MODELLING PROGRESSION OF HIV/AIDS DISEASE USING HOMOGENOUS SEMI-MARKOV PROCESSES: COHORT STUDY, NAMIBIA

A THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOSTATISTICS

OF

THE UNIVERSITY OF NAMIBIA

BY

SIMON POMBILI KASHIHALWA

200646001

APRIL 2019

SUPERVISOR: DR LILIAN PAZVAKAWAMBWA (DEPARTMENT OF STATISTICS AND POPULATION STUDIES- UNAM)

Abstract

The progression of HIV infection to AIDS and then to death can be considered as a Markovan stochastic process. Disease progression can be broken down into a finite number of intermediate states, based on CD4 counts. The four states of the Markov process of HIV/AIDS progression are commonly defined as: S1: CD4 count > 500 cells/microlitre of blood; S2: 350 < CD4 count ≤ 500 cells/microlitre of blood; S3: 200 < CD4 count ≤ 350 cells/microlitre of blood; S4: CD4 count ≤ 200 cells/microlitre of blood.

The objective of this study was to model the progression of HIV/AIDS disease of patients under ART follow-up in Namibia using homogenous semi-Markov processes, using the data obtained from MoHSS. A retrospective study design was used to obtain data on 2422 patients who were observed 11028 times. The semi-Markov model was employed to estimate the transition probabilities, transition intensity rate and sojourn time. Time homogeneous model was fitted to assess effectiveness of ART by comparing the forward transition and reverse transitions. At treatment commencement ($t = 0$), 657(27.13%) patients started ART in state 1, 683(28.19%) patients started ART in state 2, 677(27.95%) patients started ART in state 3 and 405(16.72%) patients started ART in state 4.

As expected the probabilities of transiting from good states to worse states increased with time (from state 1 to state 3 and 4 after 6 months is 0.023 and 0.004, after 12 months is 0.059 and 0.010 respectively). As time increase the probabilities of remaining in the same state is decreasing (probabilities of remaining in state 1 after 6, 12 and 18 months is 0.804, 0.698 and 0.633).

Estimated sojourn times for states 1, 2, 3 and 4 were 22 (40%), 8 (15%), 10 (18%) and 15 (27%) months respectively. As anticipated the conditional probability of staying in same state given number of month decreases with increasing time (conditional probabilities of remaining in state 2 after 6, 12 and 18 months is 0.547, 0.387 and 0.328). As expected the intensity indicates that the rate of transiting from good states to worst states is decreasing (the intensity of transiting from state 1 to 3 and 4 is $p<0.001$ and the intensity of transiting from state 2 to 3 and 4 is 0.04 and $p<0.001$).

The strongest predictor of transition from state 1 to 2 is TDF/3TC/EFV, which has a hazard ratio of 1.338 (with p value of 0.002). Patients who were prescribed TDF/3TC/EFV, are over 1.338 times more likely to transit from state 1 to state 2 than patients who did not receive TDF/3TC/EFV. A hazard ratio of 0.678 for the predictor variable female shows that female were less likely to transit from state 2 to 3 than their male counterparts. The hazard ratios of females from a bad state to a better state are more than 1, which is an indication that females are less likely to respond to treatment compared to males.

HIV can progress to AIDS without delay if there is no intervention. Early ART initiation is crucial at slowing prognosis of progressing from good states to worse states.

List of Conference(s) proceedings

Kashihalwa SP &Pazvakawambwa L (2019). Estimating Sojourn time and future clinical states of HIV patients under ART follow up in Namibia. Submitted to Frontiers in Public Health

MODELLING PROGRESSION OF HIV/AIDS DISEASE USING HOMOGENOUS SEMI-MARKOV PROCESSES: COHORT STUDY, NAMIBIA.

Presented at 3rd DELTAS Africa SSACAB Research Conference Intercontinental Hotel, Nairobi, Kenya, 10 -12th September 2018.

MODELLING PROGRESSION OF HIV/AIDS DISEASE USING HOMOGENOUS SEMI-MARKOV PROCESSES: COHORT STUDY, NAMIBIA.

Presented at Namibia Statistical Symposium, NIPAM Windhoek, Namibia, 19-20 November 2018

Table of Contents

List of tables

List of Figures

List of Acronyms

AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Treatment
ARV	Anti-Retroviral
CD4	Cluster of Differentiation 4
DF	Degree of Freedom
DNA	Deoxyribose Nucleic Acid
GEE	Generalized Estimating Equation
GLMM	Generalized Linear Mixed Model
HIV	Human Immunodeficiency Virus
HSMP	Homogenous Semi-Markov Processes
MoHSS	Ministry of Health and Social Services
MSM	Multistate Model
NAMPHIA	Namibia Population based HIV Impact Assessment
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitors
NRTI	Nucleoside Reverse Transcriptase Inhibitors
OI	Opportunistic Infection
PCP	Pneumocystis Pneumonia
RNA	Ribonucleic Acid
RTIs	Respiratory Tract Infections
SE	Standard Error
SMM	Semi-Markov Model
STI	Sexually Transmitted Infectious
TB	Pulmonary Tuberculosis
UNAIDS	United Nations Program on HIV/AIDS
WHO	World Health Organization

Acknowledgement

I would like to express my sincere gratitude to ‘Mama’ Dr Lillian Pazvakawambwa who served as my supervisor, gave me priceless guidance, relentless encouragement and meticulous supervision. Without her direction and supervision, this thesis would not have been possible.

I also thank ‘our prof” Lawrence Kazembe for his words of encouragement throughout the course. My vote of thanks also go to Dr Innocent Maposa who introduced me to the field of Biostatistics. I also thank Dr Dismas Ntirampeba (Komera Komera) and Mr Owen Mtambo for their valuable advice and guidance in using R software. My appreciation also goes to the Ministry of Health and Social Services staff for availing the data.

I forward my greatest admire and acknowledgment to the class of 2017 “MSc Biostatistics students” for the valuable, friendly and unlimited information we have shared. My sincere appreciation and thanks also go to my lecturers and the rest staff of the department of Statistics and Population Studies, UNAM, for their unreserved knowledge sharing and collaboration. My great respect and appreciation also goes to Mr C Jackson who developed the msm R package. I am thankful to my family members and last but not least Hilkka Ndeyapewa ‘WaNangha’ for being my loving life partner.

This study was supported through the DELTAS Africa Initiative SSACAB (Grant No. 107754/Z/15/Z). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and is supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author and not necessarily those of the AAS, NEPAD Agency, Wellcome Trust or the UK government.

Dedication

This thesis is dedicated to my late parents Mendofoli Velasia Pulamwenyo Hashiti and Tate Nanghambe Jeronium Yemwenemwene ya Kashihalwa. The desire to make you proud prompted me to complete this thesis.

Declarations

I, Simon Pombili Kashihalwa, 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/dissertation 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, Simon Pombili Kashihalwa, 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.

…… …… ….……

Name of Student Signature Date

CHAPTER ONE: INTRODUCTION

1.1 Background of the study

HIV/AIDS, is one of the leading causes of death in Namibia and worldwide (WHO, 2015). HIV/AIDS does not only have an enormous economic impact through lost productivity and medical care spending, but is also a major cause of disability and human suffering. It is important, therefore, to understand the natural history and etiology of HIV/AIDS (Laird, 2013). Further, since many chronic diseases are caused or made worse by modifiable factors such as diet and lifestyle, understanding factors affecting disease progression is critical. For a number of chronic diseases, the evolution is characterized by visits to clinically relevant and ordered stages. Examination of the sequence of visited stages and duration in each stage can enhance our understanding of the natural progression of the disease and how demographic and clinical factors may have an impact on disease evolution (Laird, 2013).

Infection by the Human Immunodeficiency Virus (HIV) gradually leads to the Acquired Immune Deficiency Syndrome (AIDS), and AIDS progress to death if not handled carefully. One may consider this progression of HIV infection to AIDS and then to death as a stochastic process (Goshu and Dessie, 2013). The disease evolution or progression can be broken up into a finite number of intermediate states, based on the immunological indicators namely $CD4^{+}$ count including death as one state (Janssen and Monica, 2001), which may offer a greater understanding and clarity of the evolution of disease.

Conceptually, the progression of HIV infection from an asymptomatic stage to AIDS, is associated with a gradual decline in the total number of CD4 cells in the blood. Biologically, the decrease in the total number of CD4 cells also correlates with an increase in the number of infected T cells and an increase in the amount of free virus in the blood (Martinelli *et al*., 2013). HIV/AIDS is ranked as one of the most destructive microbial scourges in human history and has posed formidable challenges to the biomedical research and public health communities of the world (WHO, 2016).

The probability that an HIV/AIDS patient transitions from one state to another depends on how long he/she has spent in that state. As time spent in each stage of the disease can't be predictable on the basis of clinical and immunological measures, this needs to be modeled by the semi-Markov stochastic process (Viladent and Ackere, 2007; Giuseppe, 2007). A semi-Markov process is defined as a stochastic process that can be in any one of possible states and that each time it enters a state it remains there for a random amount of time and then makes a transition into another state with some probability (Ross, 2007).

Stages of HIV/AIDS

The evolution of the virus in the human body, and the response of the body typically takes several years. According to World Health Organization (WHO, 2006) clinical staging of HIV/AIDS, HIV infection has four distinct stages. In order to diagnose an individual as being in a specific stage of HIV, the WHO developed a set of criteria that can be used worldwide.

Stage 1: Primary HIV infection

The first stage of HIV infection is called primary infection, patients in this stage have CD4 cell count of greater than 500 cells per cubic mm. This begins shortly after an individual is infected with HIV and lasts for a few weeks. Symptoms may include: fever (raised temperature), body rash, sore throat, swollen glands, joint aches and pains, muscle pain, which can happen because the body is reacting to the HIV virus. Cells that are infected with HIV are circulating throughout the blood system. The immune system, in response, tries to attack the virus by producing HIV antibodies. This process is called seroconversion. Timing varies but it can take up to a few months to complete.

Stage 2: The asymptomatic stage

During this stage patients have CD4 cell count between 200 and 499 cells/mm^3 during this stage of the disease, HIV continues to multiply in the body but at very low levels. People with chronic HIV infection may not have any HIV-related symptoms, but they can still spread HIV to others. Without treatment with HIV medicines, chronic HIV infection usually advances to AIDS. Symptoms can include moderate unexplained weight loss (<10% of presumed or measured body weight), recurrent respiratory tract infections (RTIs), sinusitis, bronchitis, otitis media, pharyngitis, herpes zoster, angular cheilitis, recurrent oral ulcerations, papular pruritic eruptions , seborrhoeic dermatitis, and fungal nail infections of fingers.

Stage 3: Symptomatic HIV infection

Patients in this stage have CD4 cell count of less than 200 cells/$mm^{3.}$ By the third stage of HIV infection there has been a lot of damage to the immune system. At this point, one is more likely to get serious infections or bacterial and fungal diseases. These infections are referred to as 'opportunistic infections'. The symptoms includes, severe weight loss (>10% of presumed or measured body weight), unexplained chronic diarrhoea for longer than one month, unexplained persistent fever (intermittent or constant for longer than one month) ,oral candidiasis, oral hairy leukoplakia, Pulmonary Tuberculosis (TB) diagnosed in last two years ,severe presumed bacterial infections (e.g. pneumonia, empyema, pyomyositis, bone or joint infection, meningitis, bacteraemia), acute necrotizing ulcerative stomatitis, gingivitis or periodontitis.

Stage 4: Progression from HIV to AIDS

The WHO clinical stage 4 (the severely symptomatic stage) designation includes all of the AIDS-defining illnesses. Clinical manifestations for stage 4 disease that allow presumptive diagnosis of AIDS to be made based on clinical findings alone are HIV wasting syndrome.

Pneumocystis pneumonia (PCP), recurrent severe or radiological bacterial pneumonia, extra pulmonary tuberculosis, HIV encephalopathy, CNS toxoplasmosis, chronic (more than 1 month) or orolabial herpes simplex infection, esophageal candidiasis, and Kaposi's sarcoma.

1.2 How HIV affects the human immune system

HIV first attaches to and penetrates its target cell. HIV releases ribonucleic acid (RNA), the genetic code of the virus, and an enzyme called reverse transcriptase into the cell. For the virus to replicate, its RNA must be converted into DNA using the viral RNA as a pattern. This reverses the pattern of human cells, which copy RNA from the pattern of human DNA (thus the term "retro" for "backward"). Reverse transcriptase performs this conversion from viral RNA to viral DNA. HIV mutates itself at this point, because reverse transcriptase is prone to errors during this conversion from viral RNA to viral DNA.

The viral DNA enters the cell's nucleus. Once inside, HIV finds the host cell DNA in the cell nucleus. HIV makes a copy of itself from the host cell DNA. With the help of another enzyme called integrase, the viral DNA becomes integrated with the host cell's DNA. The viral DNA often hides itself in the cell's DNA. Under the microscope, the cell's DNA appears normal, even though it is now mixed with HIV DNA (Health 24 HIV/AIDS, 2018).

Once safely hidden in the cell's DNA, the HIV can do one of two things. It can stay quietly in the cell, or it can turn on the cell's DNA and use the cell's machinery to make copies of itself. If it begins reproducing, it can make millions of new HIV. Some scientists think the virus makes a billion copies of itself a day. The HIV DNA now replicates and reproduces HIV RNA and proteins. A new virus is assembled from this RNA and short pieces of protein.

The virus buds through the membrane of the host cell, wrapping itself in a fragment of the cell membrane (envelope). To be able to infect other cells, the budded virus must mature. Another viral enzyme called protease performs this process by cutting and rearranging the viral proteins. These new viruses leave the cell and enter other CD4 cells and the process repeats itself (Kurle, 2008). HIV makes the immune system overactive, producing more and more cells.

Over time the immune system loses out and once it loses out the transition intensity and conditional probabilities of staying in the same state and transiting to better states will be very low. The sojourn time in good states will very low since the infection will be devastating. This is why without ART the CD4 count drops over time. ART blocks HIV from reproducing, and returns the immune system back to an almost normal state. The correlation between CD4 count and ART is demonstrated in the figure 1.

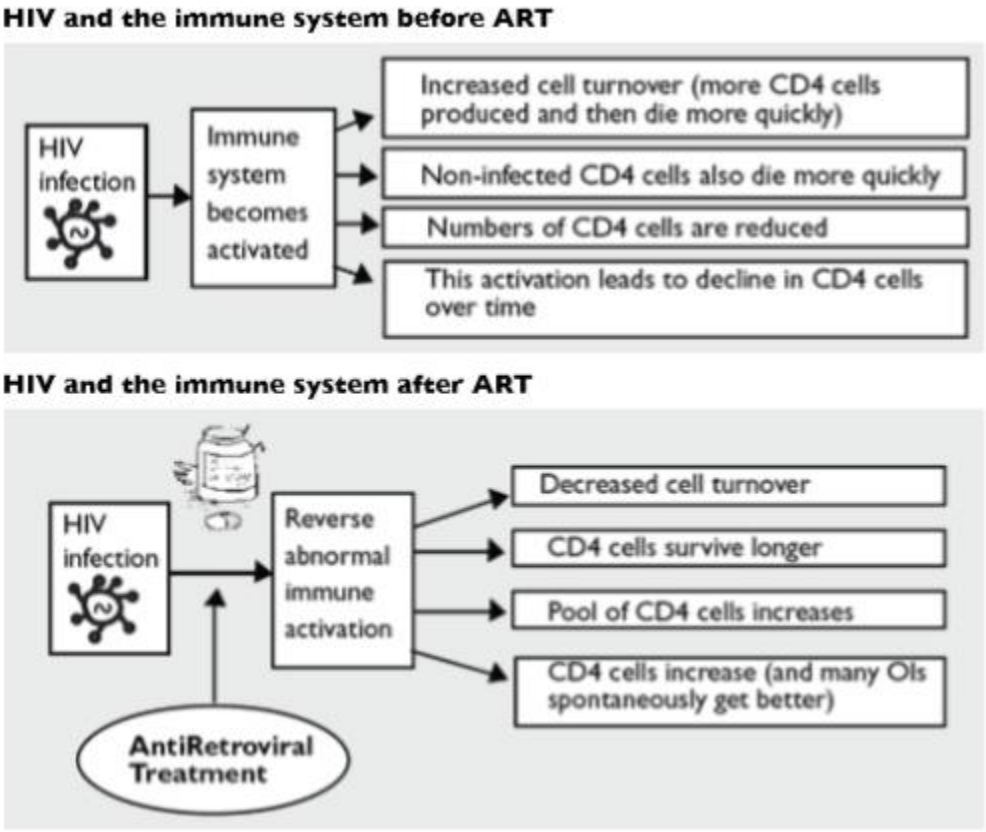

Figure 1 .Relationship between ART and CD4 count

(Source: http://i-base.info/wp-content/uploads/2016/01/Manual-2015-update-Section-1-–-pages-1-to17-pdf.pdf.pdf, 09/08/2018)

1.3 Statement of the problem

A fundamental characteristic of the immune system is its ability to expand rapidly the number of antigen-specific lymphocytes to combat pathogens. CD4 cells are cells of the immune system, which begin to deplete as the virus infects the body. These cells are considered as important biomarkers of disease progression for HIV infected individuals. Given the disease's direct relation to the immune system, CD4 cells are the primary indicator for prognostic information and a guide for antiretroviral therapy for HIV positive individuals (Février, 2011).

Ordinarily, the concern is with the number of CD4 cells for a group of patients on their first visit to the hospital to say something about the disease progress; however, since the number of CD4 cells at one time point is not very instructive to tell about the disease status; the change in the number of cells over time is a good indicator of disease condition. Antiretroviral Therapy (ART) fights against the progression of the disease by increasing the main body immune system or by decreasing RNA (Ribonucleic acid) concentration (viral load) in the blood content. Even though ART improves the immune system, there are many questions which arise about the numerical improvement (Werner *et al*., 2009). Among them how the CD4 cells count involve over time after patients initiated ART or do a change has different pattern depending on the patient's gender, educational level, and functional status etc. Generally, the research questions this study addressed are:

- How does the average progressions of CD4 cells for HIV positive patients following ART changes over time?
- Does the progression differ by patients demographic and clinical characteristics (age, sex and prescribed ART regimen)?
- What is the probability of a patient moving from one state to another state?

1.4 Objective of the study

General Objective:

The main objective of this study is to model the progression of HIV/AIDS disease of patients under ART follow-up in Namibia using homogenous semi-Markov processes.

Specific Objective

a) To explore the general average progression of CD4 cells over time.
b) To compare the progression of CD4 cells or the progression of HIV infection between groups (sex, age and prescribed ART regimen).
c) To estimate the duration a patient stay in each state.
d) To predict the future clinical state.
e) To recommend strategies for monitoring HIV/AIDS to policy makers.

1.5 Significance of the study

Although ART is delivered as part of a comprehensive care with emphasis on the reduction in HIV related morbidity and mortality, the observed changes are not sufficient enough compared to the desired goals of the ART, especially, in developing countries like Namibia where survival of HIV/AIDS patients on ART depends on a variety of factors, which may also vary greatly with economic, demographic, behavioral risk and health factors. According to the WHO (2013), global update on HIV treatment, Namibia is ranked among the top African countries in ARV delivery. Botswana is ranked the highest in Africa with 95% ARV coverage, followed by Namibia at 90 %.

In the absence of antiretroviral therapy (ART), it is well known that the progression of HIV to AIDS will instantly be devastating, the CD4 cells will gradually be depleted and the lifespan of a patient with HIV/AIDS markedly reduced.

No scientist or doctor has stepped forward to claim credit for discovering a vaccine to prevent AIDS nor is any vaccine expected for several more years, at a minimum (Duesberg, 1996).

However ART can delay the progression of the disease and improve the lifespan of patients. Therefore this study helps:-

a) To understand the importance of attending ART program , how much CD4 cells increase over time and how it delay disease progression
b) To provide confidence for regular HIV screening and avoid disease progression
c) To compare the different groups of patients how they respond to the drug concurrently

1.6 Organization of the thesis

The thesis is organized as followed. Chapter 2 provides the methodological review of Markov, semi-Markov and homogenous semi-Markov processes. Chapter 3 gives the methodology and data analysis. Chapter 4 & 5 highlight the results in line with the objective and discussions of the results. Conclusion and recommendations are presented in the last chapter.

CHAPTER TWO: METHODOLOGICAL REVIEW OF MARKOV AND SEMI MARKOV MODEL

2.1 Markov models

The progression of HIV infection to AIDS and then to death can be considered as a Markovan stochastic process. Markov model is defined as a multi-state model where the multi-state model is defined as a model for a stochastic process $(X(t), t\epsilon T)$ with a finite space (Dobson and Barnett, 2014)

$$S = \{s_1, s_2, .., s_m\}. \quad (2.1.1)$$

The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the process is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on *S*, specifies the starting state.

2.2 Multi state models

A multi-state model (MSM) is a model for a continuous time stochastic process allowing individuals to move among a finite number of states. In biomedical applications, the states might be based on biological markers (e.g. CD4 T-lymphocyte cell counts), some scale of the disease (stages of HIV infection). A change of state is called a transition, or an event. States can be transient or absorbing, if no transitions can emerge from the state (for example, death) (Meira-Machado *et al*,..2009).

2.3 Review of homogenous semi-Markov application in HIV/AIDS progression

Markov chains and semi-Markov processes are very important classes of stochastic processes with many applications in science, engineering and beyond. A Markov chain is a stochastic process, but it differs from a general stochastic process in that a Markov chain must be "memory-less". A Markov chain is a mathematical system that experiences transitions from one state to another according to certain probabilistic rules. The defining characteristic of a Markov chain is that the probability of transitioning to any particular state is dependent solely on the current state (Babu and Limnios, 2008). Homogeneous semi-Markov processes (HSMP) were introduced in the 1950s, independently by Levy (1954) and Smith (1955), with the objective of generalizing Markov processes.

In a Markov process environment, the waiting time distribution functions in each state must be exponential, whereas in a semi Markov process environment these distributions can be of any type. This study will deal with semi-Markov stochastic models applied in a clinical field. These processes turn out to be a very efficient tool for predicting the dynamic evolution of human immunodeficiency virus (HIV) infection and the probability of an infected patient's survival. This approach has the following advantages with respect to traditional epidemiological models (according to Di Biase *et al*., 2007):

- has an arbitrary number of states, linked to the seriousness of the infection;
- all transitions between states are allowed;
- consider the randomness of the evolution between all states, as well as the stochastic time spent in each state before a transition occurs;
- model parameters are directly estimated from raw data;

- all the states are interrelated, therefore any improvements are also considered;
- a large number of disease states can be considered;

Giuseppe *et al*. (2007) analyzed HIV/AIDS dynamic evolution as defined by CD4 levels from a macroscopic point of view by means of homogeneous semi-Markov stochastic processes. According to the study, an AIDS patient may be in state *j* after a time *t* given that he/she entered at time 0 (starting time) in state *i*. They found that the survival probability of AIDS patient up to month *t* given that his/her current status decreased with increasing of time. And, the survival probabilities up to month *t* increased when CD4 count increased.

HIV evolutions through two different temporal scales were evaluated using non-homogeneous Semi-Markov models by D'amico *et al*. (2009). The study was carried out from a macroscopic point of view by means of three different stochastic models. The analyses were performed through non-homogeneous semi-Markov processes. They referred 2159 subjects enrolled in Italian public structures from September 1983 to 15 January 2006. They computed conditional probabilities of survival. As a result, the probabilities of survival of AIDS patients depended on their ages, CD4 count and the elapsing of time.

A study on risk factor for progression to AIDS and mortality post-HIV infection using illness-death multistate model was carried out by Hamidi *et al.* (2017). The results indicated that using ART TB co-infection are two top most important variables in predicting cumulative incidence function for AIDS progression in the presence of competing risk, respectively.

The patient with TB had much higher predicted cumulative incidence probability. Predicted cumulative incidence probability of AIDS progression was also higher for mother to child mode of HIV transmission.

Seyoum *et al*. (2016) did a study on Predicting AIDS disease progression using longitudinal CD4 count among adult HIV/AIDS patients in Southwest Ethiopia: Application of semi-Markov process. The results indicated that of death observed from the state I, II, III, and IV was 3, 4, 15, and 40 respectively during the study period. The probability of dying was increased from the worse transition states. The probability of being found in state I after started the treatment at any other working state is higher. Reliability plot revealed that the probability of surviving 200 month, from state I, state II, state III and state IV, estimated as 0.71, 0.68, 0.63 and 0.58 respectively.

The probability of remaining at the starting CD4 count state was decreased when time increased, patients from the state I has higher probability to remain in the ART starting state.

More over Shoko and Chikobvu (2018) applied Homogenous Markov process to HIV/AIDS progression under a combination treatment therapy, South Africa. The results indicated that the strongest predictor of transition from a state of CD4 cell count greater than 750 to a state of CD4 between 500 and 750 is a negative reaction to drug therapy. Development of TB during the course of treatment is the greatest predictor of transitions to states of lower CD4 cell count.

Goshu and Asena (2017) applied semi-Markov models to the HIV/AIDS disease progression and compared two sojourn time distributions, namely, exponential and Weibull probability distributions. Using data obtained from 370 HIV/AIDS patients who were under a clinical follow up from September 2008 to August 2015, Ethiopia. The results of the study showed that within the "good" states, the transition probability of moving from a given state to the next worse state has a parabolic pattern - increases with time until it gets optimum and then declines over time.

In addition some studies like Goshu and Dessie, (2013), Masala *et al.* (2014), Zelalem (2010), Kelkile (2016) and Dessie (2014) used dynamical models to estimate the proportion of individuals changing their status at each time step. In all the papers the evolution of the HIV/AIDS was analyzed in a homogeneous Semi-Markov framework using immunological markers. They also concluded that, since the probability of a patient's transition from one state to another depended on how long he has spent in this state, the semi-Markov Models (SMM) was an appropriate to model HIV/AIDS evolution.

2.4 Semi-Markov process

The Markov assumption state that the future progress only depends on the current state not on the past states and the current state should include all relevant history. This assumption imposes restrictions on the distribution of the sojourn time in a state, which should be exponentially distributed in case of continuous-time Markov process and geometrically distributed in case of a discrete-time Markov process.

To overcome this, the Markov assumption must be relax in order to allow arbitrarily distributed sojourn times in any state and still have the Markov assumption but in a more flexible manner. The resulted process based on these two properties is called a semi-Markov process.

A semi-Markov process is concerned with the random variables that describe the state of the process at some time and it is also a generalization of the Markov process. A semi-Markov process is a process that makes transitions from state to state like a Markov process, however the amount of time spent in each state before a transition to the next state occurs is an arbitrary random variable that depends on the next state the process will enter (Ibe, 2009).

The semi-Markov chain can be described as follows;

- The initial state s_0 is chosen according to the initial distribution λ,
- Then next visited state s_1 is determined according to the transition probability matrix p
- And the chain stays in state s_0 for a time t determined by the sojourn time distribution in state i_0 before going to state s_1

2.5 Time homogenous Semi-Markov Processes

Homogeneous semi-Markov processes (HSMP) were introduced in the 1950s, independently by Levy (1954) and Smith (1955), with the objective of generalizing Markov processes. Giuseppe *et al*. (2007) defines homogenous semi-Markov process (HSMP) model as follows:

Let $X_n : \Omega \to S$ be a stochastic process with state space $S = \{S_1, S_2, \dots, S_m\}$ and $T_n : \Omega \to \mathbb{R}$ be the time of the n^{th} transition, with Ω domain of the process and $\mathbb{R}$ set of real numbers. Here the time is a random variable. The kernel $Q = [Q_{ij}]$ associated with the process and the transition probability P_{ij} of the embedded Markov chain is defined as follows:

$$Q_{ij}(t) = P[T_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i] \quad (2.5.1)$$

$$P_{ij} = \lim_{t \to \infty} Q_{ij}(t) \quad (2.5.2)$$

Define the probability that the process will leave a state *i* in a time *t* as

$$H_i(t) = P[T_{n+1} - T_n \leq t | X_n = i] = \sum_{j=1}^{m} Q_{ij}(t) \quad (2.5.3)$$

The distribution of waiting time in each state *i*, is given that the state *j* is subsequently occupied is

$$G_{ij}(t) = P[T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j], \quad (2.5.4)$$

which can be computed as:

$$G_{ij}(t) = \begin{cases} \frac{Q_{ij(t)}}{P_{ij}}, & if P_{ij} \neq 0 \\ 1, & if P_{ij} = 0 \end{cases} \quad (2.5.5)$$

For any homogenous semi-Markov process $\{X(t), t \geq 0\}$, the transition probabilities are given by (2.4.6) for which the solution should be obtained using the progression (2.4.7).

$$\emptyset_{(ij)}(t) = P[X(t) = j | X(0) = i], \quad (2.5.6)$$

$$\emptyset_{ij}(t) = (1 - H_i(t))\delta_{ij} + \sum_{l=1}^{m} \int_0^t Q_{il}(\tau)\phi_{lj}(t - \tau)\, d\tau \quad (2.5.7)$$

Here δ_{ij} represents the kronecker deltaδ.

The variables involved are the following:

m= number of states of HSMP, which 5 in this case.

T = number of periods to be examined for the transient analysis of HSMP.

P = matrix of order m of the embedded Markov chain in HSMP.

G^T= square lower-triangular block matrix order T +1 whose blocks are of order m.

Q^T = kernel of SMP.

Φ^T = block vector of order T + 1 the block which are square matrices of order m.

D^T = block vector of order T + 1 the block which are the diagonal square matrix of order m.

V^T = square lower-triangular block matrix order T + 1 whose blocks are of order m.

S^T = block vector of order T+1 the block which are the diagonal square matrix of order *m*. The diagonal element of each block *t* are $s_{ii} = \sum_{j=1}^{m} Q_{ij}(t)$.

2.6 Time non-homogenous semi-Markov model

Even though this study focused on homogenous semi-Markov model a brief description of the non-homogeneous model is given. Non homogenous semi-Markov is used when there is an underlying reason for transition rates to change with time or age. Unlike homogenous semi-Markov model, the non-homogenous semi-Markov model does not assume constant transition intensity.

2.7 Stochastic process

A first order Markov process, *X (t)* state that a stochastic process in which future knowledge about the process is provided only by the current state and is not altered with the additional knowledge of past states. This means that, the future state is independent of the past given the present state of the process (Ibe, 2009).

That is,

$$P = [X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, .., X(t_0) = x_0]$$

$$= P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}] \qquad (2.7.1)$$

In stochastic process the system enters a state, spends an amount of time called the sojourn time and then moves to another state where it spends another sojourn time, and so on.

A stochastic process changes over time in an uncertain manner and its model (that is stochastic model) has five components such as time *t*, number of observation *n*, state *S*, activity (which depends on time), transition and stochastic process (a collection of random variables *X (t)).*

The time can be either continuous or discrete parameter. The random variable in stochastic process is denoted by *X (t)* and it represents the measurement that has been observed at the particular state at a given time for the particular subject.

All the possible random variables *X (t)* of stochastic process that are assumed are collected in a state space S where

$$S=\{s_1, s_2, \ s_3, \dots, s_m\} \tag{2.7.2}$$

If *S* is discrete, the then process is called a discrete-state stochastic process. Similarly if *S* is continuous, then the process is called a continuous-state stochastic process. The set of parameters of the stochastic process is denoted by *T* and it is usually a set of times. If *T*, is a countable set then the process is called a discrete-time stochastic process. If *T*, is an interval of real numbers then the process is called continuous-time stochastic process. If the Markov process is a discrete-time Markov process then the transitions occur at fixed points in time and we consider transition probabilities and if the Markov process is a continuous-time Markov process then the transitions can occur at any point in time, with transition rates (Ibe, 2013).

To describe the Markov process let *S* defined above represent a set of states then

- The process moves successively from one state to another state having started in one of these states.
- If the process is currently in state i, then it moves to state *j* with a transition probability of P_{ij}. The probability does not depend upon which states the process was in before the current state.
- The process can remain in the state it is in and this occurs with probability, P_{ii}

- The starting state defined in *S* is specified by the initial probability distribution

The absorbing state *i* of a Markov process is the state *i* in which the process will never leave that state. In a Markov process the state that is not absorbing is called transient. The recurrence time is the first time t at which the process has returned to its initial state.

2.8 Probability transition matrix

The transition probability matrix is the *M* x *M* matrix whose entry in row *i* and column *j* is the transition probability *P (t)* and is denoted by

$$P(t) = \begin{bmatrix} P_{11}(t) & \cdots & P_{1M}(t) \\ \vdots & \ddots & \vdots \\ P_{M1}(t) & \cdots & P_{MM}(t) \end{bmatrix} \qquad (2.8.1)$$

P (t) denote transition probability matrix of a multi-state process at time *t*. In the transition matrix *P (t):*

- the rows represent now, or from(X_t)
- the columns represent next, or to(X_{t+1})
- entry (*i*,*j*) is the conditional probability that next= *j*, given that now = *i* : the probability of going from state *i* to state *j*

$$p_{ij} = P(X_{t+1} = j) | X_{t=}i \qquad (2.8.2)$$

P (t) is a square matrix (*M* × *M*), because $X_{t+1}$and X_t both take values in the same state space S (of size *M*). The transition probability matrix (2.7.1) is a stochastic matrix because for any row *i*,

$$\sum_{j=1}^{M} p_{ij} = \sum_{j=1}^{M} P(X_{t+1} = j \mid X_t = i) = \sum_{j=1}^{M} P_{\{X_t\}}(X_{t+1} = j) = 1 \qquad (2.8.3)$$

The column of *P (t)* do not in general sum to 1, if it does then the matrix *P (t)* is called doubly stochastic matrix. The entries of probability transition matrix are transition/movement probabilities of subjects through states. The transition is the movement from one state to another.

The matrix *P* is time dependent and to accentuate that is denoted by *P* (*t*). In every transition probability matrix, the probabilities must be greater than or equal to zero, and each row must sum to one.

$$P_{ij} \geq 0 \; for \; all \; i,j \; \in \{1,..,M\} \tag{2.8.4}$$

2.9 Transition intensity matrix

The intensity between two states *i* and *j*, is the rate of change of the probability P_{ij} in a very small time interval Δt. Define as

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{P\{X(t,t+\Delta t)=j \mid X(t)=i,F_t\}}{\Delta t} \qquad i \neq j \tag{2.9.1}$$

For any given time$\{t: 0 < t < T\}$ $and \; interval \; length \; \Delta t > 0$. An intensity matrix Q contains all possible intensities between the various states.

The intensities may also depend on the time of the process *t* or time-varying explanatory variables F_t . For example, an outcome containing *M* states would have the following intensity matrix, *Q*.

$$Q(\lambda) = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1M} \\ \vdots & \ddots & \vdots \\ \lambda_{M1} & \cdots & \lambda_{MM} \end{bmatrix} \tag{2.9.2}$$

The parameter λ represents independent parameters and it is a vector of length *b*.

$Q(\lambda)$ denote transition intensity matrix of a multi-state process. The transition intensity matrix is used to define the multi-state model. The transition intensity matrix again is also used to calculate the transition probability matrix.

The elements in each row of the transition intensity matrix must sum to zero and off diagonal elements must be non-negative that is

$$\sum_{j=1}^{M} \lambda_{ij} = 0 \tag{2.9.3}$$

And $\lambda_{ij} \geq 0 \; for \; i \neq j$ respectively. The elements in diagonal must be negative for all *i* is

equal to *j* that is

$$\lambda_{ij} = -\sum_{i \neq j} \lambda_{ij} \;\; for \; i = 1, \dots, M. \tag{2.9.4}$$

This implies that subjects in those states remain in their state while the off diagonals are rates in which subjects move to other states. The *Q* matrix (2.8.2) is called the transition intensity or rate matrix where each element (λ_{ij}) represent rate at which transitions are made from state *i* to state *j*.

2.10 Sojourn time

When the process enters state *i*, the time it spends there before moving to another state is called the holding time in state *i* or the sojourn time. (Zare *et al.,* 2014).The sojourn time of a process *X* in a subset of states will be an integer-valued random variable if *X* is a chain or real-valued one in the case of a continuous-time process.

The sojourn times of a continuous-time Markov process in a state *j* are independent, exponential random variables with mean

$$-\frac{1}{\lambda_{ii}} \tag{2.1.1}$$

or rate given by $-\lambda_{ii}$. The other remaining elements of the *i*th row of transition intensity matrix (2.7.1) are proportional to the probabilities governing the next state after *i* to which the individual makes a transition. The probability that the subject's next move from state *i* to state *j* is

$$-\frac{\lambda_{ij}}{\lambda_{ii}} \tag{2.1.2}$$

The sojourn time and the new state depend only on state *i* and not on the history of the system prior to time *t.* Given that the current state is *i*, the sojourn time and the new state are independent of each other.

2.11 Model assumption

Different model assumptions can be made about the dependence of the transition rates on time (Meira-Macado & Roca-Pardinas, 2011). Markov property and the homogeneity assumptions are strong assumptions which may lead to biased estimates if violated, consequently it is significant to assess and further examine a multi-state model once it has been fitted.

2.12 Markov model assumption

The Markov assumption state that the future progress only depends on the current state not on the past states. This means that the transition times from each state are independent of the history of the process prior to entry to that state. The past history of a system plays no role in its future evolution, which is usually known as the "memoryless property of a Markov process" (Barbu & Limnios ,2008).

2.13 Semi-Markov assumption

The semi- Markov assumption state that the future progress not only depends on the current state *i*, but also on the entry time into the current state *j* (Macado & Pardinas, 2011). The definition (2.8.1) under this assumption can be simplified as

$$\lambda_{ij}(t, F_t) = \lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P\{X(t,t+\Delta t)=j | X(t)=i\}}{\Delta t} \quad (2.13.1)$$

where $\lambda_{ij}(t, F_t)$is the transition rate of a multi-state process or instantaneous hazard/risk rate of progressing from state *i* to state *j* at time *t*, given the history F_t .

2.14 Time homogeneous assumption

In time homogeneous Markov models, all transition intensities are assumed to be constant as functions of time, independent of time *t* (Zare *et al.*, 2014). This assumption can be assessed with a likelihood ratio test. When intensities are treated as being time homogeneous then the dependency on time can be removed. The transition probability matrix and transition intensity matrix form the building block of Kolmogorov equations that are used to yield unique solutions for probability matrix *P* (*t*).

2.15 Kolmogorov equations

The Kolmogorov equations are used to derive the correlation among the transition intensity matrix *Q* and the transition probability matrix *P*. The transition probabilities can be calculated from the intensities by solving the Kolmogorov differential equation. The relationship between the transition intensity and probability matrix involves canonical decomposition. The canonical decomposition was discussed by Kalbeisch and Lawless (1985).

The Kolmogorov equations state that

$$\frac{\partial}{\partial t} P(t) = P(t)\, Q, \tag{2.15.1}$$

which yield a unique for *P (t)* and condition on P (0) =I,

$$P(t) = e^{Qt} = \sum_{r=0}^{\infty} \frac{(Qt)^r}{r!} \tag{2.15.2}$$

Definition (2.14.2) is only valid with time homogeneous intensities. *Q* is the transition intensity matrix therefore *P* can be found from *Q* using Kolmogorov equations.

2.16 Maximum likelihood estimation

The method of maximum likelihood estimation enables the unknown parameters in the model to be estimated. The maximum likelihood estimate is the number of transitions from state *i* to state *j* divided by number of overall transitions from state *i* to other states calculated from the transition probability matrix. Maximum likelihood estimates for a particular class of a model can be computed from transition probability matrix *P* (*t*), with (*ij*) entry through the Kolmogorov relationship $P(t) = \exp(tQ)$ (Cox and Miller, 1965).

2.17 Covariates

Explanatory variables can be included at each level of the model through generalized regressions in order to incorporate covariates (Dantony *et al.*, 2016). Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. Variables associated with transition intensities are assumed to have a multiplicative effect.

Each transition intensity can have a separate set of covariate effects. These effects are introduced as covariates in the model via the transition intensities. They are included in the model by assuming that the transition intensities are functions of the covariates of interest and are of the form

$$\lambda_{ij} = e^{z^T \beta_{ij}}, \quad i \neq j \tag{2.17.1}$$

where z is a vector of covariates and β_{ij} is the vector of regression coefficients corresponding to z.

2.18 Diagnostics for Model Assessment

The assumptions of time homogeneity and the Markov assumption are assumptions which need to be assessed as an incorrect application can lead to bias (Meira-Machado *et al*., 2009). The Markov models have the following assumptions that need to assessed or corroborated:

- Homogeneity of the transition rates through time
- Homogeneity of the transition rates across the subject population
- The Markov property or assumption

2.19 Homogeneity of the transition rates through time

The characteristic of time homogenous Markov models is that the transition intensities remain constant through time. This assumption can be tested by fitting a piecewise constant intensities model and thereafter using a likelihood ratio test as a test for time independence. The likelihood ratio test is used to test the assumption of constant rates through time and again the likelihood ratio test can be used to compare the piecewise constant model with homogenous model. As the alternative, Kalbfleisch and Lawless (1985) suggest the fitting of parametric time-dependent model

$$\lambda_{ij} = \lambda_{ij} e^{-\beta t} \tag{2.19.1}$$

The likelihood ratio test is performed on the hypothesis that

$$H_0 : \beta = 0 \ , \tag{2.19.2}$$

to assess the homogeneity of the transition rates through time. If the null hypothesis is not rejected based on p-value then we will conclude that intensities are constant across time, implying that the assumption of homogeneity of transition intensities across time is valid.

2.20 Homogeneity of the transition rates across the subject population

This assumption can be checked by including covariates and treatment indicators

$$x = (x_1, .., x_p)^T \tag{2.20.1}$$

which can be reparametrized as,

$$\lambda_{ij}^x = \lambda_{ij} e^{-\beta_{ij} x} \tag{2.20.2}$$

where *x* is a binary variable with 0 and 1 values that is equivalent of dividing the population into two groups according to its value and (i, j)=1,2, …, *M* denote transitions rates in the model.

The likelihood ratio test can be used to test

$$H_0: \beta_{ij} = 0 \tag{2.20.3}$$

This can be used to test if the transition rates differ with regard to the two population groups. An overall likelihood ratio test of homogeneity can be obtained by comparing the overall log-likelihood with the sum of the log-likelihoods obtained from the two subpopulation groups. If there is a significant difference between the two population groups then assumption of homogeneity of the transition intensities across population groups has been violated.

If no significant difference is found between the two groups in terms of transition intensities, assumption of homogeneity is valid.

2.21 Markov Assumption

The Markov assumption that the future state of a process depends only on the present state, independent of the past. A method suggested by Kay (1986) involves creating data for the exact transition times between states using interpolation. A test can then be performed on this completed dataset to test the Markov assumption.

Assuming an illness-death model allowing recovery with states: (1) healthy, (2) illness and (3) death, let *x* denote the time spent in state 2 from the most recent transition from state1. Kay (1986) proposed fitting a model where the transition intensity λ_{23} is given by

$$\lambda_{23} = \lambda_0 e^{\beta x} \tag{2.21.1}$$

after testing the hypothesis $H_0: \beta = 0$ which would assess the Markov assumption that the transition intensity to death from state 2 is unaffected by the previous sojourn time(*x*-1).

CHAPTER THREE: RESEARCH METHODS

3.1 Description of the Study Area and Population

Namibia is a country in southern Africa whose western border is the Atlantic Ocean. It shares land borders with Zambia and Angola to the north, Botswana to the east and South Africa to the south and east. Namibia has a population of 2.6 million people. HIV/AIDS in Namibia is a critical public health issue. HIV has been the leading cause of death in Namibia since 1996, but its prevalence has dropped by over 70 percent in the last 10 years (MoHSS, 2017).

UNAIDS (2017) reported that there were 210,000 people living with HIV in Namibia. Presently ART services have been rolled out countrywide and are available at all 35 district hospitals as well as at all health centers and most clinics. The data from all 35 district hospitals, health centers and clinics on CD4 counts of HIV/AIDS patients who initiated treatment between January 2008 and January 2012 and were followed till December 2017 was obtained from Ministry of Health and Social Services, Namibia. All HIV patients who measured their $CD4^+$ T cells at least once during the study period constitutes the study population.

3.2 Research Design

This study used secondary data from Ministry of Health and Social Services of HIV positive patients who initiated ART between January 2008 and January 2012 and were followed till December 2017. All patients who measured their CD4 cells at least once constitutes the study population, the study was a retrospective longitudinal study.

3.3 Procedure

The study was a retrospective longitudinal study that was done on patients under the follow up of ART in Namibia. In this study, the sampling frame is the list of all who started the ART between January 2008 and January 2012 and who have a unique ART identification number. The dynamic evolution of HIV infection was modeled as a sequence of different states based on the severity of the infection.

The data on CD4 counts of HIV/AIDS patients who initiated treatment between January 2008 and January 2012 and were followed till December 2017 was obtained from Ministry of Health and Social Services, Namibia.

The four states of the Markov process of the seriousness of HIV/AIDS sickness based on the CD4 counts of a patient are defined as: S1: CD4 count > 500 cells/microliter. S2: 350 < CD4 count ≤ 500 cells/microliter. S3: 200 < CD4 count ≤ 350 cells/microliter. S4: CD4 count ≤ 200 cells/microliter (Giuseppe *et al.*, 2007). Figure 2 shows all the immunological states a HIV infected patient can go into. All the states are inter-related.

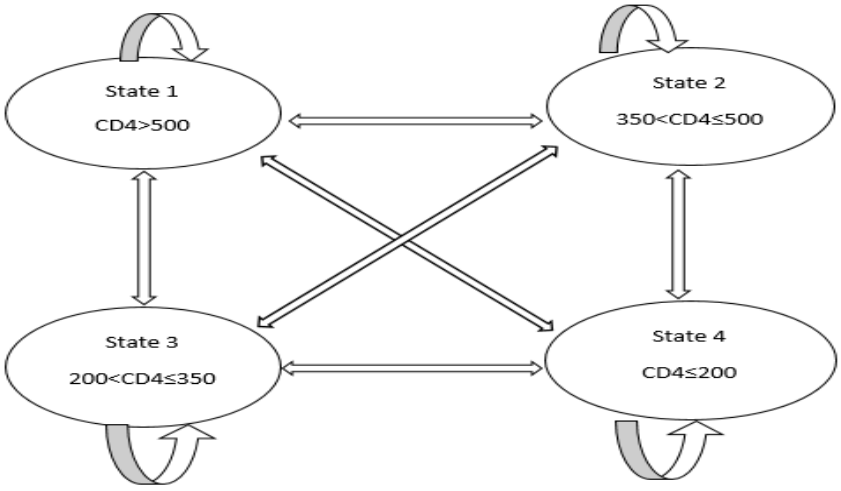

Figure 2 .Immunological state a HIV infected patient can go into

3.4 Data description

The data included the following variables: age, gender, stage of HIV infection at diagnosis, date of HIV infection and duration on ART. All registered patients with determined HIV infection and who measured their CD4 count for at least once constituted the study, irrespective of age, gender, stage of disease and date of diagnosis. Pre-processing of data was done and fields with spelling error, other irregularities and irrelevancies like outliers were corrected or removed.

3.5 Variables in the study

Table 1 summarized the variable used in the study and the way the variable were coded.

Table 1: Variable description

Variable	Coding
State(CD4)	1=CD4>500, 2=350<CD4≤500, 3=200<CD4≤350, 4= CD4≤200
Age	age of HIV patient on ART (continuous)
Age *	1= <25, 2 = 25-49, 3 = >50
Sex	Male=0 , Female=1
Prescribe ART regimen	1=AZT/3TC/EFV, 2=AZT/3TC/LPV, 3=TDF/3TC/EFV, 4=TDF/FTC/EFV, 5=TDF/3TC/NVP, 6=Other
Duration	Time on ART(in months)
Observation	1= first observation, 0=repeated observation

* Age category used for descriptive purpose

3.6 Data Analysis

Measure of frequency were used to describe the basic features of the data in the study. They provide simple summaries about the sample and the measures and were generated in excel.

Jackson (2011) developed the R package *msm*, implementing several functions for fitting continuous-time Markov to longitudinal data. The *msm* package provides several numerical outputs such as time spent in each state, transition probabilities and transition intensity rate.

Data often consist of observations of the process at arbitrary times, so that the exact times when the state changes are unobserved. The data are specified as a series of observations, grouped by patient. The msm function is given as

msm (state ~ duration, data=list (),…) (3.6.1)

where state is the observed state, duration is the time a patient has been on ART and data is a data frame list containing the variables in the model. Explanatory variables for a particular transition intensity can be investigated by modelling the intensity as a function of these variables.

Different model assumptions can be made about the dependence of the transition rates on time. These include:

1. Time homogeneous models: the intensities are constant over time, that is, independent of *t*.
2. Markov models: the transition intensities only depend on the history of the process through the current state.
3. Semi-Markov models: future evolution not only depends on the current state *i*, but also on the entry time t_i into state *i*.

Formulation of the continuous homogeneous semi Markov model is done by considering transition probabilities over narrow interval of time Δt. In this study $\Delta t = ½$ months making it appropriate to assume that transition rates over these intervals are constant (Shoko and Chikobvu, 2018). These transition rates, also known as transition intensities, are the essential concept in continuous semi-Markov processes. They can take values greater than 1, unlike transition probabilities.

At any time $t + \Delta t$, the state of an HIV-infected individual is defined based on the CD4 cell count level as follow: S1: CD4 count > 500 cells/microliter; S2: 350 < CD4 count ≤ 500 cells/microliter; S3: 200 < CD4 count ≤ 350 cells/microliter and S4: CD4 count ≤ 200 cells/microliter.

Markov models are favorable to the modelling of diseases in particular cases where the disease is grouped into a set of exhaustive and mutually exclusive health states, thereby forming a multi-state model (Abner *et al*, 2014). All analysis was done using the package 'msm' for multistate modelling in R software developed by Jackson (2011).

3.7 Research Ethics

The ethical clearance was granted by the Research and Publication Centre of University of Namibia, and research permission from the Centre for Postgraduate. Clearance was also obtained from Ministry of Health and Social Services.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Descriptive analysis

The study used data from MoHSS, with 2422 HIV patients on anti-retroviral therapy (ART) who were observed 11028 times (Table 3). Table 2 shows that 7489 (67.9%) were females and 785 (32.41%) were males , 657(27.13%) patients started ART in state 1, 683(28.19%) patients started ART in state 2, 677(27.95%) patients started ART in state 3 and 405(16.72%) patients started ART in state 4, at treatment commencement ($t = 0$).

Table 2. Proportion of male and female patients at the commencement of ART

State	Sex, n (%)		Total
	Male	Female	
1	124(5.12)	533(22.01)	657(27.13)
2	220((9.08)	463(19.11)	683(28.19)
3	246(10.16)	431(17.79)	677(27.95)
4	195(8.05)	210(8.67)	405(16.72)

Table 3 shows that the highest observation were recorded in the age category of 25-49. Female have the highest observation in all states except for state 4. The highest observed prescribed ART regimen in state 1 and 3 is TDF/3TC/NVP, the highest in state 2 is AZT/3TC/EFV and the highest in state 4 is AZT/3TC/LPV.

Table 3. Variable description

Variable	State, n = 11028 (%)				Total (n)
	1	2	3	4	
Age*					
<25	135 (50.0)	62 (23.0)	41(15.2)	32(11.9)	270
25-49	3985 (40.3)	2928 (29.6)	2135 (21.6)	837 (8.5)	9885
=>50	253 (29.0)	240 (27.5)	241 (27.6)	139 (15.9)	873
Sex					
Male	887 (25.1)	1105 (31.2)	1037 (29.3)	510 (14.4)	3539
Female	3486 (46.5)	2125 (28.4)	1380 (18.4)	498 (6.6)	7489
Prescribed ART regimen					
AZT/3TC/EFV	537 (43.2)	316 (25.4)	247 (19.9)	142 (11.4)	1242
AZT/3TC/LPV	1302 (37.8)	1057 (30.7)	796 (23.1)	290 (8.4)	3445
TDF/3TC/EFV	545 (36.1)	389 (25.8)	387 (25.7)	187 (12.4)	1508
TDF/FTC/EFV	557 (41.7)	381 (28.5)	291 (21.8)	108 (8.1)	1337
TDF/3TC/NVP	1421 (40.9)	1081 (31.1)	694 (20.0)	279 (8.0)	3475
Other	11 (52.4)	6 (28.6)	2 (9.5)	2 (9.5)	21

Note: n is number of times patients has been observed. TDF = tenofovir, AZT= azidothymidine, FTC = emtricitabine, EFV = efavirenz, 3TC = lamivudine, NVP = nevirapine, OTHER = abacavi (ABC) and stavudine (D4T)

4.2 Clinical Progression of HIV/AIDS Disease

This study considered that an infected patient can move among the immunological marker stages related to CD4 count. Patient who started treatment under any state has a likelihood to reach any other state. If there is an improvement on CD4 count, the patient has a recovery from the initial state and can transit to a better state. The transition of the patient in different state occurs at any time. Table 4 summarizes transition counts that took place for the whole period of the study.

Table 4. Transition counts

	State 1	State 2	State 3	State 4
State 1	2547	616	99	26
State 2	917	1193	414	35
State 3	211	666	928	161
State 4	41	72	299	381

Table 4 shows that, transition counts from state i to j are higher for all the values in which $i = j$. In the followed up period, 3288, 2559, 1966 and 793 transitions had already been from state 1, 2, 3, and 4, respectively. Twenty six patients transited to state 4 from state 1 while 41 left state 4 to state 1. The time homogeneous model was fitted to the data to assess the effectiveness of the treatment by comparing the forward transition and the reverse transitions.

Table 5 shows the estimated transition probability, patient from state 1, 2 and 3 transit to state 4 with probability p<0.001, p<0.001 and 0.018, respectively. Patients show improvement from state 4 to; state 3, state 2 and state 1 with probability of 0.060, 0.002 and p<0.001, respectively. Patients show improvement from state 3 to 2, from state 3 to 1 and from state 2 to 1 with probability 0.070, 0.003 and 0.071, respectively.

Table 5. Estimated Transition Probability Matrix

	State 1	State 2	State 3	State 4
State 1	0.957	0.040	p<0.001	p<0.001
State 2	0.070	0.887	0.041	p<0.001
State 3	0.003	0.071	0.909	0.018
State 4	p<0.001	0.002	0.060	0.937

"p" is the probability value

The solution of the evolution equation is presented for specific month in table 6. It represents the probability that an HIV positive patient being at time 0 in state *i* will be after *t* months, in the state *j*. Table 3, indicate the probability of a patient starting from state *i* at time zero, will do a transition after month *t* to state *j*. The conditional probability of a patient starting from state 4 at time zero, and transiting to state 3, 2 and 1 after 2 years is 0.328, 0.227 and 0.162 respectively. A patient being in state 4 at time zero, stay in same state after 2 years with probability 0.288.

The probabilities of direct transition from state 1 to state 2, state 2 to state 3 and state 3 to state 4 after 4 years are estimated to be 0.284, 0.172 and 0.077 respectively. As *t* increases, the probability of the patient transiting to a next worse state is increasing while the probability to remain in the same state is decreasing.

Table 6. The solution of the evolution equation for month *t*

Transition	Month=6	Month=12	Month=18	Month=24	Month=30	Month=36	Month=42	Months=48
1→1	0.804	0.698	0.633	0.592	0.563	0.543	0.528	0.518
1→2	0.168	0.233	0.260	0.273	0.278	0.281	0.283	0.284
1→3	0.023	0.059	0.089	0.118	0.128	0.139	0.148	0.154
1→4	0.004	0.010	0.017	0.024	0.030	0.036	0.040	0.044
2→1	0.292	0.405	0.451	0.471	0.479	0.484	0.486	0.486
2→2	0.547	0.387	0.328	0.304	0.293	0.289	0.287	0.286
2→3	0.150	0.183	0.184	0.180	0.177	0.174	0.173	0.172
2→4	0.011	0.026	0.037	0.045	0.049	0.053	0.054	0.056
3→1	0.065	0.168	0.254	0.318	0.363	0.396	0.419	0.436
3→2	0.255	0.309	0.311	0.303	0.295	0.29	0.288	0.286
3→3	0.606	0.425	0.332	0.279	0.247	0.225	0.211	0.201
3→4	0.007	0.098	0.103	0.283	0.094	0.088	0.082	0.077
4→1	0.008	0.045	0.100	0.162	0.220	0.272	0.317	0.353
4→2	0.054	0.130	0.189	0.227	0.251	0.264	0.272	0.277
4→3	0.246	0.327	0.341	0.328	0.306	0.283	0.262	0.243
4→4	0.069	0.497	0.369	0.288	0.222	0.179	0.149	0.127

4.4 Sojourn time and Transition intensity matrix

Sojourn time describes the average time an individual spends in each state in a single stay before he/she makes a transition to another state. The sojourn times of a continuous-time Markov process in a state *j* are independent, exponential random variables with mean

$$-\frac{1}{\lambda_{ii}} \qquad (4.4.1)$$

Or rate given by$-\lambda_{ii}$. A subject/patient that is currently in state 3 can make a move to state 2. The time the subject spends in state 1 before moving to state 2 (sojourn time) is

$$-\frac{1}{\lambda} = \frac{-1}{-0.04499382} = 22.225$$

Table 7. Transition intensity matrix

	State 1	State 2	State 3	State 4
State 1	-0.044	0.044	p<0.001	p<0.001
State 2	0.076	-0.122	0.045	p<0.001
State 3	p<0.001	0.078	-0.097	0.019
State 4	p<0.001	p<0.001	0.064	-0.064

The estimated intensity indicates that the rate of transiting from good states to the worst state is decreasing. The elements in each row of the transition intensity matrix (Table 7) sum to zero and off diagonal elements non-negative and the elements in diagonal is negative for all *i* equal to *j*. This implies that subjects in those states remain in their respective state while the off diagonals are rates at which subjects move to other states.

Table 8. Sojourn time

i	Estimates	SE	L	U
State 1	22.225	0.996	20.356	24.265
State 2	8.133	0.261	7.635	8.663
State 3	10.276	0.415	9.493	11.123
State 4	15.396	0.957	13.629	17.392

Results from Table 8 show estimates of sojourn time, the standard error (SE), the lower bound (L) and the upper bound (U) for each of the transient state *i*. From the results, if an individual is in state 4 (corresponding to a CD4 count below 200cell/mm^3) he/she spends 15 months in that state before making a transition to other states.

While a patient spends 22 months in state 1 before transiting to other states. These two state has the highest sojourn time mainly because a patient in state one (corresponding to a CD4 count greater than 500cell/mm^3) have high CD4 count level and the CD4 counts will take time to decrease, because the patient is responding well to treatment. While in state 4 the CD4 count will take time to improve this could be due the time taken by an individual to respond to treatment since state 4 is the worst state in HIV/AIDS progression

4.7 Hazard ratios of covariates on transition intensities

In this section the hazard ratios for each of the covariates; gender, age and prescribed ART regimen are estimated. The results show that the strongest predictor of transition from state 1 to 2 is TDF/3TC/EFV, which has a hazard ratio of 1.338. This means that patients who were prescribed TDF/3TC/EFV, this means that patients who received TDF/3TC/EFV were over 1.338 times more likely to transit from state 1 to state 2 than patients who did not receive TDF/3TC/EFV. The strongest predictor of immune deterioration from a CD4 level between 200 and 350 to a CD4 level less than or equal to 200 (3 to 4) is sex, with a hazard ratio of 2.074. This means that sex is the major cause of further immune deterioration when the immune system is too weak.

A hazard ratio of 0.678 for the predictor variable female shows that female were less likely to transit from state 2 to 3 than their male counterparts. The hazard ratios of females from a bad state to a better state are more than 1, which is an indication that females are less likely to respond to treatment compared to males.

For states which do not have intensity (i.e. (1,3)) the underlying model specifies that the patient must have passed through state 2 in between, rather than jumping straight from 1 to 3, Jackson (2011).

Table 9 shows the hazard ratio for prescribed ART regimen. The strongest predictor of immune deterioration from 3 to 4 is TDF/3TC/NVP, with a hazard ratio of 0.651.

Table 9. Hazard ratio of prescribed ART regimen

State	Baseline Hazard	Prescribed ART regimen				Other
		AZT/3TC/LPV	TDF/3TC/EFV	TDF/3TC/NVP	TDF/FTC/EFV	
(1,2)	0.043	0.956	0.489	1.137	0.928	0.124
(1,3)						
(1,4)	0.000	1.354	0.922	1.856	1.714	1.314
(2,1)	0.076	0.711	0.513	0.850	0.823	1.515
(2,3)	0.045	1.283	1.305	1.063	1.523	0.541
(2,4)	0.000	2.137	3.066	0.197	0.485	0.669
(3,1)						
(3,2)	0.077	1.218	1.150	1.168	1.488	3.045
(3,4)	0.018	0.563	0.597	0.651	0.371	0.430
(4,1)						
(4,2)	0.001	6.623	1.318	4.258	1.787	1.922
(4,3)	0.063	1.203	1.113	1.010	0.969	0.120

Note: n is number of times patients has been observed. TDF = tenofovir, AZT= azidothymidine, FTC = emtricitabine, EFV = efavirenz, 3TC = lamivudine, NVP = nevirapine, OTHER = abacavi (ABC) and stavudine (D4T). The msm did not give intensity for (1, 3), (3.1) and (4.1).

Table 10 shows the hazard ratio of age. The results show that the strongest predictor of transition from state 1 to 2 is age, which has a hazard ratio of 1.098. This means that a one unit increase in age of patients is associated with a hazard of 1.098.

Table10. Hazard ratio of age as a covariate

State	Hazard
(1,2)	1.098
(1,3)	
(1,4)	0.933
(2,1)	1.064
(2,3)	0.907
(2,4)	0.229
(3,1)	
(3,2)	0.947
(3,4)	0.848
(4,1)	
(4,2)	0.228
(4,3)	0.851

A hazard ratio of 2.568 (in table 11) for the predictor variable female shows that female were more likely to transit from state 1 to 4 than their male counterparts. The hazard ratios of females from a better state to a worse.

Table 11. Hazard ratio of sex as a covariate

State	Hazard
(1,2)	0.675
(1,3)	
(1,4)	2.568
(2,1)	1.452
(2,3)	0.678
(2,4)	1.179
(3,1)	
(3,2)	0.008
(3,4)	6.359
(4,1)	
(4,2)	5.022
(4,3)	7.204

Table 12 summarized covariates with their respective hazard ratio from state *i* to state *j*. A hazard ration of 1.684 for TDF/3TC/NVP shows that patients who were prescribed this regimen have a high risk of transiting from state 1 to state 4.

Table 12. Hazard ratio of covariates

State	Hazard ratio						
	Age	Sex	Prescribed ART regimen				
			AZT/3TC/LPV	TDF/3TC/EFV	TDF/3TC/NVP	TDF/FTC/EFV	OTHER
(1,2)	1.028	0.62	1.051	0.545	0.998	1.338	0.721
(1,3)							
(1,4)	0.927	0.838	0.841	0.494	1.684	1.002	0.302
(2,1)	1.007	1.415	0.726	0.574	0.877	0.999	0.716
(2,3)	1.043	0.854	0.935	0.974	0.864	1.089	0.632
(2,4)	0.403	0.885	1.277	0.749	1.409	1.063	0.633
(3,1)							
(3,2)	1.007	1.353	1.373	1.397	1.387	1.587	0.948
(3,4)	0.983	2.074	0.679	0.417	0.919	0.557	0.919
(4,1)							
(4,2)	0.397	1.511	1.645	0.611	1.684	1.407	0.869
(4,3)	0.949	2.065	1.539	0.733	1.302	1.636	0.681

4.8 Model comparison

A continuous-time semi-Markov model for the effects of covariates; age, sex and prescribed ART regimen is fitted as shown in table 13. Identification of covariates that have a significant effect is done by entering each covariate one after the other and performing the likelihood ratio test in comparison to the model without covariates.

A Likelihood ratio test is performed to compare the models that were fitted. The fitted time homogeneous model with prescribed ART regimen as a covariate has -2xLL = 78.106. The other fitted time homogenous models have likelihoods less than -2 x LL= 78.106. Which represents a weakening of LRT. The value of the $LRT = -2log_e\left(\frac{L_0(\theta)}{L_1(\theta)}\right)$ where $L_0(\theta)$ is the null model (without covariates) and $L_1(\theta)$ is the general model (with covariates).

Table 13. Likelihood ratio test

Model	-2 Log likelihood ratio test	df	p-value
Sex as a covariate	-131.088	9	1.00
Age as a covariate	-1721.034	9	1.00
Prescribed ART regimen as a covariate	78.106	45	0.002
All covariates	-546.99	63	1.00

The results show that the model with prescribed ART regimen as a covariates does fit better than all other the models.

4.5 Prediction of Clinical AIDS Disease Progression in Individual Patient

The probability that a patient starting from state $i \in \{1,2,3,4\}$ at time 0 enters state $j \in \{1,2,3,4\}$ after month *t* is plotted in Figure 3. Figure 3A shows the probability that a patient starting from state 1 at time 0, after month *t* enters to stage $j \in \{1, 2, 3, 4\}$. The probability of remaining in state 1 is high as compared to others, become constant after 75 months.

The probability of a patient starting from state 1 at time zero enters to state $j \in \{1, 2, 3 , 4\}$ after 108 months, are estimated to be 0.5, 0.4, 0.17 and 0.02 respectively. The conditional probability that a patient starting from state 1 at time zero, enters to state $j \in \{1, 2, 3, 4\}$ after 120 month, are estimated to be 0.5, 0.3, 0.1 and 0.02 respectively.

Figure 3B shows the probability that a patient starting from state 2 at time 0, after month *t* enters to stage $j \in \{1, 2, 3, 4\}$. The probability of remaining in state 2 is high as compared to others for the first 8 month, then start decreasing after 9 month. The probability of a patient starting from state 2 at time zero enters to state $j \in \{1, 2, 3 , 4\}$ after 108 months, are estimated to be 0.5, 0.28, 0.11 and 0.5 respectively.

The conditional probability that a patient starting from state 1 at time zero, enters to state $j \in \{1, 2, 3, 4\}$ after 120 month, are estimated to be 0.5, 0.28, 0.18 and 0.04 respectively.

Figure 3C shows the probability that a patient starting from state 3 at time 0, after month *t* enters to stage $j \in \{1, 2, 3, 4\}$. The probability of remaining in state 3 is high as compared to others for the first 20 month, and declined subsequently.

The probability of a patient starting from state 3 at time zero enters to state $j \in \{1, 2, 3, 4\}$ after 108 months, are estimated to be 0.5, 0.3, 0.2 and 0.04 respectively. The conditional probability that a patient starting from state 3 at time zero, enters to state $j \in \{1, 2, 3, 4\}$ after 120 month, are estimated to be 0.5, 0.3, 0.2 and 0.04 respectively.

Figure 3D shows the probability that a patient starting from state 4 at time 0, after month *t* enters to stage $j \in \{1, 2, 3, 4\}$. The probability of remaining in state 4 is high as compared to others for the first 12 month, declined thereafter until it increases after 85 month and decreased after 105 month again.

The probability of a patient starting from state 4 at time zero enters to state $j \in \{1, 2, 3, 4\}$ after 108 months, are estimated to be 0.5, 0.29, 0.3 and 0.04 respectively. The conditional probability that a patient starting from state 3 at time zero, enters to state $j \in \{1, 2, 3, 4\}$ after 120 month, are estimated to be 0.5, 0.29, 0.3 and 0.04 respectively.

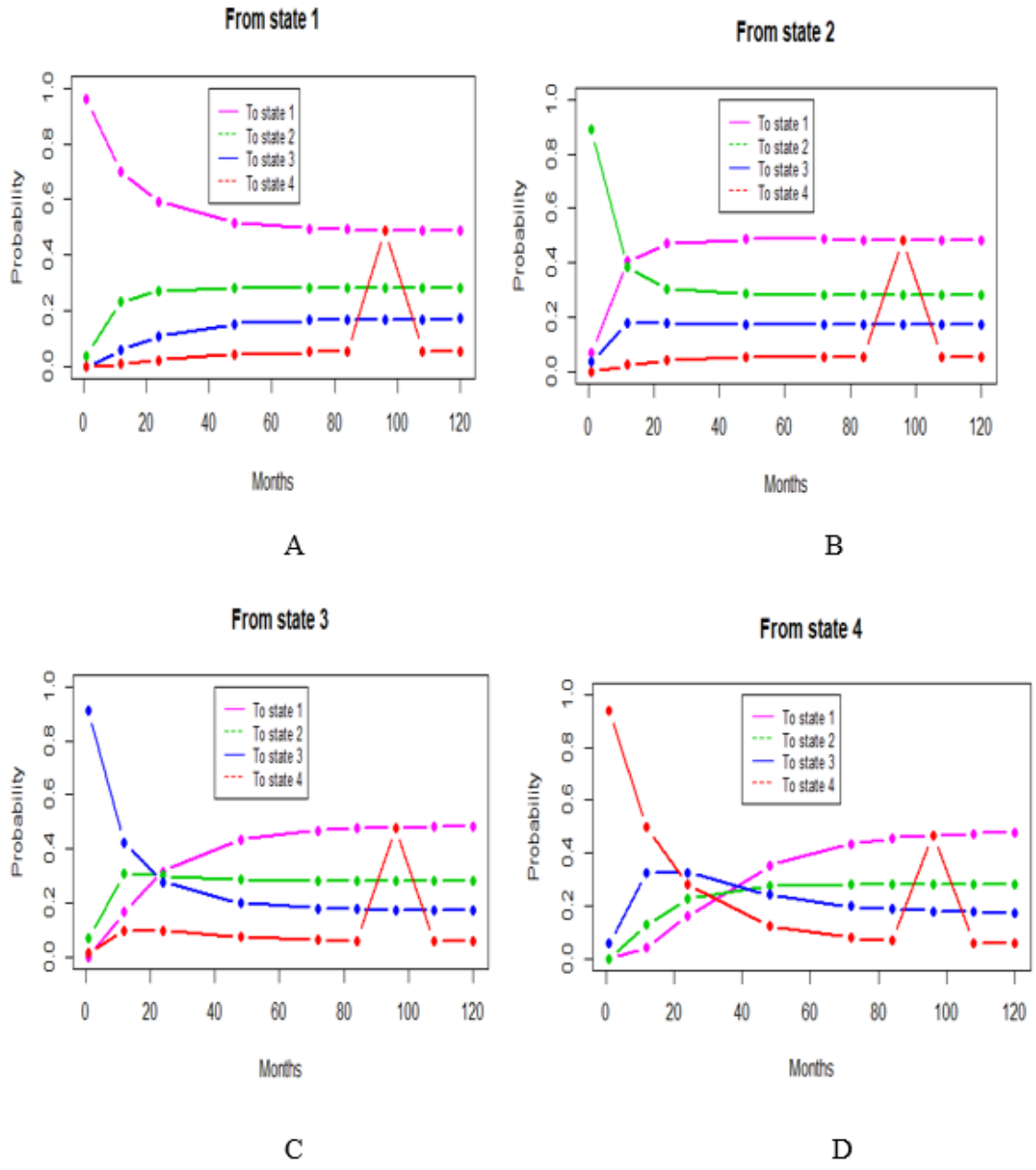

Figure 3. Conditional probabilities for each state

A)The probability that a patient at time 0 in state 1 will be in state $j \in\{1,2,3,4\}$, after month t ;

B) The probability that a patient at time 0 in state 2 will be in state $j \in\{1,2,3,4\}$, after month t ;

C) The probability that a patient at time 0 in state 3 will be in state $j \in\{1,2,3,4\}$, after month t ;

D) The probability that a patient at time 0 in state 4 will be in state $j \in\{1,2,3,4\}$, after month t .

4.8 Discussion for HIV/AIDS disease progression analysis results

This study intended to model the progression of HIV infection using longitudinally measured CD4 count for HIV positive patients initiated to ART. A continuous-time homogeneous semi-Markov model is fitted with and without covariates and comparison of these two models is done using the likelihood ratio test. Results shows that the model with prescribed ART regimen is the best model.

The probability of a patient transiting from state 1, 2 and 3 to state 4 after 24 months is 0.024, 0.045 and 0.283 respectively. Patients shows improvement from state 4 to, state 1, state 2 and state 3 with probability of 0.162,0.227 and 0.328. Similar study in Ethiopia, Zelalem (2010) has shown that probability of a patient to enter from stage IV to stage III, stage II and stage I in 2 year follow up period was 0.17, 0.9 and 0.2, respectively.

From this cohort, transitions to bad states are higher for female than for their male counterparts. This is quite pronounced on transitions from state 1 to state 4 where the hazards for females are 2.58 times that of males. This result is consistent with the findings from Shoko and Chikobvu (2018).

The hazard of covariates; sex, age and prescribed ART regimen are estimated. The results show that the strongest predictor of transition from state 1 to 2 is TDF/3TC/EFV, which has a hazard ratio of 1.338.

As time increases the conditional probability that a patient transit from state 1 to state 2, from state 2 to state 3 and from state 3 to state 4 after 48 months later is 0.284, 0.172 and 0.077 respectively. The probabilities are very small indicate that as time increases the conditional probabilities of transiting to the next worst state is very small, this support the results of Shebeshi (2011).

As time increases the probability of remaining in the same state is decreasing, this is in agreement with the results of Seyoum *et al.,* (2016) and that of Goshu and Dessie (2013).

The estimated sojourn time for state one, state two , state there and state four are, 22, 8, 10, and 15 month respectively. Comparable study in South Africa, Shoko and Chikobvu (2018), estimated the sojourn time for state one, state two, state three and state four as , 0.88, 0.88, 1.24, 1.20 and 1.57 years respectively. The sojourn time is of great interest in disease modeling as it gives an indication of how rapidly the disease is progressing. Longer sojourn times in a disease state mean a slow progressing disease and shorter sojourn times mean a rapidly progressing disease.

4.9 Limitation of the study

This study used secondary data from Ministry of Health and Social services. The data were not primarily collected for the purpose of this study, but steps were taken to make sure the data fit the description of msm data. Such steps includes, patients who were observed only once didn't form part of the study and the observation time was ordered

4.10 Delimitation of the study

The study considered all HIV infected patients who initiated ART January 2008-January 2012 and were followed till December 2017.

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

With the increasing availability of longitudinal and survival data, proper statistical methodology is needed in order to deal with various aspects related to the analysis of longitudinal data and to develop suitable statistical methods. In this dissertation, various aspects related to modelling of longitudinal data with an inclination towards its application to progression of HIV/AIDS based on CD4 count were addressed.

5.1 Conclusions

This study evaluated the progressions of HIV /AIDS infection using longitudinally measured CD4 count and its possible predictors via homogenous semi-Markov processes. Model with and without covariates have been compared using the LRT, the model with prescribed ART regimen exhibited the best fit.

The study also found that the conditional probabilities of transiting to the next worst state as time increases is very small and the probabilities of remaining in the same state as time increase is increasing. Finally the evolution of CD4 count (HIV infection) is differing by patient's baseline demographic and clinical characteristics like sex, age, WHO stages and prescribed ART regimen.

5.3 Recommendations

HIV infection is the most serious disease in the world, modeling the progression of this disease helps to identify the factors that affect the success of ART which helps to discover new vaccine. Thus further studies should be done in the area using developed and most flexible methodologies by including additional covariates like viral, functional status, educational level and marital status of patients which predict the evolution of HIV infection.

The dynamic nature of the HIV/AIDS progression is confirmed with particular findings that it is more likely to be in worse state than better one unless interventions are made. It is recommendable to keep up the ongoing ART treatment services in most effective ways with the careful considerations of recent disease status of patients.

7. References

Abner EL, Nelson PT, Schmitt FA, Browning SR, Fardo DW, Wan L, Jicha GA, Cooper GE, Smith CD, Caban-Holt AM, Van Eldik LJ & Kryscio RJ. (2014). Self-reported head injury and risk of late-life impairment and AD pathology in an AD center cohort. Dement Geriatr Cogn Disord. *Dementia and geriatric cognitive disorders*, 37(5-6), 294–306.

Adams M & Luguterah A. (2013). Longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on antiretroviral therapy (ART) in the Builsa district hospital. *European Scientific Journal*, 9(33), 1857-7881.

Barbu V S & Limnios N. (2008). *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*. Springer.

Base I. (n.d.). *The Immune system and CD4 count*. Retrieved August 9, 2018, from I-Base: www.i-base.info

Bodhade AS, Ganvir SM & Hazarey VK. (2011). Oral manifestations of HIV infection and their correlation with CD4 count. *Journal of oral Science*, 53(2), 203-211.

Brookmeyer R & Gail MH. (1994). *AIDS Epidemiology: a Quantitative Approach*. New York: Oxford University Press.

Caseiro MM , Golegã AAC, Etze A & Diaz RS. (2008). Characterization of virologic failure after an initially successful 48-week course of antiretroviral therapy in HIV/AIDS outpatients treated in Santos, Brazil. *Brazilian Journal of Infectious Diseases*, 12(3), 162-166.

Castaneda LB, Arunachalam V & Dharmaraja S. (2012). *Introduction to Probability and Stochastic Processes with Application*. John Wiley & Sons, Inc.

Cox DR & Miller HD. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hal.

D'Amico G, Di Biase G, Janssen J & Manca R. (2009). HIV EVOLUTION THROUGH TWO DIFFERENT TEMPORAL SCALES ACCORDING TO NON-HOMOGENEOUS SEMI-MARKOV MODELS. *The XIII International Conference, "Applied Stochastic Models and Data Analysis"*. LITHUANIA.

Dantony E, Elsensohn MH, Dany A, Villar E, Couchond C & Ecochard R. (2016). Estimating the parameters of multi-state models with time-dependent covariates through likelihood decomposition. *Spatial and Spatio-temporal Epidemiology*, 69, 37-43.

Dessie ZG. (2014). Multi-state model of HIV/AIDS by homogenous semi-markov process. *American Journal of Biostatistics*, 4(2), 21-28.

Di Biase G, D'Amico G, Di Girolamo A, Janssen J, Iacobelli S, Tinari N & Manca R. (2007). A Stochastic Model for the HIV/AIDS Dynamic Evolution. *Mathematical Problems in Engineering*, 2007, 1-14.

Dobson AJ & Barnett AG. (2014). *An Introduction to Generalized Linear Models*. Boca Raton: CRC Press.

Duesberg P. (1996). *INVENTING THE AIDS VIRUS*. Washington: REGNERY PUBLISHING, INC.

Février M, Dorgham K & Rebollo A. (2011). CD4+ T cell Depletion in Human Immunodeficiency Virus (HIV) Infection: Role of Apoptosis. *Viruses-MDPI*, 3(5), 586-612.

Fischl MA, Reichmann DD & Grieco MH. (1987). The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS related complex. A double blind placebo-controlled trial. *New England Journal of Medicine*, 317(4), 185–191.

Giuseppe D, D'Amico G, Girolamo A, Jansen J, Iacobelli S, Tinari N & Manca R. (2007). A stochastic model for HIV/AIDS dynamic progression. *Mathematical problems in engineering*, 2007, 1-14

Goshu AT & Asena TF. (2017). Comparison of Sojourn Time Distributions in Modeling HIV/AIDS Disease Progression. *Journal of Biometrics & Biostatistics*, 54(2), 155-174.

Goshu AT & Dessie Z. (2013). Modelling Progression of HIV/AIDS Disease Stages Using Semi-Markov Processes. *Data Science*, 11, 269-280.

Halim D. (n.d.). In H. D, *Maximum likelihood estimation for generalized semi-Markov processes. Discrete event dynamics systems: theory and applications* (pp. 73–104). Retrieved July 4, 2018, from u.math.biu.ac.il: u.math.biu.ac.il/~amirgi/CTMCnotes.pdf

Hamidi O, Tapak L, Poorlajal J & Amini P. (2017). Identifying risk factors for progression to AIDS and mortality post-HIV infection using illness-death multistate model. *Spatial and Spatio-temporal Epidemiology*, 5(4), 163-168.

Health 24 HIV/AIDS. (2018, August 8). *Health 24 HIV/AIDS*. Retrieved from Health 24: https://www.health24.com/Medical/HIV-AIDS/Management-of-HIV-AIDS/Antiretroviral-treatment-20120721

Ibe O. (2013). *Markov Processes for Stochastic Modeling*. USA: Elsevier.

Ibe OC. (2009). *Markov processes for stochastic modelling*. Elsevier Academic Press.

Jackson CH. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, 38(8), 1-29.

Jansen JN & Monica. (2001). Numerical solution of non-homogenous Semi-Markov processes in transient case. *Methodology and Computing in Applied Probability*, 3, 271-279.

Kalbfleisch JD & Lawless JF. (1985).The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80, 863–871.

Kaplan EL & Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457-485.

Kay R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometric*, 42, 855–865.

Kelkile P D. (2016). Statistical Analysis of Adult HIV/AIDS Patients and Modelling of AIDS Disease Progression. *Science Journal of Applied Mathematics and Statistics*, 4(5), 189-201.

Kurle L. (2008). *Hesperian health quides*. Retrieved August 14, 2018, from http://www.crossroadslink.org/blog/wp-content/uploads/2008/07/1-understanding-hiv.pdf

Laird A. (2013). *Modeling a Progressive Disease Process under Panel Observation*. PhD Thesis, University of Washington.

Levy P. (1954). Processus semi-Markoviens. *Congress of Mathematicians*, (pp. 416-426). Amsterdam.

Longini IM., Clark J., Gardner WS & Brundage J. (1991). The dynamics of CD4+ T lymphocyte decline in HIV infected individuals: A Markov modelling approach. *Journal of Acquired Immunodeficiency Syndromes*, 4(11), 1141–1147.

Lubyayi L. (2014). *Evolution of CD4 cell counts over time, for patients on antiretroviral therapy (ART) in Mildmay Uganda*. MSc Thesis, Hasselt University, Limburg.

Martinelli E, Vlia F, Goode D , Guerra-Perez N, Aravantinou M, Arthos J, Piataka M, Lifson J D, Blanchard J, Gettie A & Robbiani M. (2013). The frequency of high memory CD+ T cells correlates with susceptibility to rectal SIV infection. *J Acquir Immune Defic Synd*, 64(4), 325-331.

Masala G, Cannas G & Micocci M. (2014). Survival probabilities for HIV infected patients through semi-Markov processes. *Biometrical Letters*, 51(1), 13-36.

Meira-Machado L & Roca-Pardin˜as J. (2011). Analyzing Survival Data from an Illness-Death Model. *Journal of Statistical Software*, Volume 38,(Issue 3), 195-213.

Meira-Machado L,deU˜na ´Alvarez J,Cadarso-Su´arez C,& Andersen PK. (2009). Multi-state models for the analysis of time-to-event data. *Statistics in Medicine*, 18, 195–222.

Ministry of Health and Social Services. (2016). *National Guidelines for Antiretroviral Therapy*. Windhoek: MoHSS.

MoHSS. (2016). *National guidelines on ART*.

MoHSS. (2017). *National Strategic Framework for HIV and AIDS Response in Namibia 2017/18 to 2021/22*. Windhoek.

Moore RD, Keruly JC & Bartlett JG. (2012). Improvement in the Health of HIV-Infected Persons in Care: Reducing Disparities. *Clinical Infectious Diseases*, 55(9), 1242-1251.

Osmond DH. (1998, May). *University of California*, San Francisco. Retrieved June 2018, from HIV Insight: http://hivinsite.ucsf.edu/InSite?page=kb-00&doc=kb-03-01-04

Ross MS. (2007). *Introduction to Probability Models*. New York: Wiley.

Seyoum D, Wondayew S, Sisay K & Tadesse M. (2016). Predicting AIDS disease progression using longitudinal CD4 count among adult HIV/AIDS patients in Southwest Ethiopia: Application of Semi-Markov processes. *International Journal of Computational Bioinformatics and In Silico Modeling*, 5 (2), 808-814.

Shebeshi DS. (2011). *survival analysis of adult HIV/AIDS patients and stochastic modelling of AIDS disease progression: a case study of Jimma university*. MSc Thesis, JIMMA University, Addis Ababa.

Shoko C & Chikobvu D. (2018). Time -homogenous Markov process for HIV/AIDS progression under a combination treatment therapy: Cohort study, South Africa. *Theoretical Biology and Medical modelling*, 15(1), 3.

Smith WL. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society of London Series A*, (pp. 6-31). London.

UNAIDS. (2017). *UNAIDS DATA 2017. New York*: UNAIDS.

US Department of Health and Human Science. (2018, 05 22). *AIDS info*. Retrieved from AIDS info: https://aidsinfo.nih.gov

Verbeke VL & Lesaffre I. (1996). Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *Journal of the America Statistical Association*, 91 No 433, 217-221.

Viladent C & van Ackere A. (2007). HIV/AIDS modeling, a two-angle retrospective. Toward generic deterministic model for pattern II countries? University of Lausanne, Institute of Research in Management, Switzerland. *Toward generic deterministic model for pattern II countries?*

Werner M, Thuriaux P & Soutourina J. (2009). Structure-function analysis of RNA polymerases I and III. *Spatial and Spatio-temporal Epidemiology*, 19(6), 740-745.

WHO. (2006). *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV related disease in adults and children*. Geneva: WHO Press.

WHO. (2013). *Global update on HIV treatment*. Geneva: WHO Press.

WHO. (2015). *Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV*. Geneva: WHO Press.

WHO. (2016). *Global health sector strategy on HIV 2016-2021, towards ending AIDS*.

WHO. (2018). *Global Health Observatory (GHO) data*. Retrieved from World Health Organisation: http://www.who.int/gho/countries/nam/country_profiles/en/

Zare A,Mahmoodi M,Mohammad K, Zeraati H, Hosseini M & Naieni KH. (2014). Assessing Markov and Time Homogeneity Assumptions in Multi-state Models: Application in Patients with Gastric Cancer Undergoing Surgery in the Iran Cancer Institute. *Asian Pacific Journal of Cancer Prevention*, Vol 15, 441-447.

Zelalem G. (2010). *Statistical Modelling of HIV/ AIDS Progression and Survival of AIDS Patients. A Case Study of Bahir-Dar, Feleg-Hiwot Referral Hospital*. M.Sc. Thesis, Hawassa University, Hawassa.

Zewale TA. (2012). *Modeling the progression of HIV infection using longitudinally measured CD4 count for HIV positive patients following highly active antiretroviral therapy*. MSc Thesis, Jimma University.

8. APPENDICES

Appendix 1. MSM R file

```
library(foreign)

library(msm)

library(minqa)

data=as.data.frame(read.csv("C:\\MSc Biostatistics Thesis\\formula.csv"))

#TRANSITION MATRIX (COUNTS)

counts=statetable.msm(state, person, data = data)

#INITIAL QUESS FOR INTENSITY MATRIX

twoway4.q <- rbind(c(0, 0.25, 0, 0.25), c(0.166, 0, 0.166, 0.166), c(0, 0.25, 0, 0.25), c(0, 0.25, 0.25, 0))

rownames(twoway4.q) <- colnames(twoway4.q) <- c("state 1", "state 2", "state 3", "state 4"

#WORKING OUT MIN DIFFERENCES

#diffs <- aggregate(cbind(minDiff=duration)~person, FUN=function(x) min(diff(x)),data=data)

#MERGING

#MD_data1<- merge(data,diffs,by='person',all.x=T)

#SUBSETTING

#MD_data=subset(MD_data1,subset = diffs$minDiff>0.1)
```

```
INTENSITY MATRIX

Q=crudeinits.msm(state ~ duration, person, data=data, qmatrix=twoway4.q)

# WITHOUT COVARIATES

cav.msm <- msm(state ~ duration, subject = person, data = data,

        qmatrix =twoway4.q,opt.method = "bobyqa")

# WITH COVARIATES

cav1.msm <- msm(state ~ duration, subject = person, data = data,

        qmatrix =twoway4.q, covariates = ~ age + sex,opt.method = "bobyqa")

#WITH SEX AS A COVARIATE

cav2.msm <- msm(state ~ duration, subject = person, data = data,

         qmatrix =twoway4.q, covariates = ~ sex,opt.method = "bobyqa")

# WITH AGE AS A COVARIATE

cav3.msm <- msm(state ~ duration, subject = person, data = data,

         qmatrix =twoway4.q, covariates = ~ age,opt.method = "bobyqa")

cav.msm

P=pmatrix.msm(cav.msm, t = 84, ci = "normal")

PE=round(P$estimates,6)

PL=round(P$L,6)
```

```
PU=round(P$U,6)

Q=qmatrix.msm(cav.msm)

P=pmatrix.msm(cav.msm, t=48, ci = "normal")

S=sojourn.msm(cav.msm)

#EXPORTING DATA FROM R TO EXCEL

#write.csv(MD_data, "c:/MSc Biostatistics Thesis/mydata.csv")

# predicting future state plot

P=pmatrix.msm(cav.msm, t = 1, ci = "normal")

PE0=P$estimates-P$estimates

for (i in c(1,12,24,48,72,84,96,108,120))

{P=pmatrix.msm(cav.msm, t = i, ci = "normal")

PE=P$estimates

PE0=cbind(PE0,PE)

}

#PLOTS FROM STATE1 TO STATES

Month=as.data.frame(c(1,12,24,48,72,84,96,108,120))

T1=as.data.frame(PE0[1,c(5,9,13,17,21,25,29,33,37)])

T2=as.data.frame(PE0[1,c(6,10,14,18,22,26,30,34,38)])
```

```
T3=as.data.frame(PE0[1,c(7,11,15,19,23,27,31,35,39)])

T4=as.data.frame(PE0[1,c(8,12,16,20,24,28,29,36,40)])

plot(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,

    T1$`PE0[1, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,

    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",

    main = "From state 1")

lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,

    T2$`PE0[1, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,

    type = "b",col=3,lwd=2,pch=16)

lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,

     T3$`PE0[1, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,

     type = "b",col=4,lwd=2,pch=16)

lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,

     T4$`PE0[1, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,

     type = "b",col=2,lwd=2,pch=16)

legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),

      col=c(6,3,4,2), lty=1:2, cex=0.8)

#PLOTS FROM STATE2 TO STATES
```

```
Month=as.data.frame(c(1,12,24,48,72,84,96,108,120))
T1=as.data.frame(PE0[2,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[2,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[2,c(7,11,15,19,23,27,31,35,39)])
T4=as.data.frame(PE0[2,c(8,12,16,20,24,28,29,36,40)])
plot(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
    T1$`PE0[2, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
    main = "From state 2")
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
     T2$`PE0[2, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
     type = "b",col=3,lwd=2,pch=16)
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
     T3$`PE0[2, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
     type = "b",col=4,lwd=2,pch=16)
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
     T4$`PE0[2, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
     type = "b",col=2,lwd=2,pch=16)
```

```
legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
       col=c(6,3,4,2), lty=1:2, cex=0.8)
#PLOTS FROM STATE3 TO STATES
Month=as.data.frame(c(1,12,24,48,72,84,96,108,120))
T1=as.data.frame(PE0[3,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[3,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[3,c(7,11,15,19,23,27,31,35,39)])
T4=as.data.frame(PE0[3,c(8,12,16,20,24,28,29,36,40)])
plot(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
     T1$`PE0[3, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
     type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
     main = "From state 3")
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T2$`PE0[3, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
      type = "b",col=3,lwd=2,pch=16)
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T3$`PE0[3, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
      type = "b",col=4,lwd=2,pch=16)
```

```
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T4$`PE0[3, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
      type = "b",col=2,lwd=2,pch=16)
legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
       col=c(6,3,4,2), lty=1:2, cex=0.8)
#PLOTS FROM STATE4 TO STATES
Month=as.data.frame(c(1,12,24,48,72,84,96,108,120))
T1=as.data.frame(PE0[4,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[4,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[4,c(7,11,15,19,23,27,31,35,39)])
T4=as.data.frame(PE0[4,c(8,12,16,20,24,28,29,36,40)])
plot(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
     T1$`PE0[4, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
     type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
     main = "From state 4")
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T2$`PE0[4, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
      type = "b",col=3,lwd=2,pch=16)
```

```
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T3$`PE0[4, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
      type = "b",col=4,lwd=2,pch=16)
lines(Month$`c(1, 12, 24, 48, 72, 84, 96, 108, 120)`,
      T4$`PE0[4, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
      type = "b",col=2,lwd=2,pch=16)
legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
       col=c(6,3,4,2), lty=1:2, cex=0.8)
#PERFORMING LRT
lrtest.msm(cav.msm,cav1.msm)
```

Appendix 2. Frequencies and probabilities of the transitions of the states of the process

	To			
From	2547	616	99	26
	917	1193	414	35
	211	666	928	161
	41	72	299	381
	To			
From	0.77464	0.18735	0.03011	0.00791
	0.35834	0.4662	0.16178	0.01368

	0.10732	0.33876	0.47202	0.08189
	0.0517	0.09079	0.37705	0.48045

Appendix 3. Model with and without covariates plus the likelihood test

Transition intensity without covariates

Baseline

state 1 - state 1 -0.0449938 (-4.912e-02,-0.041210)

State 1 - state 2 0.0441741 (4.039e-02, 0.048315)

State 1 - state 4 0.0008198 (4.647e-04, 0.001446)

State 2 - state 1 0.0767813 (7.101e-02, 0.083018)

State 2 - state 2 -0.1229539 (-1.310e-01,-0.115431)

State 2 - state 3 0.0459415 (4.127e-02, 0.051139)

State 2 - state 4 0.0002311 (1.035e-05, 0.005163)

State 3 - state 2 0.0781898 (7.168e-02, 0.085290)

State 3 - state 3 -0.0973099 (-1.053e-01,-0.089898)

State 3 - state 4 0.0191200 (1.605e-02, 0.022774)

State 4 - state 2 0.0002844 (3.851e-08, 2.101271)

State 4 - state 3 0.0646648 (5.673e-02, 0.073704)

State 4 - state 4 -0.0649493 (-7.337e-02,-0.057495)

-2 * log-likelihood: 16854.45

Transition intensity matrix with covariates and Hazard ratio

Baseline

State 1 - state 1 -0.0478942 (-0.0523914,-0.043783)

State 1 - state 2 0.0472749 (0.0431441, 0.051801)

State 1 - state 4 0.0006193 (0.0001782, 0.002153)

State 2 - state 1 0.0779321 (0.0719117, 0.084456)

State 2 - state 2 -0.1271058 (-0.1356715,-0.119081)

State 2 - state 3 0.0449819 (0.0400486, 0.050523)

State 2 - state 4 0.0041918 (0.0024409, 0.007199)

State 3 - state 2 0.0750291 (0.0681359, 0.082620)

State 3 - state 3 -0.0992711 (-0.1083049,-0.090991)

State 3 - state 4 0.0242420 (0.0186161, 0.031568)

State 4 - state 2 0.0179987 (0.0122151, 0.026521)

State 4 - state 3 0.0826675 (0.0665342, 0.102713)

State 4 - state 4 -0.1006662 (-0.1230721,-0.082339)

Age

State 1 - state 2 1.0116 (0.9987, 1.025)

State 1 - state 4 0.9443 (0.7910, 1.127)

State 2 - state 1 1.0053 (0.9938, 1.017)

State 2 - state 3 1.0076 (0.9918, 1.024)

State 2 - state 4 1.0015 (0.9316, 1.077)

State 3 - state 2 0.9922 (0.9800, 1.005)

State 3 - state 4 1.0247 (0.9995, 1.051)

State 4 - state 2 0.9899 (0.9615, 1.019)

State 4 - state 3 0.9982 (0.9800, 1.017)

Sex

State 1 - state 1

State 1 - state 2 0.9111 (0.7411, 1.120)

State 1 - state 4 1.6372 (0.1158, 23.146)

State 2 - state 1 1.8198 (1.5151, 2.186)

State 2 - state 2

State 2 - state 3 0.8263 (0.6514, 1.048)

State 2 - state 4 5.6017 (1.2349, 25.411)

State 3 - state 2 1.0486 (0.8604, 1.278)

State 3 - state 3

State 3 - state 4 1.8906 (1.2003, 2.978)

State 4 - state 2 12.9509 (4.5874, 36.563)

State 4 - state 3 1.9900 (1.3870, 2.855)

State 4 - state 4

-2 * log-likelihood: 16828.88

PERFORMING LRT

	-2 log LR	Df	p
Model with covariates	25.56933	18	0.1100204

Appendix 4. Cases processing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3539	32.1	32.1	32.1
	1	7489	67.9	67.9	100.0
	Total	11028	100.0	100.0	

```
library(foreign)
library(msm)
library(minqa)
#data=as.data.frame(read.csv("C:\\MSc Biostatistics Thesis\\Katutura HIV data.csv"))
data<-read.table("file:///C:/MSc Biostatistics Thesis/Data/Final HIV data.csv", header=TRUE, sep=",")

str(data)
```

```
#TRANSITION MATRIX (COUNTS)
counts=statetable.msm(state, person, data = data)
counts

#INITIAL QUESS FOR INTENSITY MATRIX
twoway4.q <- rbind(c(0, 0.25, 0, 0.25), c(0.166, 0, 0.166, 0.166), c(0, 0.25, 0, 0.25), c(0, 0.25, 0.25, 0))
rownames(twoway4.q) <- colnames(twoway4.q) <- c("state 1", "state 2", "state 3", "state 4")

#WORKING OUT MIN DIFFERENCES
#diffs <- aggregate(cbind(minDiff=duration)~person, FUN=function(x) min(diff(x)),data=data)

#MERGING
#MD_data1<- merge(data,diffs,by='person',all.x=T)

#SUBSETTING
#MD_data=subset(MD_data1,subset = diffs$minDiff>0.1)

#INTENSITY MATRIX
Q=crudeinits.msm(state ~ duration, subject = person, data=data, qmatrix=twoway4.q)
Q
# WITHOUT COVARIATES
cav.msm <- msm(state ~ duration, subject = person, data = data,
          qmatrix =twoway4.q,opt.method = "bobyqa")
cav.msm


#WITH SEX AS A COVARIATE
```

```
cav1.msm <- msm(state ~ duration, subject = person, data = data,
        qmatrix =twoway4.q, covariates = ~ sex, opt.method = "bobyqa")

cav1.msm
hazard.msm(cav1.msm)
# WITH AGE AS A COVARIATE
cav2.msm <- msm(state ~ duration, subject = person, data = data,
        qmatrix =twoway4.q, covariates = ~ age,opt.method = "bobyqa")

cav2.msm

hazard.msm(cav2.msm)

# WITH regimen AS A COVARIATE
cav3.msm <- msm(state ~ duration, subject = person, data = data,
        qmatrix =twoway4.q, covariates = ~ regimen,opt.method = "bobyqa")

cav3.msm

# WITH COVARIATES
cav4.msm <- msm(state ~ duration, subject = person, data = data,
        qmatrix =twoway4.q, covariates = ~ age + sex+ regimen,opt.method = "bobyqa")
hazard.msm(cav4.msm,cl=0.95)
hazard.msm(cav4.msm,cl=0.95)

P=pmatrix.msm(cav.msm, t = 42, ci = "normal")
PE=round(P$estimates,3)
PL=round(P$L,6)
PU=round(P$U,6)
```

```
P
PE
PL
PU

Q=qmatrix.msm(cav.msm)
P=pmatrix.msm(cav.msm, t=48, ci = "normal")
S=sojourn.msm(cav.msm)

Q
P
S


#SURVIVAL ANALYSIS OF EACH COVARIATES
hazard.msm(cav.cov.msm)


qmatrix.msm(cav.cov.msm, covariates = list(age , sex ))
qmatrix.msm(cav.cov.msm, covariates = list(age , sex ))


lrtest.msm(cav.msm, cav1.msm)


cav.msm <- msm(state ~ duration, subject = person, data = data,
               qmatrix = twoway4.q, pci = 5, method = "BFGS")
```

```
lrtest.msm(cav.msm, cav.pci.msm)

hazard.msm(cav.msm )

cav.msm <- msm(state ~ years, subject = PTNUM, data = cav,
        qmatrix = twoway4.q, death = TRUE, censor = 99,
        censor.states = c(1, 2, 3))

#EXPORTING DATA FROM R TO EXCEL
#write.csv(MD_data, "c:/MSc Biostatistics Thesis/mydata.csv")
# predicting future state plot
P=pmatrix.msm(cav.msm, t = 1, ci = "normal")
PE0=P$estimates-P$estimates

for (i in c(1,6,12,18,24,30,36,42,48))
{
P=pmatrix.msm(cav.msm, t = i, ci = "normal")
PE=P$estimates
PE0=cbind(PE0,PE)
}


#PLOTS FROM STATE1 TO STATES
Month=as.data.frame(c(1,6,12,18,24,30,36,42,48))

T1=as.data.frame(PE0[1,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[1,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[1,c(7,11,15,19,23,27,31,35,39)])
T4=as.data.frame(PE0[1,c(8,12,16,20,24,28,29,36,40)])
```

```
plot(Month$`c(1,6,12,18,24,30,36,42,48)`,
    T1$`PE0[1, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
    main = "From state 1")

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
    T2$`PE0[1, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
    type = "b",col=3,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T3$`PE0[1, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
     type = "b",col=4,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T4$`PE0[1, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
     type = "b",col=2,lwd=2,pch=16)

legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
      col=c(6,3,4,2), lty=1:2, cex=0.8)



#PLOTS FROM STATE2 TO STATES
Month=as.data.frame(c(1,6,12,18,24,30,36,42,48))

T1=as.data.frame(PE0[2,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[2,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[2,c(7,11,15,19,23,27,31,35,39)])
```

```
T4=as.data.frame(PE0[2,c(8,12,16,20,24,28,29,36,40)])

plot(Month$`c(1,6,12,18,24,30,36,42,48)`,
    T1$`PE0[2, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
    main = "From state 2")

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T2$`PE0[2, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
     type = "b",col=3,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T3$`PE0[2, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
     type = "b",col=4,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T4$`PE0[2, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
     type = "b",col=2,lwd=2,pch=16)

legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
      col=c(6,3,4,2), lty=1:2, cex=0.8)


#PLOTS FROM STATE3 TO STATES
Month=as.data.frame(c(1,6,12,18,24,30,36,42,48))

T1=as.data.frame(PE0[3,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[3,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[3,c(7,11,15,19,23,27,31,35,39)])
```

```
T4=as.data.frame(PE0[3,c(8,12,16,20,24,28,29,36,40)])

plot(Month$`c(1,6,12,18,24,30,36,42,48)`,
    T1$`PE0[3, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
    main = "From state 3")

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T2$`PE0[3, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
     type = "b",col=3,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T3$`PE0[3, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
     type = "b",col=4,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T4$`PE0[3, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
     type = "b",col=2,lwd=2,pch=16)

legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
      col=c(6,3,4,2), lty=1:2, cex=0.8)


#PLOTS FROM STATE4 TO STATES
Month=as.data.frame(c(1,6,12,18,24,30,36,42,48))

T1=as.data.frame(PE0[4,c(5,9,13,17,21,25,29,33,37)])
T2=as.data.frame(PE0[4,c(6,10,14,18,22,26,30,34,38)])
T3=as.data.frame(PE0[4,c(7,11,15,19,23,27,31,35,39)])
```

```
T4=as.data.frame(PE0[4,c(8,12,16,20,24,28,29,36,40)])

plot(Month$`c(1,6,12,18,24,30,36,42,48)`,
    T1$`PE0[4, c(5, 9, 13, 17, 21, 25, 29, 33, 37)]`,
    type = "b",col=6,lwd=2,ylim=c(0,1),pch=16,xlab="Months",ylab="Probability",
    main = "From state 4")

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T2$`PE0[4, c(6, 10, 14, 18, 22, 26, 30, 34, 38)]`,
     type = "b",col=3,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T3$`PE0[4, c(7, 11, 15, 19, 23, 27, 31, 35, 39)]`,
     type = "b",col=4,lwd=2,pch=16)

lines(Month$`c(1,6,12,18,24,30,36,42,48)`,
     T4$`PE0[4, c(8, 12, 16, 20, 24, 28, 29, 36, 40)]`,
     type = "b",col=2,lwd=2,pch=16)

legend(40, 1, legend=c("To state 1", "To state 2","To state 3","To state 4"),
      col=c(6,3,4,2), lty=1:2, cex=0.8)
#PERFORMING LRT
lrtest.msm(cav.msm,cav4.msm)
```