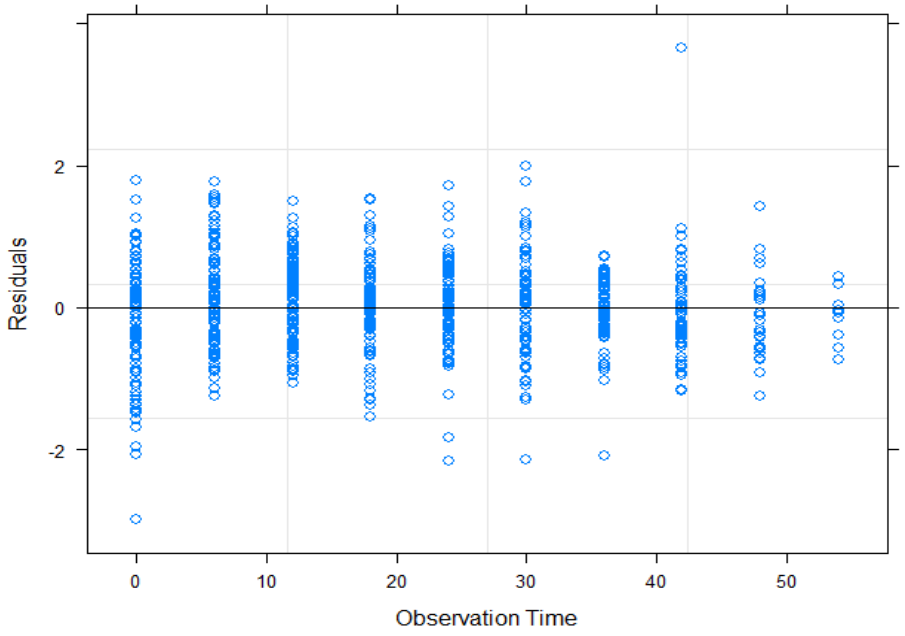


Figure 12: Residual versus fitted values for the random effects of the fitted linear mixed model.

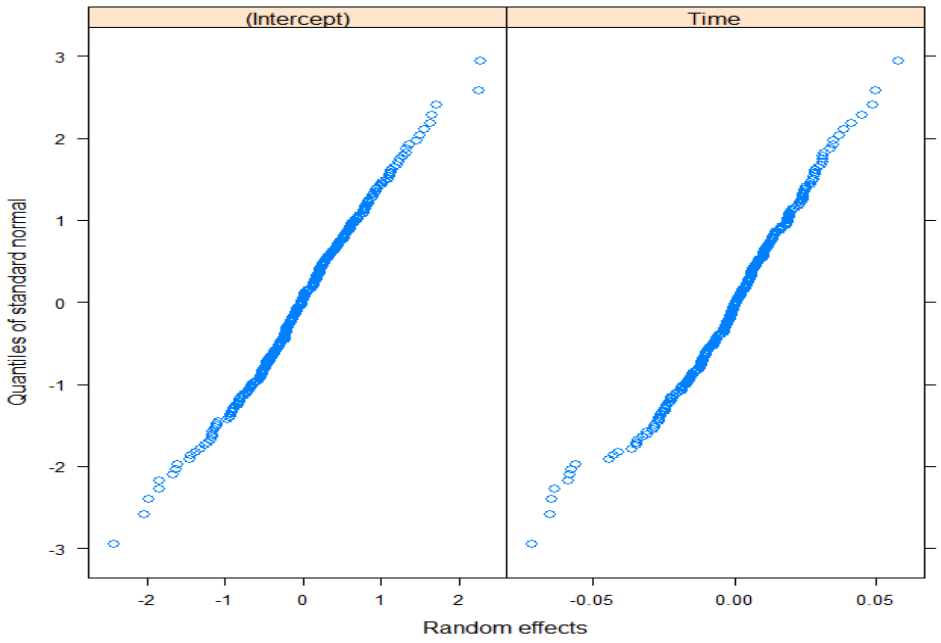


Figure 13: Standardized quantile-quantile plots for the normality of random effects of the linear mixed model.

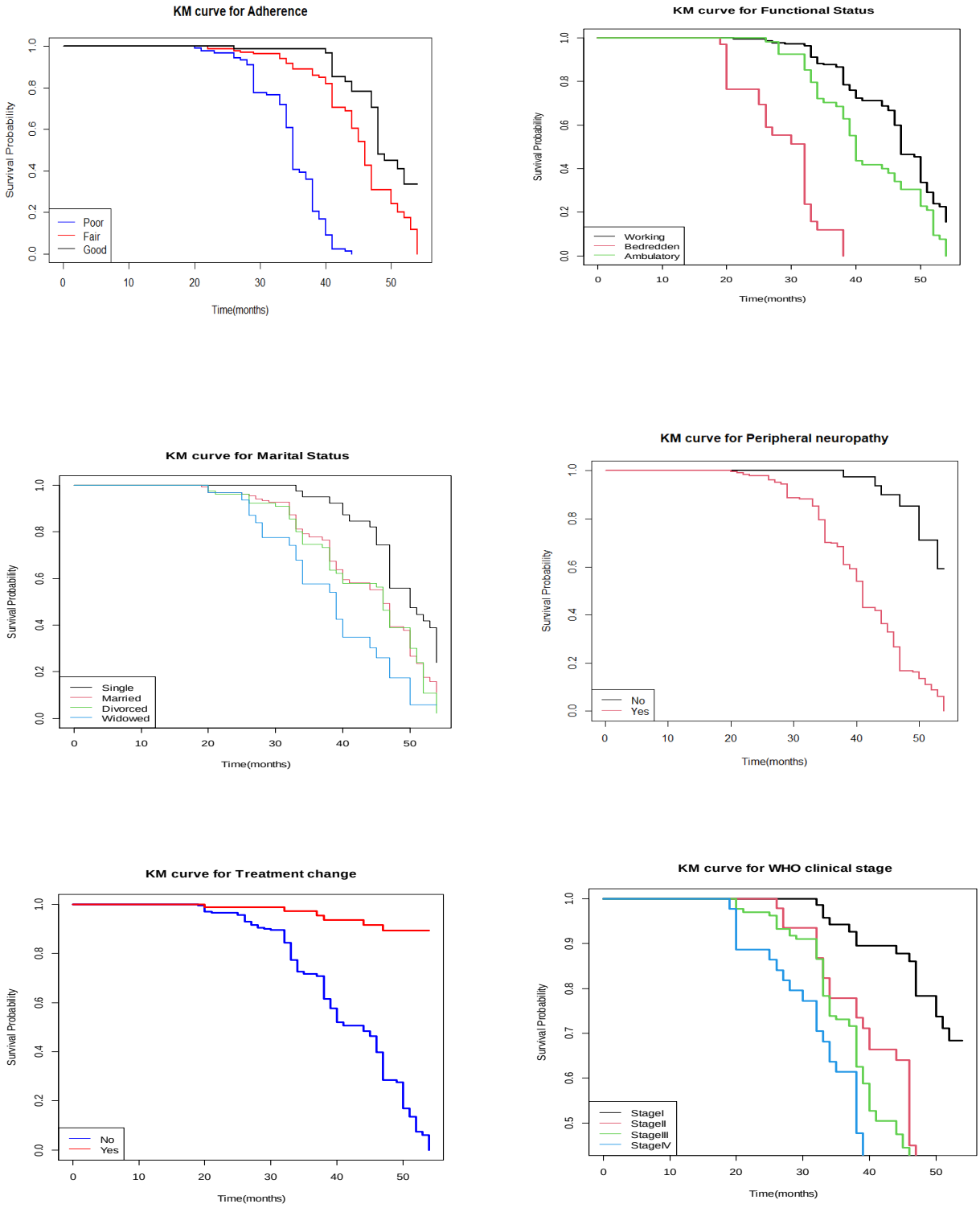


Figure 14: Estimated Kaplan-Meier curve of survival probability of HIV- infected patients for categorical covariates.

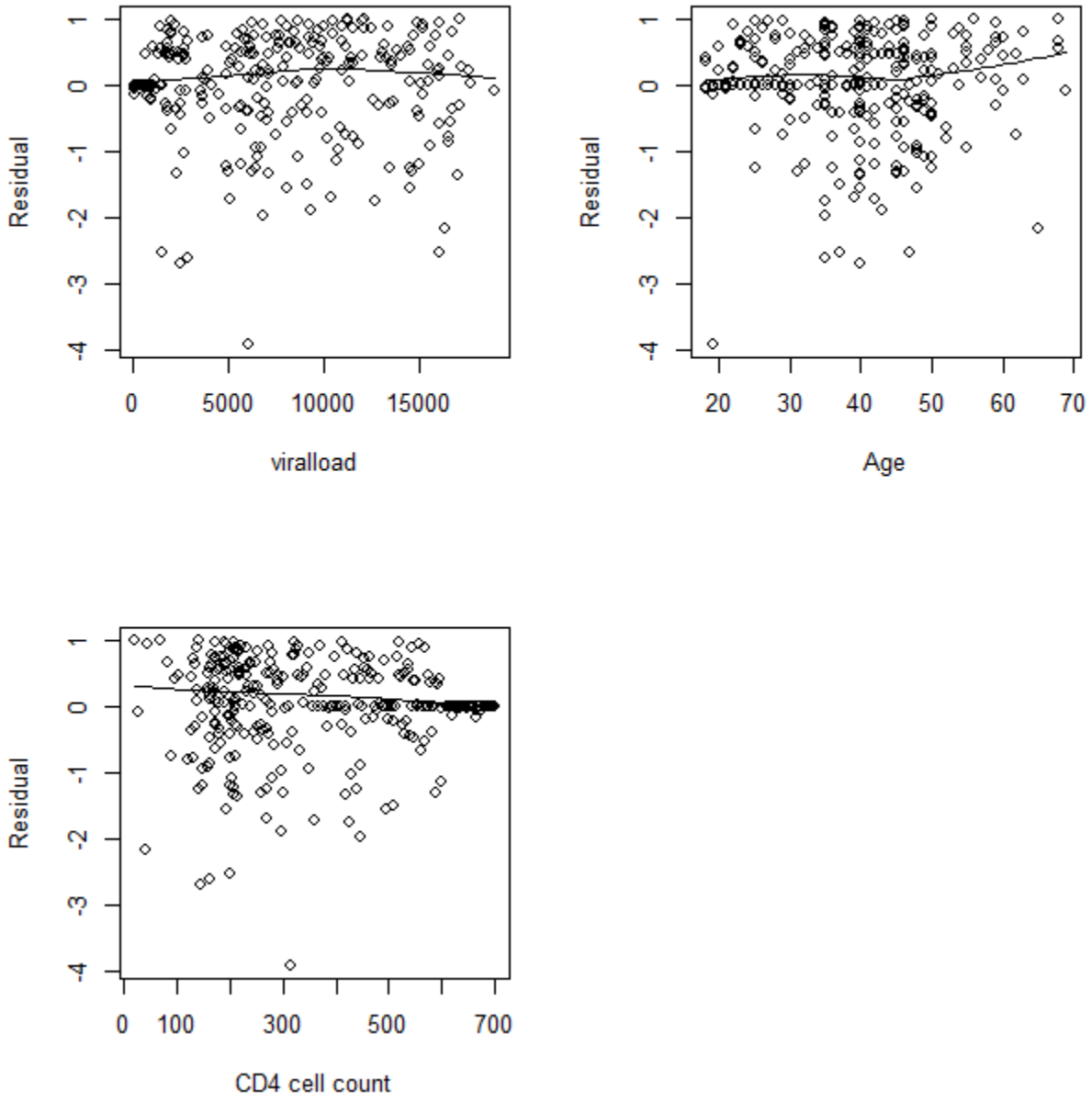


Figure 15: Martingale residual plots for continuous covariate used in Cox PH model to test the linearity structures.

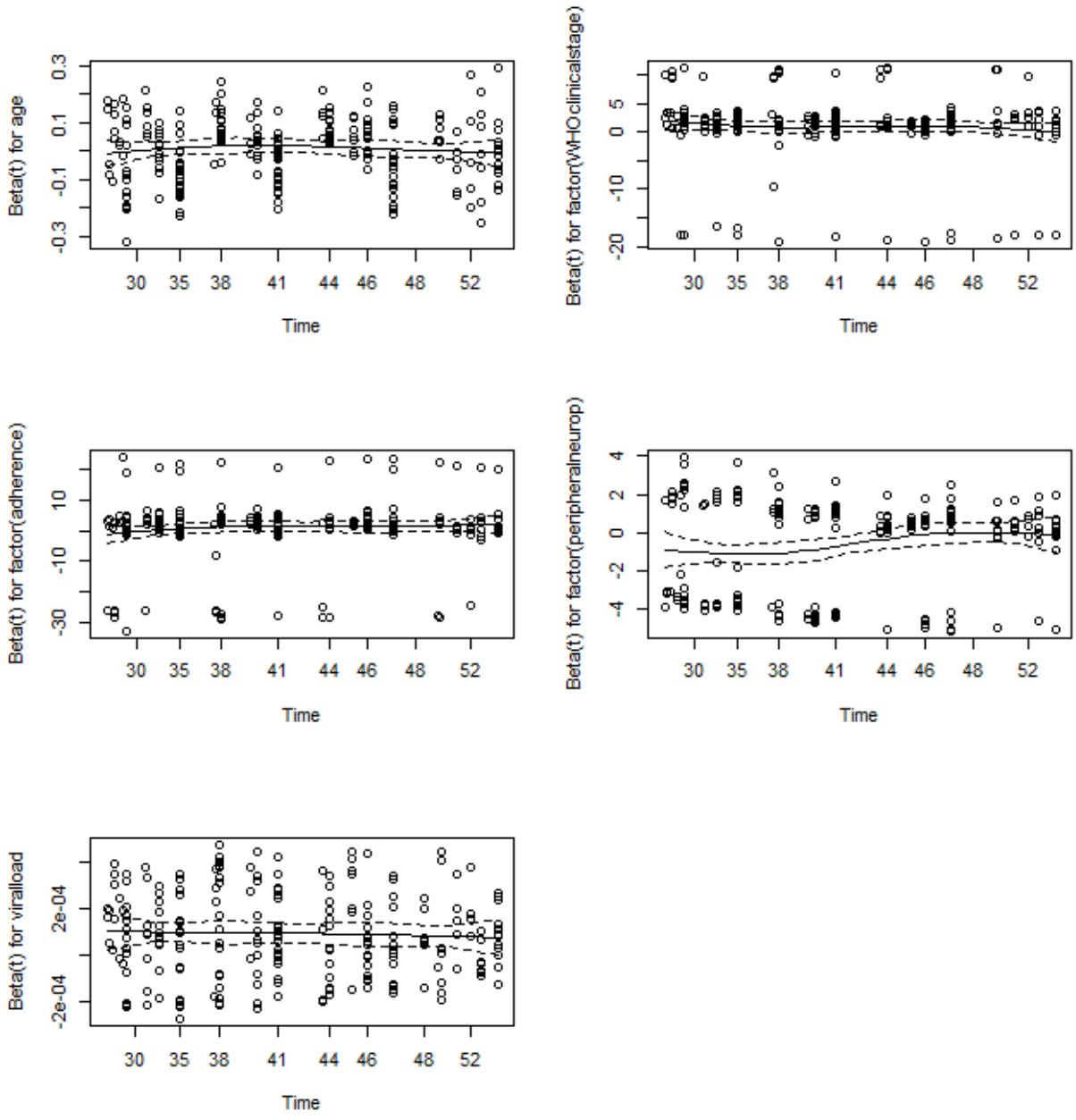


Figure 16: Schoenfeld residual plots for Cox PH assumption tests.

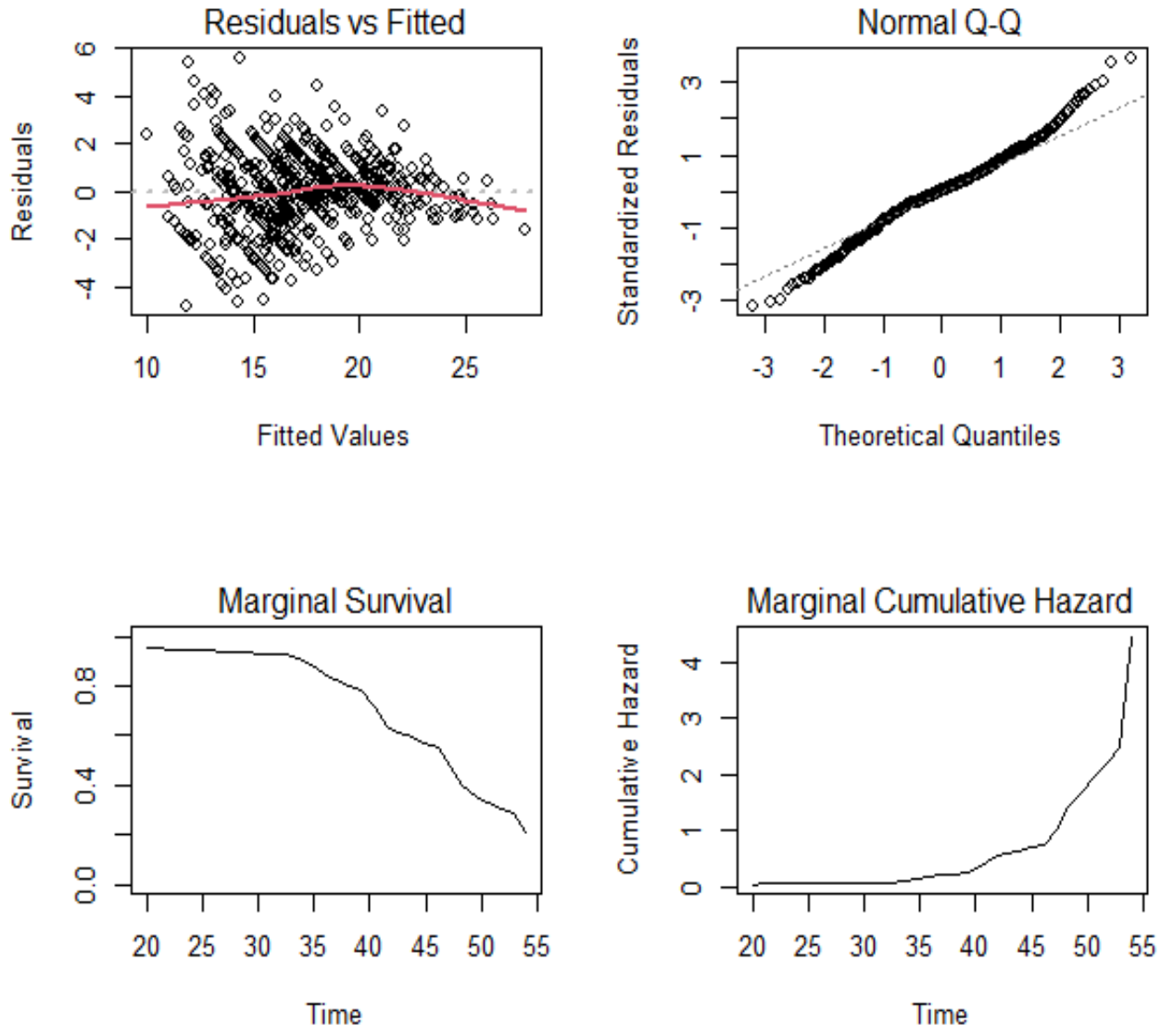


Figure 17: Default plots for joint model.

Table 12: Result of normality test for longitudinal CD4 cells count.

Follow-up Time	Observed CD4 Count		Square root transformed CD4 Count	
	Tests of Normality		Tests of Normality	
	Shapiro-wilk		Shapiro-wilk	
	Statistics	p-value	Statistics	p-value
0	0.94	<0.001	0.98	0.010
6	0.96	<0.001	0.97	0.608
12	0.96	<0.001	0.91	0.566
18	0.91	<0.001	0.82	0.684
24	0.86	<0.001	0.78	0.049
30	0.81	<0.001	0.77	0.095
36	0.83	<0.001	0.81	0.717
42	0.86	<0.001	0.89	0.319
48	0.86	<0.001	0.86	0.025
54	0.77	0.002	0.86	0.805

Table 13: Comparison of random intercept, random slope, random intercept random slope model, and random intercept and quadratic slope in random effect.

Random effect		Variance	Std.Dev	95% CI (Std.Dev)
Random intercept	Intercept	0.97	0.98	(0.86, 1.12)
	Residual	3.58	1.89	(1.83, 1.96)
Random slope	Time	0.001	0.03	(0.02, 0.03)
	Residual	4.14	2.03	(1.97, 2.10)
Random intercept- Random slope	Intercept	2.23	1.49	(1.30, 1.72)
	Time	0.001	0.04	(0.03, 0.05)
	Residual	3.30	1.82	(1.75, 1.89)
Random intercept and quadratic slope	Intercept	2.15	1.47	(1.25, 1.72)
	Time	0.02	0.13	(0.11, 0.16)
	Time^2	0.00001	0.003	(0.0024, 0.004)
	Residual	2.49	1.58	(1.52, 1.64)

Table 14: Parameter estimate of the univariable linear mixed effect models on the factor for the change in square root of CD4 cell count

Covariates	Estimate	Std	t - value	p – value	95% CI	
Age	-0.05	0.01	-3.18	0.001**	[-0.07,	-0.02]
Viral load	-0.41	0.18	-1.86	<0.001**	[-0.77,	-0.05]
Gender; Female **Ref**						
Male	-0.28	0.18	-1.51	0.131	[-0.64,	0.09]
WHO Stage; Stage1 **Ref**						
Stage2	-5.00	0.15	-32.43	<0.001***	[-5.31,	-4.70]
Stage3	-8.46	0.14	-61.64	<0.001***	[-8.73,	-8.19]
Stage4	-13.35	0.15	-90.89	<0.001***	[-13.64,	-13.1]
Marital Status; Single **Ref**						
Married	-0.16	0.26	-0.59	0.5570	[-0.68,	0.37]
Divorced	-0.36	0.29	-1.22	0.2232	[-0.93,	0.22]
Widowed	-0.35	0.37	-0.94	0.3494	[-1.09,	0.39]
Residence; Rural **Ref**						
Urban	0.19	0.20	0.97	0.3343	[-0.19,	0.59]
Functional status; Working **Ref**						
Bed redden	-11.53	0.30	-38.11	<0.001***	[-12.12,	-10.9]
Ambulatory	-9.52	0.21	-46.31	<0.001***	[-9.92,	-9.11]
Regimen type; AZT+3TC+ATV/r **Ref**						
AZT+3TC+LPV/r	-0.25	0.45	-0.54	0.5877	[-1.14,	0.64]
TDF+3TC+DTG	-2.13	0.62	-3.44	0.0007**	[-3.34,	-0.91]
TDF-3TC-EFV	0.06	0.40	0.16	0.8748	[-0.73,	0.86]
AZT+3TC+DTG+DRV/r	0.89	0.37	2.38	0.0177*	[0.16,	1.62]
Treatment change; No **Ref**						
Yes	5.63	0.27	20.81	0.0401*	[5.10,	6.17]
Peripheral neuropathy; Yes **Ref**						
No	2.53	0.24	10.55	<0.001***	[2.06,	2.99]
Adherence; Poor **Ref**						
Fair	4.58	0.19	23.76	<0.001***	[4.19,	4.95]
Good	9.82	0.16	60.88	<0.001***	[9.51,	10.1]
Education level; Not educated **Ref**						
Primary	-0.02	0.26	-0.06	0.9505	[-0.52,	0.49]
Secondary	0.04	0.26	0.16	0.8718	[-0.48,	0.56]
Tertiary	0.14	0.31	0.44	0.6607	[-0.47,	0.75]

** Indicates the significance at 25% level of significance.*

*** Indicates the significance at 5% level of significance.*

**** Indicates the significance at 1% level of significance.*

Table 15: Univariable analysis of Cox proportional hazard model on the factor affecting the survival of HIV infected patients initiating ART

Variables	Estimate	StDev	z-value	p-value	HR (95% CI) Lower	Upper
Age	0.19	0.01	15.32	<0.001***	1.21[1.18,	1.25]
Viral load	0.05	0.03	15.01	<0.001***	1.05[1.00,	1.01]
Baseline CD4	0.15	0.07	6.45	0.227	1.16[0.29	4.58]
Gender Male	0.02	0.14	0.02	0.987	1.02[0.76,	1.31]
Marital status						
Married	-0.02	0.21	-0.09	0.921	0.98[0.65,	1.48]
Divorced	0.09	0.23	0.41	0.679	1.09[0.70,	1.72]
Widowed	0.36	0.27	1.30	0.193	1.43[0.83,	2.46]
Functional status						
Bed redden	0.27	0.21	0.51	1.195	1.31[0.87,	1.97]
Ambulatory	0.47	0.16	2.94	0.003**	1.59[1.17,	2.18]
Regimen type						
AZT+3TC+LPV/r	-0.34	1.41	-0.24	0.812	0.71[0.04,	11.42]
TDF+3TC+DTG	2.79	1.22	2.28	0.022*	16.4[1.49,	181.2]
TDF-3TC-EFV	4.21	1.01	4.17	0.005**	67.6[9.33,	490.5]
AZT+3TC+DTG+DRV/r	3.90	1.01	3.87	0.010*	49.5[6.87,	356.1]
WHO clinical stage						
Stage II	1.38	0.29	4.80	<0.001***	3.98[2.27,	7.01]
Stage III	1.34	0.25	5.21	<0.001***	3.83[2.31,	6.36]
Stage IV	1.25	0.28	4.37	<0.001***	3.49[1.99,	6.14]
Peripheral neuropathy						
No	-0.87	0.14	-6.19	<0.001***	0.41[1.82,	3.17]
Treatment change Yes	-3.25	0.43	-7.59	<0.001***	0.04[0.02,	0.09]
Adherence						
Fair	-1.79	0.25	-7.23	<0.001***	0.17[0.10,	0.27]
Good	-1.07	0.23	-4.55	<0.001***	0.34[0.22,	0.54]
Education level						
Primary	-0.18	0.17	-1.05	0.294	0.83[0.59,	1.17]
Secondary	-0.30	0.18	-1.62	0.104	0.74[0.51,	1.06]
Tertiary	-0.17	0.23	-0.74	0.458	0.85 [0.54,	1.32]
Place of reside Urban	-0.05	0.14	-1.85	0.733	0.95[0.72,	1.26]

*Indicates the significance at 25% level of significance.

P-Value of **Indicates the significance at 5% level of significance.

*** Indicates the significance at 1% level of significance.