2021

Using semi-Markov process to model incremental change in HIV staging with cost effect

Andrew, Joram Malului
Strathmore Institute of Mathematical Sciences
Strathmore University

Using Semi-Markov Process to Model Incremental Change in HIV Staging with Cost Effect

Strathmore
UNIVERSITY

Joram Malului Andrew

Thesis presented in fulfillment of the academic requirement for the degree of Master of Science in Statistical Science of Strathmore University

October 2021

Declaration and Recommendation

Declaration

I declare that this thesis is my original work and has not been presented in any other university.

Signature  Date 4/21/2021

Joram Malului Andrew

Recommendation

This research has been submitted for examination with my approval as supervisor according to Strathmore University regulations.

Signature Collins Odhiambo Date 5/21/2021

Dr. Collins Odhiambo

Strathmore University

Abstract

Over the past years, parametric and non-parametric methods have been used in modelling cost and effectiveness according to one studied event or one health state. In this study we used semi-Markov model in which the distributions of sojourn times are explicitly defined. Weibull distribution was chosen and used in modelling the hazard function for each transition. Using a regression model for cost, a cumulative cost function of cost was developed enabling us to determine the estimated mean cost per patient in each state defined in the semi-Markov model. ICER was used for cost effectiveness analysis in comparing two strategies (Patients in DCM and patients not in DCM) of follow up. Using viral load, three states were defined; $VL < 200ml$, $200ml < VL < 1000ml$, $VL > 10000ml$ and an absorbing state death. The mean cost of the patients for each state 1, 2 and 3 was \$765, \$829 and \$1395 respectively. The calculated ICER ratio was \$483.8268/life-year-saved. The cost of keeping patients in state 1 (on DCM) was relatively cheaper and efficient compared to the other states.

List of contents

List of Figures

List of Tables

Acknowledgment

Throughout the writing of this thesis I have received a great deal of support and assistance.

I would first like to thank my supervisor, Dr. Collins Odhiambo, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would like to acknowledge my course mates from the university for their wonderful collaboration. In addition, I would like to thank my parents for their wise counsel and sympathetic ear. I also thank my siblings who have always stood by me and the moral support they have given me. You are always there for me. Finally, I could not have completed this dissertation without the support of my friend, victor Kiprono, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

Abbreviations

AIDS	Acquired Immune-deficiency Syndrome
ART	Anti-retroviral Therapy
CMA	Cost Minimization Analysis
DCM	Differentiated Care Model
GOF	Goodness of Fit
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
ICER	Incremental Cost Effectiveness Ratio
INB	Incremental Net Benefit
MOH	Ministry of Health
NACC	Kenya National AIDS Control Council
UNAIDS	United Nations Programme on Acquired Immune-deficiency Syndrome
WHO	World Health Organization

Chapter 1

Introduction

1.1 Background of the study

1.1.1 Introduction

Over the years HIV/AIDs infection has largely reduced in Sub-Saharan Africa (SSA), with a drop of over 30% of the number of incident infections since 2010 (UNAIDS, 2020). According to WHO (2012) and UNAIDS (2016) this decrease in burden reflects the accomplishment of a global effort focused on a region in which approximately 70% of all people living with HIV reside. Kenya is not an exemption in this great achievement. The milestones achieved in reducing HIV/AIDs infections is as a result of programs or strategies geared towards monitoring the HIV/AIDs patients. Anti-retroviral therapy (ART) is a program that highly benefits HIV-infected patients (Marseille et al., 2009). ART helps in the increase of CD4 cells, reduction of viral load, the decrease in the incidence of opportunistic infections, enhancement of well-being and functionality, and the decrease in mortality (Marseille et al., 2009). Additionally, HAART program has highly contributed to the reduction of HIV-infections. HAART creates long-term opportunities that enhance the capacity in the healthcare systems. Combination of these programs and other follow up strategies like the use of Home-Based Care (HBAC) or Differential Care Models (DCM) have effectively reduced the HIV/AIDs prevalence in Kenya by a value below 5% (UNAIDS, 2018).

The motivation of this study was to embrace the use of semi-Markov models in modelling incremental change in HIV staging with a cost effect. This would help in quantifying the cost of keeping a patient in any of the HIV/AIDs stages. This research was necessitated by decrease of financial aid in support of HIV programming from international donors of whom the country highly depends on in funding the above-mentioned Programs/s-

trategies. The willingness of donors' withdrawal and entrusting the country with the programs is because of the great achievement by the Kenyan government in reducing HIV prevalence to below 5% (Kates et al., 2019).

1.1.2 Markov and Semi-Markov models

Survival analysis using Cox proportional hazards regressions are the most common tools used for time to event data (Kryscio et al., 2013). Abner et al. (2013) argues that the methods are limited especially in cases where multiple or recurrent outcomes are of interest and when there is presence of competing risks hence making them not appropriate for all studies. They further suggested that the utilization of Markov chain models which can account for competing risks (informative censoring), report censored data, numerous outcomes, repetitive outcomes, non-constant survival probabilities, and frailty (Hillis et al., 1986, Kryscio et al., 2013). Clinical data used for this editorial was prone to censoring hence Markov chain models were used for analysis.

A stochastic method describing the transition of individuals through a finite number of the defined state at a given time is referred to as a Markov process. Hillis et al. (1986) argues that the possible movements between stages could be represented with a state diagram or transition matrix. To terminate the Markov process one state must be an absorbing state. That is, the probability of an individual in that state to transit to other states is zero. For this project, death state was the absorbing state. Markov processes may be continuous or discrete as well as time homogeneous or time nonhomogeneous.

Markov chain models allow analysts to calculate the probability and rate (or intensity) of movement associated with each transition between states within a single observation cycle as well as the approximate number of cycles spent in a particular state. When observations are made at regular intervals, the number of cycles can be interpreted as time in a state. Time spent in all states prior to absorption can be summed to estimate the total survival time. There exist two assumptions that govern the use of Markov chains: i) The first-order Markov property that is the probability of transition to another state is dictated by the current state; and ii) the time homogeneity property that overtime the transition probabilities are constant. These models have the ability to accommodate simultaneous analysis of multiple events of interest and inclusion of

competing risks through the states defined in the model, as well as consideration of individual frailty through subject-specific random effects (Salazar et al., 2007, Song et al., 2011).

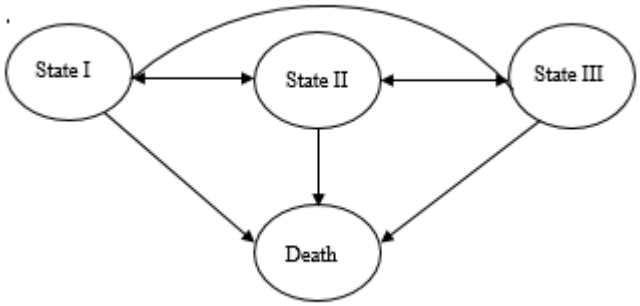
According to Abner et al. (2013), Markov chain models are the most suitable for a time-to-event analysis when using left or right censored datasets. Individuals who are lost to a follow up or withdraws from a study, never reach the absorbing state, thus they are regarded as right-censored since they can contribute information to the model regarding the transitions before being lost to study Hillis et al. (1986). This makes Markov models more advantageous over traditional survival analysis methodology since individuals are not required to enter the transition matrix in any particular state, left-censored data are also accommodated.

Interval censoring is not formally accommodated in Markov chains, which assume that transitions take place only once per observation cycle, either at the beginning or the end. In reality, transitions may take place at any time, and multiple unobserved transitions may take place between cycle assessments. Approaches such as the half-cycle correction, where transitions are assumed to occur in the middle of the observation cycle (Sonnenberg and Beck, 1993), have been proposed to mitigate bias resulting from assuming that transitions take place only at the cycle's beginning or end.

A semi-Markov model, which is a special case of Markov chain where the time spent in the current state depends on both the prior and future adjoining states (Kang and Lagakos, 2007), is highly recommended for interval censored data (Kang and Lagakos, 2007, Kryscio et al., 2013). A semi-Markov approach enables a statistical study of clinical state transitions while accounting for the competing risk of death and facilitates insights into both odds that a risk factor will affect clinical transitions, hence can be used to carry out statistical research of clinical state transitions(Kryscio et al., 2013).

The stages of the semi-Markov process were determined by the level of viral load. Viral load is the measure of the number of HIV viruses in the blood stream majorly measured in milliliters. Literature provides four different viral load stages, $VL < 200$, $200 < VL < 1000$, $VL > 10000$ and Death. The figure below shows the semi-Markov model used to show the flow of HIV/AIDs dynamics through the three stages. To define the model, an arrow will be attached to the diagram to show the probability of the transition and

the holding time it takes for that transition to occur. The holding time at each phase is assumed to match to a distribution that relies on both the current state and the entry state.



1.2 Problem Statement

The main cause of premature deaths in the limited resource setting remains to be HIV/AIDs despite the many HIV management researches done (UNAIDS, 2010). According to (UNAIDS, 2018) report, Kenya has significantly improved in responding to this deadly disease to a prevalence below 5%. This achievement has led to dwindling financial aid in support of HIV programs from global donors, causing challenges in the sustainability of Kenya's HIV response (Kates et al., 2019). Health policy makers are now faced with a challenge of resource allocation among an increasing range of prevention, treatment and care options. Identification of opportunity cost of various allocation options considered by health policy makers as an equally great challenge. Additionally, development and refinement of operational plans that minimize the cost of meeting program goals is a problem in the limited resource setting (Marseille et al., 2009). For all the reasons mentioned above and more, it is essential to analyze and understand the cost and cost-effectiveness of keeping patients within the WHO staging in the resource limited setting.

1.3 Objectives

1.3.1 Main objective

The main objective of the study was to use semi-Markov models in determining the incremental cost-effectiveness of keeping patients on World Health Organization (WHO)-state one verses higher staging on a Differentiated Care Model program in a resource limited setting.

1.3.2 Supporting objectives

1. To assess the mean total cost per state using the semi-Markov model.
2. To model the cumulative cost of each health state by combining semi-Markov modeling process and the regression approach.

1.4 Significance of the research

The acquaintance gained while researching should enable policy makers to efficiently allocate resources among an escalating range of prevention, care option, and treatment. Additionally, the knowledge would contribute in development and improvement of operational strategies that reduce the cost of treatment.

In theory, the study will demonstrate that a semi-Markov approach in assessing HIV/AIDs state cost-effectiveness would yield more accurate results that can be used for inter-state comparison.

Chapter 2

Literature review

2.1 Introduction

HIV/AIDs infection in resource limited areas is a threat that challenges the population health. Ensuring timely access to HIV/AIDs treatment and quality health care services could significantly avert the high number of deaths due to HIV/AIDs in resource limited areas. Currently, there is a lot of evidence on interventions that could help to overcome the difficulty in resource-limited areas Filmer and Scott (2008).

ART and HAART are some of the major interventions increasingly being used in monitoring HIV/AIDs patients. Differentiated Care Models are as well used as follow up strategies. These strategies have proven to provide quality care to HIV/AIDs patients.

These strategies, however, are majorly financed by donor agencies, thus continue to operate as small scale pilot schemes. Moreover, such organizations generally suffer substantial administrative and operational costs. These costs present a challenge of sustainability, and scale-up. Furthermore, there is still scarcity of information regarding their cost implications on health. Cost information is essential for facilitating health policy and decision making (Bhatia and Gorter, 2007, Ensor and Cooper, 2004).

2.2 Incremental multistage cost

According to Cohen and Reynolds (2008), health care spending is an increasingly important economic and political issue. The aim of incremental multistage cost research is to allow clinicians and policymakers to make more rational decisions regarding clinical care and resource allocation especially in resource limited areas. Incremental multistage cost analysis would significantly contribute in assessing the value of new medical strategies,

by simultaneously examining incremental health benefits in light of incremental costs

2.3 Current models

Lin et al. (1997) pioneered a non-parametric method for estimating medical costs from incomplete follow-up data. The main principle of this methodology was to divide the entire time period of interest into several intervals and then estimate the average total cost by the sum of the Kaplan-Meier estimator for the probability of dying in each time interval multiplied by the sample mean of the total costs from those who are observed to die in that interval. This method is limited due to its assumption of independent censoring.

In 2003, Austin et al. compared the performance of several broad approaches of modelling cost analysis: linear regression, linear regression with log-transformed cost, generalized linear models (GLMs), median regression, and proportional hazards models. Both survival and cost data are commonly censored; therefore, methods presented by Austin et al. (2003) were limited since they could not account for censored data.

Willan et al. (2005) presented two regression models for cost and survival data which use the inverse probability of censoring weighted (IPCW) method to account for censored data and also address the covariance between survival and cost. However, regression methods are not well suited to survival data.

Recently, Liu et al. (2007) proposed a shared random effects model for monthly medical costs and survival time. This model would account for correlation between survival time and monthly medical costs.

The above reviewed methods do not account for the whole clinical evolution of disease in survival analysis. Literature presents several methods which would work well in studying the whole clinical evolution of disease in survival analysis. Markov models are particularly useful when a decision probably involves a risk that is ongoing over time (Sonnenberg and Beck, 1993).

2.4 Markov and semi-Markov models

Sonnenberg and Beck (1993) argues that decision trees are best used in performing cost effective analysis. They presented discrete time Markov Models for medical decision making. The model eases the ability to describe all the patient's trajectories, which are applied with fixed transition rates computed from previous information.

Gardiner et al. (2006b) presented a stochastic model for statistical inference in cost effectiveness analysis. According to Gardiner et al. (2006b) it is less common for multi-state models to be actually fitted to data for use in cost-effectiveness analyses. Gardiner et al. (2006a) recently identified a multi state model (Markov model) and a regression method for estimating changes in health status and costs.

2.5 Model Assessment

2.5.1 Assessing semi-Markov sojourn time distribution

The goodness of Fit (GOF) of a statistical model describes how well it fits into a set of observations. Akaike (1969), presents Akaike information criterion (AIC) for comparing models based on goodness-of-fit. We used AIC in selecting the best model for this study.

2.5.2 Assessing cost

Cost-effectiveness analysis is a method for evaluating costs and health outcomes of interventions that allows the relative value of different interventions to be compared (Weinstein et al., 1996). While policy makers use cost-effectiveness analyses to assist in understanding what interventions might provide the best value for money (Chambers et al., 2015), cost-effectiveness analyses – and their related sensitivity analyses – also provide important additional information such as clinical, epidemiologic, and/or economic benchmarks for interventions to achieve cost-effectiveness. If an intervention is not cost-effective under current conditions, analyses can project under what conditions it might become so.

Willan and Lin (2001) identifies three broad approaches to health economic evaluation for comparing two therapies: cost minimization analysis (CMA), in which one assumes or observes no difference in cost effectiveness, incremental cost effectiveness ratio (ICER), and incremental net benefit (INB).

The main problem of CMA is it's prone to overestimation or underestimation of the value of information and the probability that treatment is cost-effective, because it introduces bias into uncertainty estimates (Dakin and Wordsworth, 2013). For this study we used ICER which corrects the shortcomings of CMA.

2.6 Conclusion

The reviewed literature in this study, points out the milestones made so far in modeling of HIV/AIDs disease using Markov and semi Markov models. According to Marseille et al. (2009) incremental cost effectiveness analysis is important in a resource limited setting since it is geared toward describing treatment cost in each state. This would shed light to the health policy makers on the opportunity cost of various allocation options, help in allocating resources among an increasing range of prevention, treatment and care options. It would also help in refining and developing operational plans that would minimize the treatment cost.

However, we are cognizant that time spent in each stage of the disease cannot be predictable on the basis of clinical and immunological measures, which would help in analyzing the cost effectiveness of the heath stages.

Models' appropriateness has been reviewed by developing transition matrices; the semi Markov stochastic process is preferred, since it's natural and flexible in modelling clinical progression. Its ability to accommodate covariates while assessing the mean total costs credits the model (Foucher et al., 2005).

In view of the dwindling financial support of HIV programming there is need to determine the cost of keeping patients on World Health Organization(WHO) stage one verses other higher stages. This study uses a semi-Markov approach in assessing the total mean, and a combination of semi-Markov and regression methods in modeling the

clinical evolution of the patients' disease and modeling of cumulative costs in each health state respectively.

Chapter 3

Methodology

3.1 The Modeling Framework

3.1.1 Semi-Markov Model

Introducing semi-Markov model according to Castelli et al. (2007), suppose S is a discrete stage space and each patient is observed for a period of time t for h successive stages. That is $X = \{X_0, X_1, \cdots, X_h\}$, where the initial and final stages are denoted by X_0 and X_h respectively. Assuming $X \in S$ and the number of possible stage to be finite, the process $X = X_h; h \geq 0$ is considered to be a semi-Markov chain with h transitions (Castelli et al., 2007). In this case entry times sequence T_n for each stage X_h after h transitions is denoted as $T = (0 = T_0, T_1, \cdots, T_h)$. The transition probabilities from a stage to another $(i \to j)$ is represented as,

$$P_{ij} = P(X_{h+1} = j | X_h = i)$$

and is homogeneous since P_{ij} doesn't depend on t. The conditional distribution function

$$G_{ij}(t) = P(T_{h+1} - T_h \leq t | X_{h+1} = j, X_h = i)$$

defines the sojourn time between two stage (i, j).

3.1.2 Distribution of Sojourn Times

The transition probability of a HIV/AIDs patient from one stage to another relies on how much time he/she spends on the stage (Goshu and Dessie, 2013). The time spent in a given stage is called waiting or sojourn time. Supposing the sojourn time for a

patient is random and follows a given distribution $G_{ij}(t)$, the following waiting time distributions will be considered in modeling and accessing the best distribution that describes the HIV/AIDs patients progression.

Exponential Distribution

The hazard function under exponential distribution is constant for a Markovian case (Mengesha et al., 2018). The distribution is defined by

$$G_{ij}(t) = 1 - exp(\lambda_{ij}), \quad t \geq 0$$

where, $\lambda_{ij} = \frac{1}{\sigma_{ij}}$ and $\sigma_{ij} > 0$. λ_{ij} is the expected time that the a patient stays in a specific stage X_h before transiting to stage j from stage i. The hazard function of exponential distribution is given by

$$\lambda_{ij}(X) = \frac{1}{\sigma_{ij}}, \quad \forall X \geq 0 \; and \; \sigma_{ij} \; j > 0$$

Weibull Distribution

The weibull distribution sojourn time is given by;

$$G_{ij}(t) = 1 - exp(-\lambda_{ij}\, t^{V_{ij}}), \quad t \geq 0$$

Weibull distribution generalizes the exponential distribution and is highly used in modeling clinical progression because of its flexibility in nature (Castelli et al., 2007, Mengesha et al., 2018). Mengesha et al. (2018) suggests a weibull distribution with two parameters $\{\sigma_{ij}, V_{ij}\}$ to take care of various shapes for monotone hazards. The Weibull distribution hazard function is defined as,

$$\lambda_{ij}(x) = V_{ij} \left(\frac{1}{\sigma_{ij}}\right)^{V_{ij}} X^{V_{ij}-1}, \quad X \geq 0$$

3.1.3 Modeling the costs and the cumulative costs in each health stage

Based on the notation of Liu, (2007) a regression approach was selected to model the mean cost in each health stage. The time consumed in each health stage is distributed into k_h intervals and a regression model was constructed for each interval.

Assuming $Y = (0, \tau]$ is the time of interest. Y is divided into K_h intervals such that $y_k = (a_k^h, a_{k+1}^h]$. The regression equation takes the form;

$$C_{k,i}^h = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \cdots + \alpha_p Z_p \tag{3.1}$$

Where $C_{k,i}^h$ is the observed cost of patient i, z_i are covariate in health stage h in the interval k and α_i are the co-responding coefficiencies of the covariates. Equation 3.1 can be reduced to;

$$C_{k,i}^h = \alpha_k^h Z_i; \quad i = 1, 2, \cdots, N; h = 1, \cdots, H \tag{3.2}$$

where;
α_k^h is a vector of unidentified regression parameters in an interval k in the health stage h, Z_i is a vector of covariates, i and h are number of patients and health stages respectively.

If a patient i within the interval k has no cost record, then $C_{k,i}^h = 0$. Taking the expectation of observed cost (3.2) above;

$$\mathbb{E}(C_{k,i}^h) = \mathbb{E}(\alpha_k^h Z_i),$$

$$\hat{C}_{k,i}^h = \hat{\alpha}_k^h Z_i$$

where $\hat{\alpha}_k^h$ is the estimator of the vector of regression parameters on the interval k for health stage h defined as follows

$$\hat{\alpha}_k^h = (\mathbf{Z}\mathbf{Z}')^{-1}\mathbf{Z}'\mathbf{C}_k^h$$

where $\mathbf{Z}$ is is a $p \times N$ matrix of the covariates $z1, \cdots, z_N$ and $\mathbf{C}_k^h$ is the $N \times 1$ vector of

the observed costs $C^h_{k,1}, \cdots, C^h_{k,N}$.

Using U^h_i to denote the total time spent by a patient i in stage h, then the cumulated cost function at time t of the patient i is expressed as;

$$C^h_i(t) = \sum_{\forall k(a^h_k \leq U^h_i; a^h_k \leq t \leq a^h_{k+1})} \hat{C}^h_{k,i} + \hat{C}^h_{k+1,i} \left[\frac{U^h_i - a^h_k}{a^h_{k+1} - a^h_k} \right]$$

The cumulative cost functions obtained are equivalent to the number of the heath stage (H) present. The variance of the estimate of C^h_i can be deduced by assuming independence of the $\hat{\alpha}^h_k$ between each interval.

3.1.4 Assessing the goodness of fit

To validate the best Markov model that describes the real data set we used Akaike information criterion (AIC) for comparison between the above discussed sojourn time distributions. According to Portet (2020), the lower the AIC value the better the model.

Mathematically the AIC is calculated as,

$$AIC = 2K - 2ln(\mathcal{L}(\hat{\alpha}_{MLE}|c))$$

Where K is the number of estimated parameters including the variance, $\mathcal{L}$ is the likelihood function, $\hat{\alpha}$ is the maximum likelihood estimate of α.

3.1.5 Assessing the incremental cost-effectiveness using Incremental cost-effectiveness ratio (ICER)

Cost-effectiveness analysis is a method for evaluating costs and health outcomes of interventions that allow the relative value of different interventions to be compared (Weinstein et al., 1996). While policy makers use cost-effectiveness analyses to assist in understanding what interventions might provide the best value for money (Chambers et al., 2015), cost-effectiveness analyses – and their related sensitivity analyses – also provide important additional information such as clinical, epidemiologic, and/or eco-

nomic benchmarks for interventions to achieve cost-effectiveness. If an intervention is not cost-effective under current conditions, analyses can project under what conditions it might become so. In assessing the incremental cost of patients on DCM and they who were not on DCM we used incremental cost-effectiveness ratio as stated by, Willan and Lin (2001). Mathematically it is defined as;

$$ICER = \frac{\overline{C}_{intervention} - \overline{C}_{control}}{\overline{E}_{intervention} - \overline{E}_{control}}$$

where $\overline{C}$ and $\overline{E}$ represent the average costs and average time levels, respectively.

Chapter 4

Results

4.1 Data description

The follow-up of patients with HIV/AIDs is still controversial. The means mobilized for postoperative monitoring come at a high cost. To retrospectively evaluate modalities, results and costs of HIV/AIDs patient follow-up, we studied clinical data from record of 738 patients obtained from 9 facilities namely Kiambu District Hospital, Ruiru Sub-District Hospital, Karuri Sub-District Hospital, Thika Level V Hospital, Kiandutu Health Center, Wangige Sub-District Hospital, Gatundu District Hospital, Igegania Sub-District Hospital and Limuru Health Center.

The available characteristics for each patient were Patient_Id, State.i, State.j time, Viral load, DCM (1=Patients on DCM, 2=Patients not on DCM), Age, Gender (1-male, 2-female), and facility location. Table 4.1 describes the proportion of patients in the sample according to their characteristics. The mean costs for the state I, State II and State III were about $765, $829 and $1395 respectively.

Table 4.1: *Distribution of patients according to their characteristics.*

	DCM follow up (Percent)-1	No follow-up (percent)-2
Sex		
1	240 (0.52)	253 (0.52)
2	226 (0.48)	232 (0.48)
Tb-coinfection		
1	66 (0.13)	33 (0.07)
2	452 (0.87)	472 (0.93)
Facility location		
1	20 (0.03)	26 (0.04)
2	20 (0.03)	27 (0.04)
3	36 (0.05)	25 (0.04)
4	137 (0.2)	11 (0.02)
5	189 (0.27)	73 (0.11)
6	146 (0.21)	140 (0.21)
7	73 (0.11)	213 (0.32)
8	42 (0.06)	108 (0.16)
9	32 (0.05)	37 (0.06)
N (per cent)	465 (0.49)	485 (0.51)

4.2 Semi-Markov Model

The objective of this study is to model the follow-up of patients with HIV/AIDs. Three health stages are defined (Figure 2): 'stage 1, 'stage 2', and 'stage 3. Note that the stage 'dead' is persistent because the probability to move out of this stage is null. Table 4.2 describes the frequency of the transitions observed in the database. To obtain the most parsimonious model, the data set was subjected to two distributions (Weibull and Exponetial) and their AIC examined.

Table 4.2: *Frequency of transitions and summaries of the sojourn times*

			Sojourn times				
Transition	n	per cent	Min	Max	Mean	Median	Std. dev
Transition 1 → censoring*	303	16.38	0.10	7.64	4.37	5.10	2.60
Transition 1 → 2	190	10.27	0.10	5.98	0.83	0.39	5.98
Transition 1 → 3	87	4.70	0.10	3.03	0.72	0.34	0.80
Transition 2 → 1	223	12.05	0.10	4.90	0.54	0.30	0.71
Transition 2 → 3	142	7.68	0.12	4.39	0.82	0.46	0.89
Transition 2 → censoring*	232	12.54	0.00	7.64	3.92	3.41	2.63
Transition 3 → 1	230	12.43	0.07	1.93	0.34	0.25	0.28
Transition 3 → 2	240	12.97	0.10	5.28	0.67	0.32	0.91
Transition 3 → censoring*	203	10.97	0.07	7.64	3.72	3.09	2.10
Total	1850	100					

Table 4.3: *Estimates of semi-Markov parameters for Exponential and Weibull distributions and the significant effects from the set of three covariates.*

Model	Transition	Distribution	DCM	sd	Age	sd.1	Gender	sd.2	pij
Weibull	1–>2	Weibull	24.80	0.01	-1.06	0.02	0.14	0.16	0.59
Weibull	1–>3	Weibull	9.59	0.02	-0.84	0.01	0.14	0.04	0.41
Weibull	2–>1	Weibull	-0.92	0.01	-1.23	0.06	0.09	0.13	0.66
Weibull	2–>3	Weibull	10.20	0.02	-0.88	0.01	-0.07	0.02	0.34
Weibull	3–>1	Weibull	78.10	0.03	-4.84	0.00	0.09	0.13	0.37
Weibull	3–>2	Weibull	13.00	0.01	-1.15	0.01	-0.09	0.13	0.63
Exponential	1–>2	Exp	6.21	0.02	-0.35	0.01	0.05	0.13	0.62
Exponential	1–>3	Exp	4.53	0.02	-0.45	0.02	-0.15	0.21	0.38
Exponential	2–>1	Exp	-1.80	0.01	-0.17	0.01	0.48	0.11	0.65
Exponential	2–>3	Exp	3.60	0.04	-0.34	0.02	0.00	0.15	0.35
Exponential	3–>1	Exp	13.40	0.00	-0.82	0.01	0.06	0.13	0.37
Exponential	3–>2	Exp	4.77	0.02	-0.48	0.01	-0.00	0.11	0.63

Table 4.4: *Measure of best model*

Model	AIC	Complexity
Weibull	717.44	33
Exponential	2395.17	27

4.3 Cost

In this section we assessed the mean costs. All treatment cost available for each patient were included. The mean cost per patient is assessed according to the method previously presented. A regression model is thus performed in this study. This helped in dropping all non-informative covariates and select those covariates that seemed to have

an effect on the mean cost. For convenience, the selected covariates were the same in each interval. To choose these covariates we plotted the cumulated costs according to the characteristics of patients observed in the database (Figure 4.1). Costs seem to be different for each level of covariate 'TB-coinfection','Follow-up (DCM)', and 'stage' and very close for sex and location facility.

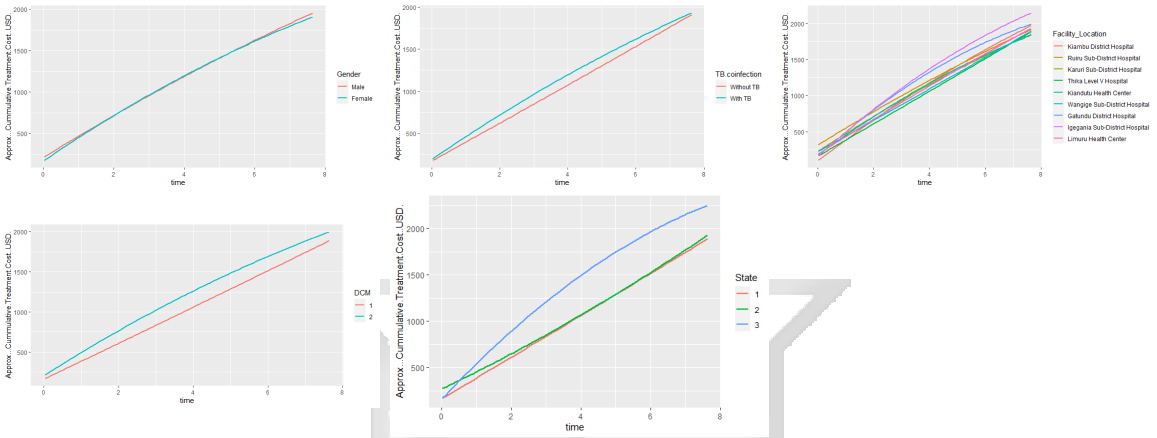


Figure 4.1: *Cumulative costs of patients staying alive at time t according to their covariates.*

Thus, there seems to be no significant difference in cost for location and sex. In addition to the graphical analysis, a regression model was performed for the total cost of the three stages for all patients. The regression parameters ($\hat{\alpha}_C$) were estimated by least-squares method. Results are shown in Table 4.5 below.

Table 4.5: *Results of the regression model on total costs for all patients within the three stages*

	Cumulative total cost		
	$\alpha(Pr(>\|t\|))$		
Covariates	stage 1	stage 2	stage 3
Intercept	-633.895 (0.00)	-662.779 (0.00)	-918.843 (0.00)
DCM	-147.369 (0.00)	-92.315 (0.00)	-93.2921 (0.00)
Age	66.0635 (0.00)	64.4413 (0.00)	81.3969 (0.00)
Gender	4.20326 (0.00)	6.4529 (0.24)	10.6762 (0.15)

A plot of the cumulative functions obtained with the $\hat{\alpha}_k$ is as shown in Figure 4.2 below. The DCM follow up only applied for patients in stage 1 as stage 2 and 3 were considered critical hence close medical care. Notably, from the graph the incremental cumulative cost for patients in stage 1 (in DCM) was significantly lower compared to the other stages (not in DCM). Though the differences in cumulative cost for patients

was significant, the cumulative cost for the three stages had an increasing trend over time. Using the ICER equation presented in the methodology the incremental cost was about $483.83.

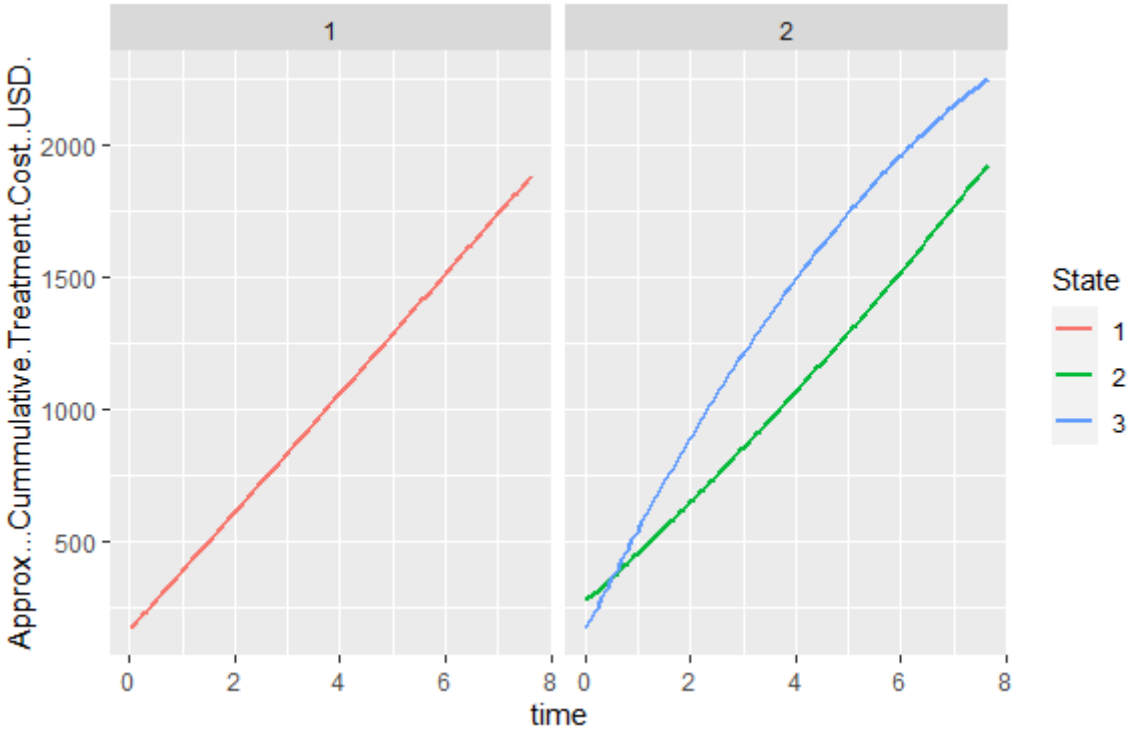


Figure 4.2: *Cumulative costs function for DCM according to patient profiles.*

Chapter 5

Discussion, Conclusion and Recommendations

Despite the many research that have been done, HIV/AIDs remains to be one of the primary contributors of premature deaths. In this paper, we presented a way of performing cost-effectiveness analysis using HIV/AIDs dataset. Understanding the cost-effectiveness of HIV/AIDs patients is very critical following the decrease in the countries prevalence level at a time when the donors' financial aid has decreased. This will simplify policy makers' ability to make decisions.

A semi-Markov model which is a special case of the Markovian models was used for this study. The semi-Markov process allows for more flexible modeling of hazard functions than the Markovian process in which the hazard functions are time constant and therefore the sojourn times are exponentially distributed.

With the existence of many distributions that can be used in describing sojourn time in semi-Markov process it is important to perform an analysis geared towards choosing the best distribution. AIC values are used in showing performance of distribution the smaller the AIC the better. The dataset was subjected to exponential and weibull distributions and best model chosen by means of evaluating their AICs.

Weibull distributions performed the best compared to Exponential hence was used for this paper in explaining the sojourn times. Costs and health states can be linked with this method. Additionally, by using this method, cost data are modeled by health states and therefore more homogeneous. A combination of semi-Markov and Regression models was used in order to access the mean total cost depending on the heath state.

On average the mean cost of managing individuals on state I, State II and State III are \$765, \$829 and \$1395 respectively. From the cumulative cost plots above, the incremental cost of stage one is relatively lower compared to state II and State III. This

difference is attributed to the follow up strategy (DCM). From this study's findings it is relatively cheap to keep a patient in State I compared to II and III.

The cost effectiveness analysis of DCM shows that putting patients on DCM is not only cheap but also more effective compared to treating patients not on DCM. Results from this study indicate that the average cost for interventions group (individuals not in DCM) was \$213.3806 with increase in time by 0.4410. Based on this the ICER was \$483.8268 for any additional time. Notably from Table 4.5, the patients cumulative cost within any stage is influenced by, age and a follow up strategy. Gender was found not to be statistically significant in influencing the total cost within a stage.

Cost data are mostly skewed, with long-tailed distribution and so the assumption of normality for the cost distributions in the cost model presented, might not always apply. Additionally, it is not verifiable that the different intervals were independent. This modeling could further be extended to using some parametric regression models, such as GLM with a gamma or log-linear distribution, which seem to be more adapted to a skewed distribution. This approach can also be extended to scenarios where interventions which could affect cost were part of the study. Lastly, additional distributions like gamma and log-normal could be considered in choosing the best sojourn time distribution.

Appendix A

R codes

```
library(readxl)
library("p3state.msm")
library("mstate")
library(flexsurv)
library(SemiMarkov)
library(tidyverse)
library(dplyr)
library(xtable)
library(stargazer)
Data <-data.frame(read_excel("SemiMakovData_WithCost.xlsx"))
head(Data)


table.state(Data)
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("W","W")
mtrans_1[2, c(1,3)] <- c("W","W")
mtrans_1[3, c(1,2)] <- c("W","W")
table.state(Data, states = states_1, mtrans = mtrans_1)



## semi-Markov model with a covariate "Age","Gender","
Age <- as.data.frame(Data$Age)
Gender <- as.data.frame(Data$Gender)
DCM<-as.data.frame(Data$DCM)
model <- semiMarkov(data = Data, cov = as.data.frame(cbind(DCM,Age,Gender
```

```
                    states = states_1, mtrans = mtrans_1)

model.name <- "Weibull"
res <- matrix((as.numeric(as.matrix(model$table.coef[,5:6]))),ncol=2)
res.coef <- cbind(res[c(1:6),],res[c(7:12),],res[c(13:18),],model$table.p
res.coef[,c(1,3,5,7)] <- signif(res.coef[,c(1,3,5,7)],digits = 3)
colnames(res.coef) <- c("DCM","sd","Age","sd","Gender","sd","pij")
res.data.frame <- rbind(data.frame(Model=rep(model.name,3),
                    Transition=c("1-->2","1-->3","2-->1","2-->3","3-->1","3
                    Distribution=rep("Weibul",3),res.coef))
AIC.sojourn <- 2*(model$minus2loglik+dim(model$param.init)[1])
AIC.sojourn

res.AIC.sojourn <- NULL
res.NP.sojourn <- NULL

res.AIC.sojourn <- c(res.AIC.sojourn,AIC.sojourn)
res.NP.sojourn <- c(res.NP.sojourn,dim(model$param.init)[1])

######################Exponent#############
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("E","E")
mtrans_1[2, c(1,3)] <- c("E","E")
mtrans_1[3, c(1,2)] <- c("E","E")
table.state(Data, states = states_1, mtrans = mtrans_1)
model <- semiMarkov(data = Data, cov = as.data.frame(cbind(DCM,Age,Gender
                    states = states_1, mtrans = mtrans_1)
model.name <- "Exponential"
res <- matrix((as.numeric(as.matrix(model$table.coef[,5:6]))),ncol=2)
res.coef <- cbind(res[c(1:6),],res[c(7:12),],res[c(13:18),],model$table.p
res.coef[,c(1,3,5,7)] <- signif(res.coef[,c(1,3,5,7)],digits = 3)
```

```
colnames(res.coef) <- c("DCM","sd","Age","sd","Gender","sd","pij")
res.data.frame <- rbind(res.data.frame,data.frame(Model=rep(model.name,3)
                                                  Transition=c("1-->2","1-->3","2-->1","
                                                  Distribution=rep("Exp",3),res.coef))
AIC.sojourn <- 2*(model$minus2loglik+dim(model$param.init)[1])
AIC.sojourn
res.AIC.sojourn <- c(res.AIC.sojourn,AIC.sojourn)
res.NP.sojourn <- c(res.NP.sojourn,dim(model$param.init)[1])

#########################Log Normal###############
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("E","E")
mtrans_1[2, c(1,3)] <- c("E","E")
mtrans_1[3, c(1,2)] <- c("E","E")
model <- semiMarkov(data = Data, cov = as.data.frame(cbind(DCM,Age,Gender
                    states = states_1, mtrans = mtrans_1)
model.name <- "Exponential"
res <- matrix((as.numeric(as.matrix(model$table.coef[,5:6]))),ncol=2)
res.coef <- cbind(res[c(1:6),],res[c(7:12),],res[c(13:18),],model$table.p
res.coef[,c(1,3,5,7)] <- signif(res.coef[,c(1,3,5,7)],digits = 3)
colnames(res.coef) <- c("DCM","sd","Age","sd","Gender","sd","pij")
res.data.frame <- rbind(res.data.frame,data.frame(Model=rep(model.name,3)
                                                  Transition=c("1-->2","1
                                                  Distribution=rep("Exp"
AIC.sojourn <- 2*(model$minus2loglik+dim(model$param.init)[1])
AIC.sojourn
res.AIC.sojourn <- c(res.AIC.sojourn,AIC.sojourn)
res.NP.sojourn <- c(res.NP.sojourn,dim(model$param.init)[1])

################AIC###########
model <- rep(c("Weibull","Exponential"),each=6)
```

```
model.AIC <- c("Weibull","Exponential")
table.AIC <- data.frame(Model=model.AIC,AIC=res.AIC.sojourn,Complexity=re
table.AIC
print(xtable(table.AIC),hline.after = c(0,2), size = "small",include.rown

subdata<-data.table(data.frame(cbind(Patient_id=Data$Patient_id,DCM=Data$I

Data1<-data.table(data.frame(cbind(Patient_id=Data$Patient_id,
                                   DCM=Data$DCM,cost=Data$Approx...Cummu
str(Data1)
# ICER
(costDif <- Data[DCM == 2, mean(Approx...Cummulative.Treatment.Cost..USD.
    Data[DCM == 1, mean(Approx...Cummulative.Treatment.Cost..USD.)])

(nDif <- Data[DCM == 2, mean(time)] -
    Data[DCM == 1, mean(time)])

costDif/nDif
```

VT OMNES
VNVM SINT

Appendix B

Approval Letter

Strathmore UNIVERSITY

10th August 2021

Mr Andrew Joram,
joram.malului@strathmore.edu

Dear Mr Andrew,

RE: Modelling Incremental Changes in HIV Staging with a Cost Effect; Case of semi-Markov Process

This is to inform you that SU-IERC has reviewed and **approved** your above **SU- master's** research proposal. Your application reference number is **SU-IERC1030/21.** The approval period is **3rd August 2021 to 2nd August 2022.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: Dr Virginia Gichuru,
Secretary; SU-IERC

Cc: Prof Fred Were, **Chairperson; SU-IERC**

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

References

Abner, E. L., Charnigo, R. J., and Kryscio, R. J. (2013). Markov chains and semi-markov models in time-to-event analysis. *Journal of biometrics & biostatistics*, (e001):19522.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, 21(1):243–247.

Austin, P. C., Ghali, W. A., and Tu, J. V. (2003). A comparison of several regression models for analysing cost of cabg surgery. *Statistics in medicine*, 22(17):2799–2815.

Bhatia, M. and Gorter, A. (2007). Improving access to reproductive and child health services in developing countries: are competitive voucher schemes an option? *Journal of International Development: The Journal of the Development Studies Association*, 19(7):975–981.

Castelli, C., Combescure, C., Foucher, Y., and Daures, J.-P. (2007). Cost-effectiveness analysis in colorectal cancer using a semi-markov model. *Statistics in medicine*, 26(30):5557–5571.

Chambers, J. D., Cangelosi, M. J., and Neumann, P. J. (2015). Medicare's use of cost-effectiveness analysis for prevention (but not for treatment). *Health Policy*, 119(2):156–163.

Cohen, D. J. and Reynolds, M. R. (2008). Interpreting the results of cost-effectiveness studies. *Journal of the American College of Cardiology*, 52(25):2119–2126.

Dakin, H. and Wordsworth, S. (2013). Cost-minimisation analysis versus cost-effectiveness analysis, revisited. *Health economics*, 22(1):22–34.

Ensor, T. and Cooper, S. (2004). Overcoming barriers to health service access: influencing the demand side. *Health policy and planning*, 19(2):69–79.

Filmer, D. and Scott, K. (2008). *Assessing asset indices*. The World Bank.

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J.-F., and Daurès, J.-P. (2005). A semi-markov model based on generalized weibull distribution with an illustration for hiv disease. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(6):825–833.

Gardiner, J. C., Luo, Z., Bradley, C. J., Sirbu, C. M., and Given, C. W. (2006a). A dynamic model for estimating changes in health status and costs. *Statistics in medicine*, 25(21):3648–3667.

Gardiner, J. C., Luo, Z., Liu, L., and Bradley, C. J. (2006b). A stochastic framework for estimation of summary measures in cost–effectiveness analyses. *Expert review of pharmacoeconomics & outcomes research*, 6(3):347–358.

Goshu, A. T. and Dessie, Z. G. (2013). Modelling progression of hiv/aids disease stages using semi-markov processes. *Journal of Data Science*, 11(2):269–280.

Hillis, A., Maguire, M., Hawkins, B. S., and Newhouse, M. M. (1986). The markov process as a general method for nonparametric analysis of right-censored medical data. *Journal of chronic diseases*, 39(8):595–604.

Kang, M. and Lagakos, S. W. (2007). Statistical methods for panel data from a semi-markov process, with application to hpv. *Biostatistics*, 8(2):252–264.

Kates, J., Wexler, A., Lief, E., et al. (2019). Donor government funding for hiv in low- and middle-income countries in 2019. *Menlo Park, CA: The Henry J Kaiser Family Foundation & UNAIDS*.

Kryscio, R. J., Abner, E. L., Lin, Y., Cooper, G. E., Fardo, D. W., Jicha, G. A., Nelson, P. T., Smith, C. D., Van Eldik, L. J., Wan, L., et al. (2013). Adjusting for mortality when identifying risk factors for transitions to mild cognitive impairment and dementia. *Journal of Alzheimer's Disease*, 35(4):823–832.

Lin, D., Feuer, E., Etzioni, R., and Wax, Y. (1997). Estimating medical costs from incomplete follow-up data. *Biometrics*, pages 419–434.

Liu, L., Wolfe, R. A., and Kalbfleisch, J. D. (2007). A shared random effects model for censored medical costs and mortality. *Statistics in medicine*, 26(1):139–155.

Marseille, E., Kahn, J. G., Pitter, C., Bunnell, R., Epalatai, W., Jawe, E., Were, W., and Mermin, J. (2009). The cost effectiveness of home-based provision of antiretroviral therapy in rural uganda. *Applied health economics and health policy*, 7(4):229–243.

Mengesha, S. K., Gebremedhn, G. A., Ferede, T., and Atsmegiorgis, C. (2018). Application of multi-state semi-markov models on hiv/aids disease progression. *i-Manager's Journal on Mathematics*, 7(3):30.

Portet, S. (2020). A primer on model selection using the akaike information criterion. *Infectious Disease Modelling*, 5:111–128.

Salazar, J. C., Schmitt, F. A., Yu, L., Mendiondo, M. M., and Kryscio, R. J. (2007). Shared random effects analysis of multi-state markov models: application to a longitudinal study of transitions to dementia. *Statistics in medicine*, 26(3):568–580.

Song, C., Kuo, L., Derby, C. A., Lipton, R. B., and Hall, C. B. (2011). Multi-stage transitional models with random effects and their application to the einstein aging study. *Biometrical Journal*, 53(6):938–955.

Sonnenberg, F. A. and Beck, J. R. (1993). Markov models in medical decision making: a practical guide. *Medical decision making*, 13(4):322–338.

UNAIDS (2016). Prevention gap report: Joint united nations programme on hiv/aids. *Geneva: UNAIDS*.

UNAIDS (2018). Kenya aids response progress report 2018. kenya. *UNAIDS*.

UNAIDS, J. (2010). Global report: Unaids report on the global aids epidemic 2010. *Geneva: UNAIDS*.

UNAIDS, n. (Retrieved July 27, 2020). *UN joint programme on HIV/AIDS (UNAIDS). Trends of new HIV infections*.

Weinstein, M. C., Siegel, J. E., Gold, M. R., Kamlet, M. S., and Russell, L. B. (1996). Recommendations of the panel on cost-effectiveness in health and medicine. *Jama*, 276(15):1253–1258.

WHO (2012). Programmatic update: antiretroviral treatment as prevention (tasp) of hiv and tb: executive summary. Technical report, World Health Organization.

Willan, A. R. and Lin, D. (2001). Incremental net benefit in randomized clinical trials. *Statistics in medicine*, 20(11):1563–1574.

Willan, A. R., Lin, D., and Manca, A. (2005). Regression methods for cost-effectiveness analysis with censored data. *Statistics in medicine*, 24(1):131–145.