Journal of Advances in Medicine and Medical Research

25(11): 1-15, 2018; Article no.JAMMR.39428

ISSN: 2456-8899
Past name: British Journal of Medicine and Medical Research, Past ISSN: 2231-0614, NLM ID: 101570965

Missclassification of HIV Disease Stages with Continuous Time Hidden Markov Models

Tatek Getachew Habtemichael[1*], Ayele Taye Goshu[1] and Gemeda Bedaso Buta[2]

[1] *School of Mathematical and Statistical Sciences, Hawassa University, Ethiopia.*
[2] *Department of Statistics, Madda Walabu University, Ethiopia.*

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMMR/2018/39428
Editor(s):
(1) Babatunde Olanrewaju Motayo, Department of Pathology, Federal Medical Center, Abeokuta, Ogun State, Nigeria and Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria.
Reviewers:
(1) S. K. Srivats, Anna University, India.
(2) Michael Chen, California State University, USA.
(3) Keneuoe Hycianth Thinyane, National University of Lesotho, Lesotho.
Complete Peer review History: http://sciencedomain.org/review-history/23897

Original Research Article

Received: 10th January 2018
Accepted 13th March 2018
Published 30th March 2018

ABSTRACT

The purpose of this study is to explore the simple Markov and Hidden Markov models with continuous time to investigate disease progression of HIV/AIDS patients under ART follow-up at Shashemene Referral Hospital, Ethiopia. The msm R package is used for the analysis. Results from the simple Markov model reveals that the disease progression of the HIV/AIDS patients considered tend to move towards the healthier than the worse state. The mean waiting time for the healthiest state is significantly higher than the other transient states. The total length of time stay in a state declines with severity of the disease stages. Analysis of the misclassification model provides transition rates of the true states. Estimation of the transition rates of the true states are found to be relatively smaller compared to those obtained by the simple Markov model. For

**Corresponding author: E-mail: tatgetas@gmail.com;*

the true states compared to observed ones, the conditional probability of moving to the healthiest state from the next worse state grows higher dramatically, while that of moving to next worst state grows slightly lower. The ART based patient care might have positive impacts on the overall progression of the disease. For covariate effects, male patient is more likely to move to worse state than the female does. But age of patient is not significant. The progression of the underlying states of the HIV/AIDS disease behaves similar to that of the generated markers observations except the turning points of the conditional probabilities. The turning points so interesting for be studied further.

Keywords: CD4 count; disease progression; markov model; misclassification; transition rate.

1 INTRODUCTION

Progression of HIV/AIDS disease of the patient can be modeled as Markov process allowing transitions of stages from one to another through time. The time of the transition are assumed random. The data may contain measurement errors. This paper focuses on the disease progression of HIV infected individuals under ART with longitudinal observations of their CD4 counts and respective follow-up times.

By categorizing the CD4 counts in to four disease stages based on WHO disease classification intervals [1] finite number of states can be defined so that progression of the disease would be modelled as Markov process. The process with continuous time leads to continuous time Markov models.

Hidden Markov model assumes that the true disease stages are hidden (unobservable). The true disease stages can only be observed indirectly through disease marker which is CD4 count. The observations generated are assumed to be conditionally independent given the true stages. Many researchers see [2, 3, 4, 5] used CD4 counts as observed markers of the disease progression of HIV-infected patients. The main parameters of interest are transition rates, misclassification matrix, and regression parameters for the covariates.

Hidden Markov model is commonly used in areas such as speech and signal processing [6]. In this paper, continuous time hidden Markov models are explored to investigate the HIV/AIDS diseases progression. Transition rates among disease stages and misclassification between the true and observed disease stages are estimated. The functional of these parameters such as waiting times and conditional transition probabilities are computed. Covariate effects are also determined.

2 DATA

The data are obtained from $n = 354$ randomly selected HIV/AIDS patients, who had been under ART follow-up at the Shashemene Referral Hospital, Ethiopia, during January 2006 to December 2012. The patients are of ages 16 years and above. The data are measurements of CD4 counts per mm^3 of blood a sample, observation times and individual specific covariates. Patient's visiting times are assumed to be irregular with different number of follow-up time points.

The five disease stages are: state 1 (CD4 count $>$ 500); state 2 (350 $<$ CD4 count $\leq$ 500); state 3 (200 $<$ CD4 count $\leq$ 350); state 4 (CD4 count $\leq$ 200); and state five (Death). The data consists of transitions from one good state to another, transitions to the absorbing state. Death is taken as the fifth state and it is an absorbing state of the process. We consider the last follow-up time of the patient as a current time. Two covariates considered are sex (0 female, 1 male) with proportions 0.585 and 0.415 respectively and age (in months) of the patient.

Bayesian joint models that combines linear mixed effects for the longitudinal outcomes (square root of CD4 cell counts) and the parametric accelerated failure time distributions for the time-to-death data have been developed and studied for this data set in [7, 8]. Here hidden states that generate the observations are considered. Using continuous time hidden Markov model, transition rates, misclassification between true

and observed disease states, waiting time in states, conditional transition probabilities are estimated. Effects of the covariates are also determined.

3 STATISTICAL MODEL

The observations are CD4 counts for HIV infected patients on ART treatment at follow-up time t. States of the Markov process are defined by the seriousness of the sickness based on the CD4 counts in cells/microliter. We consider these states as observations of the disease process and based on this observations two types of models are emerged. We assume the observation times are non informative as described by [9].

We first study the simple Markov model which is termed as a homogeneous continuous-time Markov model. Then follows the continuous time hidden Markov model for misclassification. In the simple Markov Model, we consider the observed states are exactly the same as the true states of the disease, while in the misclassification model we consider the observed states are generated by the underlying true states of the disease. In both cases, the Markov models are of five states.

3.1 Simple Markov Model

The disease progression follows a five-state Markov chain in continuous time and transitions are allowed for adjacent disease stages. It is assumed that the process allows transitions of a disease state to adjacent states and also direct transition to the absorbing state. See Fig. 1. The states of disease process modelled as a homogeneous continuous-time Markov process, with a transition intensity matrix $\boldsymbol{Q}$. The intensity represents the instantaneous risk of moving from state r to state s:

$$q_{rs}(t, \boldsymbol{X}(t)) = \lim_{t \to 0} \frac{P(\kappa_{t+\delta t} = s | \kappa_t = r)}{\delta t} \tag{3.1}$$

$$\boldsymbol{Q} = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 & q_{1d} \\ q_{21} & q_{22} & q_{23} & 0 & q_{2d} \\ 0 & q_{32} & q_{33} & q_{34} & q_{3d} \\ 0 & 0 & q_{43} & q_{44} & q_{4d} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Each row of the transition intensity $\boldsymbol{Q}$ sums to zero, so that the diagonal elements are $q_{rr} = -\sum_{r \neq s} q_{rs}$

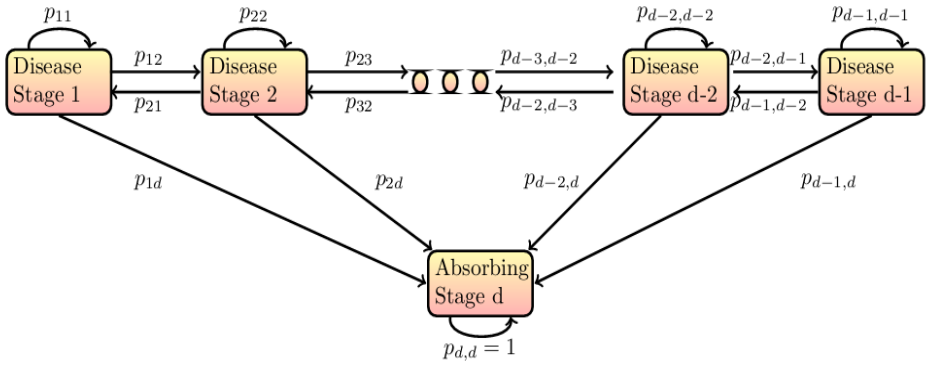

Fig. 1. General model for disease progression

3.2 Likelihood of the Observation Process

The data for individual i consist of a series of times $(t_{i1}, \ldots, t_{T_i})$ and corresponding states $(\kappa_{i1}, \ldots, \kappa_{T_i})$. These states are the values of categorization of continuous marker values of CD4 counts. The likelihood contribution of an individual [10] is given as:

$$L_i(\boldsymbol{\kappa}_i|\boldsymbol{Q}) = f(\boldsymbol{\kappa}_i|\boldsymbol{Q}) = f(\kappa_{i1},\ldots,\kappa_{iT_i}) = f(\kappa_{it_1}) \prod_{t_j=t_2}^{T_i} f(\kappa_{it_j}|\kappa_{it_{j-1}}) \tag{3.2}$$

where $\boldsymbol{P}_1 = f(\kappa_{it_1})$ is the initial state distribution and $f(\kappa_{it_j}|\kappa_{it_{j-1}})$ is the $(\kappa_{it_{j-1}},\kappa_{it_j})$ element of the transition probability matrix which is computed at time intervals (t_{j-1},t_j). The transition probability matrix $\boldsymbol{P}(u,t+u) = \boldsymbol{P}(t)$ are solved by the matrix exponential of $\boldsymbol{Q}$ scaled by the time interval, $\boldsymbol{P}(t) = exp(t\boldsymbol{Q})$.

If the last observation of an individual state is κ_{T_i} is the death, the state measured with out error and whose entry time is known, then the likelihood contribution is summed over the states $m \in \{1,2,3,4\}$ on the previous instant before death:

$$L_D = \sum_{m \neq D} P_{\kappa_{t_j},m}(t_{j+1} - t_j) q_{m,D} \tag{3.3}$$

The full likelihood is then the product of probabilities of transition between observed states, over all individuals i and observation times t :

$$L(\boldsymbol{Q}) = \prod_i L_i(\boldsymbol{\kappa}_i|\boldsymbol{Q}) \tag{3.4}$$

The likelihood $L(\boldsymbol{Q})$ is maximized in terms of $log(q_{rs})$ to compute estimates of q_{rs}, using standard optimization algorithms that is not analytic solution but use numerical methods.

Mean sojourn times will be extracted for transient states from the estimated transition rates. The initial state distribution in Equation (3.2) is not estimated from the msm Package for this model and we take the proportion of individuals in transient states at ART initiation t_1.

The estimate of the transition rate $\boldsymbol{Q}$ may be affected by the covariate and the transition probability of the model with covariate effects is P(t, $\boldsymbol{X}(t)$). The covariate can be individual specific and/or time varying and they are measured at all time points of the response. In time varying covariates considered constant in between two time intervals and have an effect in the previous time t_{j-1}. Explanatory variables can be included at each level of the model through generalized regressions. A form of a proportional hazards model described in [11, 12], where the transition intensity matrix elements q_{rs}(t) at time t to covariates $\boldsymbol{X}(t)$.

$$q_{rs}(\boldsymbol{X}(t)) = q_{rs}^{(0)} exp(\boldsymbol{\beta}'_{rs}\boldsymbol{X}(t)) \tag{3.5}$$

where $q_{rs}^{(0)}$ is the baseline intensity, $\boldsymbol{\beta}_{rs}$ is regression coefficient. Maximum likelihood estimates for baseline intensities and regression coefficients can be obtained from the msm package for R statistical software.

From the estimated transition intensity matrix $\boldsymbol{Q}$ it is of interest to estimate

- The mean sojourn times:- the average period in a single stay in each transient state. It is estimated by the diagonal entries of the transition rate $-1/q_{rr}$
- The total length of stay:- which is defined as the expected amount of time spent by a subject in each state during the study period.

3.3 Continuous Time Hidden Markov Model

The observations in longitudinal disease states that represent the underling true disease stages which are not directly observed. The underling disease model assumed to be Markov chain and observations are conditionally independent given the underling true disease states. The underling disease stages are governed by Markov transition probability matrix $\boldsymbol{P}(t)$ for a given time t, which are the matrix exponential of the transition rate $\boldsymbol{Q}$. The likelihood function linking the observations with the underling true disease state is arbitrary distribution based on observations.

3.3.1 Misclassification model

In this model the observations are $\kappa_{it}^c \in \{1,\ldots,s\}$ where s is the number of states generated by hidden states κ_{it}. The observations are assumed conditionally independent. The general description of hidden Markov model in continuous time for the misclassification model is displayed in Fig. 2.

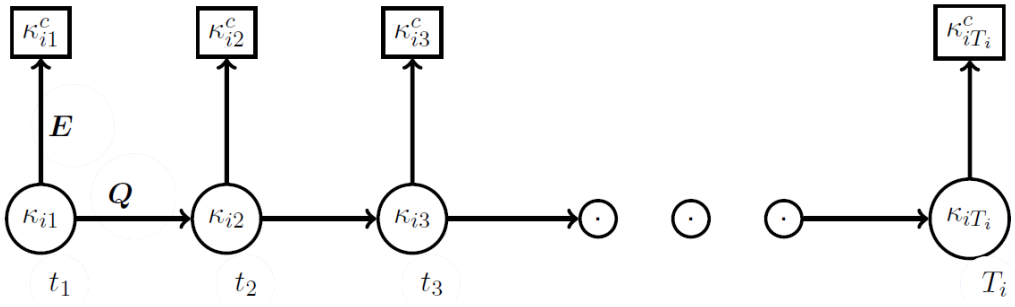

Fig. 2. General description of hidden Markov model in continuous time for misclassification model

The series of observations of an individual patient i is denoted by $\boldsymbol{\kappa}_i^c = (\kappa_{i1}^c, \ldots, \kappa_{iT_i}^c)$. The likelihood for observed categorical variable is

$$f(\boldsymbol{\kappa}_i^c) = f(\kappa_{i1}^c, \ldots, \kappa_{iT_i}^c) = \sum_{\boldsymbol{\kappa}_i} f(\kappa_{i1}^c, \ldots, \kappa_{iT_i}^c | \kappa_{i1}, \ldots, \kappa_{iT_i}) f(\kappa_{i1}, \ldots, \kappa_{iT_i})$$
$$= \sum_{\boldsymbol{\kappa}_i} \prod_{t_i=t_1}^{T_i} f(\kappa_{it_i}^c | \kappa_{it_i}) \times \left\{ f(\kappa_{i1}) \prod_{t_i=t_2}^{T_i} f(\kappa_{it_i} | \kappa_{it_i-1}) \right\} \tag{3.6}$$

where the sum is taken over all possible paths of underlying states $\boldsymbol{\kappa}_i = \kappa_{i1}, \ldots, \kappa_{iT_i}$. The observations are assumed to be independent given the series of underling states and governed by time invariant misclassification probability matrix $\boldsymbol{E}$. The e_{rr} are the diagonal elements of the misclassification probability matrix $\boldsymbol{E}$ represents the correct classification of individuals in the true disease stages. We assume that the disease stages misclassified in to adjacent disease stages so that the misclassification matrix is 3-band matrix.

$$\boldsymbol{E} = \begin{pmatrix} e_{11} & e_{12} & 0 & 0 & 0 \\ e_{21} & e_{22} & e_{23} & 0 & 0 \\ 0 & e_{32} & e_{33} & e_{34} & 0 \\ 0 & 0 & e_{43} & e_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The misclassification probability e_{rs} represents the probability that state s is observed given that the true one is r. The misclassification of the disease stages can be affected by the covariates and to investigate explanatory variables $\boldsymbol{X}(t)$ for the probability e_{rs} of misclassification as state s given underlying state r, a multinomial logistic regression model can be used [10]: with baseline state s_o.

$$log\left(\frac{e_{rs}(t)}{e_{rs_o}(t)}\right) = \boldsymbol{\beta}_{rs}' \boldsymbol{X} \tag{3.7}$$

This is the multinomial logit model with regression coefficients $\boldsymbol{\beta}_{rs}$.

Model parameters are estimated using maximum marginal likelihood method with numerical computations in the msm R package. In continuous time Markov model the state transitions are governed by the exponential of the matrix of transition rates. Fitting the Markov model in continuous time may be an advantageous to estimate rates from the data and then transform these in to probabilities of transition over a period of time.

The analysis is conducted using the msm R package [13]. The package is written to fit multi-state Markov models in continuous time with or without classification error. It is capable of estimating any form of Markov transition intensity matrix and misclassification matrix, with any number of covariates on either of transition rate and misclassification matrices.

High dimensional optimization will produce flat likelihood surfaces if there is insufficient

information in the data, leading to failure of convergence. In such cases, simplified models might be constructed with some unstable parameters fixed at values obtained from prior information. In applications to disease progression, adjacent states can often be combined to simplify the model. Some adjustments to the numerical methods may also aid convergence for weakly identifiable models [13, 10].

4 RESULTS AND DISCUSSION

From the summary of the transitions data, it is observed that there are 4 deaths among patients in state 1, 8 deaths from state 2, 4 deaths from state 3 and 5 deaths from state 4. The proportion of observed states displayed in Table 1 shows (in brackets) that there high transitions from state to itself. The lower diagonal percentages are relatively higher than the upper once. This shows the individual patients drift towards the healthy state because individuals are in ART treatment.

In our model, we assumed that the transition occurs between adjacent disease states in a short period of time and death from any state. But from the frequency table we see that there are jumps of transitions to outer diagonal elements, for example, 20 transitions from state 1 to state 3 and 8 transitions from state 1 to state 4. This indicates that an individual may have a transition to adjacent states and reach state 3 or 4 within the interval of follow-up time.

4.1 Analysis of the Simple Markov Model

In this model we consider the observed disease states as directly observed true states of the process. The parameters for this model are the transition intensity matrix Q and initial state distribution P_1. The initial state distribution for the model can be estimated from the proportion of states at t_1 for all individual starting ART treatment follow-up.

At the start, individuals are at highest risk to be in state 4 (P= 0.715) and lowest risk to be in state 1 (P=0.006). The algorithm is initiated in a variety of realistic points and results for the estimated transition rates with 95% confidence intervals are given in Table 2.

From the estimated values of the transition rate we found that the lower diagonals are relatively higher than the upper ones. The ratio of the elements in the transition rate help us to make comparisons for which transitions are more likely to occur. For example, a patient at state 2 is 10.5% more likely to be in a better state than in a worse state, and given in state 3, it is 143% more likely to be in a better state than worse state. Similarly, given being in state 4, it is highly likely to be in state 3 than to in death.

The rate of transition from any transient disease state to the next worse state (or to the death state) is the highest when the patients are in state 2. All the rates are significant at $\alpha = 5\%$ significance level.

The average waiting times (Sojourn times) is given in Table 3.

From the estimated sojourn times, mean time spent in the healthy state is the highest and mean time spent in the worst states are found to be equal.

The total length of stay, in years is declining with severity of the disease for state 1 (27.27yr), for state 2 (14.5yr), for state 3 (11.2yr), for state 4 (3.8yr).

Table 1. Frequency counts and proportions of the transitions of the observed states

from/ to	State 1	State 2	State 3	State 4	State 5
State 1	329 (0.763)	70 (0.162)	20 (0.046)	8 (0.019)	4 (0.009)
State 2	145 (0.337)	177 (0.412)	90 (0.209)	10 (0.023)	8 (0.019)
State 3	71 (0.109)	191 (0.292)	313 (0.479)	75 (0.115)	4 (0.006)
State 4	27 (0.046)	62 (0.105)	218 (0.370)	277 (0.470)	5 (0.008)

Table 2. The estimated transition intensities for the simple Markov model

Q	Estimate	95% CI
q_{11}	-0.051	(-6.35e-02, -0.041)
q_{12}	0.050	(3.98e-02, 0.062)
q_{15}	0.001	(2.57e-04, 0.005)
q_{21}	0.090	(7.69e-02, 0.106)
q_{22}	-0.175	(-2.00e-01, -0.153)
q_{23}	0.082	(6.59e-02, 0.101)
q_{25}	0.003	(1.10e-03, 0.008)
q_{32}	0.105	(9.09e-02, 0.121)
q_{33}	-0.148	(-1.68e-01, -0.131)
q_{34}	0.043	(3.39e-02, 0.055)
q_{35}	0.001	(6.01e-06, 0.038)
q_{43}	0.125	(1.09e-01, 0.145)
q_{44}	-0.126	(-1.45e-01, -0.109)
q_{45}	0.001	(4.05e-06, 0.052)

Table 3. Sojourn times estimated for the simple Markov model in months

	estimates	SE	L	U
State 1	19.601	2.189	15.748	24.396
State 2	5.717	0.386	5.008	6.526
State 3	6.750	0.425	5.966	7.637
State 4	7.972	0.571	6.928	9.173

The transition probability is computed as a matrix exponential of the transition intensity matrix. The model have the ability to compute the transition probability from the transition rate for patients at any point in time. For example, the transition probability matrix for the model for 24 months time points are given Table 4.

From the results of the 2 year transition probability, there is a highest probability to move to death given that the patient is in state 2. From the fitted transition probability we can compute the estimated survival probabilities from each state. The fitted survival probabilities is given in Fig. 3. The probabilities are declining over time.

The conditional probabilities for disease progression for 200 months are displayed in Fig. 4. The conditional probability of staying in the given state decreases with increasing time and relatively high when patients are in the healthy state. The conditional probability of the transition in to the next worst state slightly increases and reach optimum points at months 34, 11 and 12 and decreases for states $\{1, 2, 3\}$. Relatively the probability is higher from state 2 up to the intersection point with the transition from state 1. The transition to the next better state increases for all states and reach optimum at time points 34, 11 and 12 months and decrease for states $\{2, 3, 4\}$. The transitions to the next better state for individuals in state 4 is relatively higher for the first 11 month treatments. the probability increases highly and reach maximum and decrease. The turning point of the probabilities need attentions of health workers that they should give attention and special patient care for the individuals. These results are consistent with those found in [2] except there is some irregularities near the turning points.

Table 4. Transition probability at time of 24 months estimated from the transition intensity based on the simple model

from/ to	State 1	State 2	State 3	State 4	State 5
State 1	0.551	0.238	0.141	0.036	0.035
State 2	0.431	0.254	0.205	0.068	0.042
State 3	0.327	0.263	0.270	0.110	0.031
State 4	0.241	0.254	0.319	0.162	0.023
State 5	0.000	0.000	0.000	0.000	1.000

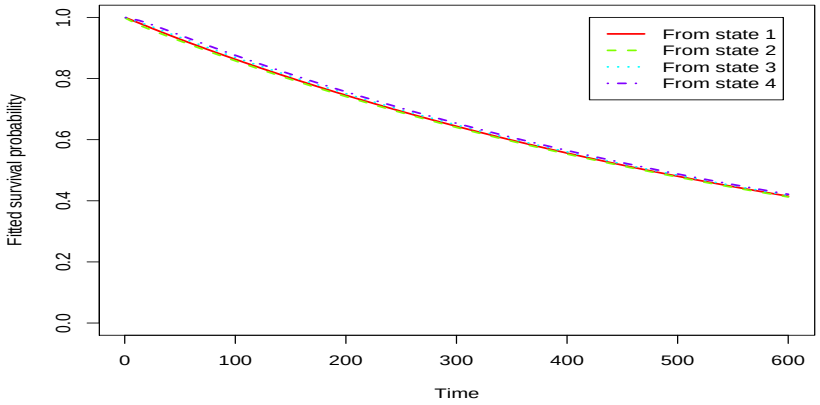

Fig. 3. Fitted survival probabilities based on the simple Markov model

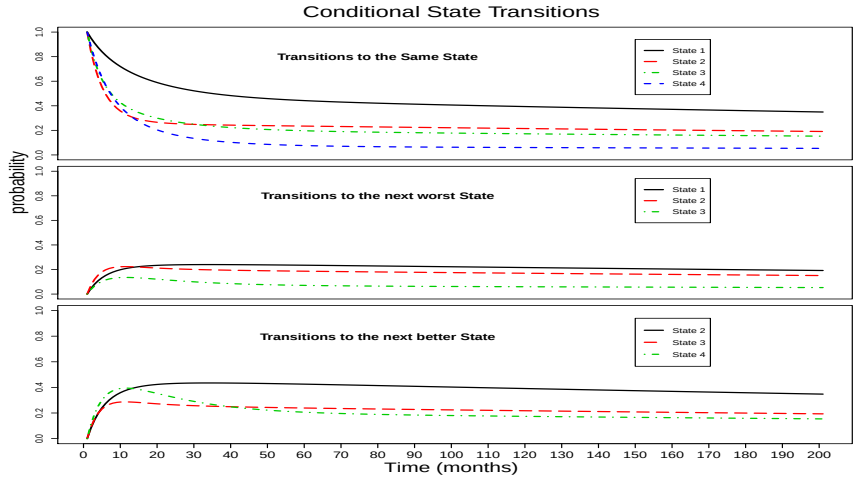

Fig. 4. The conditional probabilities of state transitions based on the simple Markov model

Probability that each state is next:- extracts the matrix of probabilities $-q_{rs}/q_{rr}$ that the next state is state s given that the current is r. The result shows that there is high transition towards the recovery than moving to worth state. This is an implication of individuals in ART treatment recover through time and their CD4 count increase through time.

This shows as an individual in state 1 is with high probability 98% will be in state 2. When an individual is in state 2 it is approximately equal probability to make a transition in to state 1 and state 3. The relative transition from state 3 is highly to state 2 with 71% and from state 4 the relative transition is 99% to move to state 3.

The observed and expected prevalences are plotted for each state in Fig. 5. The graph displays the number of individuals initially in high risk classes about 24% and 71% of them in state 3 and 4 and then declines over time. For state 1 and state 2, the prevalence starts at low rates and increase over time and then decline. For the death state the prevalence increase through out. There is some mismatch between the observed and expected prevalences in all states.

4.2 Covariate Effects on Transition Rates

The transition rates in the model can be affected by the covariate and msm package can have the ability to fit the covariate effect on the transition rate. The covariates included here are gender(male and female) and Age(continuous and varies with time) of the patient. The covariets can be individual specific or time varying. The time homogeneous assumption in the transition rate may no longer applicable due to time varying covariate. The estimated transition rate and the coefficients of the effects of each covariate for each transition rate is given Table 6.

The hazard ratio computed in the model can able to identify which covariates are significant at 5% level. The effect of the covariate gender is significant for transition intensities to the better states. The covariate age is not significant. This may be an indicative to assess the effect of age in different age groups.

The mean sojourn times in Table 8 show that females have longer mean sojourn times (20.0 with se 9.8 months) than males (17.1 with se 9.7) to stay in the healthiest state 1. The difference are insignificant.

4.3 Analysis of the Misclassification Model

Observed states are considered to be the out put of the underling disease stages and generated by some arbitrary distribution. The model parameters of the prior model is initial state distribution, the transition rate and the misclassification probability matrix. The parameter estimation is based on the maximum likelihood by numerical method, using msm package for R statistical software. The algorithm is initiated in a variety of realistic points and the estimates are given in Table 9. The results for the transition rates refer to the true states of the process that are unobserved.

Table 5. The estimated probability that each state is next with 95% confidence interval based on the simple Markov model

from/ to	State 1	State 2	State 3	State 4	State 5
State 1	0	0.98 (0.90, 0.99)	0	0	0.02 (0.00, 0.10)
State 2	0.52 (0.45, 0.58)	0	0.47 (0.40, 0.53)	0	0.02 (0.00, 0.05)
State 3	0	0.71 (0.55, 0.76)	0	0.29 (0.21, 0.35)	0.00 (0.00, 0.22)
State 4	0	0	0.99 (0.73, 0.99)	0	0.00 (0.00, 0.27)
State 5	0	0	0	0	0

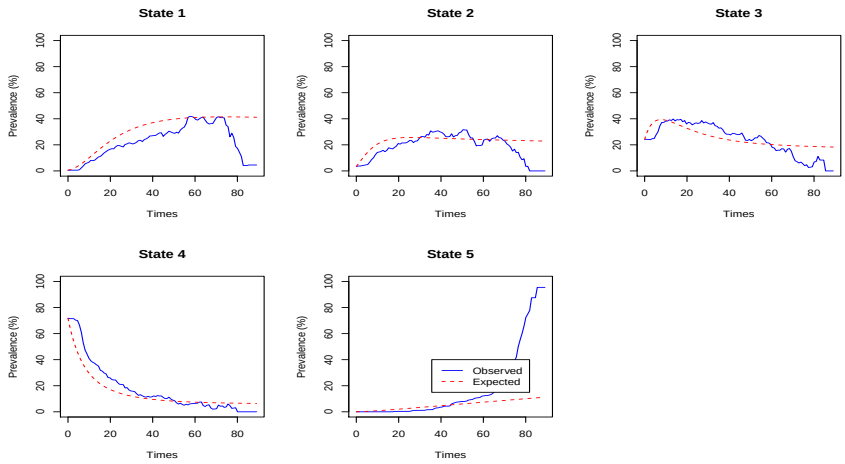

Fig. 5. The observed and expected prevalence based on the simple Markov model.

Table 6. The estimated transition intensities and coefficients of the covariates on transition rates with 95% confidence interval based on the simple Markov model

Q	Estimate	95% CI	Gender	Age in months
q_{11}	-0.052	(-6.51e-02, -0.041)		
q_{12}	0.051	(4.07e-02, 0.064)	1.185 (7.25e-01, 1.94e+00)	0.999 (0.973, 1.025)
q_{15}	0.001	(4.88e-05, 0.007)	0.109 (3.76e-04, 3.13e+01)	1.065 (0.976, 1.161)
q_{21}	0.088	(7.46e-02, 0.104)	0.690 (4.88e-01, 9.76e-01)	0.994 (0.976, 1.011)
q_{22}	-0.175	(-2.00e-01, -0.152)		
q_{23}	0.084	(6.74e-02, 0.104)	0.894 (5.80e-01, 1.38e+00)	0.985 (0.961, 1.009)
q_{25}	0.003	(7.70e-04, 0.010)	1.164 (8.41e-02, 1.61e+01)	1.025 (0.955, 1.099)
q_{32}	0.106	(9.19e-02, 0.123)	0.576 (4.23e-01, 7.83e-01)	0.995 (0.980, 1.010)
q_{33}	-0.152	(-1.73e-01, -0.134)		
q_{34}	0.045	(3.54e-02, 0.058)	0.737 (4.49e-01, 1.21e+00)	0.979 (0.955, 1.005)
q_{35}	0.0004	(4.50e-07, 0.368)	1.269 (5.27e-09, 3.05e+08)	0.958 (0.532, 1.724)
q_{43}	0.131	(1.13e-01, 0.152)	0.659 (4.83e-01, 9.00e-01)	0.993 (0.977, 1.008)
q_{44}	-0.131	(-1.52e-01, -0.113)		
q_{45}	0.001	(3.18e-06, 0.086)	3.347 (8.29e-05, 1.35e+05)	0.947 (0.664, 1.350)

The estimated initial state distribution illustrated in Table 10 shows individuals initially at high risk classes approximately 68.9% in state 4 and 28.6% for individuals in state 3. The initial state distribution in states 1 and 2 are very small. These results are approximately the same as the estimates from the simple Markov model. From the estimated values of the transition rates Q, we found that the lower diagonals are relatively higher than the upper once. This shows individual patients initially in high risk classes drift towards to healthy state through time because patients are in ART treatment. The estimated values of the transition rate in the miscalssification model is relatively small compared with the estimates from the simple model.

From the estimated values of the misclassification matrix E, state 1 and state 4 are well classified, state 2 is misclassified to state 1 with probability 0.196 and to state 3 with probability 0.144. Similarly state 3 is

misclassified to state 2 with probability 0.074 and to state 4 with probability 0.126. The highest misclassification is found to be from state 2 to state 1.

The mean time spent in the healthy state is relatively very high 14.33 years and in the worse states are found to be small and relatively equal. The mean time spent for misclassificaton model is very high compared to the estimates based on the simple model.

The estimates of the total length of stay in years is for state 1 (44.9yr), for state 2 (5.6yr), for state 3 (2yr), for state 4 (0.2yr). It is the highest for the healthy state and declines with severity of the disease stages. The total length of stay in this model for states $\{2, 3, 4\}$ is relatively very small with the estimates from the simple model.

The conditional probabilities for disease progression upto 200 months is given in the following Fig. 6. Survival probabilities are plotted for up to 600 months. In the graph the conditional probability of staying in the given state is decrease with time and relatively high when patients are in the healthy state. The conditional probability of the transition in to the next worst state is slightly increase with time and relatively state 2 the transition increase highly and decrease with time. The transition to the next better state is increase for all states and for individuals in state 4 the probability increase highly and reach maximum and decrease.

4.4 Covariate Effects on Misclassification Probabilities

The effect of the explanatory variables on the misclassification probabilities is evaluated and the estimates of the transition rate, initial distribution, miscalssification matrix and the coefficients of the covariets are give in Table 12.

Table 7. The estimated hazard ratios for the covariates for the simple Markov model

	Gender			Age in months		
Q	HR	L	U	HR	L	U
q_{12}	1.185	7.25e-01	1.94e+00	0.999	0.973	1.025
q_{15}	0.109	3.76e-04	3.13e+01	1.065	0.976	1.161
q_{21}	0.690	4.88e-01	9.76e-01	0.994	0.976	1.011
q_{23}	0.894	5.80e-01	1.38e+00	0.985	0.961	1.009
q_{25}	1.164	8.41e-02	1.61e+01	1.025	0.955	1.099
q_{32}	0.576	4.23e-01	7.83e-01	0.995	0.980	1.010
q_{34}	0.737	4.49e-01	1.21e+00	0.979	0.955	1.005
q_{35}	1.269	5.27e-09	3.05e+08	0.958	0.532	1.724
q_{43}	0.659	4.83e-01	9.00e-01	0.993	0.977	1.008
q_{45}	3.347	8.29e-05	1.35e+05	0.947	0.664	1.350

Table 8. Mean sojourn times for females and males

	Females				Males			
	estimates	SE	L	U	estimates	SE	L	U
State 1	20.035	9.829	7.659	52.408	17.134	9.694	5.653	51.935
State 2	3.533	1.084	1.936	6.447	4.439	1.525	2.264	8.705
State 3	3.723	0.978	2.225	6.228	5.734	2.371	2.549	12.895
State 4	4.996	1.423	2.859	8.729	7.105	6.789	1.092	46.235

Table 9. The estimated transition rates and misclassification probability matrix with 95% confidence intervals based on the misclassification model

Q	Estimate	95% CI	E	Estimate	95% CI
q_{11}	-0.005	(-1.66e-02, -0.002)	e_{11}	0.933	(0.877, 0.965)
q_{12}	0.004	(9.65e-04, 0.018)	e_{12}	0.067	(0.035, 0.123)
q_{15}	0.001	(1.91e-04, 0.006)			
q_{21}	0.030	(2.28e-02, 0.040)	e_{21}	0.196	(0.130, 0.286)
q_{22}	-0.053	(-7.05e-02, -0.039)	e_{22}	0.659	(0.532, 0.767)
q_{23}	0.020	(1.17e-02, 0.034)	e_{23}	0.145	(0.093, 0.218)
q_{25}	0.002	(9.92e-04, 0.006)			
q_{32}	0.063	(5.05e-02, 0.078)	e_{32}	0.074	(0.036, 0.145)
q_{33}	-0.071	(-9.00e-02, -0.057)	e_{33}	0.800	(0.650, 0.896)
q_{34}	0.009	(3.14e-03, 0.023)	e_{34}	0.126	(0.080, 0.193)
q_{35}	0.0003	(7.35e-06, 0.015)			
q_{43}	0.084	(6.71e-02, 0.105)	e_{43}	0.054	(0.019, 0.143)
q_{44}	-0.085	(-1.06e-01, -0.068)	e_{44}	0.946	(0.857, 0.981)
q_{45}	0.001	(2.68e-04, 0.005)			

Table 10. The estimated initial state occupancy probabilities for misclassification model

	Estimate	LCL	UCL
State 1	0.006	0.001	0.031
State 2	0.019	0.005	0.072
State 3	0.286	0.220	0.360
State 4	0.689	0.607	0.749

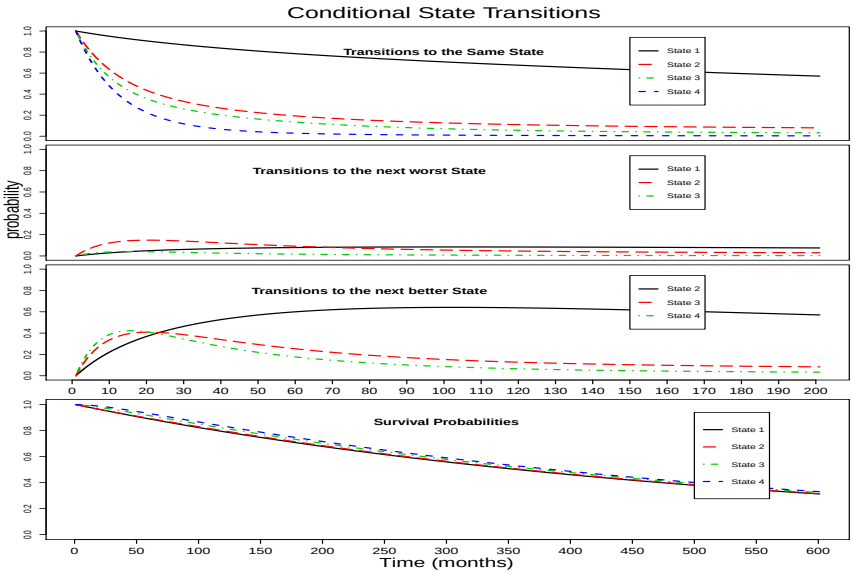

Fig. 6. The conditional probabilities of state transitions and fitted survival probabilities based on the misclassification model

Table 11. Mean sojourn times for misclassification model

	estimates	SE	L	U
State 1	172.4	88.03	63.35	469.0
State 2	17.3	2.60	12.84	23.2
State 3	13.4	1.57	10.62	16.8
State 4	11.8	1.22	9.62	14.4

Table 12. The estimated transition rate, misclassification probability matrix and coefficients of the covariates on the misclassification with 95% confidence intervals based on the misclassification model

Q	Estimate/95% CI	E	Estimate	Gender	Age in months
q_{11}	-0.005 (-1.70e-02, -0.001)	e_{11}	0.926 (0.860, 0.962)		
q_{12}	0.002 (2.56e-04, 0.023)	e_{12}	0.074 (0.038, 0.140)	2.414 (0.813, 7.168)	1.021 (0.963, 1.083)
q_{15}	0.002 (7.61e-04, 0.008)				
q_{21}	0.028 (2.09e-02, 0.039)	e_{21}	0.196 (0.121, 0.301)	0.438 (0.213, 0.899)	0.953 (0.915, 0.992)
q_{22}	-0.050 (-6.70e-02, -0.037)	e_{22}	0.652 (0.515, 0.768)		
q_{23}	0.020 (1.19e-02, 0.033)	e_{23}	0.152 (0.094, 0.237)	2.071 (0.985, 4.354)	1.008 (0.970, 1.049)
q_{25}	0.002 (3.59e-04, 0.008)				
q_{32}	0.063 (5.15e-02, 0.078)	e_{32}	0.060 (0.0248, 0.140)	0.607 (0.170, 2.166)	0.925 (0.853, 1.003)
q_{33}	-0.073 (-9.13e-02, -0.059)	e_{33}	0.800 (0.620, 0.907)		
q_{34}	0.009 (4.01e-03, 0.021)	e_{34}	0.140 (0.090, 0.210)	1.193 (0.600, 2.370)	0.989 (0.954, 1.025)
q_{35}	0.001 (4.38e-05, 0.010)				
q_{43}	0.089 (7.34e-02, 0.109)	e_{43}	0.025 (0.003, 0.200)	0.304 (0.024, 3.802)	0.889 (0.748, 1.057)
q_{44}	-0.090 (-1.09e-01, -0.074)	e_{44}	0.975 (0.800, 0.997)		
q_{45}	0.0002 (1.06e-07, 0.370)				

Table 13. Initial state occupancy probabilities for misclassification model with covariates

	Estimate	LCL	UCL
State 1	0.001	0.000	0.077
State 2	0.024	0.001	0.062
State 3	0.276	0.207	0.350
State 4	0.699	0.604	0.761

Effects of gender on misclassification probability is significant at 5% significance level. Compared to females, for males, there is more likely to misclassify a state given to worse state, for example, $1 \rightarrow 2$ 2.4 times, $2 \rightarrow 3$ 2.1 times, $3 \rightarrow 4$ 1.2 times. Age is not significant.

5 CONCLUSIONS

Disease progression of HIV/AIDS is studied for 354 patients under ART follow-up at the Shashemene Referral hospital, Ethiopia. Five states simple Markov and hidden Markov models are fitted to the data. Transition rates, misclassification probabilities, sojourn times, conditional transition probabilities and coefficients of covariates are estimated.

Results from the simple Markov model analysis reveal that at the start, individuals are at highest risk to be in severe state with probability of 0.715 and at healthy state with lowest probability of 0.006. Transition rates in the lower diagonals are relatively higher than the upper ones indicating that the progression of the disease is towards the healthier states. There is more likely to be in a better state than worse state. This is a good news that ART based patient care has positive impacts on the overall progression of HIV/AIDS disease. Among the transient states, the mean waiting time for the healthiest state is significantly high, while that of the worser states are found to be equal. Moreover, the total length of time stay in a state declines with severity of the disease.

The conditional probability of staying in the given state decreases with increasing time and relatively high when patients are in the healthy state. The conditional probability of transition in to the next worst state increases and reach optimum points at 34 months from state 1, 11 months from state 2 and 12 months from state 3 and declines then after. The conditional probability of transition in to the next better state also increases and reach optimum points at points about 0.4 and then declines. The turning points so interesting for be studied.

Results from the miscalssification model analysis reveal that, referring to the true states, a patient is initially at highest risk to be severely sick with probability of 0.69 and at lower risk to be in the better health stages. The estimated values of the transition rates of the true states is relatively smaller compared to those from the simple model.

A patient in states 1 and 4 are well classified indicating that there is high probability to be stay in those states. State 2 is misclassified to state 1 with probability 0.196 but to state 3 with probability 0.144. Sate 3 is misclassified to state 2 with probability 0.074 and to state 4 with probability 0.126. The highest misclassification is found to be from state 2 to state 1. The mean times spent in states under the misclassificaton model are respectively higher compared to the estimates found with the simple model.

The conditional probabilities under the misclassificaton model behave similar to those under the simple model for both transitions in to the next better state and in to the next worst state. However, the conditional probabilitiy of moving to healthiest state from the next worst state grows higher dramatically. And those conditional probabilities of moving to next worst state grows slightly lower. Male patients are more likely to move to worse state than the females do. Effect of age is found to be almost insignificant in this study.

We may conclude from this study that progression of the underlying states of the HIV/AIDS disease behave similar to that of the generated markers or observations except the turning points of the conditional probabilities. These findings can be used by health professionals to provide more efficient patient cares.

CONSENT

It is not applicable.

ETHICAL APPROVAL

The ethical clearance for this study has been obtained from Shashemene Referral Hospital, and Hawassa University.

ACKNOWLEDGMENT

The authors would like to thank the Shashemene referral hospital for providing us the data used in this study.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

References

1. World Health Organization. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children; 2006.
Available: http://www.who.int/hiv/pub/guidelines/hivstaging/en/ (7 August 2006)

2. Goshu AT, Dessie ZG. Modelling progression of HIV/AIDS disease stages using Semi-markov process. Journal of Data Science. 2013;11:269-280.

3. Guihenneuc-Jouyaux CG, Richardson S, Longini IM. Modeling markers of disease progression by a hidden Markov process: Application to characterizing CD4 cell decline. Biometrics. 2000;56:733-741.

4. Longini IR, Clark WS. Statistical analysis of the stages of HIV infection using a Markov model. Statistics in Medicine.1989;8:831-843.

5. Sattan AG, Longini IM, Clark WS. Markov chain with measurement error: Estimating the 'true' course of a marker of progression of human immunodeficiency virus disease. Applied Statistics.1996;45(3):275-309.

6. Rabiner LR, Fellow. A tutorial on hidden Markov models and selected applications in speech recognition. 1989;257-286.

7. Buta GB, Goshu AT, Worku HM. Bayesian joint modelling of disease progression marker and time to death event of HIV/AIDS patients under art follow-up. British Journal of Medicine & Medical Research. 2015;5(8):1034-1043.

8. Erango MA, Goshu AT, Buta GB, Dessiso AH. Bayesian joint modelling of survival of HIV/AIDS patients using accelerated failure time data and longitudinal CD4 cell counts. British Journal of Medicine & Medical Research. 2017;20(6):1-12.

9. Grüger J, Kay R, Schumacher M. The validity of inferences based on incomplete observations in disease state models. Biometrics. 1991;47:595-605.

10. Jackson, C. Multi-state modelling with r: The msm package. Version 1.6.5; 2017.

11. Jackson C, Sharples L, Thompson S, Duffy S, Couto E. Multistate Markov models for disease progression with classification error. The Statistician. 2003;52(2):193-209.

12. Marshall G, Jones RH. Multi-state Markov models and diabetic retinopathy. Statistics in Medicine. 1995;14:1975-1983.

13. Jackson, C. Multi-state models for panel data : The msm Package for r. Journal of Statistical Software. 2011;38:1-28.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/23897