PREDICTING VIRAL SUPPRESSION IN ANTI-RETROVIRAL THERAPY PATIENTS USING TRANSITION MODELS AND MIXED EFFECTS MODELS

A MINI THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS OF THE DEGREE OF

MASTER OF SCIENCE (BIOSTATISTICS)

OF

THE UNIVERSITY OF NAMIBIA

BY

ANNA-NUUSIKU ONESMUS

201303555

APRIL 2021

SUPERVISOR: PROF LAWRENCE KAZEMBE (UNIVERSITY OF NAMIBIA,

DEPARTMENT OF STATISTICS AND POPULATION STUDIES)

Abstract

Background: Acquired Immune Deficiency Syndrome (AIDS) which is caused by the human immunodeficiency virus (HIV) is a leading cause of death, a limitation to development in Africa, as well as a health threat to millions worldwide. The main purpose of antiretroviral therapy (ART) is to keep people living with HIV in good health. Progression of health in HIV/AIDS patients on ART is characterised by an increase in CD4 cell counts as well as a decrease in viral load (HIV RNA) to undetectable levels. Monitoring and evaluation (M&E) of HIV occurrence and responses allow countries to track the epidemic and their prevention and control efforts. This study assesses the use of viral load in monitoring HIV/AIDS progression and predicting viral suppression.

Objective: The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model.

Methods: The models were applied to secondary data from ART patients drawn from a retrospective cohort design. The participants of the study were aged fifteen years or older (referred to as adults), who initiated ART treatment at selected health facilities in Erongo region, Namibia, between January 2010 and December 2015, giving a total population of 8068 patients. The data was analysed in three parts, namely, the descriptive analysis, the linear mixed effects model and the transition autoregressive model. Both models assume that the future values of a variable depend on its historical values as well as on other covariates.

Results: The findings of this study points to a viral suppression prevalence rate of 71% in Erongo region. The findings provide additional support to the concept that demographic (sex, age and weight) and clinical (follow up time, adherence and clinical stage) characteristics are the determinants of viral load as well as CD4 cell counts. The Log Likelihood, AIC and BIC show that the Transition Autoregressive model is the best predicting model for viral load as well as CD4 cell counts compared to the Mixed Effects Models.

Conclusion: Based on the AIC and BIC, it was shown that the transition model and the mixed effects model have almost the same predictive accuracy at the first few visits of a patient; however, for longer time series the transition model offers somewhat better predictions. Since the transition model is convenient in practice and needs less historical information compared to the mixed effects model, HIV studies may use this model to predict the future viral load. In the final analysis, this study recommends the use of the Transition Autoregressive model in modelling change and event occurrence in longitudinal studies such as HIV studies, specifically in predicting viral suppression in the future.

Keywords: HIV/AIDS; ART; viral suppression; CD4 counts; viral load; Transition Autoregressive model; Mixed Effects models, M&E

Table of Contents

List of Tables

List of Figures

List of Abbreviations/Acronyms

AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
BIC	Bayesian Information Criterion
CI	Confidence Interval
HIV	Human Immunodeficiency Virus
ICC	Intra Class Correlation Coefficient
LB	Lower Bound
MoHSS	Ministry of Health and Social Services
UB	Upper Bound
UNAIDS	United Nations Joint Programme on AIDS
UNAM	University of Namibia
WHO	World Health Organization

Acknowledgement

First and above all, I praise God, The almighty for granting me the capability to progress successfully. This thesis appears in its current form due to the assistance and guidance of several people.

I would like to express my sincere gratitude to my supervisor Prof. Kazembe for his guidance throughout the research. His invaluable guidance, immense knowledge, patience and encouragement, despite his busy schedule have allowed me to complete the research.

I am very grateful to my colleagues, Lineekela Gabriel and Secilia Kapenda, who have been so supportive along the way of doing my Master's degree.

Finally, I want to thank my mother for always being there for me and supporting me throughout my life, her prayers brought me this far.

Dedications

With a special feeling, I dedicated this thesis to my children Salde Tuhafeni and Justy Peyavali Ndaumbwa, who I neglected most of the times in the process of conducting this study. I also dedicated this thesis to my beloved mother Maria Tangi Nangolo who always believed in me since my day one on earth.

Declarations

I, **Anna-Nuusiku Onesmus**, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, **Anna-Nuusiku Onesmus**, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit, for any person or institution requiring it for study and research; providing that The University of Namibia shall waive this right if the whole thesis has been or is being published in a manner satisfactory to the University.

……………………..……………. ……………………………… …..…………

Name of Student Signature Date

Chapter1: Introduction

1.1 Background of the study

Acquired Immune Deficiency Syndrome (AIDS) which is caused by the human immunodeficiency virus (HIV) is a leading cause of death, a limitation to development in Africa, as well as a health threat to millions worldwide (Glance and Staff, 2017 ; IHME, 2016; Jamison, 2006; Temah, 2009; Ramjee and Daniels, 2013). Although no single country has been spared the virus, the epidemic has affected certain regions of the world extremely, and sub-Saharan Africa remains by far the hardest hit region (Alleyne et al, 2006; UNAIDS, 2016). HIV prevalence is 14% in Namibia, one of the highest prevalence rates in the world (Namibia Demographic Health Survey, 2013).

AIDS was first recognized as a disease in the early 1980s (Jamison, 2006) and the first cases of HIV were reported in 1981 (UNAIDS, 2016). Since 1996, HIV/AIDS has been the leading cause of death in Namibia, the HIV/AIDS epidemic peaked in 1998 and the death from HIV/AIDS peaked in 2005, however, by 2013, death rates had more than halved for females, and almost halved for males (IHME, 2016). The Joint United Nations Programme on HIV/AIDS (UNAIDS) had estimated that about 40 million people in the world were infected with HIV in 2005, of whom about 25.8 million, or 64 per cent of the total, were in Sub-Saharan Africa (UNAIDS and WHO, 2005). Routine antenatal surveillance estimated that 1 in 5 pregnant women were infected with HIV in 2006.

In 2018 the overall HIV prevalence among adults aged 15-64 years was 12.6 per cent, NAMPHIA (2018), which was about 176,000 (95% CI, 164,000-189,000) adults living with HIV in Namibia. There are approximately 36.7 million people currently living with HIV in the world and tens of millions of people have died of AIDS-related causes since the beginning of the epidemic, UNAIDS (2016). While new cases have been reported in all regions of the world, approximately two-thirds are in sub-Saharan Africa, with 46% of new cases in Eastern and Southern Africa, UNAIDS (2016).

The use of antiretroviral (ARV) drugs in the United States and other developed countries has dramatically reduced AIDS death rates, but only a small proportion of AIDS patients in Sub-Saharan Africa receive treatment (Jamison, 2006). The main purpose of antiretroviral therapy is to keep people living with HIV in good health. In the large majority of people living with HIV, antiretroviral medication can be chosen to reduce the amount of HIV in the blood to levels that are undetectable by standard laboratory tests. Viral Suppression is defined as the event that occurs when ART reduces a person's viral load to an undetectable level (US Department of Health and Human Services, 2018). Viral suppression does not imply a person is cured; HIV still remains in the body. If ART is discontinued, viral load will reappear to a detectable level. According to WHO (2016), a viral load threshold of <1000 copies/ml defines treatment success and ART patients with viral load test results below the threshold should be considered as having reached viral suppression. However, MoHSS (2016) used a viral load threshold of <40 copies/ml to define viral suppression.

Monitoring and evaluation (M&E) of HIV occurrence and responses allow countries to track the epidemic and their prevention and control efforts (Khasakhala and Kloos, 2013). Scientists use standardized indicators to assess progress and challenges over time and make cross-national comparisons. Governments and international donors in Sub-Saharan Africa are giving increasing attention to the need to develop effective M&E programs to provide reliable data in a timely fashion that can guide and help to improve programs. The aim of M&E programs and activities is to track changes, targeting indicators and relate them to interventions. HIV/AIDS monitoring and evaluation programs play a crucial role to help advocate for, plan, monitor and evaluate policies and programmes for HIV prevention (UNAIDS, 2004). They also tailor existing widely used indicators and methods so that they are relevant to the national programmes for HIV prevention and introduce new measures specific to the determinants (risk and protective factors) which influence vulnerability and risk behaviours.

HIV transmission can be controlled by seeking HIV testing, initiating treatment, and completing the continuum of care to achieve an undetectable level of viral RNA (Detels, Wu, Wu, 2019). The US Department of Health and Human Services (2018) defined viral load as the amount of HIV in a sample of blood which is reported as the number of HIV RNA copies per millilitre of blood. Changing human behaviour in the short term is possible but for the long term is particularly challenging, but it can and must be done to stop the epidemic. The observation that individuals on treatment who are compliant with medication have their blood HIV (RNA) concentration lowered to undetectable levels has resulted in a range of new strategies to prevent transmission. Individuals who have achieved an undetectable level of viral RNA do not transmit their HIV to sexual and injection partners (Detels, Wu & Wu, 2019).

In December 2013, the Joint United Nation's Program on HIV/AIDS brought into effect a '90-90-90' strategy with the goal of bringing an end to the AIDS epidemic (UNAIDS, 2014). 90-90-90 aims to increase HIV viral suppression to 73% worldwide by 2020. According to NAMPHIA (2018), 77% of all HIV-positive adults on ART in Namibia have achieved viral suppression, surpassing the UNAIDS (2014) target of 73% by 2020. Furthermore, 90-90-90: 86.0% of people living with HIV (ages 15 to 64) in Namibia report knowing their status, 96.4% of those individuals self-report being on ART, and 91.3% of that group are virally suppressed, NAMPHIA (2018).

1.2 Statement of the problem

The 90-90-90 targets were launched in 2014 with the purpose of increasing awareness of HIV status, access to ART, and viral suppression among individuals who were diagnosed with HIV (UNAIDS, 2014). Expanding access to ART represents a priority throughout sub-Saharan Africa (UNAIDS, 2013). High rates of viral suppression are attainable at the regional and country level and studies have identified various individual, interpersonal, and institutional factors that are associated with viral suppression in general population samples (Jobanputra et al., 2015).

Regular monitoring of HIV viral load levels remains the perfect measure of effective ART. Clinicians monitor HIV RNA levels and CD4 counts response to viral suppression. The third strategy in the 90-90-90 ensures that all people receiving antiretroviral therapy (ART) will be virologically suppressed by 2020. This strategy requires countries to sustain routine viral load (VL) monitoring of all people on ART. WHO (2016) recommends viral load monitoring at 6 and 12 months after starting antiretroviral therapy and annually for patients who are stable on ART, as a direct measure of response to treatment.

The motivation to conduct this particular study emanates from the fact that despite many studies having been conducted on HIV/AIDS viral suppression, there are no or few if any studies on HIV/AIDS viral suppression in Namibia that have predicted patients' future value of viral load, and more specifically, viral load suppression. Thus, there was a need of a study that predicts viral suppression using previous measurements of viral load of the patient as well as covariates. For this reason, the purpose of this study was to compare the performance of two longitudinal models namely the transition autoregressive model and the mixed effects model in predicting viral suppression in ART patients using viral load, and deduce on the best model to be recommended to make these predictions in future.

1.3 Objectives of the study

The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model, with application to patient cohort in Erongo, Namibia.

1.3.1 Specific objectives

In order to answer the question above, the following secondary objectives were addressed:

- To estimate a mixed effects model for predicting viral load among ART patients in Erongo Region;
- To fit a transition model for predicting viral load among ART patients in Erongo Region;
- To identify the factors associated with viral load among patients on ART in Erongo region.
- To determine the best model between two longitudinal models namely the transition autoregressive model and the mixed effects model in predicting viral load and CD4 cell counts in ART patients.

1.4 Significance of the Study

The results will help in designing optimal treatment support mechanisms, hence permitting early viral suppression and improved health care outcomes. When patients get suppressed earlier, their CD4 counts are most likely to increase, reducing the risk of acquiring opportunistic infections and hence keeping them healthy, not falling ill and neither passing on the virus. An adequate prediction of viral suppression would help the government and its partners to inform policy decisions and resource allocation, whereby allowing policy makers to utilise mechanisms that define and frame the HIV/AIDS problem in a fiscal responsible manner in order to reduce the burden of the high cost of ART.

1.5 Organisation of the Thesis

The thesis introduced the topic in Chapter 1 which included the background, problem statement, objectives, significance as well as limitations of the study. Chapter 2 consists of the literatures review on the models. Chapter 3 consists of the research methodology. Chapter 4 and 5 are stand-alone papers outlining the applications of the Mixed Effects and Transition Autoregressive models respectively. They consist of the introduction, results, discussion and conclusions. Discussions, Conclusion and Recommendations made up Chapter 6. Finally, the Bibliography and Appendix which consists of the R codes.

Chapter 2: Longitudinal Models with Transition and Mixed Effects: A review of theory

2.1 Introduction

This chapter reviews the theory of two longitudinal models, namely the Transition Autoregressive model and the Mixed Effects model. Section 2.2 reviews the Mixed Effects models, it defines the Mixed Effects model, provide a review of similar studies, the mathematical definition, assumptions, model specification as well as the statistical inference on the Mixed Effects Model, which entails estimation of the population parameters and individual parameters. Section 2.3 focuses on the Transition Model; it provides a history of the transition model, a review of similar studies, the mathematical definition, assumptions, and model specification and estimation procedure.

2.2 Mixed Effects Models

The mixed effect model is a flexible instrument for studying datasets with both fixed and random effect factors, as well as numerical covariates (Baayen, 2012). They are used when there is non-independence in the data, to be specific in settings where repeated measurements are made on the same statistical unit or on clusters of related statistical units, because of their advantage of dealing with missing values. Laird and Ware (1982) first applied these models to longitudinal data, consequently, the use of linear mixed effects models (LMMs) is now increasingly common in the analysis of biological data (Harrison et al., 2018).

2.2.1 Definition

In this section we follow Helwig (2017) in defining the mixed effects models. Linear mixed effects models are regression models that take into account both variation that is explained by the independent variables of interest (fixed effects), and variation that is not explained by the independent variables of interest (random effects). The random effects basically give structure to the error term. Mixed effects models explicitly model the within-subject correlation by using random effects. They rely on parametric assumptions to model the random effects as well as sampling variability and yields biased estimates if either of these assumptions is violated.

A linear mixed-effects regression model has the form:

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ijk} + V_{i0} + \sum_{k=1}^{q} V_{ik} Z_{ijk} + e_{ij} \ . \tag{2.1}$$

for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_i\}$, where $y_{ij} \in \mathbb{R}$ is response for j-th measurement of i-th subject, $\beta_0 \in \mathbb{R}$ is fixed intercept for the regression model, $\beta_k \in \mathbb{R}$ is fixed slope for the k-th predictor, $x_{ijk} \in \mathbb{R}$ is j-th measurement of k-th fixed predictor for i-th subject. The term V_{i0} is random intercept for the $\text{i} - \text{th}$ subject, while V_{ik} is random slope for k-th predictor of i-th subject, such that $V_{i0} \sim^{iid} N(0, \sigma_0^2)$, and $V_{ik} \sim^{iid} N(0, \sigma_k^2)$ corresponding to $Z_{ijk} \in \mathbb{R}$ which is j-th measurement of k-th random predictor for i-th subject.

The error term is distributed as a zero mean and constant variance Gaussian process: $e_{ij} \sim^{iid} N(0, \sigma_e^2)$. Further it is assumed that V_{i0} and e_{ij} are mutually independent. In addition, in the random intercept model the correlation between two observations of a subject is constant and is equal to the intra-class correlation given by $ICC = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2}$ and correlation between the repeated measurements remains constant over time. Further correlation amongst the errors can be housed by permitting for a more general covariance structure (e.g., autoregressive) in the model. Furthermore, it is assumed that V_i has a bivariate normal distribution with mean zero and a diagonal covariance matrix, so that e_i and V_i are mutually independent. Finally, the rate of change in the covariates differs between subjects.

Rewriting the model in matrix form:

$$y_i = X_i\beta + Z_iV_i + e_i \ ,$$

where

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix},\ X_i = \begin{pmatrix} 1 & x_{i1}^{(1)} & \cdots & x_{i1}^{(m)} \\ 1 & x_{i2}^{(1)} & & x_{i2}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in}^{(1)} & \cdots & x_{in}^{(m)} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix},\ V_i = \begin{pmatrix} v_{i0} \\ v_{i1} \\ \vdots \\ v_{im} \end{pmatrix}, e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in} \end{pmatrix} \quad (2.2)$$

Here, y_i is the v_i vector of observations for individual i, X_i is the v_i×d design matrix (with d=m+1), β is a d-vector of fixed effects (i.e. common to all individuals of the population), v_i is a d-vector of random effects (i.e. specific to each individual) and e_i is a v_i-vector of residual errors.

2.2.2 Parameters Estimation

The model parameters are the vector of fixed effects β, the variance-covariance matrix V of the random effects and the variance σ^2 of the residual errors (assuming *i.i.d.* residual errors). Let $\theta = (\beta, \mathrm{V}, \sigma^2)$ be the set of model parameters and $y = (y_i, 1 \le i \le N)$ be the set of observations for the N individuals. The maximum likelihood (ML) estimator of θ maximizes the log-likelihood function defined as:

$$LL(\theta) = log\big(p(y;\theta)\big)$$

$$= \textstyle\sum_{i=1}^{N} log\big(p((y_i;\theta)\big)$$

$$= \textstyle\sum_{i=1}^{N} \Big\{-\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\log\big(\big|A_i\mathrm{V}A_i{}' + \sigma^2 I_{n_i}\big|\big) - \frac{1}{2}(y_i - X_i\beta)'\big(A_i\mathrm{V}A_i{}' + \sigma^2 I_{n_i}\big)^{-1}(y_i - X_i\beta)\Big\}$$

Standard errors (SE) of the parameter estimate $\hat{\theta}$ can be obtained by computing the Fisher information matrix:

$$I(\hat{\theta}) = -E\left(\frac{\partial^2}{\partial\theta\partial\theta'}\right) log\left(p(y;\hat{\theta})\right)$$

The standard errors are the square roots of the diagonal elements of the inverse matrix of $I(\hat{\theta})$.

Once the set of population parameters $\theta = (\beta, \mathrm{V}, \sigma^2)$ has been estimated, the ℓ-vector of nonzero random effects η_i can be estimated. Individual parameters for individual i are the individual coefficients$(c_{ik}, 0 \leq k \leq m)$. Individual parameters can be estimated using the conditional distribution$P(\eta_i |\, y_i; \hat{\theta})$, i.e.

$$P(\eta_i |\, y_i; \theta) = (2\pi)^{-\frac{l}{2}} |\Gamma_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\eta_i - \mu_i)' \Gamma i^{-1} (\eta_i - \mu_i)},$$

where

$$\Gamma_i = \left(\frac{A_i{}' A_i}{\sigma^2} + \mathrm{V}^{-1}\right)^{-1}; \mu_i = \frac{\Gamma_i A_i{}'(y_i - X_i\beta)}{\sigma^2}.$$

The conditional mean μ_i and the conditional variance Γ_i of η_i can therefore be estimated using these formulas and the estimated parameters $\hat{\beta}, \hat{\mathrm{V}}\ and\ \hat{\sigma}^2$:

$$\widehat{\Gamma_i} = \left(\frac{A_i{}' A_i}{\widehat{\sigma}^2} + \widehat{\mathrm{V}}^{-1}\right)^{-1}; \widehat{\mu_i} = \frac{\Gamma_i A_i{}'(y_i - X_i\widehat{\beta})}{\widehat{\sigma}^2}$$

Once the ℓ nonzero random effects have been estimated, estimation of the d individual parameters can be estimated with:

$$\hat{c}_{ik} = \begin{cases} \hat{\beta}_k & if\ \eta_{ik} = 0 \\ \hat{\beta}_k + \hat{\eta}_{ik} & otherwise \end{cases}$$

If the random component $\eta_{ik} = 0$, $\hat{c}_{ik} = \hat{\beta}_k$ is the maximum likelihood estimator of c_{ik} i.e. the parameter value that maximizes the likelihood of making the observations. For any set of explanatory variables $\left(x^{(1)}, v^{(1)}, x^{(2)}, v^{(2)}, \ldots, x^{(m)}, v^{(m)}\right)$, individual prediction of the response variable is then obtained using the individual estimated parameters:

$$\hat{y}_{ij} = \hat{\beta}_{i0} + \sum_{k=1}^{p} \hat{\beta}_k x_{ijk} + \hat{V}_{i0} + \sum_{k=1}^{q} \hat{V}_{ik} \hat{Z}_{ijk} + \hat{e}_{ij} \tag{2.3}$$

Here, we have presented a maximum likelihood estimation approach. However, other estimation methods are possible, and includes the restricted MLE, quasi-and penalized likelihood, and Bayesian inference. More details can be found in (Lu et al. 2014; Hewlig, 2017).

2.3 Transition Models

Diggle, Heagerty, Liang and Zeger (2002) defined a transition model as a generalized linear model for describing the conditional distribution of each response y_{ij} as an explicit function of past responses $y_{ij-1}, \ldots, y_{i1}$, for individual $i, i = 1, \ldots, n_j$ and covariates x_{ij}. It focuses on the case where the observation times t_{ij} are equally spaced. In simpler terms, a transition model is a dynamic regression model in which the present response of a particular subject is regressed on previous responses of that subject as well as on other covariates. Transitional models estimate the probability of changing conditions of an outcome between any two given states during follow-up visits.

2.3.1 Definition

Adopting the definition of the transition model from Diggle, Heagerty, Liang and Zeger (2002), let the history for subject *i* at visit *j* be defined as $H_{ij=} \left\{y_{ik\,,} k = 1, \ldots . j-1\right\}$. Then the transition model specifies a GLM for the conditional distribution of Y_{ij} given past responses H_{ij}

$$f(\, y_{ij} \mid H_{ij}) = \exp\left\{[\, y_{ij}\theta_{ij} - \psi\left(\theta_{ij}\,\right)]\right\} / \phi + C\,\left(Y_{ij}, \phi\right)] \tag{2.4}$$

This can be extended to include covariates x_{ij}, through the mean component (θ_{ij}). For known functions $\psi(\theta_{ij})$ and $C\ \left((Y_{ij}, \phi)\right)$, the conditional mean and variance are

$$\mu_{ij}^{c} = E(y_{ij} \mid H_{ij}) = \psi'\ (\theta_{ij}) \tag{2.5}$$

and

$$V_{ij}^{c} = Var\left((y_{ij} \mid H_{ij})\right) = \psi''\ (\theta_{ij})\ \phi. \tag{2.6}$$

Consider the transition models where the conditional mean and variance satisfy the equations $h(\mu_{ij}^{c}) = x_{ij}'\beta + \sum_{\tau=1}^{\delta} f_r\ (H_{ij}; \alpha)$ for suitable functions $f_r(\cdot)$ and $V_{ij}^{c} = \ V(\mu_{ij}^{c})\ \phi$, where h and v are known links and variance functions determined from the specific form of the density function above.

Put differently, the transition model expresses the conditional mean μ_{ij}^{c} as a function of both the covariates X_{ij} and of the past responses $Y_{ij-1}, \dots, Y_{ij-q}$. Transition models for binary responses are known as Markov Models, for which the conditional distribution of Y_{ij} given H_{ij} depends only on the q prior observations $Y_{ij-1,}$….., Y_{ij-q} , where q is the model order. Markov chains are the most useful transition Models.

The assumptions of the models are as follows:

Firstly, past responses or functions thereof are simply treated as additional explanatory variables and secondly, we assume that the past affects the present through the sum of S terms, each of which may depend on the q prior values.

A number of link functions are possible. The most common is a linear link. This applies for a linear regression with autoregressive errors for Gaussian data (Tsay, 1984), which is a Markov model. It has the form:

$$Y_{ij} \ = \ x_{ij}'\beta + \Sigma_{r=1}^{q}\ \alpha_r(Y_{ij-r} - x_{ij-r}'\beta) + Z_{ij}\ , \tag{2.7}$$

where Z_{ij} are independent, zero – mean, Gaussian processes.

This is a transition model with

$$h(\mu_{ij}^c) = \mu_{ij}^c ,$$

$$V\left(\mu_{ij}^c\right) = 1 \text{ and}$$

$$f_r = \alpha_r \ (y_{ij-r} - x_{ij-r}'\beta)$$

In other words, y_{ij} is a linear function of the earlier deviations $Y_{ij-r} - x_{ij-r}'\beta$ and other covariates X_{ij} with corresponding coefficients matrix β which are lagged, for $r = 1, \ldots, q$.

Another possible link is the logit link. An example is logistic regression model for binary response that comprises first order Markov chain (Cox, 1970; Korn and Whittemore, 1979; Zeger et al., 1985) is

$$\text{logit } P_r \left(y_{ij} = 1 | H_{ij}\right) = x_{ij}'\beta + \alpha\ y_{ij-1} \tag{2.8}$$

Here,

$$h\left(\mu_{ij}^c\right) = \text{logit } \left(\mu_{ij}^c\right) = \log\left(\frac{\mu_{ij}^c}{1-\mu_{ij}^c}\right),$$

$$\nu\left(\mu_{ij}^c\right) = \mu_{ij}^c\left(1 - \mu_{ij}^c\right) \text{ and}$$

$$f_r\left(\mathrm{H}_{ij}, \alpha\right) = \alpha_r\ \ y_{ij-r}$$

A simple extension to the model of order q has the form:

$$\text{logit } P_r\left(y_{ij} = 1 | H_{ij}\right) = x_{ij}'\ \beta_q + \sum_{r=1}^{q} \alpha_r\ y_{ij-r} \tag{2.9}$$

The notation β_q indicates that the value and interpretation of the regression coefficients changes with the Markov order q.

Lastly, the log-link can be used for count data. An example is a log-linear model where y_{ij} given H_{ij} follows a Poisson distribution. Zeger and Qaquish (1988) discussed a first order Markov chain with

$$f_1 = \alpha\{\log(y^*_{ij-1}) - x_{ij-1}{}'\beta\}, \tag{2.10}$$

such that $y^*_{ij} = \max(y_{ij}, d)$, 0<$d$<1, this leads to

$$\mu^c_{ij} = \mathrm{E}\left(y_{ij} \mid \mathrm{H}_{ij}\right) = \exp\left(x_{ij}{}'\beta\right)\left(\frac{y^*_{ij-1}}{exp\left(x'_{ij-1}\beta\right)}\right)^{\alpha} \tag{2.11}$$

where d prevents $y_{ij-1} = 0$ from being an absorbing state, and $y_{ij-1} = 0$ forces all future responses to be 0. When $\alpha > 0$, we have an increased expectation, μ^c_{ij}, when the previous outcome, Y_{ij-1}, exceeds $\exp(x'_{ij-1}\ \beta)$. Similarly, when $\alpha < 0$, a higher value at t_{ij-1} causes a lower value at t_{ij}.

Within the linear regression model, the transition model can be formulated with $f_r = \alpha_r(y_{ij-r} - x_{ij-r}{}'\beta)$ so that $\mathrm{E}(y_{ij}) = x_{ij}{}'\beta$ whatever the value of q. In the logistic and log-linear cases, it is difficult to formulate models in such a way that β has the same meaning for different assumption about the time dependence.

Under a transition model, correlation among $y_{i1,}$ ……,y_{in}, exists because the past value, $y_{i1,}$ ……,y_{ij-1}, explicitly influence the present observation, y_{ij}. The past outcomes are treated as additional predictor variables.

The marginal model with an exponential autocorrelation function can be interpreted as a transition model by writing it as:

$$y_{ij} = x'_{ij}\beta + \epsilon_{ij}$$

where $\epsilon_{ij} = \alpha\epsilon_{ij-1} + Z_{ij}$ and $\alpha = \exp(-\phi)$.

The Z_{ij} are mutually independent, $N(0, \tau^2)$ random variables and $\tau^2 = \sigma^2(1-\sigma^2)$.

By substituting $\epsilon_{ij} = y_{ij} - x'_{ij}\beta$ into the second of these equations and rearranging, we obtain the conditional distribution of y_{ij}, given the preceding response y_{ij-1}, as

$$y_{ij} | \, y_{ij-1} \sim N\{x'_{ij}\,\beta + \alpha\big(y_{ij-1} - x'_{ij-1}\,\beta\big)\tau^2\} \qquad (2.12)$$

This form treats both the explanatory variables and prior response as explicit predictors of the current outcome.

General Transition Model specification

Let $\mathrm{H}_{ij} = \{y_{i1}, \ldots, y_{ij-1}\}$ represent the past responses for the i^{th} subject.

Also let $\mu^c_{ij} = E(Y_{ij}|H_{ij})$ and $v^c_{ij} = Var(Y_{ij}|H_{ij})$ be the conditional mean and variance of Y_{ij} given the past responses and the explanatory variables. Analogous to the GLM for independent data, we assume:

$$h(\mu^c_{ij}) = x'_{ij}\,\beta^{**} + \sum_{r=1}^{s} f_r\big(H_{ij}; \alpha\big) \text{ and } v^c_{ij} = V(\mu^c_{ij})\phi$$

Here we are modelling the transition from the prior state as represented by the functions f_r, to the present response. The past outcomes after transformation by the known functions f_r are treated as additional explanatory variables.

The transition model of order q can be expressed as:

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ijk} + \sum_{r=1}^{q} \alpha_r \left(y_{ij-r} - \left(\sum_{r=1}^{q} \beta_k x_{ijk-r}\right)\right) + \epsilon_{ij} \qquad (2.13)$$

where, y_{ij} is the *j*th observation of the *i*th individual, β_0 is an unknown constant, and $(\beta_k)s$ are unknown regression coefficients, y_{ij-r} and $(x_{1ij-r}, x_{2ij-r}, x_{3ij-r})$ are *r*th lagged response and covariates, respectively and α_r is the corresponding coefficient of the *r*th lag and the residuals are assumed to be normally distributed and mutually independent with mean zero and constant variance, i.e. $\epsilon_{ij} \sim^{iid} N(0, \sigma^2_\epsilon)$. In a transition model with order q, the predicted values depend on q lagged previous observations.

Expressions of a transition model

$$Y_{ij} = \beta_0^{**} + \beta_1^{**} t_{ij} + \epsilon_{ij} \tag{2.14}$$

$$\epsilon_{ij} = \alpha \epsilon_{ij-1} + Z_{ij} \tag{2.15}$$

where Z_{ij} are independent mean-zero innovations with variance σ^2. This model can be re-expressed as

$$Y_{ij} = \beta_0^{**} + \beta_1^{**} t_{ij} + \alpha \left(y_{ij-1} - \beta_0^{**} - \beta_1^{**} t_{ij-1}\right) + Z_{ij} \tag{2.16}$$

Hence

$$E\left(Y_{ij} \middle| y_{ij-1}, \ldots, y_{i1}\right) = \beta_0^{**} + \beta_1^{**} t_{ij} + \alpha \left(y_{ij-1} - \beta_0^{**} - \beta_1^{**} t_{ij-1}\right) \tag{2.17}$$

That is, the expectation for the present response given the past depends explicitly on the previous observation. Equation (2.1) and (2.2) also imply that $E(Y_{ij}) = \beta_0^{**} + \beta_1^{**} t_{ij}$, so this form of transition model has coefficients which also have a marginal interpretation. Econometricians use autoregressive (transition) models for prediction. Some prefer the form

$$Y_{ij} = x_{ij}' \beta^+ + \alpha Y_{ij-1} + Z_{ij} \tag{2.18}$$

Here the response is regressed on the covariates and on the previous outcome itself without adjusting for its expectation. Equation (2.17) implies that $E(Y_{ij}) = \sum_{r=0}^{\infty} \alpha^r x_{ij-1}' \beta^+$, so that β^+ does not have a marginal interpretation as the coefficient for x_{ij}.

2.3.2 Parameter Estimation

The parameter estimation can be carried out by conditional maximum likelihood, assuming normality. The joint distribution of the responses, $y_{i1}, \ldots, y_{in}$, can be written in the form:

$$f\left(y_{i1}, \ldots, y_{in_i}\right) = f\left(y_{in_i} | y_{in_i-1}, \ldots, y_{i1}\right) f\left(y_{in_{i-1}} | y_{in_i-2}, \ldots, y_{i1}\right) \ldots f(y_{i2}|y_{i1}) f(y_{i1}) \tag{2.19}$$

In a first order Markov Model,

$$f\left(y_{ij} | y_{ij-1}, \ldots, y_{i1}; \beta^{**}, \alpha\right) = f\left(y_{ij} | y_{ij-1}; \beta^{**}, \alpha\right) \tag{2.20}$$

So that the likelihood contribution from person i simplifies to

$$f\left(y_{i1}, \dots, y_{in_i}; \beta^{**}, \alpha\right) = f(y_{i1}; \beta^{**}, \alpha) \prod_{j=2}^{n_i} f\left(y_{ij}|y_{ij-1}; \beta^{**}, \alpha\right) \qquad (2.21)$$

As indicated in equation (7), in a first Markov Model, the contribution to the likelihood for the i^{th} subject can be written as

$$L_i\left(y_{i1}, \dots, y_{in_i}\right) = f(y_{i1}) \prod_{j=2}^{n_i} f\left(y_{ij}|H_{ij}\right) \qquad (2.22)$$

In a Markov Model of order q, the conditional distribution of Y_{ij} is

$$f\left(y_{ij}|H_{ij}\right) = f\left(y_{ij}|y_{ij-1}, \dots, y_{ij-q}\right) \qquad (2.23)$$

So that the likelihood contribution for the i^{th} subject becomes

$$f\left(y_{i1}, \dots, y_{iq}\right) \prod_{j=q+1}^{n_i} f\left(y_{ij}|y_{ij-1}, \dots, y_{ij-q}\right) \qquad (2.24)$$

In the linear model we assume that Y_{ij} given H_{ij} follows a Gaussian distribution. If $y_{i1}, \dots, y_{iq}$ are also Multivariate Gaussian and the covariance structure for the y_{ij} is weakly stationary, the marginal distribution $f\left(y_{i1}, \dots, y_{iq}\right)$ can be fully determined from the conditional distribution model without additional unknown parameters. Hence, full maximum likelihood estimation can be used to fit Gaussian autoregressive models.

In the logistic and log-linear cases, $f\left(y_{i1}, \dots, y_{iq}\right)$ is not determined from the GLM assumption about the conditional model, and the full likelihood is unavailable. An alternative is to estimate β and α by maximizing the conditional likelihood

$$\prod_{i=1}^{m} f\left(y_{iq+1}, \dots, y_{in_i}|y_{i1}, \dots, y_{iq}\right) = \prod_{i=1}^{m} \prod_{j=q+1}^{n_i} f\left(y_{ij}|H_{ij}\right) \qquad (2.25)$$

When maximizing equation (2.25), there are two distinct cases to consider.

Case 1:

$f_r\left(H_{ij}; \alpha, \beta\right) = \alpha_r f_r(H_{ij})$ so that $h(\mu_{ij}^c) = x_{ij}'\beta + \sum_{i=1}^{s} \alpha_r f_r(H_{ij})$. Here $h(\mu_{ij}^c)$ is a linear function of both β and $\alpha = (\alpha_1, \dots, \alpha_s)$ so that the estimation proceeds as in GLMs for independent data. We simply regress Y_{ij} on the (p+s) dimensional vector of extended explanatory variables $(x'_{ij}, f_1(H_{ij}), \ldots, f_s(H_{ij}))$.

Case 2:

The second case occurs when the function of past responses include both β and α, For example the linear and log-linear models.

$$S^c(\delta) = \sum_{i=1}^{m} \sum_{j=q+1}^{n_i} \frac{\partial \mu_{ij}^c}{\partial \delta} V_{ij}^{c-1} \left(y_{ij} - \mu_{ij}^c\right) = 0, \qquad (2.26)$$

where $\delta = (\alpha, \beta)$.

This equation is a conditional analogue of a score equation. The derivative $\frac{\partial \mu_{ij}^c}{\partial \delta}$ is analogous to x_{ij} but it can depend on β and α.

The two approaches, mixed effects and transition models will be applied in subsequent chapters, with mixed models applied in Chapter 4 and transition models applied in Chapter 5.

Chapter 3: Research Methodology

3.1 Study Setting

Erongo region, located at the central coast of Namibia, is one of the 14 regions of Namibia. The region was named after Mount Erongo, a famous landmark in Namibia and in this region. The region comprises of four districts namely Swakopmund**,** Walvis Bay, Omaruru and Karibib District with eight towns. Erongo Regional Health Directorate consists of 4 hospitals, 2 health centres as well as 18 Primary Health Care Clinics. These facilities provide free HIV care, which includes testing, counselling and medications. For the purpose of this study four facilities were selected in Erongo region, namely, Kuisebmund Health Center, Hakhaseb Clinic, Omaruru District Hospital and Tamariskia Clinic.

Erongo region was labelled number one in the seven regions identified as high-burden locations for adolescent HIV in Namibia (UNICEF, 2017). According to data provided by USAID (2017), Erongo has a HIV prevalence of 12.5% and Zambezi region ranks the highest with 23.7%.Walvis Bay in Erongo region was the first town in Namibia to make Pre-exposure Prophylaxis (PrEP) available through a community based approach. Namibia has done extremely well in combating HIV/AIDS and the latest survey NAMPHIA (2018) indicated that the regions with the highest burden of the disease also have the highest suppression rates. Namphia survey also indicated that viral load suppression among the 10.6% of HIV-positive residents in the Erongo Region, has reached 75%, surpassing the global target of 73% by 2020.

3.2 Study Design

This study adopted a retrospective cohort analysis using a quantitative method. Furthermore, secondary data on HIV from selected health facilities in Erongo region was used for the purpose of this study.

3.3 Study Population

The study population comprises of patients fifteen years or elder (referred to as adults), who initiated ART treatment at selected health facilities in Erongo region between January 2010 and December 2015, with a total population of 8068 patients.

3.4 Data Description

In this retrospective cohort study, the data was collected from HIV patients who initiated ART treatment at respective health facilities in Erongo region, from January 2010 to December 2015. The data was obtained from Ministry of Health and Social Services (MoHSS) at the Centers for Disease Control and Prevention (CDC). The data was obtained in "raw" form, as extracted from the data system, meaning that most items are defined in the data set as string or character variables. The data contained the following variables: Patient's unique ART number, facility, age at ART Initiation, current age, sex, marital status, clinical stage, beginning of ART, year at ART Initiation, body weight at hospital visit, CD4 cell count at hospital visit, viral load at hospital visit, ART adherence and patients who were still on ART.

A population of 8068 patients was followed up in Erongo region, with a maximum of six observations on each patient, and a total of 25575 observations. Of which 4053 (50.2%) Patients were from Kuisebmund Health Center, 322 (4.0%) patients from Hakhaseb Clinic, 897 (11.1%) patients from Omaruru District Hospital and 2796 (34.7%) patients from Tamariskia Clinic.

3.5 Variables

Variables included in the models are: Patient Identity (Patient_ID), Follow up time (Follow_up_time), Sex, current age (AgeC), Marital status (Marital_Status), Facility, WHO stages (Clinical_Stage), CD4(CD4_cell_count), viral load(Viral_Load), Year at ART initiation (Year_at_ART_Initiation), adherence. The outcomes of interest were Viral Load and CD4. The random effects will be Patient_ID and Facility. Table 1 presents the description of the outcome and selection variables (response variables) and observed confounders (explanatory variables) for health facilities in Erongo region.

Table 1: Erongo region HIV data; description of the outcome and selection variables, and observed confounders

Variables	Definition
Selection	
Log_Viral_Load	Viral load
Log_CD4_cell_count	CD4 cell counts
Outcome	
Viral_Suppression	1 = Virally suppressed (Viral load less than or equals to 40), 2 = Virally unsuppressed
Demographic	
AgeC	Current Age
AgeC_category	1= >30, 2=30-49, 3=50+
Sex	1 = male, 2 = female
Marital_Status	1 = single, 2 = married and 3 = others
Clinical	
Patient_ID	Patient identification number
Follow_up_time	0= ART Initiation time,1 = Follow up time 1, 2 = Follow up time 2, 3 = Follow up time 3, 4 = Follow up time 4, 5 = Follow up time 5
CD4_category	1=CD4 less than 200, 2=CD4 greater than or equals to 200
Facility	1 = Kuisebmund Health Center, 2 = Hakhaseb Clinic, 3 = Omaruru District Hospital, 4 = Tamariskia Clinic
Year_at_ART_Initiation	1 = 2010, 2 = 2011, 3 = 2012. 4 = 2013, 5 = 2014 , 6 = 2015
Clinical_Stage	1 = stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4
Adherence	1 = Less than 95% adherence to ART 2= At least 95% adherence to ART
Weight_Current	Current Weight

3.6 Data Analysis

The data was analysed in three parts, namely, the descriptive analysis, the Linear Mixed Effects model and the Transition Autoregressive Model. Demographic and clinical characteristics of the data were taken into account. In an attempt to predict viral suppression in ART patients using CD4 and Viral Load, a descriptive analysis was used to compare the demographic and clinical characteristics. The patient characteristics were compared by whether their CD4 counts were above 200 or not as well as by whether they achieved viral suppression or not.

3.7 Model selection

It is crucial to compare the models when fitting the data, in order to estimate the best model. It is usually not clear which model one should use, thus visual inspection of data is very important and gives one an overall picture. There are computational methods that allow assessment of how well the model fits the data, namely, the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), they are defined as:

$$\mathrm{AIC} = -2\mathrm{LL} + 2\mathrm{p} \qquad (3.35)$$

$$\mathrm{BIC} = 2\mathrm{LL} + \mathrm{p}\log(\mathrm{N}) \qquad (3.36)$$

Where, $-2\mathrm{LL}$ is twice the negative log-likelihood value for the model and is called the Deviance, which is a measure of sum of squared errors, p denotes the total number of parameters in the model and N is the total number of observations used to fit the model. The smaller the information criterion, the better the model, smaller values of AIC and BIC reflect an overall better fit. When fitting the model, one tries to go from a larger to a smaller information criterion. In a Mixed Effects Regression Analysis there are 3 models to be compared, namely, the Random Intercept, the Random Slope as well as the model with both Random Intercept and Slope. In the

Transition Autoregressive model, models are compared depending on which order of the autoregressive yields the best results.

Likelihood Ratio tests also tests whether one model is significantly better than the other, tests random effect for statistical significance as well as test covariance structure improvement.

Applying the LR test: Obtain the -2LL from both the restricted model and unrestricted model, subtract the latter from the larger former, use a chi-square with degrees of freedom= the difference in the number of parameters and finally, look this up and determine whether or not it is statistically significant.

3.8 Research ethics

This research was performed with the approval of UNAM Research Ethical Committee and the Ministry of Health and Social Services (MoHSS) of the Republic of Namibia. Ethical clearance was obtained from the Research and Publication Office (RPO).

Chapter 4: Application of Mixed Effects Models to Predict Viral Suppression in a cohort of ART patients in Erongo, Namibia

4.1 Introduction

Viral load measures are often subjected to some upper and lower detection limits depending on the quantification assays; hence, the responses are either left or right censored, may be mismeasured or missing at times of interest (Lachos, Bandyopadhyay, & Dey, 2011; Wu, & Zhang, 2018). Consecutive CD4 cell counts on the same subject may also be correlated. To address these incomplete data (censored or missing) problems as well as correlation, statistical models such as mixed effects model that can take these into account are usually be employed since they model the time-varying covariates based on observed covariate data and then use the fitted model to predict the censored, missing or mismeasured covariates (Wu & Zhang, 2018).

However, a model based on the observed data may not fully cater for the censored covariate values, due to possibly different data structures of values below detection limits. For instance, censored viral loads below the detection limit and those above detection limit (observed values) may behave very differently due to possibly different disease status for suppressed viral loads. Furthermore, mixed effects models allows for cluster-specific inference, thus widely used for modelling clustered data when there are large variations between clusters (Wu & Zhang, 2018). Moreover, random effects are introduced in the nonlinear model to account for between-individual variations and within-individual correlations among repeated measurements, yielding the nonlinear mixed effects (Wu & Zhang, 2018).

Zhang and Wu (2010) proposed a unified mixed-effects model, which models population characteristics and individual variations semi-parametrically, so that all existing and recently proposed models are its special cases. They also discussed and generalized the existing and recently proposed models, including uniexponential, multiexponential and biexponential semiparametric models. They employed AIC and BIC to compare these models and found that for the complete viral load data, the recently proposed biexponential semiparametric nonlinear mixed-effects model is the best.

Moreover, Morrell, Pearson, Carter and Brant (1995) used a piecewise nonlinear mixed-effects model to describe longitudinal changes in prostate-specific antigens in men before their prostate cancers were detected clinically. They found that nonlinear mixed-effects model provides a better understanding of PSA changes in men with prostate cancer and recommend mixed-effects models in a variety of research settings where it is necessary to estimate the unknown time of an event.

Furthermore, Fu, Zeng, Tang, Sharma, and Li (2012) used a linear mixed model and dummy variable model methods to construct compatible single-tree biomass equations at different scales. They compared the linear mixed model with random effect parameters, the dummy model with site-specific (local) parameters and population average model. The population average model showed the worst fit and prediction performance. They found that the linear mixed model is more flexible than the dummy variable model when constructing generalized single-tree biomass models or compatible biomass models at different scales. They recommended the linear mixed model to be applied to develop other types of generalized single-tree models such as basal area growth and volume models.

In addition, Yehia et al., (2015) used a multinomial logistic regression to determine the probability of patients' transitioning between retention/suppression categories from 2010 to 2011 and 2010 to 2012, adjusting for age, sex, race/ethnicity, HIV risk factor, insurance status, CD4 count, and use of ART. Using data from a large, U.S. multisite cohort, they followed individual patients over a three-year period to determine how retention in care and viral suppression statuses interact and change over time.

Similarly, Nance et al., (2018) used joint longitudinal models, i.e., survival models and linear mixed models to examine associations between viral suppression in people living with HIV (PLWH) receiving antiretroviral therapy (ART) or continuous viral load (VL) levels and demographic factors, substance use, adherence, and ISTI use.

Mixed effects models have been used widely in HIV/AIDS related studies. Lachos, Bandyopadhyay, and Dey (2011) used linear and nonlinear mixed-effects models to develop a Bayesian framework for censored linear (and nonlinear) models replacing the Gaussian assumptions for the random terms with normal/independent (NI) distributions. Furthermore, Zingoni, Chirwa, Todd, and Musenge (2019) used bayesian mixed effects multistate model to describe the spatial viral suppression and viral rebound trajectories among ART patients. In addition, Bandyopadhyay, Castro, Lachos, and Pinheiro, (2015) addressed the issues of questionable normality assumptions especially in presence of skewness and heavy tails simultaneously under a Bayesian paradigm through joint modeling of the response and covariate processes using an attractive class of skewnormal independent densities. Last but not least, Huang, Dagne, Zhou and Wang (2015) extended joinpoint modelling methods to investigate a piecewise linear mixed-effects (PLME) model with asymmetric distributions for longitudinal studies under a Bayesian framework and provide statistical estimates of the joinpoints and the regression coefficients (slopes). Finally, Wu, Hu and Wu, (2008) considered a nonlinear mixed-effects model for the longitudinal process and the Cox proportional hazards model for the time-to-event process.

The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model, with application to patient cohort in Erongo, Namibia. The previous chapter concentrated on the research methodology. This chapter reflects on the application of mixed effects models, i.e., it gives more light on the definition of the linear mixed effects model, results, discussion and conclusion.

4.2 The Linear Mixed Effects Regression Analysis

A longitudinal study measures the same subject over time. Longitudinal data analysis allows studying changes over time, which in this case is change in viral load over the follow up periods. This section adopts the methodology of Magnusson (2015).

Longitudinal two-level model formulations

The two-level models with repeated measures on individuals in different treatment groups were fitted, to begin with.

Unconditional model

Level 1

$$Y_{ij} = \beta_{0j} + R_{ij} \qquad (4.1)$$

, where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$, β_{0j} is the fixed intercept for j-th predictor and R_{ij} is the error term

Leve1 2

$$\beta_{0j} = \gamma_{00} + U_{0j} \qquad (4.2)$$

, where, U_{0j}~N$(0, \tau_{00}^2)$, γ_{00} is the overall intercept and U_{0j} is the random intercept

Unconditional growth model

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + R_{ij} \qquad (4.3)$$

, where, R_{ij}~N$(0, \sigma^2)$, t_{ij} is the time with intercept-slope covariance

Leve1 2

$$\beta_{0j} = \gamma_{00} + U_{0j} \qquad (4.4)$$

$$\beta_{1j} = \gamma_{10} + U_{1j} \qquad (4.5)$$

, where U_{1j} is the random slope

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} \\ \tau_{01} & \tau_{10}^2 \end{bmatrix}\right)$, and $\tau_{01}$are the covariances

Conditional growth model

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + R_{ij} \qquad (4.6)$$

, where, R_{ij}~N$(0, \sigma^2)$

Leve1 2

$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j}$ (4.7)

, where $\mathrm{T}X_j$ is the treatment allocation

$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j}$ (4.8)

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} \\ \tau_{01} & \tau_{10}^2 \end{bmatrix}\right)$

Conditional growth model: dropping random slope

Level 1

$Y_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + R_{ij}$ (4.9)

, where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j}$ (4.10)

, where, $U_{0j} \sim \mathrm{N}(0, \tau_{00}^2)$

$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j$ (4.11)

Conditional growth model: dropping random intercept

Level 1

$Y_{ij} = \beta_{0j} + \beta_{1j}t + R_{ij}$ (4.12)

, where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$ and t is the time

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j \qquad (4.13)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \qquad (4.14)$$

, where, $U_{1j} \sim \mathrm{N}(0, \tau_{10}^2)$

Conditional growth model: dropping intercept-slope covariance

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t + R_{ij} \qquad (4.15)$$

, where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j} \qquad (4.16)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \qquad (4.17)$$

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & 0 \\ 0 & \tau_{10}^2 \end{bmatrix}\right)$

Three-level models formulations

Here, different three-level models were covered. The form of clustering was clustering due to health-care providers, which are facilities.

Conditional three-level growth model

This is a nested model. This formula expands to a main effect of facilities and an interaction between facilities and patient identification number (which is the subject level effect).

Level 1

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} t_{ijk} + R_{ijk} \tag{4.18}$$

, where, $R_{ijk} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0jk} = \gamma_{00k} + U_{0jk} \tag{4.19}$$

$$\beta_{1jk} = \gamma_{10k} + U_{1jk} \tag{4.20}$$

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N \left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} \\ \tau_{01} & \tau_{10}^2 \end{bmatrix} \right)$

Leve1 3

$$\gamma_{0jk} = \delta_{000} + \delta_{001} \mathrm{T} X_j + V_{0k} \tag{4.21}$$

$$\gamma_{1jk} = \delta_{100} + \delta_{101} \mathrm{T} X_j + V_{1k} \tag{4.22}$$

where, $\begin{pmatrix} V_{0k} \\ V_{1k} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \varphi_{00}^2 & \varphi_{01} \\ \varphi_{01} & \varphi_{10}^2 \end{bmatrix}\right)$

Subject level randomization (Facility crossed effect)

In the conditional three-level growth model, facilities only provided one type of treatment (nested design). Sometimes facilities would be a crossed effect. If it's a randomized trial then this design presents a subject level randomization, whereas in conditional three-level growth model, randomization was at the facility level.

Level 1

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} t_{ijk} + R_{ijk} \qquad (4.23)$$

, where, $R_{ijk} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k} \mathrm{T} X_{jk} + U_{0jk} \qquad (4.24)$$

$$\beta_{1jk} = \gamma_{10k} + \gamma_{11k} \mathrm{T} X_{jk} + U_{1jk} \qquad (4.25)$$

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} \\ \tau_{01} & \tau_{10}^2 \end{bmatrix}\right)$

Leve1 3

$$\gamma_{00k} = \delta_{000} + V_{0k} \qquad (4.26)$$

$$\gamma_{10k} = \delta_{100} + V_{1k} \qquad (4.27)$$

$$\gamma_{01k} = \delta_{010} + V_{2k} \tag{4.28}$$

$$\gamma_{11k} = \delta_{110} + V_{3k} \tag{4.29}$$

where, $\begin{pmatrix} V_{0k} \\ V_{1k} \\ V_{2k} \\ V_{3k} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}, \begin{bmatrix} \varphi_{00}^2 & 0 & 0 & 0 \\ 0 & \varphi_{10}^2 & 0 & 0 \\ 0 & 0 & \varphi_{20}^2 & 0 \\ 0 & 0 & 0 & \varphi_{30}^2 \end{bmatrix}\right)$

Level 3 variance-covariance matrix

Hypothesizing that the Facilities with patients that report worse symptoms at ART initiation have better outcomes (more room for improvement) and allowing separate covariances in each treatment group, the variance-covariance matrix is updated at level 3.

$$\begin{pmatrix} V_{0k} \\ V_{1k} \\ V_{2k} \\ V_{3k} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}, \begin{bmatrix} \varphi_{00}^2 & \varphi_{01} & 0 & 0 \\ \varphi_{01} & \varphi_{10}^2 & 0 & 0 \\ 0 & 0 & \varphi_{20}^2 & \varphi_{23} \\ 0 & 0 & \varphi_{23} & \varphi_{30}^2 \end{bmatrix}\right)$$

One could also estimate all six covariances at level 3. For example, we could look at if facilities that are more successful with treatment A are also more successful with Treatment B, i.e. $cov(V_{2k}, V_{3k}) = \varphi_{23}$, and so forth. The full unstructured level 3 variance-covariance matrix would be:

$$\begin{pmatrix} V_{0k} \\ V_{1k} \\ V_{2k} \\ V_{3k} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}, \begin{bmatrix} \varphi_{00}^2 & \varphi_{01} & \varphi_{02} & \varphi_{03} \\ \varphi_{01} & \varphi_{10}^2 & \varphi_{12} & \varphi_{13} \\ \varphi_{02} & \varphi_{12} & \varphi_{20}^2 & \varphi_{23} \\ \varphi_{03} & \varphi_{13} & \varphi_{23} & \varphi_{30}^2 \end{bmatrix}\right)$$

Partially nested models

Partially nesting occurs when there is nesting in one group but not the other, for example, comparing a treatment group to a wait-list condition. Subjects in the wait-list would not be nested, but subjects in treatment group would be nested within Facilities.

Level 1

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk}t_{ijk} + R_{ij} \text{ , where, } R_{ij} \sim \mathrm{N}(0, \sigma^2) \quad (4.30)$$

Leve1 2

$$\beta_{0jk} = \gamma_{00} + \gamma_{01k}\mathrm{T}X_{jk} + U_{0jk} \quad (4.31)$$

$$\beta_{1jk} = \gamma_{10} + \gamma_{11k}\mathrm{T}X_{jk} + U_{1jk} \quad (4.32)$$

where, $\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \tau_{00}^2 & 0 \\ 0 & \tau_{10}^2 \end{bmatrix}\right)$

Leve1 3

$$\gamma_{01k} = \delta_{010} + V_{0k} \quad (4.33)$$

$$\gamma_{11k} = \delta_{110} + V_{1k} \quad (4.34)$$

where, $\begin{pmatrix} V_{0k} \\ V_{1k} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \varphi_{00}^2 & 0 \\ 0 & \varphi_{10}^2 \end{bmatrix}\right)$

More on level 1 specification

Heteroscedasticity at Level 1

Extending the two levels model and allowing for different level 1 residual variance in the treatment groups yields:

$$\left(R_{ij}|\mathrm{T}X = 0\right) \sim N(0, \sigma_0^2)$$

$$\left(R_{ij}|\mathrm{T}X = 1\right) \sim N(0, \sigma_1^2)$$

Adding another grouping factor such as time, and fitting a model with heteroscedastic level 1 residuals for each time point in each treatment group, For instance, for *i* = 0, 1, 2 yields:

$$\left(R_{ij}|\mathrm{T}X = 0\right) \sim N\left(\begin{matrix}0\\0,\\0\end{matrix}\begin{bmatrix}\sigma_{00}^2 & 0 & 0\\ 0 & \sigma_{01}^2 & 0\\ 0 & 0 & \sigma_{02}^2\end{bmatrix}\right)$$

$$\left(R_{ij}|\mathrm{T}X = 1\right) \sim N\left(\begin{matrix}0\\0,\\0\end{matrix}\begin{bmatrix}\sigma_{10}^2 & 0 & 0\\ 0 & \sigma_{11}^2 & 0\\ 0 & 0 & \sigma_{12}^2\end{bmatrix}\right)$$

4.3 Results

This section presents the results in two parts, namely, the descriptive analysis and the linear mixed effects regression analysis. Demographic and clinical characteristics of the data were taken into account.

4.3.1 Descriptive Analysis

Demographic and clinical characteristics of ART patients aggregated by viral load suppression were presented in Table 2. The results from the study indicated 72% of measured viral loads for females showed viral suppression whilst 69% of viral loads showed suppression for men. The results further indicated that 71% of measured viral loads of the population aged 30-49 years showed viral suppression compared to less than a half (29%) viral loads of the population in the same age group. Equally important young people under the age of 30 were found to have low viral suppression, in comparison to other age groups. There is not much variation between the married people (72%) and single people (70%) in terms of viral load suppression.

Results further indicated that 81% of the patients who were at least 95% adherent to medication were virally suppressed. It is also worth stating that none of the patients who were at most 95% adherent to medication were virally suppressed.

Table 2: Demographic and clinical characteristics by CD4 and viral suppression

Variables	CD4 cell counts						Viral suppression					
	<200		200+		Total		Suppressed		Unsuppressed		Total	
	Number	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**	**Number**	**%**
Facility												
KHC	2724	24.0	8622	76.0	11346	100	9073	79.1	2392	20.9	11465	100
HC	59	7.7	710	92.3	769	100	441	57.9	321	42.1	762	100
ODH	829	18.8	3590	81.2	4419	100	882	52.5	797	47.5	1679	100
TC	1483	17.3	7082	82.7	8565	100	5544	63.8	3143	36.2	8687	100
Sex												
Male	2371	21.4	8720	78.6	11091	100	6808	68.5	3135	31.5	9943	100
Female	2724	19.4	11284	80.6	14008	100	9132	72.2	3518	27.8	12650	100
Age												
< 30 years	163	19.3	680	80.7	843	100	456	61.9	281	38.1	737	100
30-49 years	4215	20.4	16488	79.6	20703	100	13494	71.5	5384	28.5	18878	100
≥ 50 years	717	20.2	2836	79.8	3553	100	1990	66.8	988	33.2	2978	100
Marital Status												
Single	4386	20.5	17040	79.5	21426	100	13320	69.9	5743	30.1	19063	100
Married	337	22.6	1153	77.4	1490	100	1028	71.6	407	28.4	1435	100
Others	20	22.2	70	77.8	90	100	61	70.1	26	29.9	87	100
Clinical stage												
Stage-I	4760	20.0	19059	80.0	23819	100	15587	71.7	6142	28.3	21729	100
Stage-II	158	27.0	428	73.0	586	100	152	38.4	244	61.6	396	100
Stage-III	129	26.7	355	73.3	484	100	129	41.3	183	58.7	312	100
Stage-IV	38	29.5	91	70.5	129	100	38	36.2	67	63.8	105	100
Year at ART initiation												
2010	969	21.6	3513	28.9	4482	100	2450	60.9	1574	39.1	4024	100
2011	1322	20.0	5278	80.0	6600	100	3972	66.2	2029	33.8	6001	100
2012	871	21.3	3225	78.7	4096	100	2461	66.7	1228	33.3	3689	100
2013	796	21.8	2850	78.2	3646	100	2369	72.2	911	27.8	3280	100
2014	612	16.7	3054	83.3	3666	100	2761	83.2	557	16.8	3318	100
2015	525	20.1	2084	79.9	2609	100	1927	84.5	354	15.5	2281	100
Follow up time												
Initial	1752	22.2	6154	77.8	7906	100	3944	51.2	3764	48.8	7708	100
Time 1	1094	17.2	5275	82.8	6369	100	4494	74.4	1546	25.6	6040	100
Time 2	942	19.7	3840	80.3	4782	100	3681	84.1	698	15.9	4379	100
Time 3	707	22.4	2454	77.6	3161	100	2330	86.8	355	13.2	2685	100
Time 4	379	21.3	1402	78.7	1781	100	1064	85.5	180	14.5	1244	100
Time 5	221	20.1	879	79.9	1100	100	427	79.5	110	20.5	537	100
Adherence												
≥ 95%	3694	19.1	15642	80.9	19336	100	15940	81.2	3691	18.8	19631	100
<95%	869	29.7	2058	70.3	2927	100	0	0.0	2962	100	2962	100

Note: KHC-Kuisemund Health Centre, HC-Hakhaseb Clinic, ODH- Omaruru District Hospital and TC-Tamariskia Clinic

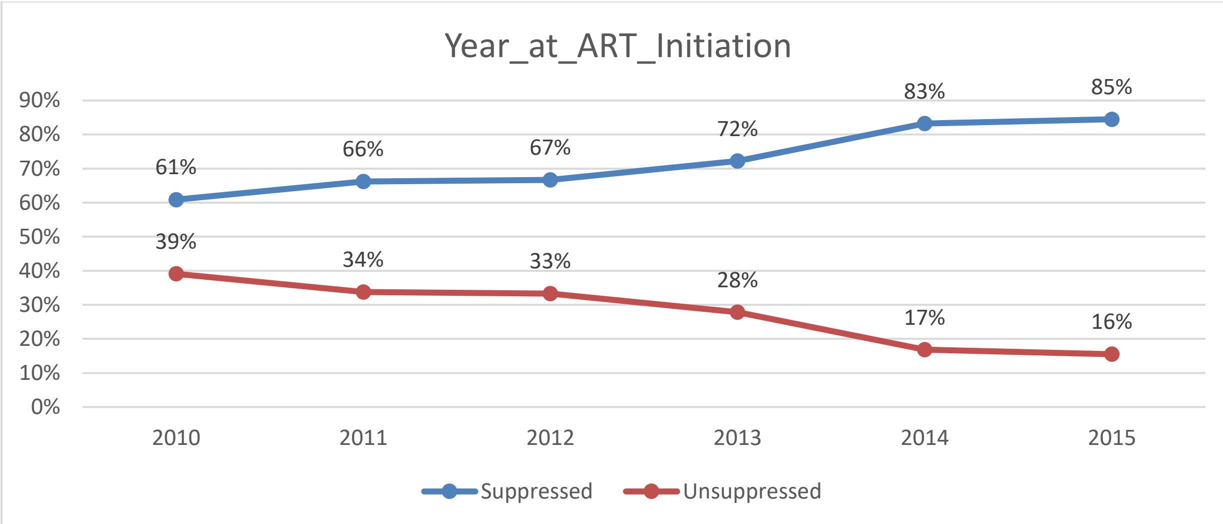


Figure 1: Trend analysis of viral suppression between 2010 and 2015

A trend analysis was conducted between the year 2010 and 2015. Results show that since 2010 the percentage of viral suppression have been increasing with a notable rise observed from72% 2013 to 85% in 2015. While at the same time unsuppression rate also declined from 28% in 2013 to 16% in 2015. This is a sign that clients are tested for HIV at the earlier stage after infection, while the virus did not multiply gradually in the blood.

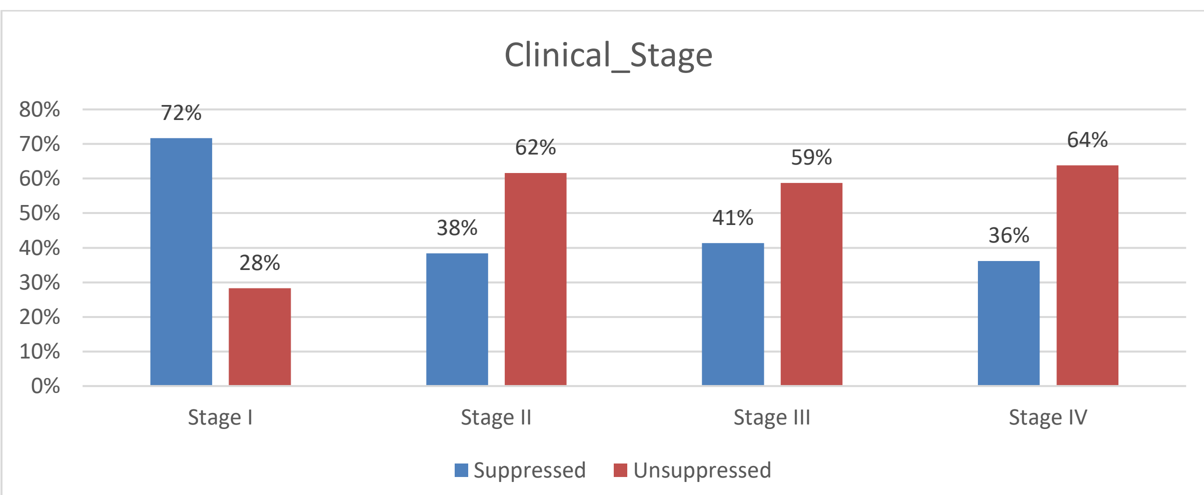


Figure 2: Viral Load Suppression plotted against Clinical Stage

The WHO system for adults sorts patients into one of four categorized WHO clinical stages ranging from stage I (asymptomatic) to stage IV (AIDS). Patients are allocated to a particular stage when they show at least one clinical condition in that stage's criteria. Figure 2 shows that viral suppression is higher i.e 72% at WHO clinical stage I, as these patients are showing no symptoms yet and their CD4 cell counts are still higher, making it easier to fight the virus. Viral suppression is lower as from stage II to IV as these patients now show symptoms and their CD4 cell counts may be lower, making it difficult to fight the virus.

4.3.2 Linear Mixed Effects Model

This section presents the results of the Mixed Effects model of the factors associated with viral suppression, using CD4 and Viral Load as response variables. Tables 3 and 4 shows the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for 3 different Mixed Effects Models using Viral Load and CD4 cell counts as a response variables. The models fitted were: Random Intercept, Random Slope as well as the mixture of Random Intercept and Slope. The Mixed Effects Model with lowest AIC and BIC values and highest log likelihood, in which viral load and CD4 were used as outcome variables points out the random slope model had the lowest AIC and BIC values as well highest log likelihood for the used outcome variables (Viral Load and CD4 counts).

Table 3: Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for the Mixed Effects Models using Viral Load

Model	**Log Likelihood**	**Likelihood Ratio Test**	**AIC**	**BIC**
Random Intercept	-9510.8	-	19048	19144
Random Intercept and Slope	-9502.6	0.0009665(***)	19037	19155
Random slope	-9355.0	< 2.2e-16 (***)	18744	18870

Note: This table presents the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criteria (AIC and BIC) from the Mixed Effects Model using Viral Load as a response variable.

Table 4: Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for the Mixed Effects Models using CD4

Model	**Log Likelihood**	**Likelihood Ratio Test**	**AIC**	**BIC**
Random Intercept	-21554	-	43134	43237
Random Intercept and Slope	-21546	0.0006462(***)	43123	43250
Random slope	-21504	< 2.2e-16 (***)	43043	43177

Note: This table presents the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criteria (AIC and BIC) from the Mixed Effects Model using CD4 as a response variable.

Table 5 presents the parameter estimates of the Random slope Mixed Effects model which was shown to be the best model. This model used viral load as a response variable. The results indicated that viral load had a significant association with follow up time, adherence, weight, age, clinical stage and adherence at different follow up time points. More specifically, follow up time, weight, adherence and adherence at different follow up time points had negative significant associations, while, age and clinical stage had positive associations. It is also worth mentioning that sex, marital status and weight at different follow up time points were not associated with viral load. The fixed effects appear to be weak, but CIs are narrow. On average the Random effect is strong, but there is considerable variation between subjects.

Table 5: Summary of Fixed Effects (with Viral Load as a response)

Fixed Effects	Unstandardized Coefficients					95% CI for estimate	
	Estimate	**Standard error**	**t-value**	**F-value**	**P-value**	**LB**	**UB**
Intercept	0.052	0.071	0.729	-	-	-0.087	0.190
Follow_up_time (0= ART Initiation time,1 = Follow up time 1, 2 = Follow up time 2, 3 = Follow up time 3, 4 = Follow up time 4, 5 = Follow up time 5)	-0.115	0.025	-4.605	2955.567	<0.001(***)	-0.164	-0.066
Sex (1 = male, 2 = female)	-0.008	0.014	-0.558	12.934	0.577	-0.034	0.019
Marital_Status (1 = single, 2 = married and 3 = others)	-0.023	0.023	-1.002	19.990	0.316	-0.068	0.022
Adherence (1 = Less than 95% adherence to ART 2= At least 95% adherence to ART)	-1.945	0.016	125.196	21169.235	<0.001(***)	1.914	1.975
Weight_Current	-0.003	0.001	-6.033	54.481	<0.001(***)	-0.004	-0.002
AgeC	0.002	0.001	2.252	7.515	0.024(*)	0.0002	0.003
Clinical_Stage (1 = stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4)	0.090	0.013	7.018	50.221	<0.001(***)	0.065	0.116
Follow_up_time:adherence	-0.037	0.010	-3.635	13.538	<0.001(***)	-0.058	-0.017
Follow_up_time:Weight_Current	0.0001	0.0003	0.375	0.141	0.708	-0.0005	0.0007

Note: (.) significance at 10%, (*) significance at 5%, (**) significance at 1%, (***) significance at 0.1%

Table 6: Random Effects (with Viral Load as a response)

Group	Name	Variance	Standard Deviation	ICC
Patient_ID (Intercept)		1.904e-01	0.436336	
	Follow_up_time	1.581e-02	0.125751	-0.67
Facility (Intercept)		4.299e-03	0.065566	
	Follow_up_time	6.445e-05	0.008028	1
Residual		1.665e-01	0.408070	

Table 7 presents the parameter estimates of the Random Slope Mixed Effects model with CD4 cell counts as the response variable. The results indicated that CD4 cell counts had a significant association with follow up time, sex, adherence, clinical stage and adherence at different follow up time points. More specifically, follow up time and clinical stage had negative significant associations, while sex, adherence and adherence at different follow up time points had positive associations. It is also worth mentioning that marital status, weight, age and weight at different follow up time points were not associated with CD4 cell counts. The fixed effects appear to be weak, but CIs are narrow. On average the Random effect is strong, but there is considerable variation between subjects.

Table 7: Summary of Fixed Effects (CD4 cell counts as response)

Fixed Effects	Unstandardized Coefficients					95% CI for estimate	
	Estimate	Standard error	t-value	F-value	P-value	LB	UB
Intercept	6.242	0.087	71.922	-	-	6.072	6.412
Follow_up_time (0= ART Initiation time,1 = Follow up time 1, 2 = Follow up time 2, 3 = Follow up time 3, 4 = Follow up time 4, 5 = Follow up time 5)	-0.158	0.028	-5.743	9.662	<0.001(***)	-0.212	-0.104
Sex (1 = male, 2 = female)	0.030	0.011	2.699	17.076	0.007(**)	0.008	0.052
Marital_Status (1 = single, 2 = married and 3 = others)	-0.019	0.019	-1.012	0.563	0.312	-0.056	0.018
Adherence (1 = Less than 95% adherence to ART 2= At least 95% adherence to ART)	0.315	0.019	-16.887	227.630	<0.001(***)	-0.352	-0.278
Weight_Current	0.001	0.001	1.725	3.286	0.092	-0.0001	0.002
AgeC	-0.001	0.001	-1.165	3.089	0.244	-0.002	0.0005
Clinical_Stage (1 =	-0.078	0.016	-4.978	26.729	<0.001(***)	-0.108	-0.047

stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4)							
Follow_up_time:adherence	0.121	0.014	8.605	75.392	<0.001(***)	0.093	0.148
Follow_up_time:Weight_Current	-0.0002	0.0003	-0.745	0.554	0.457	-0.0008	0.0004

Note: (.) significance at 10%, (*) significance at 5%, (**) significance at 1%, (***) significance at 0.1%

Table 8: Random Effects (with CD4 as a response)

Group	**Name**	**Variance**	**Standard Deviation**	**ICC**
Patient_ID				
(Intercept)		0.0404899	0.20122	
	Follow_up_time	0.0136758	0.11694	-0.68
Facility				
(Intercept)		0.0162429	0.12745	
	Follow_up_time	0.0002793	0.01671	1.00
Residual		0.4576075	0.67647	

4.4 Discussion

The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model. To be specific, to investigate the prevalence of viral suppression among ART patients, to identify the factors associated with viral suppression among ART patients as well as to inspect the predictive performance of the transition and the mixed effects model.

The findings of this study points to a viral suppression prevalence rate of 71% in Erongo region. 71% of all HIV-positive adults on ART in Erongo have achieved viral suppression, which is just 2% less than the UNAIDS (2014) target of 73% by 2020. Of the measured viral loads that showed viral suppression, 72% were females, 71.5% were patients aged 30 to 39 years, 72% were at clinical stage I, 73% had CD4 above 200 copies per millilitre and 81% were at least 95% adherent to medication.

The Random slope Mixed Effects model results indicated that viral load had a significant association with follow up time, adherence, weight, age, clinical stage and adherence at different follow up time points. More specifically, as the value of follow up time, weight, adherence and adherence at different follow up time points increases, the mean viral load decreases significantly. While as the value of age and clinical stage increases, the mean of viral load also tends to increase significantly. The results concurred with the findings of Jobanputra et al. (2015) who pointed out that Children, adolescents and those with advanced disease were most likely to have high viral loads and least likely to achieve viral suppression, while also differing slightly with Cescon et al.(2011) who found that older age, male sex and having an AIDS diagnosis at baseline predicted increased likelihood of suppression. Adherence to ART medication has proven to be the main determinant of viral load; the mean of viral load reduces by 1.945 given a one-unit increase in adherence while holding other variables in the model constant. It is also worth mentioning that sex, marital status and weight at different follow up time points were not significantly associated with viral load.

The Random slope Mixed Effects model results further indicated that CD4 counts had a significant association with follow up time, sex, adherence, clinical stage and adherence at different follow up time points. More specifically, as the value of follow up time and clinical stage increases, the mean CD4 count decreases significantly, while, as the value of sex, adherence and adherence at different follow up time points increases, the mean CD4 count also tends to increase significantly. It is also worth mentioning that marital status, weight, age and weight at different follow up time points were not significantly associated with CD4 cell counts.

4.5 Conclusions

This chapter aimed to predict viral suppression, using CD4 cell counts and viral load, in antiretroviral therapy (ART) patients in Erongo region using the Mixed Effects models. The findings of this study points to a viral suppression prevalence rate of 71% in Erongo region. The findings provide additional support to the concept that demographic (sex, age and weight) and clinical (follow up time, adherence and clinical stage) characteristics are the determinants of viral suppression. Adherence to ART medication has proven to be the main determinant of viral suppression. Marital status has proven not to be a determinant of either viral suppression or increase in CD4 cells count.

Chapter 5: Application of Transition Models to Viral Load

5.1 Introduction

A time series model in the form of autoregressive (AR) model, which assumed that the future values of a variable depend on its historical values, was first introduced by Yule in 1926 (Sen & Baharumshah, 2002). Time series analysis has since then been noticed as a powerful forecasting tool. With new developments in Methodology, the concept of non-linearity was introduced into the analysis of time series. Furthermore, Terasvirta (1994) suggested the Smooth Transition Autoregressive (STAR) model which includes the Logistic STAR (LSTAR) models and Exponential STAR (ESTAR) models as a modelling tool in Econometrics.

STAR offers an alternative to the modelling of time series variables that exhibit non-linearities. In review of the literature, it was found that the application of the STAR model in practice is still very limited, and its forecasting performance particularly with reference to viral suppression has yet to be determined.

The nonlinear STAR models have been advocated for their power to model state-dependent features of finance and economics. Zhou (2010) used the STAR model to model the industrial production index of Sweden. He carried out the linearity test against the STAR model, determined the delay parameter and made a choice between the LSTAR model and the ESTAR model. His finding was that the STAR model is better performing than the linear autoregressive model.

Moreover, Sen and Baharumshah (2002) compared the Performances of Non-linear Smooth Transition Autoregressive and Linear Autoregressive Models in forecasting the Ringgit-Yen rates. They found that the STAR model outperforms the AR model, its linear competitor.

However, Umer, Sevil, and Sevil (2018) assessed the performance of STAR model on travel and leisure stock index. They compared the performance of STAR and linear autoregressive (AR) models using monthly returns of Turkey and FTSE travel and leisure index from April 1997 to August 2016 and found that nonlinear LSTAR model cannot improve the out-of-sample forecast of linear AR model and therefore concluded that there is little to be gained from using LSTAR model in the prediction of travel and leisure stock index.

In health literature, Nasserinejad et al. (2013) predicted haemoglobin levels in whole blood donors using the transition models and the mixed effects models. They predicted the future haemoglobin level after a visit to the blood bank and concluded that transition models and the mixed effects models provide a much better prediction compared to a multiple linear regression model.

Markov chain models have been used widely to study the progression of diseases such as HIV infection. Chakraborty, Hossain, and Latif (2019) used three-state continuous time Markov chain model to compare the transition hazards, transition probabilities, and the state-specific duration of stay for different subgroups of HIV-infected individuals by gender, race/ethnicity, age, CD4 count, place of residence, and antiretroviral treatment regimen prescribed at the beginning of the study period using South Carolina state wide HIV surveillance data. In addition, Shoko and Chikobvu (2019) used a stochastic Markov approach to multistate modelling to investigate and compare the use of either CD4 cell count or viral load markers in controlled trials. Their aim was to determine whether CD4 cell count or viral load count can be used to model HIV/AIDS progression. Furthermore, Shoko, Chikobvu & Bessong (2018) used a continuous-time-homogeneous Markov process to model the progression of HIV/AIDS patients that is a Markov model to estimate mortality due to HIV/AIDS using viral load levels-based states and cd4 cell counts using principal component analysis approach.

Moreover, Bradley, Mattson, Beer, Huang, & Shouse (2016) used trend analysis to assess trends during 2009–2013 in antiretroviral therapy (ART) prescription and viral suppression among adults receiving HIV clinical care in the United States. Last but not least, Conway, Perelson & Li (2019) used a stochastic mathematical model to gain insight into post-treatment dynamics, specifically characterizing the dynamics of short term viral rebounds (< 60 days).

The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model, with application to patient cohort in Erongo, Namibia. The previous chapter concentrated on the application of mixed effects Models. This chapter focuses on the application of transition models, i.e., it gives more light on the definition of the transition autoregressive model, results, discussion and conclusion.

5.2 The Transition Model

A longitudinal study measures the same subject over time. Longitudinal data analysis allows studying changes over time, which in this case is change in viral load over the follow up periods. This section adopts the methodology of Magnusson (2015).

The first-order Autoregressive AR(1) residuals

For $T = 1, 2, 3 \dots, N_1$ time points, The level 1 variance-covariance matrix is as follows;

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{\mathrm{T}-1} \\ \rho & 1 & \rho & \cdots & \rho^{\mathrm{T}-2} \\ \rho^2 & \rho & 1 & \ddots & \rho^{\mathrm{T}-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{\mathrm{T}-1} & \rho^{\mathrm{T}-2} & \rho^{\mathrm{T}-3} & \cdots & 1 \end{pmatrix}$$

This is a level 1 residual structure and the error term $R_{ij} \sim \mathrm{N}(0, \Sigma)$.

Heterogenous AR(1)

We can also extend the level 1 variance-covariance matrix from above, to allow for different residuals at each time point.

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho & \sigma_0\sigma_2\rho^2 & \cdots & \sigma_0\sigma_i\rho^{\mathrm{T}-1} \\ \sigma_1\sigma_0\rho & \sigma_1^2 & \sigma_1\sigma_2\rho & \cdots & \sigma_1\sigma_i\rho^{\mathrm{T}-2} \\ \sigma_2\sigma_0\rho^2 & \sigma_2\sigma_1\rho & \sigma_2^2 & \ddots & \sigma_2\sigma_i\rho^{\mathrm{T}-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_{\mathrm{T}}\sigma_0\rho^{\mathrm{T}-1} & \sigma_{\mathrm{T}}\sigma_1\rho^{\mathrm{T}-2} & \sigma_{\mathrm{T}}\sigma_2\rho^{\mathrm{T}-3} & \cdots & \sigma_{\mathrm{T}}^2 \end{pmatrix}$$

Changing the functional form of time (Modelling Change and Event Occurrence)

All of the above methods assume a linear change. This section covers modelling methods for nonlinear change at level 1. This is based on the two-level model, but could also be easily combined with the three-level models outlined above as well as any other multi-level model.

In this section; β_{0j} is the fixed intercept for j-th predictor, R_{ij} is the error term, γ_{00} is the overall model intercept, U_{0j} is the random intercept and U_{1j} , U_{2j} and U_{3j} are the random slope, t_{ij} is the time with intercept-slope covariance and t is the time, $\mathrm{T}X_j$ is the treatment allocation and $\tau_{01}$are the covariances.

Quadratic trend

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{1ij} + \beta_{2j}t_{1ij}^2 + R_{ij} \qquad (5.1)$$

, where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j} \qquad (5.2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \qquad (5.3)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\mathrm{T}X_j + U_{2j} \qquad (5.4)$$

where, $$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11}^2 & \tau_{12} \\ \tau_{20} & \tau_{12} & \tau_{22}^2 \end{bmatrix}\right).$$

Piecewise growth curve

Segmenting the time trend into different pieces has got more to do with simple dummy coding of regression variables, than any specifics of lme or lmer. To fit a piecewise growth model, simply replace time with two dummy variables time1 and time2, which represent the different time periods. A common scenario is that the first piece represents the acute treatment phase, and piece 2 represents the follow-up phase.

Coding scheme 1: separate slopes

Time	0	1	2	3	4	5
Time 1	0	1	2	2	2	2
Time 2	0	0	0	1	2	3

Coding scheme 2: incremental/decremental slope

Time	0	1	2	3	4	5
Time 1	0	1	2	3	4	5
Time 2	0	0	0	1	2	3

These two coding schemes only differ in the interpretation of the regression coefficients. In scheme 1 the two slope coefficients represent the actual slope in the respective time period. Whereas in scheme 2 the coefficient for time 2 represents the deviation from the slope in period 1, i.e. if the estimate is 0 then the rate of change is the same in both periods.

The model is specified as:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j} t_{1ij} + \beta_{2j} t_{2ij} + R_{ij} \qquad (5.5)$$

where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \mathrm{T} X_j + U_{0j} \qquad (5.6)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \tag{5.7}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\mathrm{T}X_j + U_{2j} \tag{5.8}$$

where $\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} & \tau_{02} \\ \tau_{01} & \tau_{10}^2 & \tau_{12} \\ \tau_{02} & \tau_{12} & \tau_{20}^2 \end{bmatrix}\right)$

Dropping the correlation between time piece 1 and 2

Sometimes you might want to fit a model with a correlation between the random intercept and time piece 1, but no correlation between time piece 2 and the other effects. This would change the level 2 variance-covariance from above to:

$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} & 0 \\ \tau_{10} & \tau_{11}^2 & 0 \\ 0 & 0 & \tau_{22}^2 \end{bmatrix}\right)$$

Adding a quadratic effect

Extending the two-part piecewise growth model to allow for non-linear change during one or both of the pieces, one could extend the model to allow for quadratic change during piece 1 and the model would look as follows:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{1ij} + \beta_{2j}t_{1ij}^2 + \beta_{3j}t_{2ij} + R_{ij} \tag{5.9}$$

where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j} \tag{5.10}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \tag{5.11}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\mathrm{T}X_j + U_{2j} \tag{5.12}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}\mathrm{T}X_j + U_{3j} \tag{5.13}$$

where,

$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ U_{2j} \\ U_{3j} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} , \begin{bmatrix} \tau_{00}^2 & \tau_{01} & \tau_{02} & \tau_{03} \\ \tau_{10} & \tau_{11}^2 & \tau_{12} & \tau_{13} \\ \tau_{20} & \tau_{21} & \tau_{22}^2 & \tau_{23} \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_{33}^2 \end{bmatrix} \right)$$

and $R_{ij} \sim N(0, \sigma^2)$.

The high order would then be:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}t_{1ij} + \beta_{2j}t_{1ij}^2 + \beta_{3j}t_{2ij} + \cdots + \beta_{(n-1)j}t_{(n-2)ij}^q + \beta_{nj}t_{(n-1)ij} + R_{ij} \tag{5.14}$$

Where, $R_{ij} \sim \mathrm{N}(0, \sigma^2)$

Leve1 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathrm{T}X_j + U_{0j} \tag{5.15}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\mathrm{T}X_j + U_{1j} \qquad (5.16)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\mathrm{T}X_j + U_{2j} \qquad (5.17)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}\mathrm{T}X_j + U_{3j} \qquad (5.18)$$

$$\vdots$$

$$\beta_{nj} = \gamma_{n0} + \gamma_{n1}\mathrm{T}X_j + U_{nj} \qquad (5.19)$$

where,

$$\begin{pmatrix} U_{0j} \\ U_{1j} \\ \vdots \\ U_{nj} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00}^2 & \tau_{01} & \dots & \tau_{0j} \\ \tau_{10} & \tau_{11}^2 & \dots & \tau_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n0} & \tau_{n1} & \cdots & \tau_{nn}^2 \end{bmatrix} \right)$$

and $R_{ij} \sim N(0, \sigma^2)$.

Hypothesis tests

The lmer does not report *p*-values or degrees of freedom; however, there are other packages that calculate *p*-values, such as the Wald test, Likelihood ratio test, profile confidence intervals, parametric bootstrap, Kenward-Roger degrees of freedom approximation as well as the Shattertwaite degrees of freedom approximation.

5.3 Results

This section presents the results in two parts, namely, the descriptive analysis and the linear mixed effects regression analysis.

5.3.1 Descriptive Analysis

A population of 8068 patients was studied, of which 50.2% of patients were from Kuisebmund Health Center, 4.0% from Hakhaseb Clinic, 11.1% from omaruru District Hospital and 34.7% from Tamariskia Clinic, all in Erongo region. Demographic and clinical characteristics were described in table 10. 4553 (56.4%) were female, 7534 (93.4%) were single, 7603 (94.2%) were at baseline clinical stage-I, 6736 (83.5%) were adults aged 30-49, 1977 (24.5%) initiated ART in 2011and 7011 (86.9%) were at least 95% adherent to medication. The median and interquartile (IQR) of current body weight and the baseline CD4 cell count were 64 (56 – 74.2), and 339 (221 – 480), respectively.

Table 9: Demographic and clinical characteristics

Characteristics	Subjects (N = 8068(100%))
Health Facility	
Kuisebmund Health Center	4053 (50.2%)
Hakhaseb Clinic	322 (4.0%)
Omaruru District Hospital	897 (11.1%)
Tamariskia Clinic	2796 (34.7%)
Sex	
Male	3515 (43.6%)
Female	4553 (56.4%)
Marital Status	
Single	7534 (93.4%)
Married	492 (6.1%)
Others	42 (0.5%)
Clinical Stage	
Stage I	7603 (94.2%)
Stage II	205 (2.5%)
Stage III	193 (2.4%)
Stage IV	67 (0.8%)
Current age category	
<30	292 (3.6%)
30-49	6736 (83.5%)
50+	1040 (12.9%)
Year at ART Initiation	
2010	1296 (16.1%)
2011	1977 (24.5%)
2012	1216 (15.1%)
2013	1090 (13.5%)
2014	1350 (16.7%)
2015	1139 (14.1%)
Adherence	
≥ 95%	7011 (86.9%)
<95%	1057 (13.1%)

Note: This table presents the demographic and clinical characteristics of 8068 patients who initiated ART during the period 2010-2015 in Erongo region.

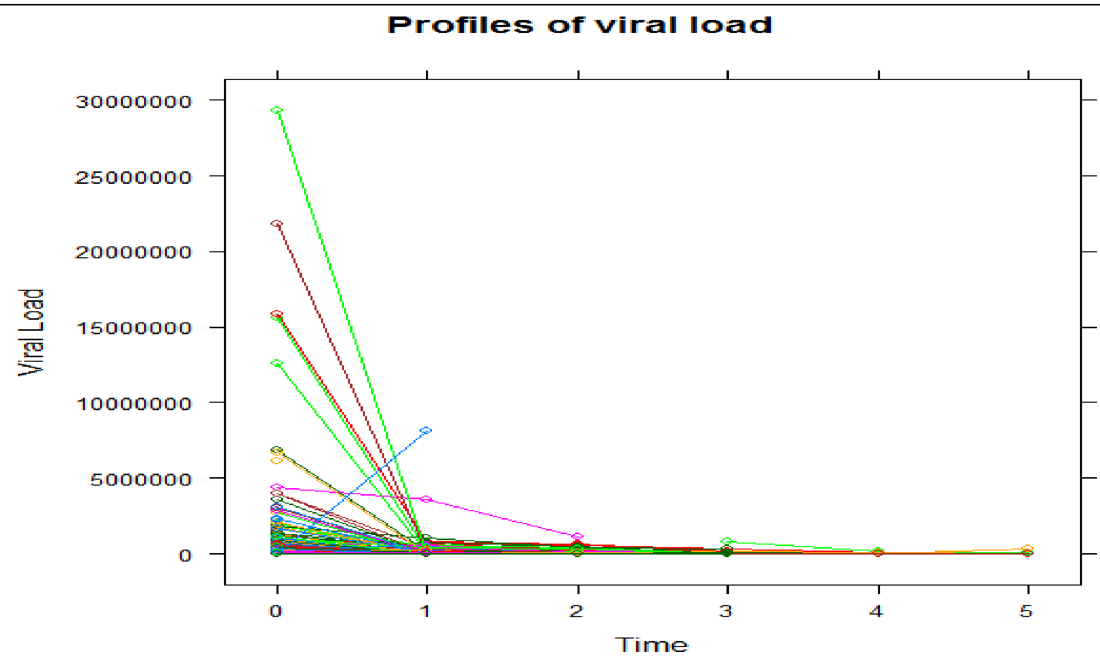
Table 10: Correlation of the number of visits (Follow up time)

	time0	**time1**	**time2**	**time3**	**time4**	**time5**
time0	**1**	**0.265**	**0.194**	**0.18**	**0.17**	**0.011**
time1	**0.265**	**1**	**0.889**	**0.68**	**0.56**	**0.077**
time2	**0.194**	**0.889**	**1**	**0.74**	**0.62**	**0.075**
time3	**0.176**	**0.676**	**0.74**	**1**	**0.85**	**0.125**
time4	**0.167**	**0.558**	**0.622**	**0.85**	**1**	**0.195**
time5	**0.011**	**0.077**	**0.075**	**0.12**	**0.19**	**1**

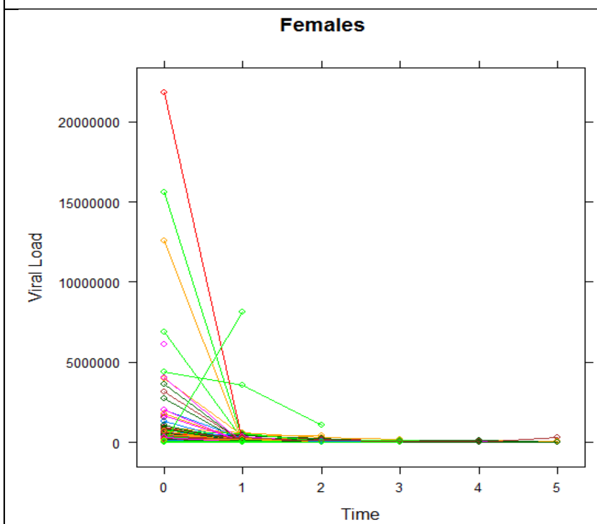
The correlation matrix above shows that there is a very strong positive correlation between ART visiting time 2 and time one, with r=0.889, another strong positive correlation is observed between ART visiting follow-up 4 and follow-up 3. In addition, the ART initiation visit time have the least strength in association with follow-up time 5.

Figure 3**:** Profile of viral load for successive ART visits to the health facilities of a population of male and female patients.

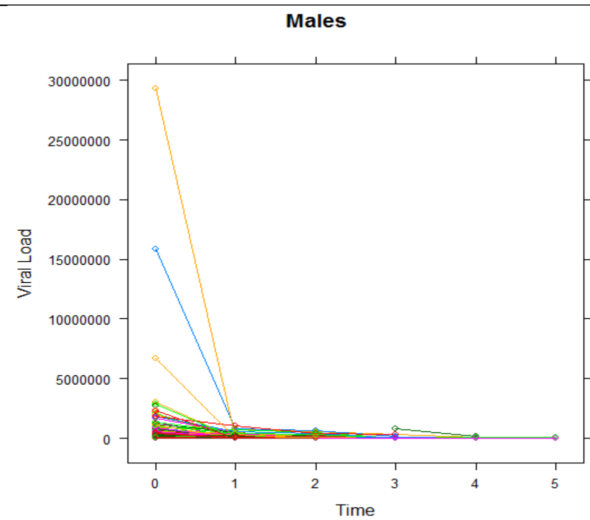
Individual profile plot: Viral load by ART visits



Female Viral load by ART visits

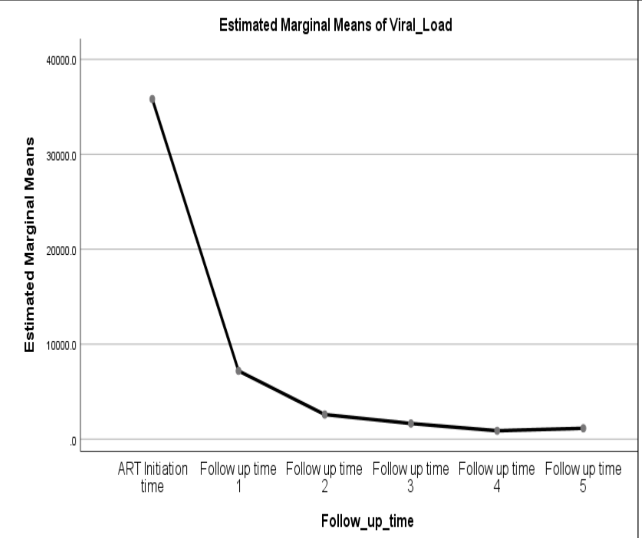
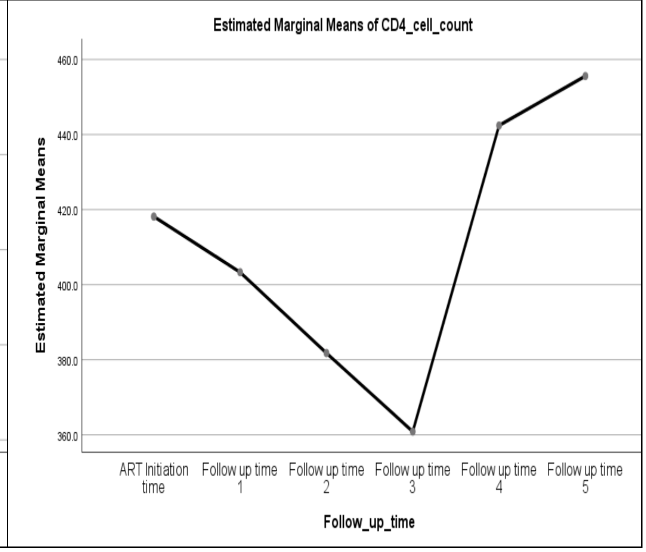


Male Viral load by ART visits



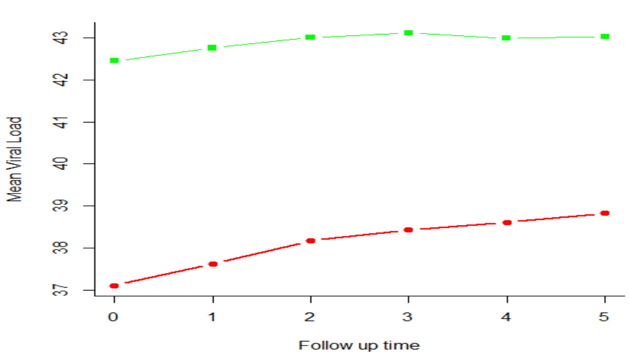
It is depicted in the profile above that, at ART initiation period, most patient have irregular viral load below 5000 000 copies/ml. It shows that a high number of patients' viral loads gradually drop to less than 1000000 copies/ml between the initiation time and first follow up. The viral loads stay at this stage as from this first visit through visit 5 while patient is on ART. It is also depicted that female viral load is between 0 and 2000 000 copies/ml at the initiation time, the viral load reduces gradually at first visit to less than 1000000 copies/ml. Nonetheless, there are female patients that experience an increasing VL at their first visit after initiation; this may be attributed to adherence issues and resistance of medication by the immune system. It is also showing that some female patients do not get virally suppressed until their second follow-up after initiation. In addition, most male patients' viral loads drop to less than 1000000 copies/ml at first visit after initiation. It is clearly seen in the chart that some male participants have as high viral load as 3000 000 copies/ml, at the initiation stage, however their viral load drops and it keeps dropping through their visit 5.

Figure 4**:** Estimated marginal mean viral load and CD4 cell counts for successive ART visits to the health facilities of a population of male and female patients.

<table>
<tr><th>Estimated marginal means of Viral Load</th><th>Estimated marginal means of CD4 cell counts</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

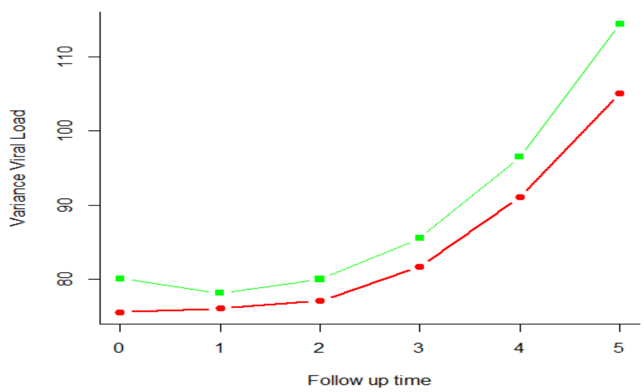
This chart shows that viral load average decreases with follow-up time, while CD4 on average decreases with constant trend from the initiation period through visit three and then start rising rapidly in the fourth follow up as most of the patients may be virally suppressed at this point.

Figure 5: Sex Average plot



Results on sex and viral load, the plot show that female (green) had a high viral load from the onset of ART initiation and remained high through the follow up period ranging between 40% and 50%. On the other had male had low viral load suppression from the onset, however it picked up with progressing follow ups.

Figure 6: Gender variance plot



It is depicted in the plot above that VL divert away from the true mean with about 80 copies/ML for female(green) and 30 copies/ML for males(red). The VL variation from the VL mean stays constant below 100 copies/ML until the third visit, and then it gradually increases after the third visit, for both male and female.

5.3.2 The Transition Model

This section presents the results of the Transition Autoregressive model of the factors associated with viral suppression, using CD4 and Viral Load as response variables. Tables 10 and 11 shows the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for 5 different orders of Transition (Autoregressive) Models using Viral Load and CD4 cell counts as a response variables. The transition models are denoted as models AR(1), AR(2), AR(3), AR(4) and AR(5), where the number indicates the order of the transition model. The Autoregressive model with lowest AIC and BIC values and highest log likelihood, in which viral load and CD4 was used as outcome variables points out the AR(3) and AR(4) respectively.

Table 11: Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for the Transition (Autoregressive) Models using Viral Load

Model	**Log Likelihood**	**Likelihood Ratio Test**	**AIC**	**BIC**
AR(1)	-9361.2	-	18752	18863
AR(2)	-9221.8	< 2.2e-16 (***)	18488	18650
AR(3)	-9148.2	< 2.2e-16 (***)	**18358**	**18587**
AR(4)	-9224.7	< 2.2e-16 (***)	18533	18844
AR(5)	-9323.6	< 2.2e-16 (***)	18757	19163

Note: This table presents the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criteria (AIC and BIC) from the Transition (Autoregressive) Model using Viral Load as a response variable.

Table 12: Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criterions (AIC and BIC) for the Transition (Autoregressive) Models CD4

Model	**Log Likelihood**	**Likelihood Ratio Test**	**AIC**	**BIC**
AR(1)	-21542	-	43113	43232
AR(2)	-21482	< 2.2e-16 (***)	43007	43181
AR(3)	-21498	0.0001528(***)	43058	43303
AR(4)	-21430	< 2.2e-16 (***)	**42944**	**43276**
AR(5)	-21522	< 2.2e-16 (***)	43155	43590

Note: This table presents the Log Likelihood, Likelihood Ratio Test, Akaike and Bayesian information criteria (AIC and BIC) from the Transition (Autoregressive) Model using CD4 as a response variable.

Table 13 presents the parameter estimates of the Transition Autoregressive model of order three (AR[3]) which was shown to be the best model. This model used viral load as a response variable. The results indicated that viral load had a significant association with a linear follow up time, quadratic follow up time, adherence, weight, age and clinical stage. More specifically, linear follow up time, adherence and weight had negative significant associations, while quadratic follow up time, age and clinical stage had positive associations. It is also worth mentioning that cubic follow up time, sex and marital status were not associated with viral load. The fixed effects appear to be weak, but CIs are narrow (Table 13). On average the random effect is strong, but there is considerable variation between subjects (Table 14).

Table 13: Summary of Fixed Effects (with Viral Load as a response)

Fixed Effects	Unstandardized Coefficients					95% CI for estimate	
	Estimate	**Standard error**	**t-value**	**F-value**	**P-value**	**LB**	**UB**
Intercept	0.129	0.066	1.968			0.0005	0.258
γ_1	-2.508	0.223	-11.245	1457.101	< 2.2e-16 (***)	-2.946	-2.071
γ_2	0.274	0.146	1.879	617.144	0.060(.)	-0.012	0.561
γ_3	0.009	0.021	0.453	4.817	0.651	-0.031	0.050
Sex (1 = male, 2 = female)	-0.005	0.013	-0.409	10.196	0.682	-0.032	0.021
Marital_Status (1 = single, 2 = married and 3 = others)	-0.025	0.023	-1.098	20.184	0.272	-0.070	0.020
Adherence (1 = Less than 95% adherence to ART 2= At least 95% adherence to ART)	-1.900	0.013	142.009	20455.721	< 2.2e-16 (***)	1.874	1.926
Weight_Current	-0.003	0.0004	-6.755	49.008	1.427e-11(***)	-0.004	-0.002
AgeC	0.002	0.0008	2.291	6.947	0.022(*)	0.0002	0.003
Clinical_Stage (1 = stage 1, 2 = stage 2, 3 = stage3, and 4 = stage 4)	0.081	0.013	6.374	40.634	1.836e-10(***)	0.056	0.105

Note: (.) significance at 10%, (*) significance at 5%, (**) significance at 1%, (***) significance at 0.1%

Table 14: Summary of Random Effects (with Viral Load as a response)

Group	**Name**	**Variance**	**Standard**		**ICC**	
Patient_ID		0.280328	0.52946			
	Period	43.17009	6.57040	-0.55		
	Period	9.784388	3.12800	0.37	-0.95	
	Period	0.156793	0.39597	-0.32	0.90	-0.99
Facility		0.004260	0.06527			
	Period	0.069128	0.26292	0.99		
	Period	0.035753	0.18909	-0.75	-0.83	
	Period	0.000623	0.02496	0.65	0.74	-0.99
Residual		0.108704	0.32970			

Table 15 presents the parameter estimates of the Transition Autoregressive model of order four (AR(4)) which was shown to be the best model. This model used CD4 cell counts as a response variable. The results indicated that CD4 cell counts had a significant association with a linear follow up time, quadratic follow up time, sex, adherence and clinical stage. More specifically, quadratic follow up time and clinical stage had negative significant associations, while linear follow up time, adherence and sex had positive associations. It is also worth mentioning that cubic follow up time, quartic follow up time, sex, marital status, weight and age were not associated with CD4 cell counts. The fixed effects appear to be weak, but CIs are narrow. On average the Random effect is strong, but there is considerable variation between subjects (Table 15).

Table 15: Summary of Fixed Effects (with CD4 as a response)

Fixed Effects	Unstandardized Coefficients					95% CI for estimate	
	Estimate	Standard error	t-value	F-value	P-value	LB	UB
Intercept	6.117	0.085	72.209			5.951	6.283
γ_1	1.234	0.561	2.200	4.650	0.028(*)	0.135	2.334
γ_2	-1.201	0.538	-2.231	5.622	0.026(*)	-2.255	-0.146
γ_3	0.287	0.182	1.577	6.385	0.115	-0.070	0.643
γ_4	-0.022	0.020	-1.111	1.358	0.267	-0.061	0.017
Sex (1 = male, 2 = female)	0.030	0.011	2.688	16.495	0.007(**)	0.008	0.052
Marital_Status (1 = single, 2 = married and 3 = others)	-0.019	0.019	-0.977	0.635	0.328	-0.056	0.019
Adherence (1 = Less than 95% adherence to ART 2= At least 95% adherence to ART)	0.209	0.016	-13.283	188.221	< 2.2e-16(***)	-0.240	-0.178
Weight_Current	0.0006	0.0004	1.532	2.700	0.125	-0.0002	0.001
AgeC	-0.001	0.0006	-1.547	3.245	0.122	-0.002	0.0003
Clinical_Stage (stage 1-4)	-0.077	0.016	-4.937	24.376	7.927e-07(***)	-0.108	-0.046

Note: (.) significance at 10%, (*) significance at 5%, (**) significance at 1%, (***) significance at 0.1%

Table 16: Summary of Random Effects (with CD4 as a response)

Group	Name	Variance	Standard Deviation	ICC			
Patient_ID							
(Intercept)		0.1136422	0.33711				
	Period1	5.7737149	2.40286	-0.80			
	Period2	2.4573121	1.56758	-0.58	0.84		
	Period3	1.0887101	1.04341	0.52	-0.91	-0.89	
	Period4	0.0198591	0.14092	-0.43	0.87	0.78	-0.98
Facility							
(Intercept)		0.0175645	0.13253				
	Period1	0.5518793	0.74289	-0.63			
	Period2	0.3415215	0.58440	0.50	-0.72		
	Period3	0.0319263	0.17868	0.11	0.29	-0.80	
	Period4	0.0004391	0.02095	-0.53	0.02	0.46	-0.90
Residual		0.3984944	0.63126				

5.4 Discussion

The main objective of the study was to compare the performance of the transition autoregressive model and the mixed effects model in forecasting viral suppression in ART patients using CD4 counts and viral load and deduce on the best model to be recommended to make these predictions in future. To be specific, to investigate the prevalence of viral suppression among ART patients, to identify the factors associated with viral suppression among ART patients as well as to inspect the predictive performance of the transition and the mixed effects model.

A population of 8068 patients was studied, of which 50.2% of patients were from Kuisebmund Health Center, 4.0% from Hakhaseb Clinic, 11.1% from omaruru District Hospital and 34.7% from Tamariskia Clinic, all in Erongo region. Demographic and clinical characteristics were described in table 9. 4553 (56.4%) were female, 7534 (93.4%) were single, 7603 (94.2%) were at baseline clinical stage-I, 6736 (83.5%) were adults aged 30-49, 1977 (24.5%) initiated ART in 2011and 7011 (86.9%) were at least 95% adherent to medication. The median and interquartile (IQR) of current body weight and the baseline CD4 cell count were 64 (56 – 74.2), and 339 (221 – 480), respectively.

The Transition Autoregressive model of order three (AR(3)) results indicated that viral load had a significant association with a linear follow up time, quadratic follow up time, adherence, weight, age and clinical stage. More specifically as the values of linear follow up time, adherence and weight increases the mean viral load decreases significantly, while, as the values of quadratic follow up time and clinical stage increase, the mean of viral load also tends to increase significantly. Adherence to ART medication has proven to be the main determinant of viral load, the mean of viral load reduces by 1.90 given a one-unit increase in adherence while holding other variables in the model constant. It is also worth mentioning that cubic follow up time, sex and marital status were not significantly associated with viral load.

The Transition Autoregressive model of order four (AR(4)) results indicated that CD4 cell counts had a significant association with a linear follow up time, quadratic follow up time, sex, adherence and clinical stage. More specifically, as the value of quadratic follow up time and clinical stage increases, the mean CD4 count decreases significantly. While, as the value of linear follow up time, adherence and sex increases, the mean CD4 count also tends to increase significantly. It is also worth mentioning that cubic follow up time, quartic follow up time, marital status, weight and age were not significantly associated with CD4 cell counts.

5.5 Conclusions

This chapter aimed to predict viral suppression, using CD4 cell counts and viral load, in antiretroviral therapy (ART) patients in Erongo region using the Transition Autoregressive model. The findings provide additional support to the concept that demographic (sex, age and weight) and clinical (follow up time, adherence and clinical stage) characteristics are the determinants of viral suppression. Adherence to ART medication has proven to be the main determinant of viral suppression. Marital status has proven not to be a determinant of either viral suppression or increase in CD4 cells count.

Chapter 6: Discussion, Conclusions and Recommendations

6.1 Introduction

The purpose of this study was to fit a prediction model for viral suppression, using CD4 cell counts and viral load, in antiretroviral therapy (ART) patients in Erongo region using transition models and mixed effects models. Chapter four and five concentrated on the applications of the Mixed Effects Model and the Transition Model respectively. The final chapter of this mini thesis is devoted to the vital discussion in line with the study objectives, the conclusions drawn from the discussion as well as the recommendations made in line with the conclusions.

6.2 Discussion

The main objective of the study was to compare the performance of the Transition Autoregressive model and the Mixed Effects model in predicting viral load in ART patients using CD4 counts and viral load and deduce on the best model to be recommended to make these predictions in future. To be specific, to investigate the prevalence of viral suppression among ART patients, to identify the factors associated with viral suppression among ART patients as well as to inspect the predictive performance of the transition and the mixed effects model.

The findings of this study points to a viral suppression prevalence rate of 71% in Erongo region. The findings provide additional support to the concept that demographic (sex, age and weight) and clinical (follow up time, adherence and clinical stage) characteristics are the determinants of viral suppression. Adherence to ART medication has proven to be the main determinant of reduction in viral load as it has a strong negative relationship (1.945 for Mixed Effects and 1.9 for Transition Model), while, marital status has proven not to be a determinant of either viral suppression or CD4 cells count. Quiet different, Rangarajan, Colby, Giang, Nguyen, Tou, Danh, & Nguyen (2016) stressed that gender does not impact viral suppression and recommended routine viral load testing, adherence monitoring, counselling, simplification of ART regimens as well as necessitating treatment support interventions to address social isolation and stigma among people living with HIV.

The random slope mixed effects model and the transition autoregressive model of order three (AR[3]) results indicated that viral load had a significant association with follow up time, adherence, weight, age, clinical stage and adherence at different follow up time points. More specifically, as the value of follow up time, weight, adherence and adherence at different follow up time points increases, the mean viral load decreases significantly, this is because the aim of treatment is to reduce viral load with adherence to medication overtime and maintaining a healthy lifestyle. While as the value of age and clinical stage increases, the mean of viral load also tends to increase significantly, this is because as people get older they tend to have weak immune systems, making it difficult to fight the virus and the clinical stages ranges from stage I (asymptomatic) to stage IV (AIDS). It is also worth mentioning that sex, marital status and weight at different follow up time points were not significantly associated with viral load.

The random slope mixed effects model and the transition autoregressive model of order four (AR(4)) indicated that CD4 cell counts had a significant association with follow up time, sex, adherence, clinical stage and adherence at different follow up time points. More specifically, follow up time and clinical stage had negative significant associations, this is because the clinical stages ranges from stage I (asymptomatic) to stage IV (AIDS). An explanation for CD4 cell count to be negatively associated with follow up time is that CD4 cell counts give an indication of the health of the immune system. The aim of treatment is to reduce viral load, so that CD4 count can increase and the body's ability to fight infections can improve. However, it is not always a case, some people's CD4 cell counts do not increase, or rise very slowly, even though their viral load is undetectable, this is because CD4 cell count may go up and down in response to different factors such as genetic make-up, lifestyle (exercise, healthy eating, lack of sleep, smoking), quality of health care and infections such as the flu or an outbreak of herpes. Sex, adherence and adherence at different follow up time points had positive associations this is because females are more likely to adhere to medication and maintaining a healthy lifestyle, unlike males. CD4 and viral load are generally negatively associated. When CD4 cell counts are high, viral load will be low and when CD4 is low, viral load is high. Over time viral load decreases and CD4s are supposed to increase for better outcomes. It is also worth mentioning that marital status, weight, age and weight at different follow up time points were not associated with CD4 cell counts.

Although studies of viral suppression have previously been conducted, the simultaneous modelling of viral suppression using CD4 and Viral load was neglected, CD4 cell counts do not indicate viral suppression, however it is vital to model change in CD4 cell counts as they may show improvement in health and indirectly decreasing viral loads. Both the Mixed Effects and the Transition Autoregressive models revealed that for patients stable on ART, viral load is the most reliable measure of response to treatment as well as viral suppression and not CD4 cell counts. This was shown in the models in the sense that decrease in viral load with follow up time did not necessarily lead to increment in CD4 cell counts. This finding is in line with the findings of Ford, Meintjes, Vitoria, Greene, & Chiller (2017) who stressed that for patients who are stable on ART and where HIV viral load testing is available, CD4 cell count was discouraged as a way to monitor the response to treatment.

This paper presented the transition models with different numbers of Autoregressive terms and a random slope Mixed Effects model, as plausible models to account for the dependence among subsequent Viral load levels within a patient and as models to predict the future Viral load level. Based on the results it was shown that the transition model and the mixed effects model have almost the same predictive accuracy at the first few visits of a patient; however, for longer time series the transition model offers somewhat better predictions.

Both the transition and the mixed effects models make use of the data of a subject's previous observations for making predictions. In the transition model only the last q observations are used for the prediction of the current response. However, the mixed effects model uses all previous observations. Therefore, the mixed effects model requires more historical information than the transition model. Since the transition model is convenient in practice and needs less historical information compared to the mixed effects model, HIV studies may use this model to predict the future viral load.

The concerning finding of this study points to the unpopularity of the Transition Autoregressive model in research, specifically the statistics literature, however, the Mixed Effects model is widely applied. The approach of using transition or autoregressive models is quite novel in the statistics research, however in other fields such as econometrics, Transition Autoregressive modeling is a very well-known technique for tackling correlated financial phenomena and time series problems.

This study was however mainly aimed at drawing the attention of the researchers to the predictive performance of the Transition model and Mixed Effects models in modeling change and event occurrence in longitudinal studies such as HIV studies.

The Log Likelihood, AIC and BIC show that the Transition Autoregressive model is the best predictor of viral load as well as CD4 cell counts. This implies that the Transition Autoregressive model has a better predictive performance than the Mixed Effects model. This finding ties with that of Nasserinejad, de Kort, Baart, Komárek, van Rosmalen, & Lesaffre (2013) who concluded that the transition model provides a somewhat better prediction than the mixed effects model, especially at high visit numbers and that the transition model offers a better trade-off between sensitivity and specificity when varying the cut-off values for eligibility in predicted values.

This paper does not claim that the final models are optimal; further research is needed to arrive at a better prediction model. First, the data set used in this paper is unbalanced in the sense that the time intervals between follow up times vary considerably, even though this was not taken into account. Secondly, there are more factors that are possibly associated with viral load, however not investigated in this study.

6.3 Conclusion

The main purpose of this study was to estimate prediction models and determine factors that influence viral load and CD4 cell counts in ART patients using transition autoregressive model and the mixed effects model. CD4 and viral load indicate a negative association. However, Both the Mixed Effects and the Transition Autoregressive models revealed that for patients stable on ART, viral load is the most reliable measure of response to treatment as well as viral suppression and not CD4 cell counts.

The Log Likelihood, AIC and BIC show that the Transition Autoregressive model is the best predictor of viral load as well as CD4 cell counts. This implies that the Transition Autoregressive model has a better predictive performance than the Mixed Effects model.

6.4 Recommendations

This study recommends the use of the Transition Autoregressive model in modeling change and event occurrence in longitudinal studies such as HIV studies, specifically in predicting viral suppression in the future. The study further recommends similar studies to consider socio-economic factors in addition to the factors considered in this study. In addition, this study covered only a few selected facilities in Erongo region; an extension to the whole Namibia at large is recommended.

Bibliography

Alleyne, G., Breman, J., Claeson, M., Evans, D., Jamison, D., Jha, P., ... & Musgrove, P. (2006). Disease control priorities in developing countries. World Bank/WHO.

Baayen, R. H. (2012). Mixed-effects models. *The Oxford Handbook of Laboratory Phonology*, 668-677.

Bailey, R. C., Moses, S., Parker, C. B., Agot, K., Maclean, I., Krieger, J. N., ... & Ndinya-Achola, J. O. (2007). Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. *The lancet*, *369*(9562), 643-656.

Bandyopadhyay, D., Castro, L. M., Lachos, V. H., & Pinheiro, H. P. (2015). Robust joint non-linear mixed-effects models and diagnostics for censored HIV viral loads with CD4 measurement error. *Journal of Agricultural, Biological, and Environmental Statistics*, *20*(1), 121-139.

Bradley, H., Mattson, C. L., Beer, L., Huang, P., & Shouse, R. L. (2016). Increased antiretroviral therapy prescription and HIV viral suppression among persons receiving clinical care for HIV infection. *AIDS (london, England)*, *30*(13), 2117.

C. R. Henderson; Oscar Kempthorne; S. R. Searle; C. M. von Krosigk (1959). "The Estimation of Environmental and Genetic Trends from Records Subject to Culling". Biometrics. International Biometric Society. 15 (2): 192–218. doi:10.2307/2527669. JSTOR 2527669

Cescon, A. M., Cooper, C., Chan, K., Palmer, A. K., Klein, M. B., Machouf, N., ... & Lima, V. D. (2011). Factors associated with virological suppression among HIV-positive individuals on highly active antiretroviral therapy in a multi-site Canadian cohort. *HIV medicine*, *12*(6), 352-360.

Chakraborty, H., Hossain, A., & Latif, M. A. (2019). A three-state continuous time Markov chain model for HIV disease burden. *Journal of Applied Statistics*, *46*(9), 1671-1688.

Coleman A. K. (2014) *A Perspective on World AIDS Day from Namibia*. Retrieved at 15h00 on 25 August 2016 from http://pinkribbonredribbon.org/2014/12/

Columbia University. Summary sheet: preliminary findings: Namibia population-based HIV impact assessment: NAMPHIA 2018. New York: ICAP, 2018. Available at: https://phia.icap.columbia.edu/wp-content/uploads/2018/10/33462%E2%80%A2NAMPHIA-SS_A4_B.v41.pdf (accessed 22 March 2019).

Conway, J. M., Perelson, A. S., & Li, J. Z. (2019). Predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers. *PLoS computational biology*, *15*(7), e1007229.

Dessie, Z. G., Zewotir, T., Mwambi, H., & North, D. (2020). Modelling immune deterioration, immune recovery and state-specific duration of HIV-infected women with viral load adjustment: using parametric multistate model. *BMC Public Health*, *20*(1), 1-13.

Detels, R., Wu, J., & Wu, Z. (2019). Control of HIV/AIDS can be achieved with multi-strategies.

Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S.L. (2002). *Analysis of longitudinal data*. Oxford University Press. Retrieved from https://epdf.tips/analysis-of-longitudinal-data-second-edition.html

Fisher, RA (1918). "The correlation between relatives on the supposition of Mendelian inheritance". Transactions of the Royal Society of Edinburgh. 52 (2): 399–433. doi:10.1017/S0080456800012163

Ford, N., Meintjes, G., Vitoria, M., Greene, G., & Chiller, T. (2017). The evolving role of CD4 cell counts in HIV care. *Current Opinion in HIV and AIDS*, *12*(2), 123-128.

Fu, L. Y., Zeng, W. S., Tang, S. Z., Sharma, R. P., & Li, H. K. (2012). Using linear mixed model and dummy variable model approaches to construct compatible single-tree biomass equations at different scales-A case study for Masson pine in Southern China. *Journal of Forest Science*, *58*(3), 101-115.

GLANCE, A., & STAFF, C. (2017). CDC in Namibia. Available at https://www.cdc.gov/ globalhealth/ countries/namibia/pdf/Namibia_FactSheet.pdf (Accesses 28 February 2020)

Golub, S. A., Operario, D., & Gorbach, P. M. (2010). Pre-exposure prophylaxis state of the science: empirical analogies for research and implementation. *Current hiv/aids Reports*, *7*(4), 201-209.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... & Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794.

Helwig, N. E. (2017). Linear Mixed-Effects Regression. *Minneapolis and Saint Paul: University of Minnesota (Twin Cities)*.

Huang, Y., Dagne, G. A., Zhou, S., & Wang, Z. (2015). Piecewise mixed-effects models with skew distributions for evaluating viral load changes: a Bayesian approach. *Statistical methods in medical research*, *24*(6), 730-746.

Institute for Health Metrics and Evaluation (IHME). *Namibia: State of the Nation's Health: Findings from the Global Burden of Disease*. Seattle, WA: IHME, 2016.

Institute for Health Metrics and Evaluation (IHME). *Namibia: State of the Nation's Health: Findings from the Global Burden of Disease*. Seattle, WA: IHME, 2016.

Jamison, D. T. (Ed.). (2006). *Disease and mortality in sub-Saharan Africa*. World Bank Publications.

Jobanputra, K., Parker, L. A., Azih, C., Okello, V., Maphalala, G., Kershberger, B., ... & Ellman, T. (2015). Factors associated with virological failure and suppression after enhanced adherence counselling, in children, adolescents and adults on antiretroviral therapy for HIV in Swaziland. *PloS one*, *10*(2), e0116144.

Joint United Nations Programme on HIV/AIDS, & Joint United Nations Programme on HIV/Aids. (2014). 90-90-90: an ambitious treatment target to help end the AIDS epidemic. *Geneva: Unaids*.

Joint United Nations Programme on HIV/AIDS. (2004). Guide to Monitoring and Evaluating National HIV/AIDS Prevention Programmes for Young People.

Joint United Nations Programme on HIV/AIDS. (2014). Reduction of HIV-related stigma and discrimination. *UNAIDS Information Production Unit. Retrieved August*, *8*, 2016.

Kalman RE: A new approach to linear filtering and prediction problems. *Trans ASME–JBasic Eng* 1960, 82:35–45.

Khasakhala, A. A., & Kloos, H. (2013). Monitoring and evaluation of HIV/AIDS prevention programs. In *Vulnerabilities, Impacts, and Responses to HIV/AIDS in Sub-Saharan Africa* (pp. 212-231). Palgrave Macmillan, London.

Lachos, V. H., Bandyopadhyay, D., & Dey, D. K. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. *Biometrics*, *67*(4), 1594-1604.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.

Lu, T., & Wang, M. (2014). Nonlinear mixed-effects HIV dynamic models with considering left-censored measurements. *Journal of Statistical Distributions and Applications*, *1*(1), 13.

Magnusson, K. (2015). Using R and lme/lmer to fit different two-and three-level longitudinal models.

Malinverni, S., Gennotte, A. F., Schuster, M., De Wit, S., Mols, P., & Libois, A. (2018). Adherence to HIV post-exposure prophylaxis: A multivariate regression analysis of a 5 years prospective cohort. *Journal of Infection*, *76*(1), 78-85.

Marks, G., Crepaz, N., & Janssen, R. S. (2006). Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *Aids*, *20*(10), 1447-1450.

Ministry of Health and Social Services. (2016). National Guidelines for Antiretroviral Therapy (5th ed.). Retrieved from https://aidsfree.usaid.gov/sites/default/files/na_national_guidelines_art.pdf

MoHSS (2009b). *HIV/AIDs in Namibia: Behavioral and Contextual Factors Driving the Epidemic.Windhoek: Measure Evaluation*

Morrell, C. H., Pearson, J. D., Carter, H. B., & Brant, L. J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association*, *90*(429), 45-53.

Namibia Demographic and Health Survey (2006/7). Windhoek, Ministry of Health and Social Services.

Nance, R. M., Delaney, J. C., Simoni, J. M., Wilson, I. B., Mayer, K. H., Whitney, B. M., ... & Christopoulos, K. A. (2018). HIV viral suppression trends over time among HIV-infected patients receiving care in the United States, 1997 to 2015: a cohort study. *Annals of internal medicine*, *169*(6), 376-384.

Nasserinejad, K., de Kort, W., Baart, M., Komárek, A., van Rosmalen, J., & Lesaffre, E. (2013). Predicting hemoglobin levels in whole blood donors using transition models and mixed effects models. *BMC medical research methodology*, *13*(1), 62.

National Planning Commission (NPC).2003. The 2001 Population and Housing Census: Basic Analysis with Highlights. Windhoek, Namibia: Central Bureau of Statistics.

Opportunistic Infections Project Team of the Collaboration of Observational HIV Epidemiological Research in Europe (COHERE) in EuroCoord. (2012). CD4 cell count and the risk of AIDS or death in HIV-Infected adults on combination antiretroviral therapy with a suppressed viral load: a longitudinal cohort study from COHERE. *PLoS medicine*, *9*(3), e1001194.

R Development Core Team: R. a language and environment for statistical computing. [http://www.R-project.org]

Raja, K., Chandrasekar, C., Krishnarajasekhar, O. R., & Manoharan, G. (2013). Zero CD4 count: A case of discordant CD4 response in a patient with well suppressed viral load. *Indian journal of medical microbiology*, *31*(3), 298.

Ramjee, G., & Daniels, B. (2013). Women and HIV in sub-Saharan Africa. *AIDS research and therapy*, *10*(1), 30.

Rangarajan, S., Colby, D. J., Le Truong Giang, D. D. B., Nguyen, H. H., Tou, P. B., Danh, T. T., ... & Nguyen, N. Q. (2016). Factors associated with HIV viral load suppression on antiretroviral therapy in Vietnam. *Journal of virus eradication*, *2*(2), 94.

Sen, L. K., & Baharumshah, A. Z. (2002). Performances of Non-linear Smooth Transition Autoregressive and Linear Autoregressive Models in Forecasting the Ringgit-Yen Rate. *Pertanika Journal of*, 131.

Shoko, C., & Chikobvu, D. (2019). A superiority of viral load over CD4 cell count when predicting mortality in HIV patients on therapy. *BMC infectious diseases*, *19*(1), 169.

Shoko, C., Chikobvu, D., & Bessong, P. O. (2018). A Markov model to estimate mortality due to HIV/AIDS using viral load levels-based states and CD4 cell counts: a principal component analysis approach. *Infectious diseases and therapy*, *7*(4), 457-471

Temah, C. T. (2009). What drives HIV/AIDS epidemic in sub-Saharan Africa?. *Revue d'économie du développement*, *17*(5), 41-70.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, *89*(425), 208-218.

Teräsvirta, T., & Yang, Y. (2014). Specification, estimation and evaluation of vector smooth transition autoregressive models with applications.

The Nambia Ministry of Health and Social Services (MoHSS) and ICF International. 2014. *The Namibia Demographic and Health Survey 2013.* Windhoek, Namibia, and Rockville, Maryland, USA: MoHSS and ICF International.

The Nambia Ministry of Health and Social Services (MoHSS) and ICF International. 2014. *The Namibia Demographic and Health Survey 2013.* Windhoek, Namibia, and Rockville, Maryland, USA: MoHSS and ICF International.

Tsay, R. (1984). Regression models with time series errors. *Journal of the American Statistical Association,* **79,** 118-24.

Umer, U. M., Sevil, T., & Sevil, G. (2018). Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index. *The Journal of Finance and Data Science*, *4*(2), 90-100.

UNAIDS (Joint United Nations Programme on HIV/AIDS). 2000. *AIDS: Palliative Care*. UNAIDS Technical Update. Geneva: UNAIDS.

UNAIDS and WHO (World Health Organization). 2005. *AIDS Epidemic Update: December 2005*. Geneva: UNAIDS and WHO.

UNAIDS, Condoms [updated 2016-10-03; cited 2019-01-19]

UNAIDS, Global HIV & AIDS statisitics-2018 fact sheet [updated 2018-12-01; cited 2019-04-19]

UNAIDS. *2016 Core Epidemiology Slides*; 2016. UNAIDS. AIDSinfo website; accessed June 2016, available at: http://aidsinfo.unaids.org/.

UNICEF (2017), Report On Phase 2 of The Namibia "All In" Country Assessment 2016- Strengthening the Adolescent Component of National HIV Programmes through Country Assessments. UNICEF, Namibia. Retrievedfrom:https://www.unicef.org/namibia/na.Namibia_ALLIN_Adolescent_HIV_Programme_Report_Phase_2_2017.pdf

US Department of Health and Human Services. (2018). AIDSinfo. *Online at https://aidsinfo.nih.gov/understanding-hiv-aids/glossary/*

Werner, R. N., Gaskins, M., Nast, A., & Dressler, C. (2018). Incidence of sexually transmitted infections in men who have sex with men and who are at substantial risk of HIV infection–A meta-analysis of data from trials and observational studies of HIV pre-exposure prophylaxis. *PloS one*, *13*(12).

World Health Organization. (2016). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach*. World Health Organization.

Wu, L., & Zhang, H. (2018). Mixed Effects Models with Censored Covariates, with Applications in HIV/AIDS Studies. *Journal of Probability and Statistics, 2018*.

Wu, L., Hu, X. J., & Wu, H. (2008). Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics*, *9*(2), 308-320.

Yehia, B. R., Stephens-Shields, A. J., Fleishman, J. A., Berry, S. A., Agwu, A. L., Metlay, J. P., ... & Gaur, A. H. (2015). The HIV care continuum: changes over time in retention in care and viral suppression. *PloS one*, *10*(6).

Zeger, S.L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics,* **44,** 1019-31.

Zhang, J. T., & Wu, H. (2010). Modeling HIV dynamics using unified mixed-effects models. *American Journal of Mathematical and Management Sciences*, *30*(1-2), 83-109.

Zhou, J. (2010). Smooth Transition autoregressive Models: A study of the industrial production index of Sweden.

Zingoni, Z. M., Chirwa, T. F., Todd, J., & Musenge, E. (2019). Spatial Heterogeneity of Viral Suppression and Viral Rebound Patterns among ART Patients in Zimbabwe from 2004 to 2017: A Bayesian Mixed Effects Multistate Model. *International Journal*, *8*, 99.

Appendix

Appendix A: R codes for the Mixed Effects model

```
require(ggplot2)

require(GGally)

require(reshape2)

require(lme4)

require(compiler)

require(parallel)

require(boot)

require(lattice)
```

Viral Load

Random intercept

```
Model1<-lmer(Log_Viral_Load~ Follow_up_time +Sex + Marital_Status + adherence*
Follow_up_time + AgeC + Clinical_Stage + Weight_Current * Follow_up_time+(1 |
Patient_ID)+ (1| Facility), REML=FALSE, data= Erongo, na.action=na.omit)

summary(Model1)
```

Random slope

```
Model2<-lmer(Log_Viral_Load~ Follow_up_time+Sex + Marital_Status+ adherence +
Weight_Current+ adherence* Follow_up_time + AgeC + Clinical_Stage + Weight_Current *
Follow_up_time+(1+ Follow_up_time |Patient_ID)+ (1+ Follow_up_time|Facility),
REML=FALSE, data= Erongo, na.action=na.omit, control=lmerControl(check.conv.singular =
.makeCC(action = "ignore",  tol = 1e-4)))
```

```
summary(Model2)

se <- sqrt(diag(vcov(Model2)))

(tab <- cbind(Est = fixef(Model2), LL = fixef(Model2) - 1.96 * se, UL = fixef(Model2) + 1.96 *
+    se))

require(car)

Anova(Model2)

anova(Model2)

anova(Model1,Model2, Model3)

require(lmtest)

lrtest(Model1, Model2, Model3)
```

Random intercept and slope

```
Model3<-lmer(Log_Viral_Load~ Follow_up_time +Sex + Marital_Status + adherence +
Weight_Current+ adherence* Follow_up_time + AgeC + Clinical_Stage + Weight_Current *
Follow_up_time+(1 | Patient_ID) +(1| Facility) +(1+Facility|Patient_ID), REML=FALSE, data=
Erongo, na.action=na.omit, control=lmerControl(check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
 summary(Model3)
```

CD4

Random intercept

```
Model4<-lmer(log(CD4_cell_count) ~ Follow_up_time +Sex + Marital_Status + adherence*
Follow_up_time + AgeC + Clinical_Stage + Weight_Current * Follow_up_time+(1 |
Patient_ID)+ (1| Facility), REML=FALSE, data= Erongo, na.action=na.omit)

summary(Model4)
```

Random slope

```
Model5<-lmer(log(CD4_cell_count) ~ Follow_up_time+Sex + Marital_Status+ adherence +
Weight_Current+ adherence* Follow_up_time + AgeC + Clinical_Stage + Weight_Current *
Follow_up_time+(1+ Follow_up_time |Patient_ID)+ (1+ Follow_up_time|Facility),
REML=FALSE, data= Erongo, na.action=na.omit, control=lmerControl(check.conv.singular =
.makeCC(action = "ignore",  tol = 1e-4)))
summary(Model5)

se <- sqrt(diag(vcov(Model5)))

(tab <- cbind(Est = fixef(Model5), LL = fixef(Model5) - 1.96 * se, UL = fixef(Model5) + 1.96 *

+    se))

require(car)

Anova(Model5)

anova(Model5)

anova(Model4,Model5, Model6)
```

Random intercept and slope

```
Model6<-lmer(log(CD4_cell_count) ~ Follow_up_time +Sex + Marital_Status + adherence +
Weight_Current+ adherence* Follow_up_time + AgeC + Clinical_Stage + Weight_Current *
Follow_up_time+(1 | Patient_ID) +(1| Facility) +(1+Facility|Patient_ID), REML=FALSE, data=
Erongo, na.action=na.omit, control=lmerControl(check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
 summary(Model6)
```

Appendix B: R codes for the Transition Autoregressive model

```
library("lme4")
library("nlme")
library("plyr")
library(ggplot2)
SampleData<-read.csv(file.choose(), head=T)
#===============Individual profile plot===========
with(ERONGO, plot(Follow_up_time, Viralload, bty='l', main = 'Profiles of viral load',
xlab='Time', ylab='Viral Load'))
d_ply(ERONGO, 'Patient_ID', function(x) lines(x$Follow_up_time, x$Viral_Load, lty=3,
col='#666666'))
with(Males, plot(Follow_up_time, Viralload, bty='l', main = 'Males', xlab='Time', ylab='Viral
Load'))
d_ply(Males, 'Patient_ID', function(x) lines(x$Follow_up_time, x$Viral_Load, lty=3,
col='#666666'))
```

```
with(Females, plot(Follow_up_time, Viralload, bty='l', main = 'Females', xlab='Time',
ylab='Viral Load'))
d_ply(Females, 'Patient_ID', function(x) lines(x$Follow_up_time, x$Viral_Load, lty=3,
col='#666666'))
SampleData<-read.csv(file.choose(), head=T)
#====================Overall mean plot====================
Ind<-aggregate(ERONGO[, 4], list(ERONGO$Follow_up_time), mean)
Y<-Ind[,2]
X<-Ind[,1]
plot(X,Y, type="b", bty='l', lwd=1,  xlab=' Follow up time ', ylab='Mean Viral Load')

#==========================================================
SampleData<-read.csv(file.choose(), head=T)
SampleData1<-read.csv(file.choose(), head=T)

#==============Gender Average plot====================#
Data<-aggregate(ERONGO[, 4], list(ERONGO$Follow_up_time), mean)
Data1<-aggregate(ERONGO1[, 4], list(ERONGO1$Follow_up_time), mean)

Y<-Data[,2]
X<-Data[,1]
```

```
Y1<-Data1[,2]

X1<-Data1[,1]

plot(X,Y, type="b", col="green", bty='l', lwd=1, pch=15,

ylim=range(c(Y,Y1)),xlim=range(c(X,X1)), xlab='Follow up time', ylab='Mean Viral Load')

lines(X1, Y1, type="b", col="red", lwd=2, pch=19)

legend(35,80000,c("Female","Male"), box.lty=0, title = "Gender:", lwd=c(1,2),

col=c("green","red"), pch=c(15,19))


#======================Gender variance plot==============#


Var1<-aggregate(ERONGO[, 4], list(ERONGO$Follow_up_time), var)

Var2<-aggregate(ERONGO1[, 4], list(ERONGO1$Follow_up_time), var)

Y11<-Var1[,2]

X11<-Var1[,1]

Y22<-Var2[,2]

X22<-Var2[,1]

plot(X11,Y11, type="b", col="green", bty='l', lwd=1, pch=15, ylim=range(c(Y11,Y22),

na.rm=TRUE),xlim=range(c(X11,X22)), xlab='Follow up time', ylab='Variance Viral Load')

lines(X22, Y22, type="b", col="red", lwd=2, pch=19)

legend(35,1.2e+11,c("Female","Male"), box.lty=0, title = "Gender:", lwd=c(1,2),

col=c("green","red"), pch=c(15,19))


plots in color
```

```
library(lattice)

xyplot(Viral_Load ~ Follow_up_time, groups = Patient_ID, type= "b",main = 'Profiles of viral
load', xlab='Time', ylab='Viral Load',data= ERONGO)

xyplot(Viral_Load ~ Follow_up_time, groups = Patient_ID, type= "b",main = 'Males',
xlab='Time', ylab='Viral Load',data= Males)

xyplot(Viral_Load ~ Follow_up_time, groups = Patient_ID, type= "b",main = 'Females',
xlab='Time', ylab='Viral Load',data= Females)
```

OVERALL MEAN PLOT(SPSS)

Covariance/correlation structure

```
#============MATRICES==========================#
library(tidyr)
Data_wide<-spread(DATA, Follow_up_time, Viral_Load) #LONG-WIDE FORMAT
head(Data_wide)
Dataset2<- subset(Data_wide, select = c(time0:time5))
head(Dataset2)
(varx <- var(Dataset2, use = "complete.obs")) #var-cov matrix
print(cor(Dataset2, use = "complete.obs"),digits=2) #correlation matrix
(covx <- cov(Dataset2, use = "complete.obs")) #var-cov matrix
Erongo<-read.csv("C:\\Users\\spekie\\Desktop\\THESIS 2019\\Datasets\\ERONGO DATA.csv")
attach(Erongo)
```

```
PERIOD1 <- Follow_up_time

PERIOD2 <- Follow_up_time^2

PERIOD3 <- Follow_up_time^3

PERIOD4 <- Follow_up_time^4

PERIOD5 <- Follow_up_time^5

PERIOD6 <- Follow_up_time^6
```

Viral load

```
Modela<-lmer(Log_Viral_Load~PERIOD1 +Sex + Marital_Status+ adherence +
Weight_Current+ AgeC + Clinical_Stage +(1+ PERIOD1|Patient_ID)+ (1+ PERIOD1|Facility),
REML=FALSE, data= Erongo, na.action=na.omit,
control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
Modelb<-lmer(Log_Viral_Load~ I(PERIOD1/10)+ I(PERIOD2/10)+Sex + Marital_Status+
adherence + Weight_Current+ + AgeC + Clinical_Stage +(1+ I(PERIOD1/10)+
I(PERIOD2/10)|Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)|Facility), REML=FALSE,
data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore",
check.conv.singular = .makeCC(action = "ignore",  tol = 1e-4)))
Modelc<-lmer(Log_Viral_Load~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+Sex +
Marital_Status+ adherence + Weight_Current+ AgeC + Clinical_Stage + (1+ I(PERIOD1/10)+
I(PERIOD2/10)+ I(PERIOD3/10)|Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)+
I(PERIOD3/10)|Facility), REML=FALSE, data= Erongo, na.action=na.omit,
control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
```

```
summary(Modelc)

se <- sqrt(diag(vcov(Modelc)))

(tab <- cbind(Est = fixef(Modelc), LL = fixef(Modelc) - 1.96 * se, UL = fixef(Modelc) + 1.96 *

+     se))

require(car)

Anova(Modelc)

anova(Modelc)

anova(Modela, Modelb, Modelc, Modeld, Modele)

require(lmtest)

lrtest(Modela, Modelb, Modelc, Modeld, Modele)

Modeld<-lmer(Log_Viral_Load~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+
I(PERIOD4/10)+Sex + Marital_Status+ adherence + Weight_Current+ AgeC + Clinical_Stage
+(1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)|Patient_ID)+ (1+
I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)|Facility), REML=FALSE,
data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore",
check.conv.singular = .makeCC(action = "ignore",  tol = 1e-4)))

Modele<-lmer(Log_Viral_Load~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+
I(PERIOD4/10)+ I(PERIOD5/10) +Sex + Marital_Status+ adherence + Weight_Current+ AgeC
+ Clinical_Stage +(1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)+
I(PERIOD5/10) |Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+
I(PERIOD4/10)+ I(PERIOD5/10) |Facility), REML=FALSE, data= Erongo, na.action=na.omit,
control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
```

CD4

Model1<-lmer(log(CD4_cell_count)~PERIOD1 +Sex + Marital_Status+ adherence + Weight_Current+ AgeC + Clinical_Stage +(1+ PERIOD1|Patient_ID)+ (1+ PERIOD1|Facility), REML=FALSE, data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action = "ignore", tol = 1e-4)))

Model2<-lmer(log(CD4_cell_count)~ I(PERIOD1/10)+ I(PERIOD2/10)+Sex + Marital_Status+ adherence + Weight_Current+ + AgeC + Clinical_Stage +(1+ I(PERIOD1/10)+ I(PERIOD2/10)|Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)|Facility), REML=FALSE, data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action = "ignore", tol = 1e-4)))

Model3<-lmer(log(CD4_cell_count)~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+Sex + Marital_Status+ adherence + Weight_Current+ AgeC + Clinical_Stage + (1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)|Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)|Facility), REML=FALSE, data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action = "ignore", tol = 1e-4)))

Model4<-lmer(log(CD4_cell_count)~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)+Sex + Marital_Status+ adherence + Weight_Current+ AgeC + Clinical_Stage +(1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)|Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)|Facility), REML=FALSE,

```
data= Erongo, na.action=na.omit, control=lmerControl(check.nobs.vs.nRE="ignore",
check.conv.singular = .makeCC(action = "ignore",  tol = 1e-4)))
summary(Model4)
se <- sqrt(diag(vcov(Model4)))
(tab <- cbind(Est = fixef(Model4), LL = fixef(Model4) - 1.96 * se, UL = fixef(Model4) + 1.96 *
+     se))
require(car)
Anova(Model4)
anova(Model4)
anova(Model1, Model2, Model3, Model4, Model5)
require(lmtest)
lrtest(Model1, Model2, Model3, Model4, Model5)
Model5<-lmer(log(CD4_cell_count)~ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+
I(PERIOD4/10)+ I(PERIOD5/10) +Sex + Marital_Status+ adherence + Weight_Current+ AgeC
+ Clinical_Stage +(1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+ I(PERIOD4/10)+
I(PERIOD5/10) |Patient_ID)+ (1+ I(PERIOD1/10)+ I(PERIOD2/10)+ I(PERIOD3/10)+
I(PERIOD4/10)+ I(PERIOD5/10) |Facility), REML=FALSE, data= Erongo, na.action=na.omit,
control=lmerControl(check.nobs.vs.nRE="ignore", check.conv.singular = .makeCC(action =
"ignore",  tol = 1e-4)))
```