

# Semi-Markov Modelling of HIV/AIDS Disease Progression

Tilahun Asena (✉ [feredetilahun14@gmail.com](mailto:feredetilahun14@gmail.com))

Arba Minch University

Ayele Goshu

Kotebe Metropolitan University

Mebratu Senbeta

Arba Minch University

Derbachew Teni

Arba Minch University

---

## Research Article

**Keywords:** AIDS Diseases Progression, Waiting Time Distribution, Covariate effect, Transition Probability, Semi-Markov

**Posted Date:** January 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-135933/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Semi-Markov Modelling of HIV/AIDS Disease Progression

Tilahun Asena<sup>1\*</sup>, Ayele Goshu<sup>2</sup>, Mebratu Senbeta<sup>3</sup>, Derbachew Teni<sup>1</sup>

<sup>1</sup>Department of Statistics, Arba Minch University, Arba Minch, Ethiopia

<sup>2</sup>Department of Statistics, Kotebe Metropolitan University, Ethiopia

<sup>3</sup>Department of Economics, Arba Minch University, Arba Minch, Ethiopia

\*Email of corresponding author: [feredetilahun14@gmail.com](mailto:feredetilahun14@gmail.com)

## Abstract

**Background:** HIV/AIDS epidemic continues to be the main challenge in the world.

According to United Nations Program on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) reports of 2013, 35 million people were living with HIV worldwide, with 2.1 million new infections and with 1.5 million deaths occurred each year. Among these, 24.7 million lived in sub-Saharan Africa with 1.5 million new infections and 1.1 million AIDS deaths.

**Method:** The main objective of this study is finding factors affecting HIV/AIDS disease progression. This study was conducted to investigate the effect of factors on HIV/AIDS disease progression. Patient follow-up data is obtained at Yirgalim General Hospital. A sample of 370 Patient data from a follow-up cohort is obtained at Yirgalim General Hospital. Multivariate generalized hazard regression model was employed to investigate the disease progression using both time independent and time dependent covariates.

**Result:** The study revealed that the risk of transition differs by patient's body mass index. Increase in the body mass index reduces the risk of transiting into the next worst states. The effects of sex, weight, age and body mass index of patients are significantly associated with

AIDS disease progression. The risk of transition differs by patient's body mass index. Increase in the body mass index reduces the risk of transiting into the next worst states. The effect of sex, weight, age and body mass index of patients are significantly associated with AIDS disease progression. The results further revealed that the semi-Markov model with Weibull waiting time distribution has smaller log likelihood and AIC values compared to a semi-Markov model with exponential waiting time distribution.

**Conclusion:** Transition probabilities are highly dependent on the choice of waiting times. We recommend that while choosing waiting time distributions for semi-Markov models one should consider appropriate distributions as waiting time distribution effect have a significant change on the estimated model parameters. In addition, this study recommends that concerned bodies should look at deferent contributing factors of AIDS diseases progression in addition to the ART services administered for slowing the current level of high diseased population in the country.

**Key words:** AIDS Diseases Progression, Waiting Time Distribution, Covariate effect, Transition Probability, Semi-Markov

## 1. Background

HIV/AIDS epidemic continues to be the main challenge in the world. According to United Nations Program on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) reports of 2013, 35 million people were living with HIV worldwide, with 2.1 million new infections and with 1.5 million deaths occurred each year. Among these, 24.7 million lived in sub-Saharan Africa with 1.5 million new infections and 1.1 million AIDS deaths [1]. The burden of HIV/AIDS in sub-Saharan Africa and in Ethiopia in particular remains high. Continuous-time multi-state models are widely used for categorical response data in the context of the natural history of chronic diseases [2, 3].

Inference is difficult when the process is only observed at discrete time points, with no information about the times or types of events between observation times. Studies show that mortality among HIV-infected individuals depends on their CD4 cell counts and other associated risk factors. CD4 cell counts and risk factors accelerate or decelerate the state transitions of the AIDS disease [4]. Several studies are mainly concerned with estimation of transition and survival probabilities using Markov models without considering effects of covariates on the disease progression [5, 6, and 7]. The probability of dying increases in the worse transition states. The probability of being in healthy state after he/she started the treatment is higher as compared with any other working state [4]. Some factors may increase/decrease the hazard rate of transition between disease states. When patients get older and infected with TB, AIDS disease transition rates to death state increases [8]. In semi-Markov Models, we are able to choose a parametric distribution with more freedom than in traditional Markov Chain. Parametric semi-Markov models are powerful for studying chronic diseases and estimating factors associated with transitions between different stages of disease progression. Literature shows there exists a growing interest to apply parametric semi-Markov models for disease progression [9].

The development of R packages such as the `smm` package [10] for disease modelling is motivated by the interest to include time-varying characteristics of individuals to transition rates through a parametric approach. Epifani et al. [11] used Bayesian estimation for a parametric Markov renewal model applied to seismic data. Their work focused on the development of methodology for Bayesian inference on a parametric Semi-Markov process, from the elicitation of the prior distribution to the computation of posterior summaries. The aim of this study is to find factors affecting HIV/AIDS disease progression using parametric semi-Markov model with interval censoring.

The method used in this analysis is formalized in 2. In 3 we describe our main result. In 4 we discuss our results and describe our findings of this study by comparing with the methods presented in 2. In 5, we conclude with a discussion.

## 2. Methods

Data for this study were obtained from Yirgalim General Hospital. Yirgalim General Hospital is located 300 km south of Addis Ababa in Yirgalim town of Sidama zone, the Southern Nation's Nationalities Region. The Hospital was inaugurated in 1968. The ART clinic started follow-up on HIV patient in 2000. We adopted a simple random sampling procedure to select sample of HIV patients from the lists of patients who follows ART between September 2008 and August 2015. Accordingly, the following sample size determination formula [12] is used:

$$n = \frac{n_o}{1 + \frac{n_o - 1}{N}} \quad \text{where } n_o = \left( z_{\frac{\alpha}{2}} \right)^2 \frac{Pq}{d^2} \quad (1)$$

where  $z_{\frac{\alpha}{2}}$  is the value of a standardized normally distributed variable at which the upper area

under the curve is  $\frac{\alpha}{2}$ , where  $\alpha$  is significance level. For  $\alpha = 0.05$ ,  $z_{\frac{\alpha}{2}} = 1.96$ . The term p

represents proportion of death among HIV/AIDS patients. The value of P used here is obtained from the previous comparable study conducted by Goshu and Dessie [5] on data taken from Felege Hiwot Referral Hospital which is  $p = 0.134$ . The degree of precision d selected for this study was taken to be 0.03. With total number of  $N = 1570$  HIV/AIDS patients at the Yirgalem General Hospital, the sample size for this study is estimated to be 375 patients.

Referring to the CDC [13] immunological classification of HIV/AIDS infected patients; we have five states defined as follows. The first four states are the good states and the last state is bad state or death state. Thus, based on the seriousness of the cases we have the following states:

SI:  $CD4 \text{ T cells count} \leq 500 \times 10^6 \text{ T cells/L}$

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

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

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

D: Death.

We assume the good states (state I, state II, state III and state IV) communicate with each other, and they communicate with the absorbing state, which is death. The number of patients finally analyzed was 365. The covariates included are Age, Sex, Place of residence, Religion, Marital status, Educational status, Occupational status, Drug use or not, Knowledge of ART service, TB status (positive or negative) and Post opportunistic infection. Due to the complex form of the waiting time distributions to the number of covariates under study our optimization methods fails to reach convergence. Thus, factors were added to the model- one at a time in univariate model and all factors are included simultaneously in multivariate model. Likelihood Ratio tests with covariates were performed to assess its goodness-of-fit. A Semi-Markov model is a statistical model with the same structure as a Markov model except that the sojourn time distribution is flexible in semi Markov rather than an exponential distribution in Markov.

Suppose  $S = (1,2,3,4,5)$  is the state space. We define the following random variables.

$$X_n : \Omega \rightarrow S \quad T_n : \Omega \rightarrow N^+ \quad (2)$$

$X_n$  represents the state at the  $n^{\text{th}}$  transition and  $T_n$  represents the chronological time of the  $n^{\text{th}}$  transition. Let  $N(t)$  be the counting process associated to the point process  $(T_n, n \in N)$  defined for any time  $t \geq 0$  by :

$$N(t) = \sup\{n, T_n < t\} \quad (3)$$

The random variable  $N(t)$  represents the number of transitions occurred in the interval of time  $(0, t]$ . Let us define  $(T_n)_{n \in N}$  as the duration process by :

$$T_0 = 0 \quad \Delta T_{n+1} = T_{n+1} - T_n$$

Here  $\Delta T_{n+1}$  represents the duration time spent in state  $X_n$ . The process  $(X_n, T_n)_{n \in N}$  is called a Markov renewal process if

$$\begin{aligned} P(X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_0, \dots, X_n, X_{n-1} = i, T_0, \dots, T_n) \\ = P(X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i, ) \end{aligned} \quad (5)$$

and for  $i \neq j$ . The Semi-Markov kernel is given by

$$Q_{ij}(t) = P(X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i) \quad (6)$$

The component of  $Q_{ij}(t)$ , namely  $t$ , represents duration time of the process [9].

$$P_{ij}(t) = \lim_{t \rightarrow \infty} Q_{ij}(t) \quad i, j \in S \quad t \in N^+ \quad (7)$$

Equation 7 represents the probability of a patient making its next transition to state  $j$ , given that he/she entered state  $i$  at time  $t$  and  $\mathbf{P}(t) = \lim Q_{ij}(t)$  is the 5 by 5 matrix of transition probabilities of the embedded Markov chain  $(X_n)_{n \in N}$

The probability density function of the waiting time in state  $i$  before passing to state  $j$  is given by:

$$f_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T_{n+1} - T_n \leq t \mid X_{n+1} = j, X_n = i)}{\Delta t} \quad (8)$$

From density function, we can derive the cumulative probability function,  $(F_{ij}(t))$  and the survival function  $(S_i(t))$  of waiting time in state  $i$  as defined by:

$$F_{ij}(t) = P(T_{n+1} - T_n \leq t \mid X_{n+1} = j, X_n = i) = \frac{Q_{ij}(t)}{P_{ij}(t)} \quad (9)$$

$$S_i(t) = 1 - P(T_{n+1} - T_n \leq t \mid X_n = i) = \sum_{j \in S} P_{ij}(1 - F_{ij}(t)) \quad (10)$$

In semi-Markov modelling one can choose any eligible waiting (sojourn) time distribution. Thus, we considered both exponential distribution and Weibull distribution as the waiting time distribution. For exponential distribution, the hazard function is constant. The hazard function of the waiting time is given by:

$$\alpha_{ij}(t) = \frac{1}{\sigma_{ij}} \quad \forall t \geq 0, \forall \sigma_{ij} > 0$$

Further, the Weibull distribution generalizes the exponential distribution by using two parameters, which is more flexible and well adapted to various shapes. Its hazard function is defined by:

$$\alpha_{ij}(t) = \frac{\nu_{ij}}{\sigma_{ij}} \left( \frac{t}{\sigma_{ij}} \right)^{\nu_{ij}-1} \quad , \forall \nu_{ij} \geq 0 \quad \forall t \geq 0, \forall \sigma_{ij} > 0$$

Where  $\sigma_{ij}$  is the rate of transition from state  $i$  to state  $j$  and  $\nu$  is the shape parameter of the Weibull probability distribution.

The hazard function of the Semi-Markov process, which represents the probability of transition towards state  $j$  between time  $t$  and  $t + \Delta t$ , given that the process is in state  $i$  for duration  $t$  can be derived as follows:

$$\lambda_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(X_{n+1} = j, t < T_{n+1} - T_n < t + \Delta t \mid X_n = i, T_{n+1} - T_n > t)}{\Delta t}$$



$$= \frac{p_{ij} f_{ij}(t)}{S_{i.}(t)} = \frac{p_{ij} S_{ij}(t) \alpha_{ij}(t)}{S_{i.}} \quad i, j \in S, i \neq j \quad (11)$$

$$\lambda_{ii}(t) = - \sum_{i \neq j} \lambda_{ij}(t)$$

Equation (11), can be interpreted as the subject's risk of progressing from state  $i$  to state  $j$  after having stayed in state  $i$  for duration  $t$  [15].

Explanatory variables can be included at each level of the model through generalized regressions. A proportional hazards model is used to relate the transition intensities  $(q_{ij}(t))$  at time  $t$  to covariates  $z(t)$  at that time.

$$q_{ij}[t, z(t)] = q_{ij} \exp[\beta_{ij}^T z(t)]$$

The parametric model allows to incorporate covariates in the distribution of sojourn times using a proportional-hazards regression model [16].

### 3. Results and Discussions

The aim of this study is to find factors affecting HIV/AIDS disease progression using a parametric semi-Markov model with interval-censored data. We use the R package semi-Markov to analyses the data and estimate the parameters of the semi-Markov model. Table 1 shows comparison of waiting time distribution for fitting progression of the disease. Based on the result the log likelihood of exponential waiting time distribution ( $-2 * \log\text{-likelihood} = 5159.791$  and  $\text{AIC}=5167.791$ ) is higher than the log likelihood of Weibull sojourn time distribution ( $-2 * \log\text{-likelihood} = 5056.362$  and  $\text{AIC}=5064.362$ ). Therefore, the Weibull waiting time distributions are preferred for HIV/AIDS disease progression. Moreover, differences in the transition estimates from state to state in both waiting time distributions are obtained. Exponential waiting time distribution generates wider confidence intervals as

compared with Weibull waiting time distributions. The Weibull waiting time distribution is preferable as compared with the exponential waiting times distribution for fitting our data.

Table 1: Comparison of estimated transition intensities under Markov Model with Exponential and Weibull distribution

Transition	Exponential	Weibull	
	Estimate (CI) for $\sigma$	Estimate (CI) for $\sigma$	Estimate (CI) for $\nu$
1->2	25.042(20.02,30.06)	28.462(24.84 ,32.09)	1.232(1.09,1.37)
1->3	26.311(10.32,42.30)	19.311 (14.59 ,24.03)	1.791(1.29,2.29 )
1->4	15.916(1.58, 33.42)	14.441(4.35 ,24.53)	1.364(0.55 ,2.18)
1->5	86.317(85.62 ,87.01)	26.075(18.58 ,33.57)	2.517(1.35 ,3.69 )
2->1	18.005(15.46 ,20.55)	18.136(16.09 ,19.93)	1.333(1.21 ,1.46 )
2->3	18.403(14.68 ,22.13)	14.235(15.78 ,20.49)	1.471(1.27 ,1.67 )
2->4	13.845(6.98 ,20.71)	14.235(10.13, 18.34)	1.614(1.15 ,2.08)
2->5	19.543(6.05 ,33.04)	17.77 (16.14 ,84.35)	0.906(0.51 ,1.30 )
3->1	16.516(11.51 ,21.53)	50.244(16.14, 84.35)	1.17(0.93 ,1.41 )
3->2	14.371(12.04 ,21.53)	14.602(12.92 , 16.28)	1.436(1.27 ,1.60 )
3->4	16.611(10.49 ,22.73)	17.017(12.26 , 21.78)	1.372(1.02,1.72)
3->5	25.244(11.53 ,38.95)	29.414(15.64 ,43.19)	1.326(0.82 ,1.83 )
4->1	9.716(3.87,15.57)	10.176(5.05 , 15.31)	1.326(0.82 ,1.83 )
4->2	8.871(5.54,12.20)	9.157(5.88 ,12.43)	1.086(0.81 ,1.37)
4->3	10.367(7.9,12.83)	11.003(8.96 , 13.04)	1.291(1.07 ,1.51 )
4->5	20.526(12.02,29.03)	20.124(10.63 ,29.62)	0.87 (0.61 ,1.13 )
-2log-like	5202.282	5056.362	
AIC	5167.791	5064.362	

Univariate regression modelling is used to identify significant factors associated with the effect of covariates for AIDS progression using generalized hazard regression modelling and we test the effect by Wald test statistics ( $H_{\{0\}}: \theta_{ij} = 1$ ). The null hypotheses are the nullity of distribution parameters and the regression coefficients are equal to 1. In this model, each factor was added to the model one at a time. The regression coefficient  $\beta$ , the p-value of the Wald test when testing the absence of effect ( $H_0: \beta = 0$ ) is also provided. Moreover, results regarding each distribution parameter  $\sigma$  (or  $\nu$  or  $\theta$ ) the p-value of the Wald test when testing  $H_0: \sigma = 1$  are provided. In this model, the regression coefficients are interpreted in terms of relative risk. The effect of covariates and the

proportional hazard assumption is evaluated by representing the hazard rates. The proportional hazard regression model is considered to study the effect of covariates. According to the univariate analysis covariates sex, age, weight and body mass index are found to be significant. However, TB co-infection, religion, educational status, place of residence, occupational status and opportunistic infections are not associated with increasing or decreasing the risk of developing the progression of the HIV/AIDS disease.

Results in Table 2 are the multivariable regression coefficients associated with the effect of covariates for progression to the death state analyzed using generalized hazard regression model and Weibull waiting time distribution. The estimate of coefficient associated to the transition from state one to death state for sex is significantly different from 0 ( $\beta = -1.351$ ; CI: (-2.68, -0.02);  $p < 0.05$ ). It means that male patient increases the risk of progression of leaving the first state to enter the death state compared with female patient who are under follow up. The estimated coefficient associated to the transition from state two to death for sex is significantly different from 0 ( $\beta = 3.732$ ; CI: (2.04, 5.43);  $p < 0.05$ ).

Table 2: Estimates of the Multivariable Generalized Hazard Regression Coefficients from the starting state to death

Variable	Transition	Coef	95% CI	P_Value
Sex: Male (Female , ref)	1->5	-1.351	(-2.68 , -0.02)	<0.05
	2->5	3.732	(2.04 , 5.43)	<0.05
	3->5	0.002	(-1.41 , 1.41)	0.4401
	4->5	-0.447	(-1.49 , 0.59)	0.091
Age: >35 years ( $\leq 35$ years , ref)	1->5	1.270	(1.24 , 2.09)	<0.05
	2->5	0.830	(-0.49 , 2.15)	0.1121
	3->5	-0.708	(-2.10 , 0.68)	0.3451
	4->5	-0.954	(-2.25 , 0.34)	0.7201
Weight: >60kg ( $\leq 60$ kg , ref)	1->5	0.044	(0.01 , 0.08)	<0.05
	2->5	-0.079	(-0.12 , -0.04)	<0.05
	3->5	0.004	(-0.03 , 0.04)	0.7913
	4->5	0.006	(-0.02 , 0.03 )	0.231
BMI: >25 ( $\leq 25$ , ref)	1->5	-0.420	(-5.33 , -0.18)	<0.05
	2->5	-0.630	(-2.41 , 1.15)	0.331

	3->5	-0.215	(-1.60 , 1.17)	0.0801
	4->5	0.793	(-0.19 , 1.77)	0.1561

This means that male patients increase the risk of progression for leaving the second state to enter to the death state as compared with female patient. However, the estimated coefficient associated to the transition from state three to death ( $\beta = 0.002$ ; CI: (-1.41, 1.41);  $p > 0.05$ ) and from state four to death ( $\beta = -0.447$ ; CI: (-1.49, 0.59);  $p > 0.05$ ) for the variable sex are not significant. This describes that for AIDS progression, the effect of gender difference is significant at the early stage (stage one and stage two) but does not have an effect at the later stages (stage three and stage four) to progress to the death state.

The estimate of coefficient associated to the transition from state one to death state for age is significantly different from 0 ( $\beta = 1.270$ ; CI: (1.24, 2.09);  $p < 0.05$ ). It means that patient age  $> 35$  years increases the risk of progression of leaving the first state to enter the death state compared with patient age  $< 35$  years. Whereas, the estimated coefficient associated to the transition from state two to death ( $\beta = 0.830$ ;  $p > 0.05$ ), state three to death ( $\beta = -0.709$ ;  $p > 0.05$ ) and state four to death ( $\beta = -0.954$ ;  $p > 0.05$ ) for age are not significant. The effect of age is important only during the early stage of the disease. The estimate of coefficient associated to the transition from state one to death state for weight of patient is significantly different from 0 ( $\beta = 0.044$ ; CI: (0.01, 0.08);  $p < 0.05$ ). It means that patient weight  $> 60$  kg decreases the risk of progression of leaving the first state to enter the death state compared with patient weight  $\leq 60$  kgs. Similarly, the estimated coefficient associated to the transition from state two to death for weight is significantly different from 0 ( $\beta = -0.079$ ; CI: (-0.12, -0.04);  $p < 0.05$ ). This means that patients weight  $> 60$  kg decreases the risk of progression for leaving from the second state to enter to the death state as compared with patient weight  $< 60$

kg. However, effect of weight is not significant for transitioning from state three to death and from state four to death.

The estimate of coefficient associated to the transition from state one to death state for BMI is significantly different from 0 ( $\beta = -0.420$ ; CI: (-5.33, -0.18);  $p < 0.05$ ). It means that BMI > 25 increases the risk of progression of leaving the first state to enter the death state compared with BMI  $\leq 25$  who are under follow up. However, body mass index is not significant for transition in to death at state two, three and four.

In Table 3, we considered multivariable generalized hazard regression model to investigate disease progression to the next higher healthier state when covariates, sex, age, weight, and BMI of patients are simultaneously included.

Table 3: Estimates of the Multivariable Generalized Hazard Regression Coefficients Between the Good States.

Variable	Transition	Coef	95% CI	P_Value
Sex: Male (Female , ref)	1->2	0.361	(0.03 , 0.69 )	<0.05
	2->3	-1.755	(-2.97 , -0.54)	<0.05
	3->4	1.178	(0.48 , 1.87 )	<0.05
Age: >35 years ( $\leq 35$ years , ref)	1->2	-0.033	(-0.42 , 0.36)	0.8625
	2->3	0.009	(-0.39 , 0.41)	0.075
	3->4	0.714	(0.01 , 1.41)	<0.05
Weight: >60kg ( $\leq 60$ kg , ref)	1->2	0.015	(0.01 , 0.02)	<0.05
	2->3	0.044	(0.02 , 0.07)	<0.05
	3->4	-0.057	(-0.10 , -0.02)	<0.05
BMI: >25 ( $\leq 25$ , ref)	1->2	-6.605	(-11.66 , -1.55)	<0.05
	2->3	-0.313	(-0.91 , 0.29)	0.3078
	3->4	0.658	(-0.27 , 1.59)	0.1659

The estimate of coefficient associated to the transition from state one to state two for BMI is significantly different from 0 ( $\beta = 0.361$ ; CI: (0.03, 0.69);  $p < 0.05$ ). It means that male patient decreases the risk of progression of entering to the next higher state as compared with

females when covariates sex, age, weight and BMI are included simultaneously. Similarly, the estimate of coefficient associated to the transition from state two to state three ( $\beta = -1.755$ ; CI: (-2.97, -0.54);  $p < 0.001$ ), state three to state four ( $\beta = 1.178$ ; CI: (0.48, 1.87);  $p < 0.001$ ) for are significantly different from 0. This depicts that the effect of gender is significant for a patient to progress to the next higher disease state.

The estimate of coefficient associated to the transition from state one to state two ( $\beta = 0.015$ ; CI: (0.01, 0.02);  $p < 0.05$ ) and state two to state three ( $\beta = 0.02$ ; CI: (0.02, 0.07);  $p < 0.001$ ) and state three to state four ( $\beta = -0.057$ ; CI: (-0.10, -0.02);  $p < 0.05$ ) for weight are all significantly different from 0. This shows that patients weight is important to understand the process of evolution of the disease. The estimated coefficient associated to the transition from state one to state four ( $\beta = -11.66$ ; CI: (-11.66, -1.55) ) for BMI is significantly different from 0. It means that BMI  $> 25$  decrease the risk of progression of leaving state one and entering to state four as compared with BMI  $\leq 25$  when all covariates are included simultaneously. However, BMI of patients is not important to understand the process of evolution of the disease from state two to state three and state three to state four.

## 4. Discussions

This study was intended to assess the effects of covariates on the evolution of HIV/AIDS disease progression using parametric Semi-Markov model on the data obtained from follow-up of patients at Yirgalem General Hospital during 2008-2015. Both univariate and multivariable generalized hazard regression models were used to assess the effects of covariates on the evolution of the disease process. The results obtained are discussed as follows.

Concerning the comparisons of exponential waiting time distribution with Weibull waiting time distribution, estimates of Weibull waiting time distributions with similar values of Log likelihood and AIC were more preferred than estimates of exponential waiting time distribution when similar covariates are included in the model. Most likely these may be due to the incorporation of other shape parameter in Weibull waiting time distributions and multi-modal hazard functions for the Weibull probability distribution may maintain a good approximation. However, the real advantage of using the Weibull it allows a reasonable degree of flexibility with few additional parameters as stated in [3, 17 and 18]. Unlike the studies of [4, 5, 8, 9] this showed that the estimated parameters of the disease progression are highly dependent on the waiting time distribution considered in the hypothesized model.

Unlike the study of [19] fitting hitting times under a continuous Markov model through simulation results were far from expected. The most striking of these results is the absence of waiting time distributions between the states, which suggests that after reach the steady state there is no evidence of a relation between these two diseases states. This further suggests that assuming parametric waiting time distributions has the advantage to switch non-overlapping states and absorbing states. Unlike our study, Mandel proved that goodness-of-fit tests for the Markov assumption could be conducted by embedding the model in a larger model and using the likelihood ratio test. Although the tests are impartial, it should be interpreted as a crude approximation as the distributions of the waiting times are unknown in each state.

The results in this article contrasts with the method proposed by [2, 3] in which a phase-type distributions are used directly as the sojourn waiting time) distributions in the semi-Markov model. Although there exists state misclassification in the phase type approximation, the direct approximation of the waiting time distributions recovery enables the semi-Markov

model to estimate the transition times of the censored observations. The findings of the study using generalized hazard regression model shows that sex, age, weight and body mass index of the HIV/AIDS patient are found to be significantly associated with the AIDS disease progression. The risk of patients moving from state I to state II for patients aged greater than 35 years of old were greater than those of patients aged 35 years of old and less as [4]. This might be because of two reasons. First, when patients become older their immunity for the disease decreases therefore this will increase the progression in latter stages. Second, old patients are more exposed with different activities like drinking alcohols, smoking cigarettes than in the younger ages. Similarly, the risk of transition differs based on their Gender, weight and Body Mass Index (BMI). The estimated survival probabilities for HIV infected patients can also be used for comparing them with respect to certain categories such as gender, age group or type of antiretroviral therapy. Similar to our studies [6] study allowed him to deduce certain correlations between survival probabilities and specific factors such as the patient's gender and/or age. As a further improvement unless our study his methodology was used to compare the efficiency of different antiretroviral therapies.

Finally, Inference for semi-Markov models under direct incorporation of waiting time distribution permitted us to estimate and assess the effects of different covariates on the progression of the diseases. The effect of choosing appropriate waiting time distribution has significant effect on the estimated model parameters.

## **5. Conclusions**

This study aims to assess the effects of covariates on AIDS disease progression using 370 follow-up data obtained from Yirgalim General Hospital. Univariate regression analysis using generalized hazard regression revealed that effects of sex, age, weight and body mass index are significantly associated with AIDS progression. Other factors such as the effects of



TB co-infection, religion, educational status, place of residence, occupational status and opportunistic infections are not associated with increasing or decreasing the risk of developing the progression of the disease. We also assessed the effect of waiting time distribution on disease progression when covariates are included in the model. The study also found that gender differences are very important during the early stage of the AIDS progression and it becomes unimportant at the later stages (stage three and stage four). Our final multivariable model revealed that the effect of these covariates (sex, age, weight and body mass index) is different at different disease stages for HIV/AIDS progression.

This study confirmed that AIDS disease progression was affected by the choice of sojourn time distributions. The results further revealed that the semi-Markov model with Weibull waiting time distribution has smaller log likelihood and AIC values compared to a semi-Markov model with exponential waiting time distribution. This reveals that, in the setting of semi-Markov model with Weibull waiting time distributions are better for providing statistically valid estimates of the effects of covariates for AIDS diseases progression. Thus, this study recommends that while choosing waiting time distributions for semi-Markov models one should consider appropriate waiting time distribution. In addition, the study recommends that concerned bodies should look at different contributing factors of AIDS diseases progression in addition to the ART services administered for slowing the current High level mortality and progression of the disease in the country.

## Declarations

## Ethics Approval and Consent to Participate

All participants and parents/legal guardian of minors and illiterate respondents gave written informed consent for the participation during the survey. The ethics committee and research

advisory board of the Yirgalim General Hospital have approved this consent along with the entire study. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and national research committee.

## Consent for Publication

Not applicable

## Availability of Data and Materials

The datasets generated and analyzed during the present study are available from the corresponding author on reasonable request.

## Competing interests

The authors declare that they have no competing interests.

## Funding

Not applicable

## Authors' Contributions

This work was carried out in collaboration between all authors. TA and AG conceived the idea of the paper. TA analyzed the data and wrote the first draft of the paper. TA, AG, MS and DT revised further to improve the paper. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank Yirgalim General Hospital administration for permitting us to use the data analyzed in this study. Our special appreciation goes to Hawassa and Arba Minch Universities for the financial supports for this study.

## Authors' Information

Tilahun Asena, PhD

Arba Minch University, Department of Statistics, 21, Arba Minch, Ethiopia

**Ayele Goshu, PhD**

Department of Statistics, Kotebe Metropolitan University, Ethiopia

Mebratu Senbeta, PhD

Arba Minch University, Department of Economics, 21, Arba Minch, Ethiopia

**Derbachew Teni, PhD**

Department of Statistics, Arba Minch University, Arba Minch, Ethiopia

Corresponding author: **Tilahun Asena**

## Abbreviations

ARI	Acute Respiratory Infection
AIC/BIC	Akaike/Bayesian Information Criteria
CSA	Central Statistics Agency of Ethiopia
DIC	Deviance Information Criterion
EA	Enumeration Areas
EDHS	Ethiopian Demographic Health Survey
ICC	Intra class Correlation Coefficient
MCMC	Markov Chain Monte Carlo
OR	Odds Ratio
WHO	World Health Organization

## References

- [1] UNICEF (2015). Committing to Child Survival: A Promise Renewed – Progress Report 2014. UNICEF. [www.unicef.org](http://www.unicef.org).
- [2] Mortal, G. B. D. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 385(9963), 117-171.
- [3] Williams BG, Gouws E, Boschi PC, Bryce J, Dye C (2002) Estimates of Worldwide distribution of child death from acute respiratory infections. *Lancet infec Dis* 2(1): 25-32.
- [4] Montasser, N., Helal, R., & Rezaq, R. (2012). Assessment and classification of acute respiratory tract infections among Egyptian rural children. *British Journal of Medicine and Medical Research*, 2(2), 216.
- [5] Liu L, Johnson HL, Cousens S, Perin J, Scott S, et al. (2012). Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. *Lancet* 379: 2151–61
- [6] Nair, H., Simões, E. A., Rudan, I., Gessner, B. D., Azziz-Baumgartner, E., Zhang, J. S. F., & Baggett, H. C. (2013). Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *The Lancet*, 381(9875), 1380-1390.
- [7] Sonogo, M., Pellegrin, M. C., Becker, G., & Lazzerini, M. (2015). Risk factors for mortality from acute lower respiratory infections (ALRI) in children under five years of age in low and middle-income countries: a systematic review and meta-analysis of observational studies. *PloS one*, 10(1), e0116380.
- [8] UNICEF. (2016). A Fair Chance for Every Child: The State of the World's Children 2016.

- [9] Chatterjee, S. (2016). A study of epidemiological factors related to acute respiratory infection in under five children attending the immunization clinic of Calcutta National Medical College and Hospital. *Internet J Pulm Med*, 7(2), 1-6.
- [10] Federal Democratic Republic of Ethiopia Ministry of Health. Health Sector Transformation Plan, August 2015. PAGE 24.
- [11] A Survey of the last ten kilometers project (L10K), (2013). Integrated Community Case Management (iCCM) Survey in Amhara, SNNP and Tigray regions.
- [12] CSA, I. (2016). Central statistical agency (CSA) [Ethiopia] and ICF. *Ethiopia demographic and health survey, Addis Ababa, Ethiopia and Calverton, Maryland, USA*.
- [13] Acquah, H. D. (2013). Bayesian logistic regression modelling via markov chain Monte Carlo algorithm. *Journal of Social and Development Sciences*, 4(4), 193-197.
- [14] Kawo, K. N., Asfaw, Z. G., & Yohannes, N. (2018). Multilevel Analysis of Determinants of Anemia Prevalence among Children Aged 6–59 Months in Ethiopia: Classical and Bayesian Approaches. *Anemia*, 2018.
- [15] Gelman, A., B., C. J., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. (2014). *Bayesian data analysis*. New York, NY, USA: Chapman & Hall.
- [16] Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2000). *Bayesian data analysis*, Chapman & Hall/CRC, USA.
- [17] Ibrahim J.G., Chen M.H. (2000). *Power Prior Distributions for Regression Models*, Statistical Science, 15 (1), 46-60.
- [18] Geberetsadik, A., Worku, A., & Berhane, Y. (2015). Factors associated with acute respiratory infection in children under the age of 5 years: evidence from the 2011 Ethiopia Demographic and Health Survey. *Pediatric health, medicine and therapeutics*, 6, 9.

- [19] Harerimana, J. M., Nyirazinyoye, L., Thomson, D. R., & Ntaganira, J. (2016). Social, economic and environmental risk factors for acute lower respiratory infections among children under five years of age in Rwanda. *Archives of Public Health*, 74(1), 19.
- [20] Snijders T.A.B., and Bosker, R.(1999). Multilevel Analysis. Sage, Thousand Oaks. CA
- [21]Goldstein, H. (2003). Multilevel Statistical Models. Arnold, London.
- [22] Rasbash J, Charlton C, Browne WJ, *et al.* (2012). MLwiN Version 2.26, Centre for Multilevel Modelling, University of Bristol.
- [23] Browne WJ, Charlton C, Rasbash J, Kelly M, Pillinger R (2016) MCMC estimation in MLwiN version 2.36. Centre for multilevel modelling. University of Bristol, Bristol
- [24] Browne, W. J., & Rasbash, J. (2009). *MCMC estimation in MLwiN*. Bristol: Centre of Multilevel Modelling, University of Bristol.
- [25] Rehman, A., Shaikh, B. T., & Ronis, K. A. (2014). Health care seeking patterns and out of pocket payments for children under five years of age living in Katchi Abadis (slums), in Islamabad, Pakistan. *International Journal for Equity in Health*, 13(1), 1.
- [26] Jabessa, S. (2015). Multilevel Analysis of Acute Respiratory Infection Symptoms among under Five Children in Ethiopia. *Journal of Biometrics & Biostatistics*, 6(4), 1.
- [27] Keter, P., Mbakaya, C., Gikunju, J., & Mutai, J. (2015). Knowledge, Perceptions and Practices of Mothers in Relation to Childhood Pneumonia at Kapsabet District Hospital in Nandi County, Kenya. - Google Search. Retrieved August 16, 2016.
- [28] Noordam, A. C., Carvajal-Velez, L., Sharkey, A. B., Young, M., & Cals, J. W. (2015). Care seeking behavior for children with suspected pneumonia in countries in sub-Saharan Africa with high pneumonia mortality. *PloS one*, 10(2), e0117919.

- [29] Fatmi, Z., & White, F. (2002). A comparison of 'cough and cold' and pneumonia: risk factors for pneumonia in children under 5 years revisited. *International Journal of Infectious Diseases*, 6(4), 294-301.
- [30] Bbaale, E. (2011). Determinants of diarrhoea and acute respiratory infection among under-fives in Uganda. *The Australasian medical journal*, 4(7), 400.
- [31] Dadi, A.F., Kebede, Y. and Birhanu, Z. (2014) Determinants of Pneumonia in Children Aged Two Months to Five Years in Urban Areas of Oromia Zone, Amhara Region, Ethiopia. *Open Access Library Journal*, 1: 1044..
- [32] Azad, K. M. A. K. (2009). Risk factors for acute respiratory infections (ARI) among under-five children in Bangladesh. *Journal of Scientific Research*, 1(1), 72-81.
- [33] Tilahun Ferede Asena. Multilevel modelling of modern contraceptive use among Rural and Urban population of Ethiopia. *American Journal of Mathematics and Statistics*. 2013;3(1):1-16.
- [34] Briggs, D. (2003). *Making a difference: Indicators to improve children's environmental health* (pp. 1-17). Geneva: World Health Organization.
- [35] Ghimire, M., Pradhan, Y. V., & Maskey, M. K. (2010). Community-based interventions for diarrhoeal diseases and acute respiratory infections in Nepal. *Bulletin of the World Health Organization*, 88, 216-221.