

## Comparison of Sojourn Time Distributions in Modeling HIV/AIDS Disease Progression

Tilahun Ferede Asena<sup>1</sup>, Ayele Taye Goshu<sup>2</sup>

<sup>1</sup>School of Mathematical and Statistical Sciences, Hawassa University, Hawassa, 05, Ethiopia  
e-mail: feredetilahun14@gmail.com / tilahun.ferede@amu.edu.et

<sup>2</sup>School of Mathematical and Statistical Sciences, Hawassa University, Hawassa, 05, Ethiopia  
e-mail: ayele\_taye@yahoo.com

### SUMMARY

An application of semi-Markov models to AIDS disease progression was utilized to find best sojourn time distributions. We obtained data on 370 HIV/AIDS patients who were under follow-up from September 2008 to August 2015, from Yirgalim General Hospital, Ethiopia. The study reveals that within the “good” states, the transition probability of moving from a given state to the next worst state has a parabolic pattern that increases with time until it reaches a maximum and then declines over time. Compared with the case of exponential distribution, the conditional probability of remaining in a good state before moving to the next good state grows faster at the beginning, peaks, and then declines faster for a long period. The probability of remaining in the same good disease state declines over time, though maintaining higher values for healthier states. Moreover, the Weibull distribution under the semi-Markov model leads to dynamic probabilities with a higher rate of decline and smaller deviations. In this study, we found that the Weibull distribution is flexible in modeling and preferable for use as a waiting time distribution for monitoring HIV/AIDS disease progression.

**Key words:** HIV/AIDS, semi-Markov model, sojourn time distributions, transition probability

### 1. Introduction

Disease modeling and mapping are becoming important in the biomedical sciences for investigation of the future status of individual patients. A Markov model can describe the natural course of disease progressions. Such models are particularly used in medical applications where stages or levels of diseases are represented by states in the model. Both semi-Markov and hidden Markov models

are promising and exciting tools for modeling a wealth of biological and biomedical data. In pathology, individuals who are suffering from a particular disease pass through a series of disease stages, moving either to worse or to better states. The continual passage of a disease from state to state can be modeled by assuming these disease stages to be states of a stochastic model. Disease progression modeling is the modeling of the progression of a target disease with computational methods, and is an important technique that can help with the early detection and management of chronic diseases. By characterizing the entire disease progression trajectory, disease progression modeling also facilitates improved prognosis, drug development, and clinical trial design. Examples of staging systems that are based on disease progression exist for diseases such as cancer, HIV, TB, Alzheimer's, diabetes, and chronic obstructive pulmonary diseases. In recent decades homogeneous semi-Markov processes have been utilized by Corradi et al. (2004) and Janssen and Monica (2001). Other studies include D'Amico et al. (2009) and Goshu and Dessie (2013).

Progression between the stages of a disease can be characterized by the use of well-defined states to represent the various stages. Many diseases have a long preclinical phase during which the disease progresses and eventually becomes clinically recognized. It is then important to know what factors accelerate the progression and when this progression occurs. Disease progression modeling is found to be very important because slowing or preventing disease progression may be a more effective strategy for reducing morbidity than trying to prevent onset. HIV/AIDS is a continuum of progressive damage to the immune system, from the time of infection to the manifestation of severe immunological damage by the opportunistic infections, neoplasms, wasting, or low CD4 lymphocyte count that define AIDS (Teni and Asena, 2015).

This study deals with the comparison of sojourn times for semi-Markov models of longitudinal data, which refers to data on individuals measured repeatedly at different times. Semi-Markov models for disease progression have been utilized by numerous researchers in the past two decades. Among them, early papers on disease progressions using semi-Markov studies include Joly and

Commenges (1999), Foucher et al. (2005), Mathieu et al. (2005), Mandel (2010), Fonteijn et al. (2011), Masala et al. (2014) and Goshu and Dessie (2013). A comprehensive introduction to semi-Markov processes is Janssen and Manca (2016), whereas they were first introduced by Levy (1954). Goshu and Dessie (2013) and Dessie (2014a, 2014b) used semi-Markov models for disease progression modeling in Ethiopia.

According to the study of Goshu and Dessie (2013) on the progression of HIV/AIDS disease stages at Felege Hiwot referral hospital, Ethiopia, which also used semi-Markov processes, the probability of being in a better state is non-zero, but smaller than the probability of being in a worse state. Their results suggest that patient transition from one state to another depends on how long the patient has been in the state. They also report that the survival probability of an HIV/AIDS patient depends on his/her current disease state: the lower is the CD4 count, the higher is the risk of transitioning to a worse health state or death state.

The World AIDS report (2013) shows that 19 million of the 35 million people living with HIV today do not know that they have the virus. The report highlights that efforts to increase access to ART are working. In 2013, an additional 2.3 million people gained access to the life-saving medicines. This brings the global number of people accessing ART to nearly 13 million by the end of 2013. Based on past scale-up, UNAIDS projects that as of July 2014 as many as 13,950,296 people were accessing ART. By ending the epidemic by 2030, the world would avert 18 million new HIV infections and 11.2 million AIDS-related deaths between 2013 and 2030. Infection by the human immunodeficiency virus (HIV) gradually evolves to the acquired immune deficiency syndrome (AIDS), and AIDS evolves to death if not handled carefully. Therefore, we consider the progression of HIV infection to AIDS and then to death as a stochastic process, splitting the progression into various states of the disease based on an immunological indicator, namely CD4+ count, and including death as one state, following Janssen and Monica (2001). The division of HIV/AIDS progression into five disease states is based on Center for Disease Prevention and Control (CDC) (1992). In this study five states will be considered: state one (CD4

count>500); state two ( $350 < \text{CD4 count} < 500$ ); state three ( $200 < \text{CD4 count} < 350$ ); state four ( $\text{CD4 count} < 200$ ); and state five (Death).

It is believed that accurate and detailed modelings of HIV/AIDS disease progressions are crucially important for reliable early diagnosis and the determination of effective treatments (Titman and Sharples, 2010). Models of disease progression are among the core tools of modern medicine for early disease diagnosis, for determining treatment and for explaining symptoms to patients. Thus, the aim of this study is to analyze the progression of the disease in order to provide theoretical support for decision-making on health intervention, prognostics and prevention and to establish the best sojourn time distributions for monitoring HIV/AIDS disease progression.

## **2. Statistical Methods and Data**

### **2.1. Markov Model**

The Markov assumption is restrictive and not necessary realistic. In recent times semi-Markov models have been highly utilized to model the progression of HIV/AIDS disease stages: see for instance Titman and Sharples (2010), Goshu and Zelalem (2013), Dessie (2014a, 2014b). It is also known that the environment of the semi-Markov process is much richer than the Markov chain, and complete application of semi-Markov processes is the right choice to model the progression of disease states under the homogeneous discrete time scheme (Foucher et al., 2005; Goshu and Dessie, 2013). Time-homogeneous semi-Markov models assume that the trajectory of the process depends only on the amount of time spent in the current state, allowing the sojourn times in each state to have an arbitrary distribution rather than only exponential distribution.

### **2.2. The Semi-Markov Model**

Semi-Markov models were studied in detail by Janssen and Monica (2001), Corradi et al. (2004) and Foucher et al. (2005). In a semi-Markov model we have two important quantities: the first is the state of the transition, denoted by  $X_n$ , and

the second is the time of the transition  $T_n$ . In defining these two important variables we are able to consider the randomness of the state along with the randomness of the time elapsed in each state. Thus we define the two random variables simultaneously as:

$$X_n : \Omega \rightarrow S, \quad T_n : \Omega \rightarrow \mathfrak{R} \quad (1)$$

$X_n$  represents the state at the  $n^{\text{th}}$  transition of the Markov process with state space  $S = \{S_1, S_2, \dots, S_m\}$ , and  $T_n$  represents the time of the  $n^{\text{th}}$  transition. The kernel  $Q = [Q_{ij}]$  and the transition probabilities  $P_{ij}$  embedded in the Markov process are:

$$Q_{ij}(t) = P(X_{n+1} = j, \quad T_{n+1} - T_n \leq t \mid X_n = i) \quad (2)$$

$$P_{ij} = \lim_{t \rightarrow \infty} Q_{ij}(t) \quad i, j \in S, \quad t \in \mathfrak{R} \quad (3)$$

Furthermore, it is necessary to introduce the probability that the process will leave state  $i$  in time  $t$ , which is given as:

$$H_i(t) = P[T_{n+1} - T_n \leq t \mid X_n = i] = \sum_{j=1}^m Q_{ij}(t) \quad (4)$$

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

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

This can be computed as:

$$G_{ij}(t) = \begin{cases} \frac{Q_{ij}(t)}{P_{ij}} & \text{if } P_{ij} \neq 0 \\ 1 & \text{if } P_{ij} = 0 \end{cases} \quad (5)$$

The transition probabilities are given by (5), for which the solutions should be obtained using the progression or evolution equation (6).

$$\phi_{ij}(t) = P[X(t) = j \mid X(0) = i] \quad (6)$$

$$\phi_{ij}(t) = (1 - H_i(t))\delta_{ij} + \sum_{l=0}^m \int_0^t Q_{il}(\tau)\phi_{ij}(t-\tau)d\tau \quad (7)$$

Corradi et al. (2004) proved a numerical solution for equation (7) that converges to the discrete time HSMP described as an infinite countable linear system by:

$$\phi_{ij}^h(kh) = d_{ij}^h(kh) + \sum_{l=1}^m \sum_{\tau=1}^k v_{ij}^h(\tau h)\phi_{ij}^h((k-\tau)h) \quad (8)$$

where  $h$  represents the discretization step, and

$$d_{ij}^h(kh) = \begin{cases} 0 & \text{if } i \neq j \\ 1 - H_i^h(kh) & \text{if } i = j \end{cases}$$

$$v_{ij}^h(kh) = \begin{cases} 0 & \text{if } k = 0 \\ Q_{ij}^h(kh) - Q_{ij}^h((k-1)h) & \text{if } k > 0 \end{cases}$$

The evolution equation in (8) can be written in matrix form as:

$$\Phi^h(kh) - \sum_{\tau=1}^k V^h(\tau h)\Phi^h((k-\tau)h) = D^h(kh) \quad (9)$$

$$\mathbf{V}^T \Phi^T = \mathbf{D}^T \quad (10)$$

Goshu and Zelalem (2013) implemented this algorithm by writing R code. The same R code is used for the computations in this study. Given the solutions of the evolution equation  $\hat{\phi}_{ij}(t)$  for transition probabilities, the reliability function is computed as:

$$R_i(t) = \sum_{j \in U} \hat{\phi}_{ij}(t) \quad (11)$$

For the algorithm, we refer the reader to the Appendix.

### 2.3. Sojourn Time Distributions

We suppose that the sojourn or waiting time in a given state is random and has a distribution. It is assumed that the process spends some time in a given state, and the random time has distribution  $G(t)$ . Two sojourn distributions are considered here for modeling the progression of the disease. The first is the exponential distribution  $G_{ij}(t) = 1 - \exp(-\lambda_{ij}t)$ ,  $t \geq 0$ , where  $\lambda_{ij} = 1/\sigma_{ij}$  is the expected time

that the process spends in state  $i$  before it enters state  $j$  from  $i$ , with scale parameter  $\sigma_{ij} > 0$ . The second distribution we consider is the Weibull distribution  $G_{ij}(t) = 1 - \exp(-\lambda_{ij}t^{\nu_{ij}})$ ,  $t \geq 0$ , with scale parameter  $\sigma_{ij} > 0$  and shape parameter  $\nu_{ij} > 0$ . When the shape is equal to 1, the Weibull becomes an exponential distribution.

#### 2.4. Data

The data considered for this study were taken from Yirgalem General Hospital, which is located 300 km south of Addis Ababa in the town of Yirgalem, Sidama zone, in the Southern Nation's Nationalities and Peoples regional state. HIV patients have enrolled for ART follow-up in the hospital since 2000. We adopted a simple random sampling procedure to collect our data from the lists of patients who were under ART follow-up from 2008 to 2015. The following sample size determination formula (Cochran, 1977) is used:

$$n = \frac{Z_{\alpha/2}^2 \frac{p(1-p)}{d^2}}{1 + \frac{1}{N} \left[ \frac{p(1-p)}{d^2} - 1 \right]} \quad (12)$$

where  $Z_{\alpha/2}$  is computed on a standard normal distribution for the significance level  $\alpha = 0.05$ , which gives  $Z_{\alpha/2} = 1.96$ . The term  $p$  represents the proportion of deaths among HIV/AIDS patients. It was obtained from the previous comparable study Goshu and Dessie (2013) on data taken from Felege-Hiwot Referral Hospital, from which  $p = 0.1043$ . The degree of precision  $d$  selected for this study was taken to be 0.03. From a total number of  $N = 1570$  HIV/AIDS patients at the Yirgalem General Hospital, the sample size for this study was calculated to be 375 patients.

Referring to the CDC (1992) immunological classification of HIV/AIDS infected patients, we have five states, where the first four states are good states and the last is the bad or death state. Thus, based on the seriousness of cases, we have the following states:

SI: CD4 count  $> 500 \times 10^6$  cells/L

SII:  $350 \times 10^6 \text{ cells/L} < \text{CD4 count} \leq 500 \times 10^6 \text{ cells/L}$

SIII:  $200 \times 10^6 \text{ cells/L} < \text{CD4 count} \leq 350 \times 10^6 \text{ cells/L}$

SIV:  $\text{CD4 count} \leq 200 \times 10^6 \text{ cells/L}$ .

D: Death.

The death state D is considered an absorbing state, while the other four “good” states all communicate with each other.

### 3. Results and Discussion

We study the progression of the HIV/AIDS disease and health risk factors and its relationship with the well-being of HIV/AIDS patients in order to provide theoretical support for decision-making on health intervention, prognostics and prevention, using flexible waiting time distributions. For comparison purposes we use an exponential distribution and a two-parameter Weibull distribution as waiting time distributions in semi-Markov stochastic processes. Semi-Markov models explicitly define distributions of waiting times, giving an extension of continuous time and homogeneous Markov models based implicitly on exponential distributions. To see the difference in waiting time distribution of disease progressions, we obtain results for both distributions. Finally, we compare the two models of HIV/AIDS disease progression and estimate the transition probabilities using these waiting time distributions in stochastic processes. The data analyzed in this study were collected at Yirgalem General Hospital during September 2008 to August 2015, when follow-up visits took place every six months at known and fixed time points, although the transition between levels of the state space could occur at any time. Frequencies and estimated transition probabilities between the states are summarized from the data and displayed in Table 1. Each patient was tracked throughout the study period for changes of status of the disease under ART follow-up. Among these patients, 69 (18.4%) died.

The model we consider is a continuous time Markov model. It helps us to compute transition probabilities and mean waiting times over time. The solution

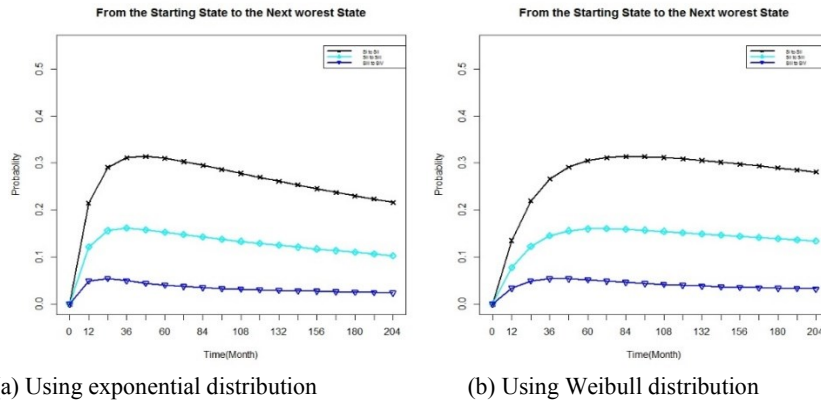


**Table 1:** Transition probability matrix computed from the progression data

State	I	II	III	IV	D
I	171 (0.438462)	169 (0.433333)	32 (0.082051)	7 (0.017949)	11 (0.028205)
II	240 (0.495868)	76 (0.157025)	132 (0.272727)	22 (0.045455)	14 (0.028926)
III	59 (0.166667)	183 (0.516949)	52 (0.146893)	41 (0.115819)	19 (0.053672)
IV	12 (0.079470)	30 (0.198675)	77 (0.509934)	7 (0.046358)	25 (0.165563)

for the transition probabilities  $\phi_{ij}(t)$  at time  $t$  is obtained using the defined algorithm given  $m = 5$  states,  $T = 204$  months, and transition probability matrix  $\mathbf{P}$  as given in Table 1. For the Weibull sojourn time distribution, the shape parameter is taken to be 1.5 (the choice of the shape parameter is based on the natural properties of the two-parameter Weibull distribution, i.e. when the shape parameter is greater than 1, the disease progression increases with time. This happens when there is an “aging” process or parts are more likely to fail as time goes on; this means that the hazard function is a monotonically increasing function of the progression). The scale parameters are estimated from the data.

All analysis is based on the algorithm defined. The results are plotted in Figures 1–7. The plots in Figure 1 display the several conditional probabilities that a patient will be in state  $j$  at time  $t$  given that she/he is initially in state  $i$  – among the good states. These progressions are computed for SI to SII, SII to SIII and SIII to SIV, for a specific HIV/AIDS patient.



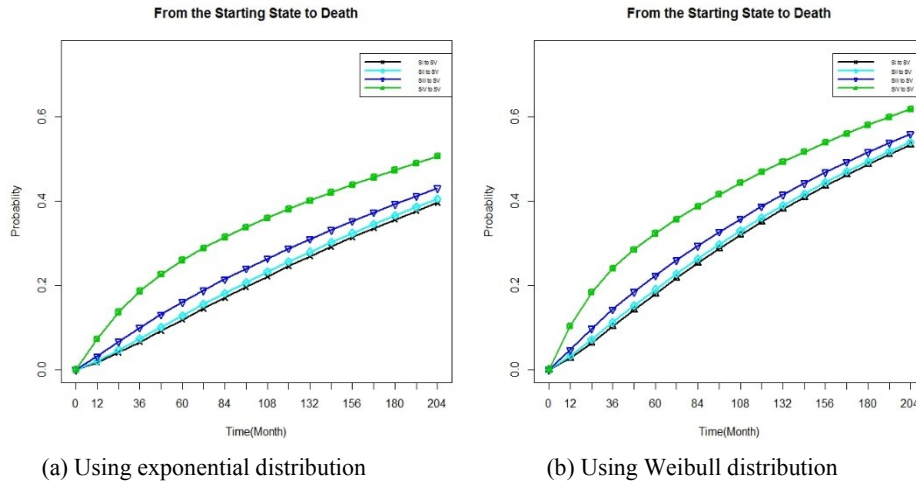
**Figure 1.** Conditional probability that a patient will be in state  $j$ ,  $j \in \{SI, SII, SIII, SIV\}$  after  $t$  months given that she/he is currently in state  $i \in \{SI, SII, SIII, SIV\}$ . The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution.

The plots in the left panel are for exponential sojourn time, and those in the right panel for Weibull sojourn time distributions. Figure 1(a) is plotted for the exponential waiting time distribution. The parabolic curves of the probability reach peaks at the time-probability points (48, 0.314) for state I to II, (36, 0.161) for state II to III, and (12, 0.054) for state III to IV. The peaks may indicate the time when a patient will be at highest risk of passing to a worse state. Moreover, the transition probability from SII to SIII is lower than the others. It is interesting to observe that, within the good states, the transition probability from a given state to the next worst state increases with time, reaches a maximum at a certain time and then decreases with increasing time, when the exponential waiting time distribution is considered.

Figure 1(b) is plotted for the Weibull waiting time. The progressions from state SI to SII, SII to SIII and SIII to SIV for a specific HIV/AIDS patient result in a similar parabolic curve to that obtained for exponential waiting time distribution, with peak points (48, 0.314), (72, 0.160) and (36, 0.054) on the time-probability axes. Interestingly, we observed a marked increment in progression from SI to SII, SII to SIII and SIII to SIV for a specific HIV/AIDS patient when considering a Weibull waiting time distribution as compared to an exponential waiting time distribution. This fact suggests a relationship between clinical outcomes and the duration of waiting times, which could explain the increase or decrease in transition probabilities of disease stages of patients in the cohort. In addition, for Weibull as compared to the exponential model, the conditional probability of remaining in a good state before moving to another good state grows faster at the beginning, reaches a peak, and then declines faster over a long period of time.

Figure 2 displays the conditional probabilities that an HIV/AIDS patient who is currently in a given state SI, SII, SIII, SIV will be in the death state after  $t$  months. These transitions are plotted from SI to D, SII to D, and SIII to D and SIV to D. Based on the exponential waiting time distribution, the probability of dying after 204 months is 0.396 for a patient who is in the first stage, 0.404 for one who is in the second stage, 0.4298 for one who is in the third stage and

0.5057 for one who is in the fourth stage of the disease. The probability of dying will increase by 9 percent for a specific HIV/AIDS patient in state three compared with an HIV/AIDS patient in state one. Similarly the probability of dying will increase by 20 percent for a specific HIV/AIDS patient in state four compared with an HIV/AIDS patient in state one.



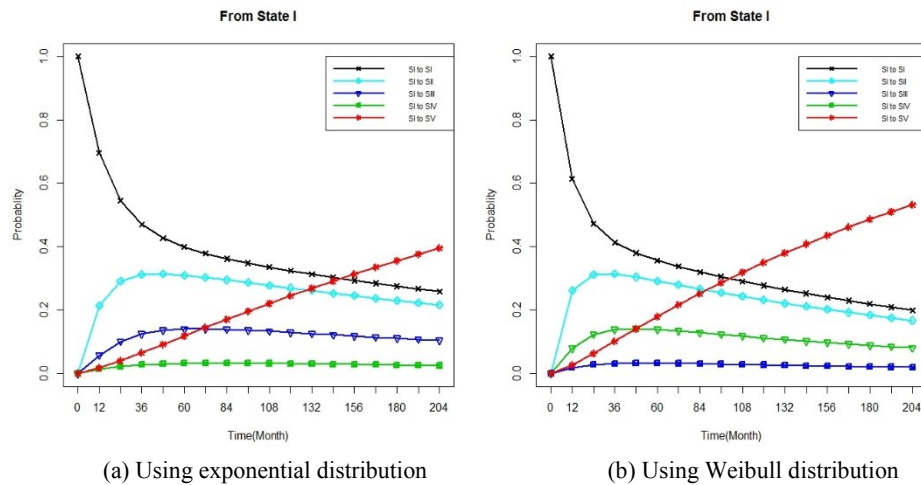
**Figure 2:** The conditional probability that a patient will be in death state after  $t$  months given that she/he is initially in state  $i \in \{SI, SII, SIII, SIV\}$ . The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution

For the Weibull distribution model, the probability of dying after 204 months is 0.213 for a patient who is in the first stage, 0.224 for one who is in the second stage, 0.256 for one who is in the third stage and 0.353 for one who is in the fourth stage of the disease. Each plot is an increasing parabolic curve over time with no peak point. The probability of dying will increase by 5 percent for a specific HIV/AIDS patient in state three compared with an HIV/AIDS patient in state one. Similarly the probability of dying will increase by 66 percent for a specific HIV/AIDS patient in state four compared with an HIV/AIDS patient in state one. This can be interpreted as indicating that the probability that an HIV/AIDS patient in any one of the good states will reach the death state is increasing with time. Moreover, a patient who is in the fourth state has the highest

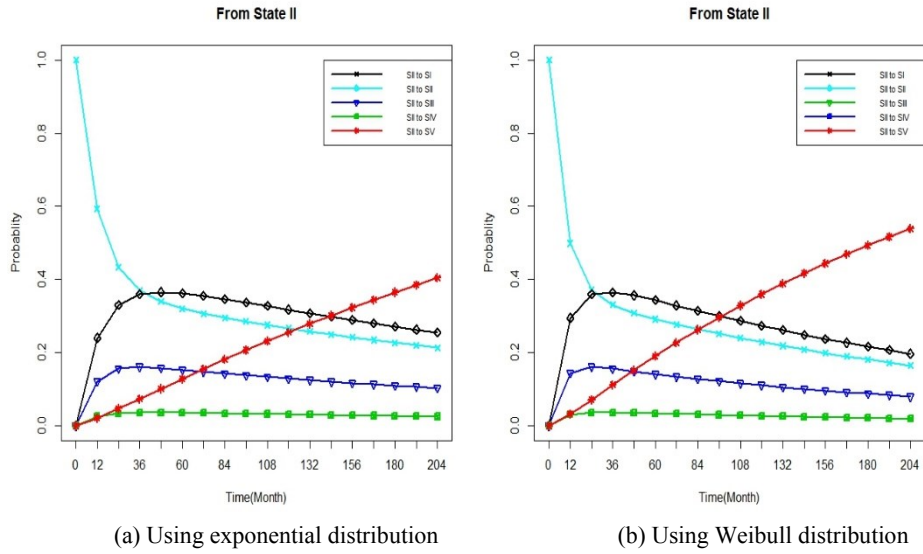
probability of dying after any given  $t$  months, while the probability for one who is in the first state is the lowest throughout the time considered.

In comparison with the Weibull waiting time, the exponential waiting time gives higher probabilities that an HIV/AIDS patient in any one of the good states will reach the death state. This result also shows that over a long time, the Weibull distribution leads to a smaller deviation of probabilities than the exponential case. Special medical intervention may be required to slow down the rate of decline. For instance, at time 204 months, the differences between the probabilities are smaller for the Weibull than for the exponential distribution.

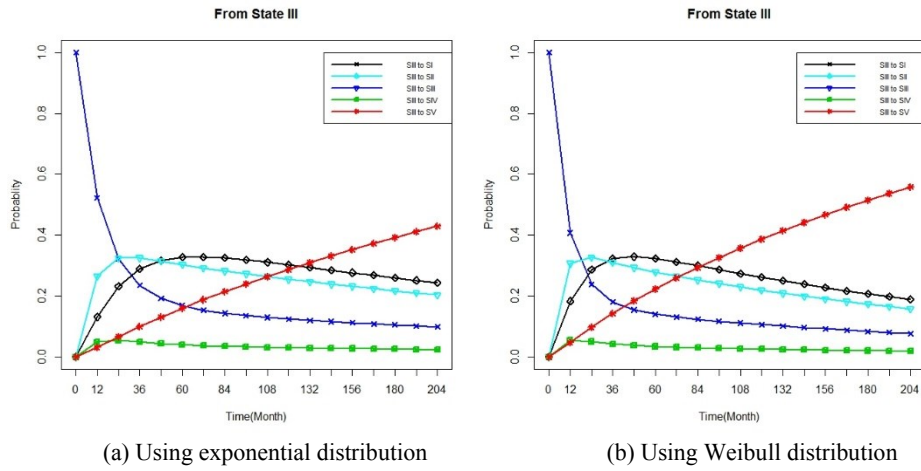
Figures 3–7 display conditional probabilities of a patient's transitioning between disease states given his/her current status, using exponential and Weibull waiting time distributions. The results show that the probabilities of being in the same state for both waiting time distributions across the study period decrease with time. However, the probabilities are much smaller for exponential waiting time distributions than for Weibull waiting time distributions, demonstrating that low probabilities (small rates) tend to be those with low waiting time rates and, consequently, slow progression.



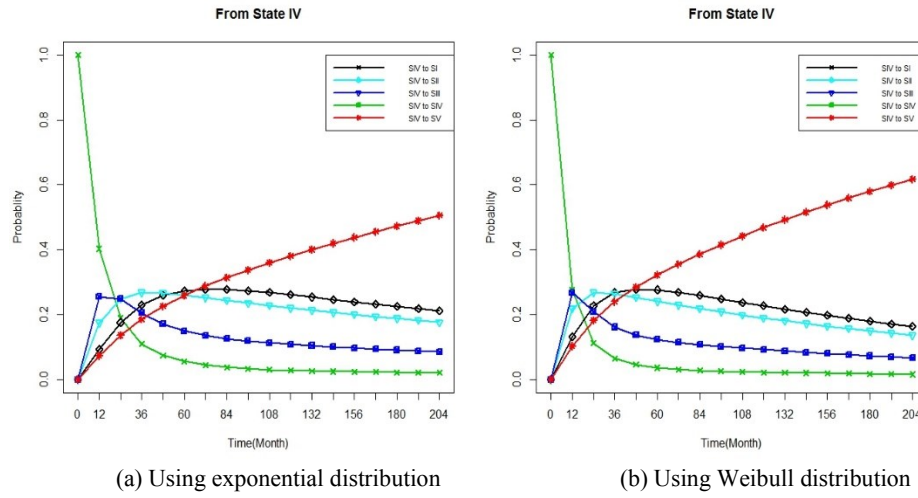
**Figure 3.** Conditional probability of being in the next state  $j$  after month  $t$  given the starting state I. The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution



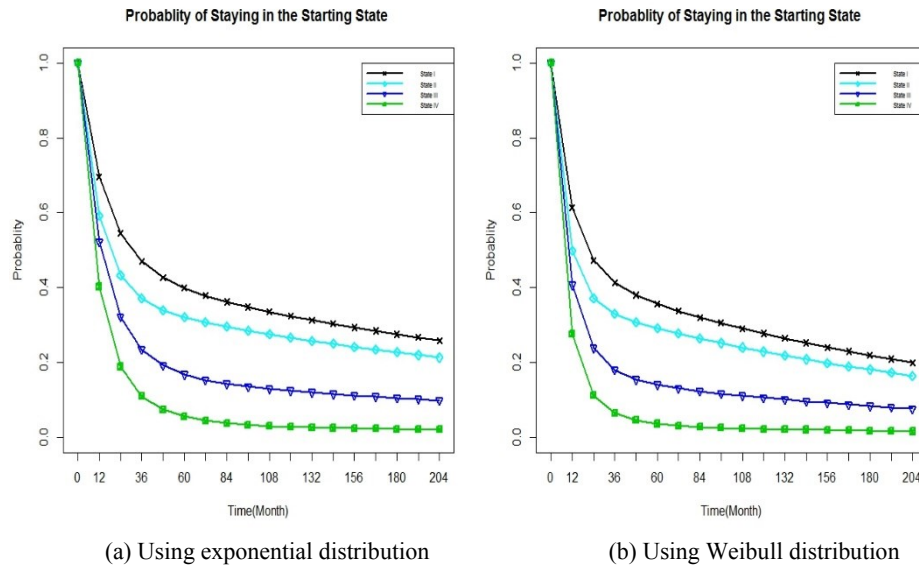
**Figure 4:** Conditional probability of being in the next state  $j$  after  $t$  months given a starting state II. The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution



**Figure 5:** Conditional probability of being in the next state  $j$  after  $t$  months given a starting state III. The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution



**Figure 6:** Conditional probability of being in the next state  $j$  after  $t$  months given a starting state IV. The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution



**Figure 7:** The conditional probability that a patient remains in a good disease state for at least  $t$  months. The plot in the left panel is for exponential sojourn time and that in the right panel for Weibull sojourn time distribution

Conversely, it demonstrates that high probabilities (high rates) tend to be those with high waiting time rates and early progression. Thus, physicians can monitor a patient's disease status at each state and take corrective actions as needed as early as possible, if Weibull waiting time distributions are considered. Therefore, considering the Weibull waiting time distribution is far preferable to the exponential waiting time distribution in terms of the early detection and management of diseases.

In Figure 7, we compute the probability of staying in same state. It is interesting to observe that the conditional probability of staying in the same state decreases with increasing time for both waiting time distributions. This result indicates that an HIV/AIDS patient in a specific good state of the disease will stay in that state with a non-zero probability.

The patient is more likely to be in a good state than a worse state; for example, the probabilities of being in state I, II, III and IV at time 24 months are about 0.545, 0.433, 0.321 and 0.187 respectively under the exponential model. This result also indicates that for an HIV/AIDS patient in a specific state of the disease, the probability of being in the same state decreases over time. With the good (living) states, the results show that the probability of being in a better state is non-zero, but is less than the probability of being in worse states.

#### 4. Discussion

This study was intended to model the progression of HIV/AIDS to predict the future clinical state and survival probability of a patient in order to provide theoretical support for decision-making in health intervention, prognostics and prevention, using flexible waiting time distributions. Accordingly, different probability plots are produced from the semi-Markov model based on the data on CD4+ counts of patients under ART follow-up during September 2008 to August 2015. The results showed that high deviations are observed when exponential waiting time distribution is used in modeling HIV/AIDS disease progression, as

compared with Weibull waiting time distribution, which leads to lower deviations.

The high deviation observed for exponential waiting time distribution led to a higher growth rate, and the lower deviations for Weibull waiting time distribution result in a low growth rate. The study revealed that within the good states, the transition probability from a given state to the next worst state increases with time, reaches a maximum at a certain time, and then decreases with increasing time, when an exponential waiting time distribution is considered. This is consistent with the findings of Goshu and Dessie (2013). Furthermore, for a Weibull waiting time distribution, the conditional probability of remaining in a good state before moving to another good state grows faster at the beginning, reaches a peak, and then declines faster over a long period of time.

We obtained similar findings as in Goshu and Dessie (2013), that the probability that an HIV/AIDS patient in any one of the good states will reach the death state is increasing over time. Moreover, a patient who is in the fourth state has the highest probability of dying after any given  $t$  months, while the probability for one who is in the first state is the lowest throughout the time considered. This is because of the fact that the fourth state is the most advanced sickness stage, and a patient at this stage is more likely to move to the absorbing state. Such patients need effective medical care in order to divert the disease progression to the healthier states.

In this paper, we study the disease progression of AIDS without considering risk factors or covariates, which may affect the progression. While exclusion of these factors is one of the limitations of this study, past studies have confirmed that AIDS disease progression is affected by covariates such as opportunistic infections, treatments, etc. (Foucher et al., 2005; Fonteijn et al., 2011; Mandel, 2010). Thus, we anticipate that if those factors are included the trends of disease progression seen in Figures 2–7 might change. The results also show that high deviations are observed when an exponential waiting time distribution is used in modeling HIV/AIDS disease progression, as compared with Weibull waiting time distribution, which leads to lower deviations. The observed high deviation in the



case of exponential waiting time distribution results in a higher growth rate, and the lower deviations in the case of Weibull waiting time distribution result in a low growth rate.

In conclusion, as the expected AIDS disease progression is dependent on risk factors and the use of appropriate mathematical models, our study has provided a theoretical support for decision-making in health intervention, prognostics and prevention, and this in turn reduces the costs of diagnosis and therapy.

## 5. Conclusions

This study was intended to model the progression of HIV/AIDS to predict a patient's future clinical state and probability of survival. The semi-Markov model is used to model the progression using varying waiting time distributions. Thus, the following conclusions are drawn from this study. Within the good states, the transition probability from a given state to the next worst state increases with time, reaches a maximum at a certain time and then decreases with increasing time, when exponential and Weibull waiting time distributions are considered. However, the probability becomes narrower for exponential waiting time distribution as compared with the Weibull distribution. This indicates that patients will suffer at some time in their follow-up period, and thus it is recommended that physicians can help patients with the early detection and management of their disease symptoms and intervene during their follow-up and alter their medication systems. We observed a marked increment in progression from a good state to the next consecutive worst state for a specific HIV/AIDS patient when Weibull waiting time distribution was considered, compared with exponential waiting time distribution. In the Weibull case, as compared to the exponential case, the model shows that the conditional probability of remaining in a good state before moving to another good state grows faster at the beginning, reaches a peak, and then declines faster over a long period of time. The probabilities of being in the same state across the study period, for both waiting time distributions, decrease over time. Comparatively, however, the probability of remaining in the same good

state declines faster for the Weibull waiting time distribution than for the exponential distribution.

The findings indicate that the probability of staying in same good state of the disease declines over time, with higher values for healthier states. The survival probabilities are all decreasing with increasing time, for both waiting time distributions. Moreover, the Weibull distribution under the semi-Markov model leads to dynamic probabilities with higher rate of decline and smaller deviations. The Weibull distribution is flexible in modeling, and so it is preferable for use as a waiting time distribution for semi-Markov modeling and monitoring of HIV/AIDS disease progression over time.

It would be desirable and recommended that the study be further enhanced by studying the effects of covariates, such as opportunistic infections and treatments, on sojourn times by using the Cox proportional model. It is also recommended to explore non-parametric alternatives to the exponential and Weibull distributions and compare the results. Thus, we recommend that further study should be conducted by incorporating time-dependent covariates and others proxy factors of AIDS disease progression.

### **Acknowledgments**

The authors would like to thank Zelalem Getahun Dessie for permitting us to use the R function developed for the semi-Markov algorithm. Our special appreciation goes to Hawassa University and Arba Minch University for financial support.

### **REFERENCES**

- Centre for Disease Control & Prevention (1993): Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recommendations and Reports.
- Cochran W.G. (1977): Sampling Technique. New York, John Wiley & Sons Inc.
- Corradi G., Janssen R., Monica R. (2004): Numerical treatment of homogeneous semi-Markov processes in transient case: a straightforward approach. Methodology and Computing in Applied Probability 6: 233-246.
- D'Amico G., Janssen J., Giuseppe D., Monica R. (2009): HIV progression through two different temporal scales according to non-homogeneous semi Markov models. In:

- Proceedings of the 13th International Conference on Applied Stochastic Models and Data Analysis, Sakalauskas L, Skiadas C (eds.), pp. 473-476.
- Dessie Z.G. (2014): Modeling of HIV/AIDS dynamic evolution using non- homogeneous semi-Markov process. 3: 2-11.
- Dessie Z.G. (2014): Multi-State Models of HIV/AIDS by Homogeneous Semi-Markov Process. American Journal of Biostatistics 4: 21-28.
- Fontejn H.M., Matthew J.C., Marc M., Josephine B., Manja L., et al. (2011): An Event-Based Disease Progression Model and its Application to Familial Alzheimer's Disease. Springer-Verlag Berlin Heidelberg 6801: 748-759.
- Foucher Y., Mathieu E., Saint-Pierre P., Durand J.F., Daurès J.P. (2005): A semi Markov model based on generalized Weibull distribution with an illustration for HIV disease. Biometrical Journal 47: 825-833.
- Goshu A.T., Dessie Z.G. (2013): Modeling Progression of HIV/AIDS Disease Stages Using Semi-Markov Processes. Journal of Data Science 11: 269-280.
- Janssen J., Manca R. (2006): Applied Semi-Markov Processes. New York.
- Janssen J., Monica N. (2001): Numerical solution of non-homogenous semi Markov processes in transient case. Methodology and Computing in Applied Probability 3: 271-279.
- Joly P., Commenges D. (1999): A penalized likelihood approach for a progressive three-state model with censored and truncated data application to AIDS. Biometrics 55: 887-890.
- Levy P. (1954): Processes semi-markoviens. Proceedings of the International Congress of Mathematicians, The Netherlands 3: 416-426.
- Mandel M. (2010): Estimating disease progression using panel data. Biostatistics. Oxford University Press 1-13.
- Masala G., Cannas G., Micocci M. (2014): Survival probabilities for HIV infected patients through semi-Markov processes. Biometrical Letters 51: 13 -36.
- Mathieu E., Yohann F., Pierre D., Jean-Pierre D. (2007): Parametric and non-homogeneous semi-Markov process for HIV Control. PNHSM Process for HIV Control. Methodology and Computing Applied Probability 9(3): 389-397.
- Teni D.A., Asena T.F. (2015): Modelling of AIDS Disease Associated Risk Factors and Mortality in Gamo Gofa Zone, Ethiopia. World Journal of Medical Science 12(4): 424-437.
- Titman A.C., Sharples L.D. (2010): Semi-Markov Models with Phase-Type Sojourn Distributions. Biometrics 66: 742-752.

## APPENDIX

### Algorithm and R Codes

For solving the evolution equation (10), Corradi et al. (2004) proposed the following algorithm. Given a fixed epoch  $T$ , and matrices  $G^T$  and  $P$ , the algorithm numerically solves the linear systems for the unknown matrix  $\Phi^T$ . The variables involved are the following:

$m$  = number of states of the process.

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

$P$  = matrix of order  $m$  of the embedded Markov process.

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

$Q^T$  = kernel of the Markov process.

$\Phi^T$  = block vector of order  $T+1$  the blocks of which are square matrices of order  $m$ .

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

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

$S^T$  = block vector of order  $T+1$  the blocks of which are diagonal square matrices of order

$m$ . The diagonal element of each block at time  $t$  is given by  $s_{ii} = \sum_{j=1}^m Q_{ij}(t)$ .

The algorithm is:

(a) Read  $m, T, P, G$

(b) Construct  $Q^T, V^T, D^T$

$V_{(0)} = I; Q_{(0)} = 0; S_{(0)} = I; D_{(0)} = I;$

for  $t = 1$  to  $T$

$Q_{(t)} = P * G_{(t)}$

for  $i = 1$  to  $m$

$s_{ii(t)} = Q_{i.}(t) \bullet 1$

end for

$V_{(t)} = Q_{(t)} - Q_{(t-1)}$

$D_{(t)} = D_{(0)} - S_{(t)}$

end for

(c) Given  $\Phi_{(0)} = D_{(0)}$ , Solve for  $\Phi^T$

for  $t = 1$  to  $T$

$\Phi_{(t)} = D_{(t)}$

for  $s = 1$  to  $t$

$\Phi_{(t)} = \Phi_{(t)} + V_{(s)} \bullet \Phi_{(t-1)}$

end for

end for

(d) Return the results  $\Phi^T, Q^T$