A MARKOV CHAIN ANALYSIS OF HIV PROGRESSION IN KEY POPULATIONS; A CASE STUDY OF HIV IN HOMABAY COUNTY

BY

ODIENY, CHRISTINE AKINYI

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN APPLIED STATISTICS

SCHOOL OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

MASENO UNIVERSITY

Declaration

This project is my own work and has not been presented for a degree award in any other institution.

ODIENY,CHRISTINE AKINYI

Admission Number: MAT/MSC/00037/2016

Signature:.....Date:.....

This project has been submitted for examination with my approval as the university supervisor.

Supervisor

Department of Statistics and Actuarial Science

Maseno University

Name: Dr. APAKA, BONIFACE RANGITTA

Signature:.....Date:.....

Acknowledgement

My gratitude goes to God who renewed my strength to continue despite the challenges I went through. Much thanks to my project supervisor who ensured this work is done through his support and completed successfully, not forgetting the input from the entire department. My family as well, may God cheer us to fuel more successes.

Dedication

To my family, I dedicate this project to you for enduring and sacrificing your time and support to help me achieve this. To the Ministry of health, this project will assist in decision making on various cores of HIV/AIDS.

Abstract

Key populations account for the highest number of individuals with HIV infections. They include female sex workers, men who have sex with men and people who inject drugs. While many studies have been done regarding HIV/AIDS and its spread, the disease continue to affect newer populations and remain prevalent among those who have already experienced it. Our study focused on HIV progression in key populations in Homabay county.The overall objective of this study was to apply Markov chains in analyzing HIV progression in the key populations. We drew the following specific objectives: 1. To estimate the probability of moving from susceptible state to infected HIV state in the key populations; 2. To estimate the expected time to absorption. We also estimated the probability of absorption of a patient and the probability of moving from one state to the other. This study utilized secondary data from Key Population size Estimates 2019 and primary data from HIV care and treatment register from Homabay county referral hospital.The data was for the whole population in the category.The population mean estimates per spot were as follows: Female sex workers (FSW) 8.1, men who have sex with men(MSM) 6.1 and People who Inject drugs (PWID) 7.6.We estimated parameters using Maximum Likelihood Estimator and calculated 99% confidence interval of the parameters. A Discreet Time Markov chain analysis was used to model the progression of the disease among key populations in the county.The data was then simulated in R markdown to estimate the number of individual in each state at time t. Findings from the study show that the mean time to absorption of susceptible individuals is 264 months,66 months after confirmed infection with HIV and 63 months when infected with AIDS. These findings will be informative to the ministry of health and donor partners in the managing HIV/AIDS in the key populations.

TABLE OF CONTENTS

List of Figures

CHAPTER ONE: INTRODUCTION

Adult HIV prevalence is 4.9% in Kenya with substantial geographic variation ranging from 0.1% in Garissa County to 20.7% in Homabay county. [2]. Lake region areas register high number of new infection and death due to HIV/AIDS .Fishing being a major economic activity in Homabay,pose challenges in management of the virus. Fishermen are characterized by high mobility which is associated with increased risk of antiretroviral therapy (ART) non adherence, lost to follow-up, deterioration in CD4 count, HIV-related death, development of drug resistance and general non continuity of HIV care. It also makes it hard to enroll in HIV care programs and to maintain adherence to HIV treatment [13]. Key populations also account for the highest infection rate in the county. They include men who have sex with men, female sex workers, people who inject drugs, adolescent girls and young women. This population are highly vulnerable because of their behavior that dispose them to the infection.They account for the highest infection rate across the world. Other factors contributing to the heavy burden include; The gender inequality especially, lack of access to health services and poverty compounded with lack of exposure among residents leading to negative perception on HIV test and diagnosis. This increases delayed diagnosis, delayed treatment, under-recognition and further worsens the transmission.Luo culture with polygamy, sexual cleansing rituals, and wife inheritance,has created a fertile ground for culturalistic discourses on the spread of HIV among the Luos themselves [6].Although current treatments, such as highly active antiretroviral therapy (HAART), have successfully improved the survival and life expectancy in HIV/AIDS population and reduced HIV incidences, these gains may be reversed if a concerted effort is not made to reduce HIV transmission among Key Populations.Our study will focus on the disease progression specifically on the key populations in Homabay county.

Mathematical and epidemiological models have been used to study the disease dynamics. However these models have not been able to estimate other disease metrics that are important for modeling, such as the probability of remaining in healthy state, probability of moving from healthy state to infected, and probability of reaching one state from another. This study seek to address those metrics using discreet time Markov chain. The findings will help in determining the degrees of vulnerability for the key populations

and inform more resource allocation for designing efficient, economical and timely health policies targeting the high risk populations and high prevalent areas.

1.2 Statement of the problem

Several epidemiological models have been proposed in studying infectious diseases. These models have helped in understanding the spread of these diseases however these models are not able to capture many other metrics such as probability of staying healthy, of being infected, and of dying.This study seek to address those problems through the finite Markov chains in discreet times.

1.3 Statement of objectives

1.3.1 General objectives

To apply Markov model in analyzing HIV progression in patients

1.3.2 Specific Objectives

1.To estimate their expected time to absorption, using their sojourns in the different states.

2.To estimate the probability of reaching one state from another.

1.4 Significance of the study

The findings of this study will inform Ministry of Health and partner's interventions towards successful management of HIV/AIDS in the vulnerable regions and key populations . The general public are also benefiting from the findings since they will be able to understand the longevity of the virus in humans when measures are put in place. The projections will also help the victims understand their life expectancy when living with the virus and inform the measures to be taken.

1.5 Basic concepts and definitions Random Variable

Is a function that maps a set of all possible outcomes of an experiment into real numbers. Random variables are considered countable when the number of variables being considered can be counted.

Stochastic Processes

A stochastic process is a random process; that is, a change in the state of some system over time whose course depends on chance and for which the probability of a particular course is defined. It is basically a family of random variables $X_t : t \in T$ defined on a given probability space, indexed by the time variable t, where t varies over an index set T.
In general these are experiments in which the outcomes of events depend on the previous outcomes.

Markov Chains

A Markov process is defined as a type of stochastic process in which a system changes in a random manner between different states.
A Markov chain is an integer-time process $X_n, n \geq 0$ for which the sample values for each random variable $X_n, n \geq 1$ lie in a countable set S and depend on the past only through the most recent random variable X_{n-1}.More specifically for all positive integers, n and for all $i, j, k, ..., m$ in S,

$$Pr\left(X_n = j | X_{n-1} = i, X_{n-2} = k, ...X_0 = m\right) = \left(PrX_n = j | X_{n-1} = i\right) \tag{1}$$

Furthermore, $Pr\left(X_n = j | X_{n-1} = i\right)$ depends only on i and j not n and is denoted by

$$Pr\left(X_n = j | X_{n-1} = i\right) = P_{ij} \tag{2}$$

The initial state X_0 has an arbitrary probability distribution. A finite state Markov chain is a Markov chain in which S is finite.

Homogeneous Markov chain

A Markov chain is called homogeneous if and only if the transition probabilities are independent of the time t.

Discrete-Time Markov model

In probability, a discrete-time Markov chain is a sequence of random variables, known as a stochastic process, in which the value of the next variable depends only on the value of the current variable, and not any variables in the past.Transitions between states are assumed to follow a stochastic Markov process, i.e, transitions to the next state depend only on the current state occupied by a patient. The previous states occupied by an individual do not matter; which is the memoryless property of the Markov models.
A discrete time Markov chain is a discrete-time stochastic process $X_n, n \geq 0$ satisfying the following properties:

The state space is countable (often labeled with a subset of N).
For all states 0, j there is a given probability p_{ij} such that:

These transitions are described using the transition probabilities $p_{ij}t$
and transition intensity.
In general, the discreet time process means you are not continuously monitoring the state of the people in the system. It would really get overwhelming if we had to ask every hour "Are you sick?". Therefore it makes more sense to monitor the individual states on discreet basis e.g a month rather than the continuous basis. This forms the basis of the use of discreet time Markov chains.

Continuous Time Markov Chain

A continuous time Markov chain is a continuous stochastic process in which for each state, the process will change state according to exponential random variable and then move to a different state as specified by the probabilities if the stochastic matrix.

1.5.1 Classification of States

Recurrent State For finite-state Markov chains, a recurrent state is a state i that is accessible from all states that are accessible from i. i is recurrent if $i \rightarrow j$ implies that $j \rightarrow i$ A state i in a finite-state Markov chain is recurrent if there is no possibility of going to state $j$from which there can be no return.If a Markov chain ever enters a a recurrent state, it returns to that state eventually with a probability of 1.

Transient state

A state i is transient if there is some j that is accessible from i but from which there is no possible return. Each time the system returns to i, there is possibility of going to j; eventually this possibility will occur with no further returns to i. For finite Markov chains, either all states in a class are transient or all are recurrent.

Period of a state i

The period of a state i denoted di is the greatest common divisor of those values of n for which P_{ii}^{n} 0. If the period is 1, the state is aperiodic, and if the period is 2 or more, the state is periodic. For any Markov chain with either a finite or countably infinite number of states, all states in the same class have the same period.

Ergodic chain For a finite state Markov chain, an ergodic class of states i a class that is both recurrent and aperiodic. A markov chain consisting of entirely one ergodic class is called an ergodic chain.

Irreducible Markov Chain

A Markov chain is said to be irreducible if it has only one communicating class. That is, all states communicate with each other.

Reducible Markov chain

Markov chain is called reducible if and only if there are two or more communicating classes.

Population: Population is defined as the total group of people, objects or events under a study.

Sample:Sample is defined as groups elements or objects chosen from a population. It is the subset of a population and should be representative of the population elements.

CHAPTER TWO: LITERATURE REVIEW

2.1 HIV in global populations

HIV has been a burden affecting populations spanning across the world with 36.9 million people living with the virus [7] . Many families have been left with the burden of raising orpans in abject poverty since the pandemic inception in Kenya in the 1980s.This burden has socio economic and psychological impacts . The disease still affects the productivity of infected individuals and therefore their families face food insecurity alongside other financial constraints [4]. In 2018 alone, there were nearly 770,000 AIDS-related deaths worldwide, with mortality rates remaining high in many settings despite the widespread provision of antiretroviral therapy (ART)[10]. Kenya's HIV epidemic remains substantial as Kenya is one of the top five most-burdened countries globally in spite of successes in expanding HIV testing and connecting people living with HIV to treatment [5].

2.2 Applications of Markov chains

Mathematical models have been extensively used in research and modeling of chronic diseases such as HIV/AIDS research. They have played an important role in enabling a deeper understanding of the metrics of the viruses and factors associated with the progression in different stages.Markov chains has been applied in the field of public health and medicine. In public health it has been applied in modeling the spread of infectious diseases and comparing infection rates in different regions. Markov Chains provide support for problems involving decision on uncertainties through a continuous period of time. The greater availability and access to processing power through computers allow that these models can be used more often to represent clinical structures. Markov models consider the patients in a discrete state of health, and the events represent the transition from one state to another. The possibility of modeling repetitive events and time dependence of probabilities and utilities associated permits a more accurate representation of the evaluated clinical structure. These templates can be used for economic evaluation in health care taking into account the evaluation of costs and clinical outcomes, especially for evaluation of chronic diseases [11]. There are several Markov chain models that have been

used to model chronic illness like HIV/AIDS. In HIV in particular, it quantifies the probabilities of patients progressing between various states (Susceptibele, Infected, Removed). With the aim of contributing to the epidemiological surveillance of HIV/AIDS several researchers have used Markov models.

[3] Used CD4 based Markov model to determine factors associated with progression between different stages of the disease for individuals on ART. [1] used a continuous time homogeneous Markov model to analyze HIV/AIDS progression in patients. Their findings indicated that inclusion of both viral load levels and CD4 cell counts, in monitoring and management in time homogeneous Markov models help in the prediction of mortality in HIV/AIDS patients on ART. Higher CD4 cell counts improve the health and consequently survival of HIV/AIDS patients. [8] used discreet time Markov chain to analyze HIV progression amongst the races in the United States.In their study, they predicted the number of blacks living with HIV/AIDS Verses the number of whites but did not capture the probability of moving from different states. Their study was a focused based study to establish mortality trend based on the racial disparity. [12] Applied a Markov model for the effects of virological failure on HIV/AIDS progression in tuberculosis co-infected patients receiving antiretroviral therapy in a rural clinic in northern South Africa.A continuous-time non-homogeneous Markov model was used to model the progression of HIV/AIDS in patients on combination ART (cART). Their model confirmed that virological failure, coupled with developing active TB while on cART, increases mortality rates irrespective of patient CD4+ count status. It also suggests that while TB at the time of cART initiation does not increase the risk of viral rebound, development of active TB after cART initiation does increase this risk. [9]Applied statistical analysis of the stages of HIV infection using a Markov model.They used a staged Markov model to estimate the distribution and mean length of the incubation period for acquired immunodeficiency syndrome (AIDS). [14] applied Markov chains to the study of wildlife disease dynamics.He noted that disease specialists and policy analysts are unfamiliar with the mathematical/statistical language of disease models, translation of probability statements into meaningful terms for disease research and control may be challenging. Markov chain models are powerful tools, applicable to the study of disease dynamics that allow straightforward calculations of eas-

ily interpretable metrics of interest including probabilities of infection/recovery, expected times to initial infection, duration of illness and life expectancies for susceptible and infected individuals.

This forms the basis of our study in analyzing HIV progression in the key population to estimate the metrics.

CHAPTER THREE: METHODOLOGY

3.1 Model Process

We define four discreet states of a Markov model; Susceptible, InfectedHIV, InfectedAIDS and Removed. These states follow Kermack McKendrick S-I-R model. If $X_i = 0, 1, 2, 3$ represent the number of individuals at any state from underlying diseases at any time t,then clearly, X_t is a stochastic process with states $0, 1, 2, 3$. Therefore the first order time homogeneous Markov dependency can be modelled as ;

$$P\left(X_n = i_n | X_{n-1} = i_{n-1}, ..., X_i = i_1, X_0 = i_0\right) = P\left(X_n = i_n | X_{n-1} = i_{n-1}\right) \tag{3}$$

Then the transition probability for (P_{ij}), for j=0,1,2,3 is given by the following matrix.

$$(P_{ij}) = \begin{vmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & p_{33} \end{vmatrix} \tag{4}$$

Where,

$\sum_{j=0}^{3} P_{ij} = 1; i = 0, 1, 2, 3.$

3.2 Defining States of the Markov Model.

1.The Susceptible state(S) is comprised of individuals who are in healthy state and not exposed before to the virus.

2.The InfectiveHIV state I comprises of infected individuals and carriers of the disease.

3.The InfectiveAIDS state I comprises of HIV individuals who developed AIDS.

4.The removed state R comprises of individuals who died from the disease.

These states are represented in the figure 1 of the appendix .

The states 0 and 1 are both transient states.This is because upon leaving those states, there is no positive probability of returning to the original state. State 1 and 2 are recurrent states since the two states have equivalence class as well as communicates. State 3

which is known as Removed state is an absorbing state. Individual at this state remain dead with probability of one. The chain is not ergodic.The chain is also not irreducible because not all states belong to the same class.

3.2 Model Assumptions

1. The current state of an individual is dependent only on the state of the individual at the previous time step.
2. No individuals from the removed states can be susceptible or infected.
3. Transitional probabilities are independent of time and remain constant over time.
4. Successive transitions, confirmed coinfections of the disease or other medical complications were not taken into consideration.
5. Removed state comprise of subjects who died from disease complications.

3.3 Parameters of the Markov Chain

P_{ii}: is the probability of remaining in state i. P_{ij} ;is the transition probability from state i to state j, $i \neq j$

The parameter P_{01} is referred to as the discreet time force of infection.

3.4 Estimating Transition Probabilities

The maximum likelihood estimation(MLE) was used to estimate the transition probabilities for the parameters with their respective standard errors. The number of infected individuals who recovered were zero. The transition events are independent of one another. The likelihood of the transition probability P_{ij}, follows a binomial model.

$$L\left(P_{ij}|N,x\right) = \binom{N_i}{x_{ij}} P_{ij}^{xij}\left(1-P_{ij}\right)^{N_i - x_{ij}} \tag{5}$$

Where the N_{ij} is the number of observed transition that starts from state i to state j and

$\Sigma_j P_{ij} = 1$.

Intuitively, the maximum likelihood estimator was the value of p which maximizes the distribution. From the equation (5) the assumption of the constant transition probabilities over a period, the transition probability matrix is estimated as a binomial proportion given

as :

$$\hat{P_{ij}} = \frac{\sum_j x_{ij}}{N_i} \tag{6}$$

where x_i denotes the i^{th} observation for the random variable.

The standard error of the sampling distribution of the ML estimate was given as

$$\hat{s.e}\left(P_{ij}\right) = \sqrt{\frac{\hat{P_{ij}}\left(1-\hat{P_{ij}}\right)}{N_i}} \tag{7}$$

We also constructed the 99 percent confidence interval of the estimates.

3.5 Estimating Disease Metrics

The probability that a susceptible individual becomes infected for the first time between $m-1$ and m steps for states $i, j = 0$ from the transitional probability matrix S-I-R. is given as

Let f_{ij}^n=probability that the first passage time from state i to state j is equal to n
The first passage time probabilities satisfy a recursive relationship:
1.f_{ij}^1=p_{ij}^1=p_{ij}

2 f_{ij}^2=$\sum_{k\neq j} p_{ik} f_{kj}^1$

3. f_{ij}^n=$\sum_{k\neq j} p_{ik} f_{kj}^{n-1}$

Therefore

$$f_{01}^m = P\left(X_{n+m}=0, X_{n+m-1}=1, ..., X_{n+1}=1 | X_n=1\right) = P_{00}^{m-1} P_{01} \tag{8}$$

There is no recovery for this model.All individuals get absorbed in the long run.

The expected time to infection is computed as,

$$E\left(T_{ij}\right) = \frac{\sum_{m=1}^{\infty} m f_{ij}^m}{P_r\left(i \to j\right)} = \frac{1}{1-P_{ii}} \tag{9}$$

for $i, j = 0, 1, i \neq j$, where the numerator $\sum_{m=1}^{\infty} m f_{ij}^m$ is the expected value of the first passage time from state i to state j and the denominator

$$P_r\left(i \to j\right) = \frac{P_{ij}}{1-P_{ii}} \tag{10}$$

This gives the lifetime probability of transitioning from state i to state j. And therefore moving between states of the model is given by the two equations:

3.6 Absorbing Markov chains

A state in a Markov chain is said to be an absorbing state if the process will never leave that state once it is entered. That is, if the state i is an absorbing state, then $P_{ii} = 1$. A Markov chain is said to be an absorbing Markov chain if it has at least one absorbing state and if any state in the chain, with a positive probability, can reach an absorbing state after a number of steps. It is certain that when one starts at a given point, he will still reach the absorbing state as $n \to \infty$.

In order to be an absorbing Markov chain, it is not sufficient for a Markov chain to have an absorbing state. It must also have the property that all non-absorbing states must eventually reach an absorbing state. Suppose that the absorbing Markov chain has t non-absorbing states and r absorbing states. The transition probability matrix $\mathbf{P}$ can be written in the following format:

$$\mathbf{P} = \begin{vmatrix} Q & R \\ 0 & I \end{vmatrix}$$

Where,

Q is a t$\times t$ matrix representing the transition probabilities from the non-absorbing states into the non-absorbing.

R is a t $\times r$ matrix representing the transition probabilities from the non-absorbing states into the absorbing states.

0 is an r $\times t$ matrix consisting of zeros in its entries

and

I is the $r \times r$ identity matrix

3.6.1 Fundamental Matrix

The fundamental matrix can be computed from the transition probability matrix of an absorbing chain. Given the transition probability matrix **P**, decompose **P** . Recall that Q consists of the transition probabilities from the transient states to the transient states. Derive $\mathbf{I} - \mathbf{Q}$ where **I** is the identity matrix of the same dimension as **Q**. The following

three theorems capture more important properties of absorbing Markov chains.

3.6.2 Other Properties of Absorbing Markov chain

In an absorbing Markov chain with transition probability matrix $\mathbf{P}$, the matrix $I - Q$ has an inverse W, i.e. $W = (I - Q)^{-1}$. The matrix W has the following properties: $W = I + Q + Q^2 + Q^3 + \cdots$.

An entry W_{ij} in the matrix W is the mean time spent in the transient state j given that the process starts at the transient state i. More specifically,

$W_{ij} = E[\text{number of times state } j \text{ is visited} | X_0 = i]$.

For an absorbing Markov chain with transition probability matrix $\mathbf{P}$, the matrix $W = (I - Q)^{-1}$ is called the fundamental matrix of the absorbing Markov chain. Once the fundamental matrix is obtained by taking the inverse of $I - Q$, we can use it to solve two problems, stated in the following two theorems.

3.6.3 Mean time to absorption.

In an absorbing Markov chain with transition probability matrix $\mathbf{P}$, consider the fundamental matrix.

$$W = (I - Q)^{-1} \tag{11}$$

The following calculation is of interest. Given that the process starts in the transient state i, consider the row of the matrix W that corresponds to state i. The sum of all entries of W on that row is the mean time spent in transient states given that the process start in state i. The sum is the mean time to absorption.

To estimate the probability of a person ever reaching any state from the other, from (15); Define W as the as the entries of the inverse of the matrix $I - Q$ and f as the probability of moving ever reaching from state $i \to j$.

$$f_{ij} = \frac{W_{ij}}{W_{jj}} \tag{12}$$

And

$$f_{jj} = 1 - \frac{1}{W_{jj}} \tag{13}$$

3.6.4 Probabilities of Absorption

In an absorbing Markov chain with transition probability matrix $\mathbf{P}$, consider the fundamental matrix $W = (I - Q)^{-1}$. The following calculation is of interest. Consider the matrix R in the decomposition of $\mathbf{P}$. The matrix $W \times R$, the product of W and R, gives the probabilities of absorption. More specifically, the row in W corresponding to transient state i multiplying the column in R corresponding to the absorbing state k results in the probability of being absorbed into state k given that the process start in state i. Even more specifically, assuming that the transient states are $0, 1, 2, \cdots, t-1$, the following gives the probability of absorption into state k given that the process starts in state i.

$$\sum_{j=0}^{t-1} \Bigg(E[\text{number of times state } j \text{ is visited} | X_0 = i] \times P_{jk} \Bigg) \cdots \tag{14}$$

This is simplified as,

$$(I - Q)^{-1} \times R \tag{15}$$

The state space contain both transient and recurring (absorbing) states, but we still need to obtain a steady state solution.

3.7 Steady state and P^n for large n

The matrix P^n (i.e., the matrix of transition probabilities raised to the nth power) is important for a number of reasons; The i, j element of this matrix is $P_{ij}^n = P\left(X_n = j | X_0 = i\right)$ if memory of the past dies out with increasing n, then we would expect the dependence of P_{ij}^n on both n and i to disappear as $n \to \infty$. This means P^n should converge as $n \to \infty$. For an m state Markov chain, there exist a determinant $(P - \lambda I)$ which is a polynomial of degree m in λ. It thus has roots (eigenvalues which are distinct). Each eigenvalue λ_i has an eigen vector v.Then we define $P = U^{-1}\Lambda U$. Where U is a a matrix whose rows are corresponding to the eigenvectors, Λ is a diagonal matrix of λ_{is}.

At n^{th} step,

$$P^n = U^{-1}\Lambda^n U \tag{16}$$

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1 Data Collection

The data used in this study was secondary data from the HIV care and treatment register from the county referral hospital and Key Population Size Estimate Report(KPSE)reports 2019. Data was given in sum total of the key population size. Patients in the infected HIV state accounted for the number who were removed or absorbed in the process. This population included those who are still healthy and those living with HIV/AIDS. Our study site was Homabay county. We can refer to the raw data of individuals at any state at any time t . X_t satisfies the Markov chain model with state space $S = (0, 1, 2, 3)$ used in this study in the 2, of the appendix. A

4.2 Estimating the Transition Probabilities

Parameter	Probability Estimate	S.E	99% Confidence Interval
P_{00}	0.75	0.005833	0.735-0.7650
P_{01}	0.23	0.005669	0.2154-0.2446
P_{02}	0.0196	0.001867	0.0148-0.00244
P_{03}	0.0004	0.000269	-0.0003-0.0011
P_{11}	0.995	0.001154	0.9920-0.9980
P_{12}	0.003	0.000895	0.0007-0.0053
P_{13}	0.002	0.000731	0.0001-0.0039
P_{22}	0.52	0.063967	0.3552-0.6848
P_{23}	0.48	0.063967	0.3152-0.6448

Table 1: State probabilities

The transition probability matrix is as below:

$$P = \begin{pmatrix} 0.75 & 0.23 & 0.0196 & 0.0004 \\ 0 & 0.995 & 0.003 & 0.002 \\ 0 & 0 & 0.52 & 0.48 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{17}$$

4.2 Probability of first Transition

The probability that a susceptible individual becomes infected is estimated using (8) Our study period is two years and the the time steps are semi annual(6 months). The tabulated probability of an individual moving from susceptible state to infected state is as follows:

Time	Probability $P_{00}^{m-1}P_{01}$
0	0.31
6	0.041
18	0.0017
24	0.0003077

Table 2: probability of moving from susceptible to infection HIV at time t

From the graph (**??**) the probability of first infection is very low due to the interventions being done like,pre-exposure prophylaxis, post exposure prophylaxis, use of protection and free testing and counseling services. It is also observed that the first time step of exposure, the probability of infection is relatively high. This means after exposure to HIV you can become infectious in the first 6 months.

4.3 Probability of infection at any time

The probability of infection at any time period gives the probability of infection over the time period for susceptible individuals.It is obtained using (12). This is obtained at 0.92. It shows that the disease is infectious for all the infected individuals.

4.4 Expected time to infection

The expected time to infection for the highly susceptible is calculated using (11). It is estimated at 4 months in our study.This means individual exposed to HIV will react to antibodies at 4 moths after the infection, though this may vary based on immunity levels and individuals.

4.5 Mean time to absorption

The mean time to absorption will estimate the life expectancy of individual at each state before absorption. This will be estimated using equation (13).

Define states as at (19).

We need to define matrix Q since we know the identity matrix.

This is obtained by decomposing P and deleting rows and columns corresponding to the matrix zeros and identifying Q.

Therefore decomposing matrix P we obtain:

$$P = \begin{pmatrix} 0.995 & 0.003 & 0 & 0.0196 \\ 0 & 0.52 & 0 & 0.48 \\ 0.23 & 0.02 & 0.75 & 0.0004 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{18}$$

Extracting the matrix R and the matrix of zeros, from the 3rd column and the 4th row, we obtain the matrix Q for our calculation.

$$Q = \begin{pmatrix} 0.995 & 0.003 & 0.0196 \\ 0 & 0.52 & 0.48 \\ 0.23 & 0.02 & 0.0004 \end{pmatrix} \tag{19}$$

I is a 3by3 identity matrix. Therefore $(I-Q)^{-1}$ is given as,

$$(I-Q)^{-1} = \begin{pmatrix} 261.662 & 1.691 & 1.340 \\ 61.662 & 2.524 & 1.340 \\ 61.662 & 0.441 & 1.340 \end{pmatrix} \tag{20}$$

The matrix resulting matrix contain the long run average number of visits in each state.The time period is in months. Implying that averagely it takes 262 months for a susceptible individual to stay uninfected. and averagely it takes 2 months for a person with HIV to stay in that state before becoming infectious.

Expected time to absorption from all states

Initial State	Life expectancy
Susceptible	261.662+1.691+1.340=264.693 months
InfectedHIV	61.662+2.524+1.340=65.526 months
InfectedAIDS	61.662+0.441+1.340=63.4461 months

Table 3: Mean time to absorption

It is easy to see that the probability of absorption in the long run is 1. Just to show this, The probability of absorption is calculated from equation (17)

$$T = \begin{pmatrix} 1.00 \\ 1.00 \\ 1.00 \end{pmatrix} \tag{21}$$

4.6 Probability of ever reaching any state from another

We now estimate the matrix F of probability of ever reaching any state from another using equation (14) and (15) from matrix W of $(I-Q)^{-1}$. The table below gives the results:

	Susceptible	Infected HIV	Infected AIDS	Dead
Susceptible	0.996	0.670	1.0	1.0
Infected HIV	0.236	0.604	1.0	1.0
Infected AIDS	0.236	0.175	0.254	1.0
Dead	0.00	0.00	0.00	1.0

Table 4: Probability matrix F of ever reaching one state from another

This matrix shows that when you are in the key population, the probability of remaining susceptible is 0.996 and the probability of infection is 0.670 and the probability of getting AIDS and eventually dying is 1; which is a sure event. This matrix shows the importance of avoiding long term effects of the disease after infection. Like when you start when living with AIDS, the probability of going back to living with HIV is low at 0.175.And equally if the disease progresses to AIDS the probability of staying with AIDS is low at 0.254 before being absorbed. This speaks the need to maintain very low viral load after infection.

4.7 Steady state solution

The steady state solution will allow for assessment of other possible intervention towards creating a lasting solution. We will calculate this using equation (19) This will require us to obtain the eigenvalues and the corresponding eigenvector then compute the steady state.To obtain the eigenvalues we use the equation,

$(A-\lambda I)=0$

The eigen values of matrix P is given by ,

$\lambda = 0.52, 0.75, 0.995, 1$

Implying that the matrix of Λ^n is given by

$$\Lambda^n = \begin{pmatrix} 0.52^n & 0 & 0 & 0 \\ 0 & 0.75^n & 0 & 0 \\ 0 & 0 & 0.995^n & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The corresponding eigenvectors is obtained using Newton Raphson method by the atozmath calculator as

When $\lambda = 0.52$,

$$v1 = \begin{pmatrix} -0.00237986 \\ -0.00631579 \\ 1 \\ 0 \end{pmatrix}$$

When $\lambda = 0.75$,

$$v2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

When $\lambda = 0.995$,

$$v3 = \begin{pmatrix} 0.93877551 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

When $\lambda = 1$,

$$v4 = \begin{pmatrix} 0.9296 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Therefore the matrix U is combined to be

$$U = \begin{pmatrix} 0.00238 & 1 & 0.9388 & 0.9296 \\ 0.00636 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Its inverse is given by;

$$U^{-1} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 1 & -0.9388 & 0.00359 & 0.00561 \\ 0 & 1 & -0.00636 & -0.99364 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Therefore the steady state $P^n = U^{-1}\Lambda^n U$ where U is a matrix of eigenvectors is given by,(C)

CHAPTER FIVE: CONCLUSIONS AND RECOMMENDATIONS

The goal of this study was to apply Markov chain principles in analyzing HIV progression in discreet times. The findings reveal the importance of using discreet time Markov chain in defining progession in different states. As a result we have discovered that starting at AIDS infections increases the likelihood of absorption than when you start susceptible. Also interventions help to increase the life expectancy post infection stages.For instance, if one start infected with HIV, there is still probability of lowering the viral loads to lowest levels and living with AIDS before absorption.The results show that more considerations should be incorporated in managing the virus in the key populations by examining their risk taking behavior and exploring more interventions. However the limitations of this work is that the model was constructed from estimates of KPSE reports and reference with the county referral data due to resources. Future studies can utilize entirely primary information for better projections.

REFERENCES

[1] Chikobvu, D., & Shoko, C. (2018). A markov model to estimate mortality due to hiv/aids using cd4 cell counts based states and viral load: A principal component analysis approach. *Biomedical Research-tokyo*, *29*, 3090–3098.

[2] Conan, N., Badawi, M., Chihana, M. L., Wanjala, S., Kingwara, L., Mambula, C., Ngugi, C., Okomo, G., Opollo, V., Salumu, L., et al. (2021). Two-fold increase in the hiv viral load suppression rate along with decreased incidence over six years in ndhiwa sub-county, kenya. *Tropical Medicine & International Health*.

[3] Dessie, Z. G. (2014). Multi-state models of hiv/aids by homogeneous semi-markov process.

[4] Fedinard, A. O., & Ogolla, M. (2018). Socio-economic effects of hiv/aids on household food security in rangwe sub county, homa-bay county, kenya. *International Journal of Physical and Social Sciences*, *8*(11), 116–129.

[5] Kimanga, D. O., Ogola, S., Umuro, M., et al. (2014). Prevalence and incidence of hiv infection, trends, and risk factors among persons aged 15–64 years in kenya: Results from a nationally representative study. *Journal of acquired immune deficiency syndromes (1999)*, *66*(Suppl 1), S13.

[6] Kovaĉiĉ, V., & Amondi, J. (2011). Part 3 cultural stereotypes and the health seeking behaviour of hiv/aids patients in homa bay, kenya. *AIDS*, *12*, 15.

[7] Lasry, A., Medley, A., Behel, S., Mujawar, M. I., Cain, M., Diekman, S. T., Rurangirwa, J., Valverde, E., Nelson, R., Agolory, S., et al. (2019). Scaling up testing for human immunodeficiency virus infection among contacts of index patients—20 countries, 2016–2018. *Morbidity and Mortality Weekly Report*, *68*(21), 474.

[8] Lee, S., Ko, J., Tan, X., Patel, I., Balkrishnan, R., & Chang, J. (2014). Markov chain modelling analysis of hiv/aids progression: A race-based forecast in the united states. *Indian journal of pharmaceutical sciences*.

[9] Longini Jr, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F., & Hethcote, H. W. (1989). Statistical analysis of the stages of hiv infection using a markov model. *Statistics in medicine*, *8*(7), 831–843.

[10] Mahy, M., Marsh, K., Sabin, K., Wanyeki, I., Daher, J., & Ghys, P. D. (2019). Hiv estimates through 2018: Data for decision-making. *AIDS (London, England)*, *33*(Suppl 3), S203.

[11] Sato, R. C., & Zouain, D. M. (2010). Markov models in health care. *Einstein (São Paulo)*, *8*, 376–379.

[12] Shoko, C., Chikobvu, D., & Bessong, P. (2020). A markov model for the effects of virological failure on hiv/aids progression in tuberculosis co-infected patients receiving antiretroviral therapy in a rural clinic in northern south africa. *South African Medical Journal*, *110*(4), 313–319.

[13] Tanser, F., Bärnighausen, T., Vandormael, A., & Dobra, A. (2015). Hiv treatment cascade in migrants and mobile populations. *Current Opinion in HIV and AIDS*, *10*(6), 430–438.

[14] Zipkin, E. F., Jennelle, C. S., & Cooch, E. G. (2010). A primer on the application of markov chains to the study of wildlife disease dynamics. *Methods in Ecology and Evolution*, *1*(2), 192–198.

APPENDICES

Appendix A

Data

Here goes the raw data of the experiment.

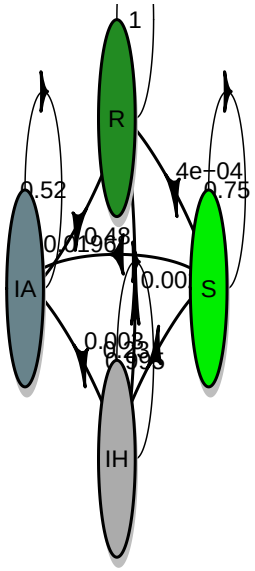
Groups	Total	Susceptible	InfectedHIV	InfectedAIDS	Removed
Susceptible	4375	2949	1263	151	12
Infected HIV	1700	0	1482	202	16
Infected AIDS	85	0	0	78	7

Table A.1: Number of individuals at any state at the end of the project period

Appendix B

Figures

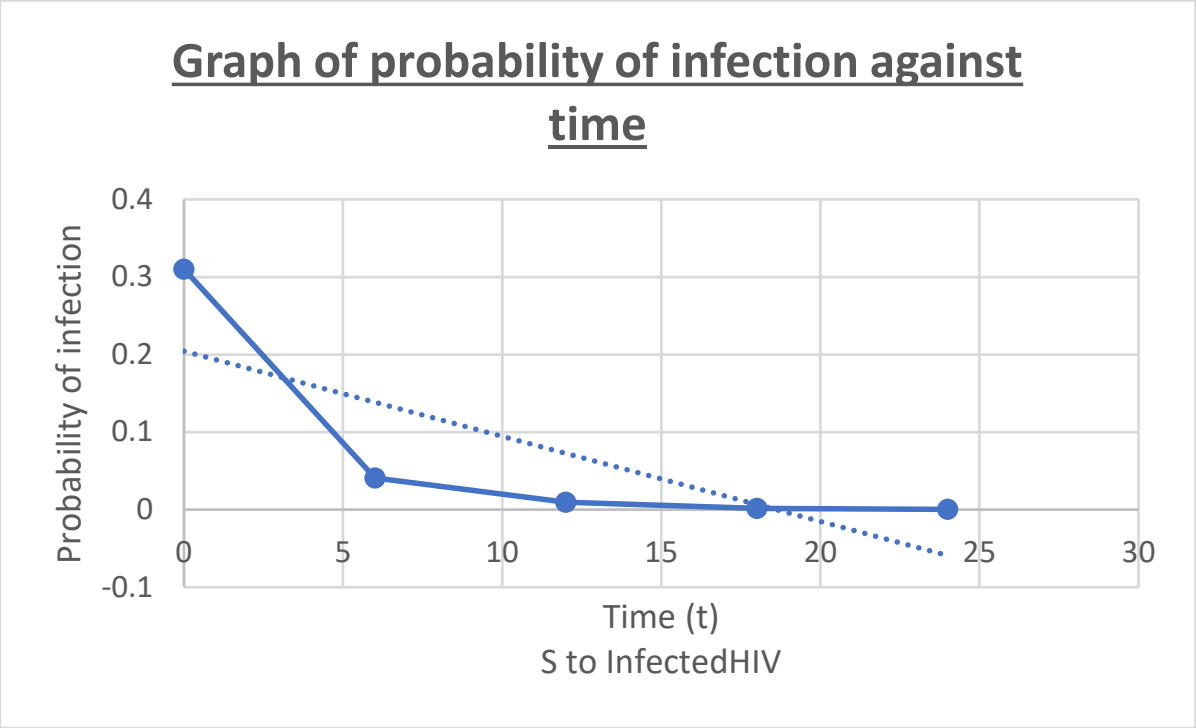
Figure B.1: States of Transition



The figure below illustrate the probability of first transition from susceptible to infected.

The graphical representation is as below:

Figure B.2: Probability of moving from susceptible to infected HIV



Appendix C

Steady state matrix P^n

$$\begin{pmatrix} 0.995^n & 0 & 0 & 0.995^n - 1 \\ 0.52^n\,(0.00238) + 0.75^n\,(-0.000597) + 0.95^n\,(0.00359) & 0.52^n & 0.52^n\,(0.9388) + 0.75^n\,(-0.9388) & 0.52^n\,(0.9296) + 0.75^n\,(-0.9388) + 0.995^n\,(0.00359 \\ 0.75^n\,(0.00636) + 0.995^n & 0 & 0.75^n & 0.75^n + 0.995^n\,(-0.00639) - 0.99364 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{C.1}$$