Journal of Data Science **11**(2013), 269-280

Modelling Progression of HIV/AIDS Disease Stages Using Semi-Markov Processes

Ayele Taye Goshu[1*] and Zelalem Getahun Dessie[2]
[1]*Hawassa University and* [2]*Bahir Dar University*

Abstract: The aim of this study is to model the progression of HIV/AIDS disease of an individual patient under ART follow-up using semi-Markov processes. Recorded hospital data were obtained for a cohort of 710 patients at Felege-Hiwot referral hospital, Ethiopia, who have been under ART follow-up from June 2005 to August 2009. States of the Markov process are defined by the seriousness of the sickness based on the CD4 counts in cells/microliter. The five states considered are: state one (CD4 count > 500); state two ($350 <$ CD4 count ≤ 500); state three ($200 <$ CD4 count ≤ 350); state four (CD4 count ≤ 200); and state five (Death). The first four states are named as good or alive states. The findings obtained from the current study are as follows: within the good states, the transition probability from a given state to the next worse state increases with time, gets optimum at a time and then decreases with increasing time. This means that there is some period of time when such probability is highest for a patient to transit to a worse state of the disease. Moreover, the probability of dying decreases with increasing CD4 counts over time. For an HIV/AIDS patient in a specific state of the disease, the probability of being in same state decreases over time. Within the good states, the results show that probability of being in a better state is non-zero, but less than the probability of being in worse state. At any time of the process, there is more likely to be in worse state than to be in better one. The conditional probability of staying in same state until a given number of month decreases with increasing time. The reliability analysis also revealed that the survival probabilities are all declining over time. This implies that patient conditions should be improved with ART to improve the survival probability.

Key words: AIDS progression, Markov chain, stochastic, survival.

1. Introduction

Infection by the human immunodeficiency virus (HIV) gradually evolves to the acquired immune deficiency syndrome (AIDS), and AIDS evolves to death

*Corresponding author.

if not handled carefully. One may consider this progression of HIV infection to AIDS and then to death as a stochastic process. By splitting the progression into various states of the disease based on the immunological indicators namely CD4+ count including death as one state (Janssen and Monica, 2001).

The HIV/AIDS patients are being treated by the drug antiretroviral therapy (ART). It is found to reduce mortality and improve quality of life of the patients. The effect, however, varies from country to country (Braitstein *et al.*, 2006). Egger (2007) indicated that several predictors of mortality for HIV/AIDS patients though on ART. These include CD4 count, viral load, total lymphocytes, body mass index and adherence.

The probability that an HIV/AIDS patient transition from one state to another depends on how long he/she has spent in that state. As time spent in each stage of the disease can't be predictable on the basis of clinical and immunological measures, this needs to be modeled by the semi-Markov stochastic process (Viladent and Ackere, 2007; Giuseppe *et al.*, 2007). A stochastic process that can be in any one of possible states and that each time it enters a state it remains there for a random amount of time and then makes a transition into another state with some probability defines semi-Markov process (Ross, 2007). A special case is that if the amount of time that the process spends in each state before making a transition is identically a unit, the semi-Markov process becomes just a Markov chain.

Numerical analyses of the homogeneous semi-Markov process are dealt by Corradi *et al.* (2004) and Janssen and Monica (2001). Other more readings include D'Amico *et al.* (2009), Davidov and Zelen (2000), and Satten and Sternberg (1999).

In this paper, we present the results of modelling of the progression of HIV/AIDS so as to predict the future clinical state and survival probability of a patient. These are: the conditional probability that an HIV/AIDS patient given that he/she is in a known state of the disease, after a period of time, be in the next subsequent worse state of the disease; the conditional probability that a patient is staying in the same disease state until a specific time t; and the probability that an HIV/AIDS patient survives for a specific time given his/her starting state of the disease.

2. Homogenous Semi-Markov Process Model

Homogenous semi-Markov process (HSMP) was initiated in the 1950s (Levy, 1954). Some of the recent studies on the semi-Markov processes include Corradi *et al.* (2004) and Giuseppe *et al.* (2007). Derivation of the HSMP model is given in the Appendix. In this paper, a computer program for solving the evolution equations is developed in the R statistical software version 2.6.2.

3. Data

In the current paper, the data on CD4 counts of HIV/AIDS patients on follow-up during 2005-2009 were obtained from Felege-Hiwot referral hospital at Bahir Dar, Ethiopia. We identify four states of the Markov process of the seriousness of HIV/AIDS sickness based on the CD4 counts of a patient as in Giuseppe *et al.* (2007). The states are defined as:

SI : CD4 count $>$ 500 cells/microliter.
SII : 350 $<$ CD4 count $\leq$ 500 cells/microliter.
SIII : 200 $<$ CD4 count $\leq$ 350 cells/microliter.
SIV : CD4 count $\leq$ 200 cells/microliter.
D : Death.

Among the states of the semi-Markov process, the death state D is considered to be an absorbing state − meaning that once a patient is in the death state she/he will never be in the others states and rather stays there forever. The state D is categorized as "bad" and the others SI, SII, SIII and SIV as "good" states.

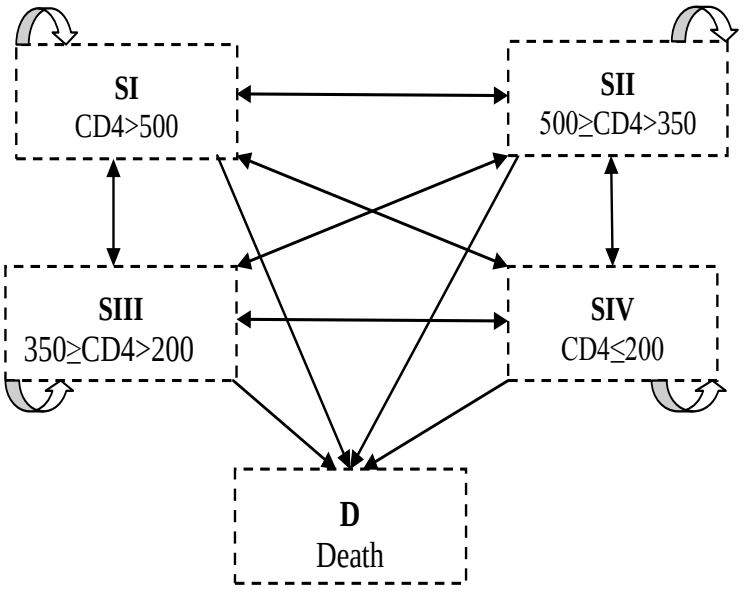

Figure 1: Communication between the states of the process

Frequencies and estimated transition probabilities of between the states are summarized from the data and displayed in Table 1. The solutions for the transition probabilities $\phi_{ij}(t)$ at time t using the algorithm are obtained with $m = 5$ states, $T = 200$ months, transition probability matrix $\boldsymbol{P}$ as given in Table 1.

Table 1: Frequencies and probabilities of the transitions of the states of the process

State	SI	SII	SIII	SIV	D
SI	244 (0.6025)	92 (0.2272)	35 (0.0864)	25 (0.0617)	9 (0.0222)
SII	108 (0.1949)	260 (0.4693)	112 (0.2022)	55 (0.0993)	19 (0.0343)
SIII	32 (0.0531)	103 (0.1708)	299 (0.4959)	143 (0.2371)	26 (0.0431)
SIV	12 (0.0211)	54 (0.0951)	99 (0.1743)	343 (0.6039)	60 (0.1056)
D	0	0	0	0	114 (1.0)

4. Results and Discussion

Results of the modelling are displayed in Figures 2-5.

First, transitions within the "good" sates are considered. The conditional probability that an HIV/AIDS patient who is currently in a given state $i \in \{\text{SI, SII, SIII}\}$ will be in the subsequent "worse" state after t months is displayed in Figure 2(a). Such progressions are from SI to SII, SII to SIII and SIII to SIV. Each plot is parabolic curve with optimal/peak points (42, 0.27), (60, 0.24) and (65, 0.27) in the time-probability axis. The peaks may indicate there is time when a patient will be at highest risk of being at worse state. Moreover, the transition probability from SII to SIII is the lowest as compared to the others. It is interesting to find out that, within the good states, thetransition probability from a given state to the next worse state increases with time, gets optimum at a time and then decreases with increasing time.

Second, transitions to the "bad" or death state are considered. The conditional probability that an HIV/AIDS patient who is currently in a given state $i \in \{\text{SI, SII, SIII, SIV}\}$ will be in the "bad or death state after t months is displayed in Figure 2(b). These are from SI to D, SII to D, and SIII to D and SIV to D. The probability of dying after 200 months is 0.39 for a patient who is in the first stage, 0.44 for one who is in the second stage, 0.48 for one who is in the third stage and 0.54 for one who is in the fourth stage of the disease. Each plot is an increasing parabolic curve over time with no optimal/peak point. This can be interpreted as the probability that an HIV/AIDS patient with any one of the good states will be in 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 that of one who is in first state is the lowest probability throughout the time.

Third, the conditional probability of a patient making changes in disease states given his/her current status is computed and displayed in Figure 3. The results show that the probabilities of being in state $j \in \{\text{SI, SII, SIII, SIV, D}\}$

(a) (b)

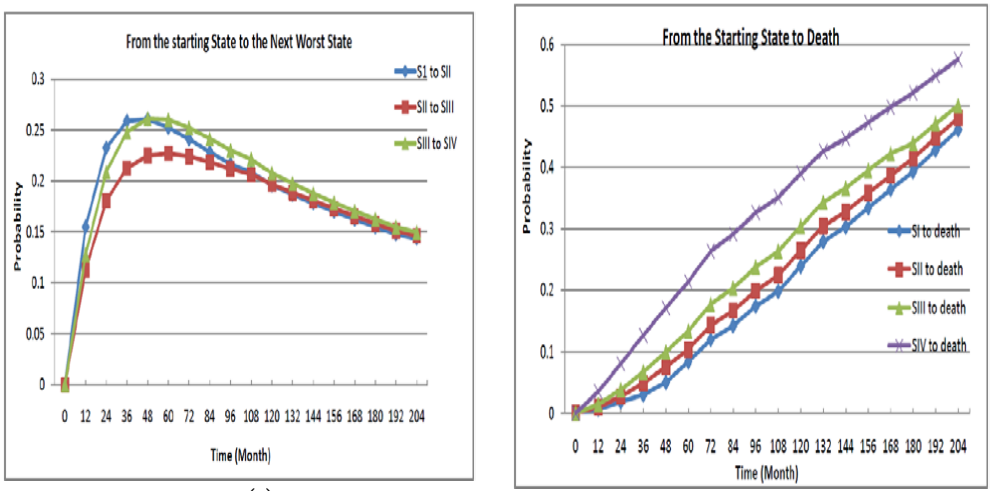

Figure 2: Conditional probabilities that a patient will be in state j after t months given that she/he is currently in state i (a) $i \in$ {SI, SII, SIII}, $j \in$ {SII, SIII, SIV} (b) $i \in$ {SI, SII, SIII, SIV}, $j \in$ {D}

after a month t given that he/she entered at time 0 in state $i \in$ {SI, SII, SIII, SIV}. The results plotted can be interpreted as follows. For an HIV/AIDS patient in a specific state of the disease, the probability of being in same state decreases over time. With the good or alive states, the results show that probability of being in a better state is non-zero, but less than the probability of being in worst states. That is, for a patient there is more likely to be in worse state than to be in better one, relatively speaking.

Fourth, the probability of staying in same state is computed. See Figure 4. The conditional probability that a patient stays in state one, two, three and four for at least 24 months are 0.14, 0.19, 0.21 and 0.24 respectively. It is increasing with increasing seriousness of the disease. Within the good states, it is more likely for a patient to stay in a worse state than in a better one. Of course, the death state is an absorbing state, i.e., once a patient enters the death state he/she stays in same state forever. It is also interesting to find out that the conditional probability of staying in same state until a given number of month decreases with increasing time. So there is a possibility of changing from one state to another which is a non-zero probability.

Fifth, the survival probability is computed. See Figure 5. The survival probability that a patient who is at the first state of the disease will be alive until 120 months is about 0.78, while that of one who is in the fourth state of the disease about 0.62. Thus a patient in first state is better survivor than that who is in the fourth state – of course second and third states also. Other results in the figure reveal that the survival probabilities are all decreasing with increasing time.

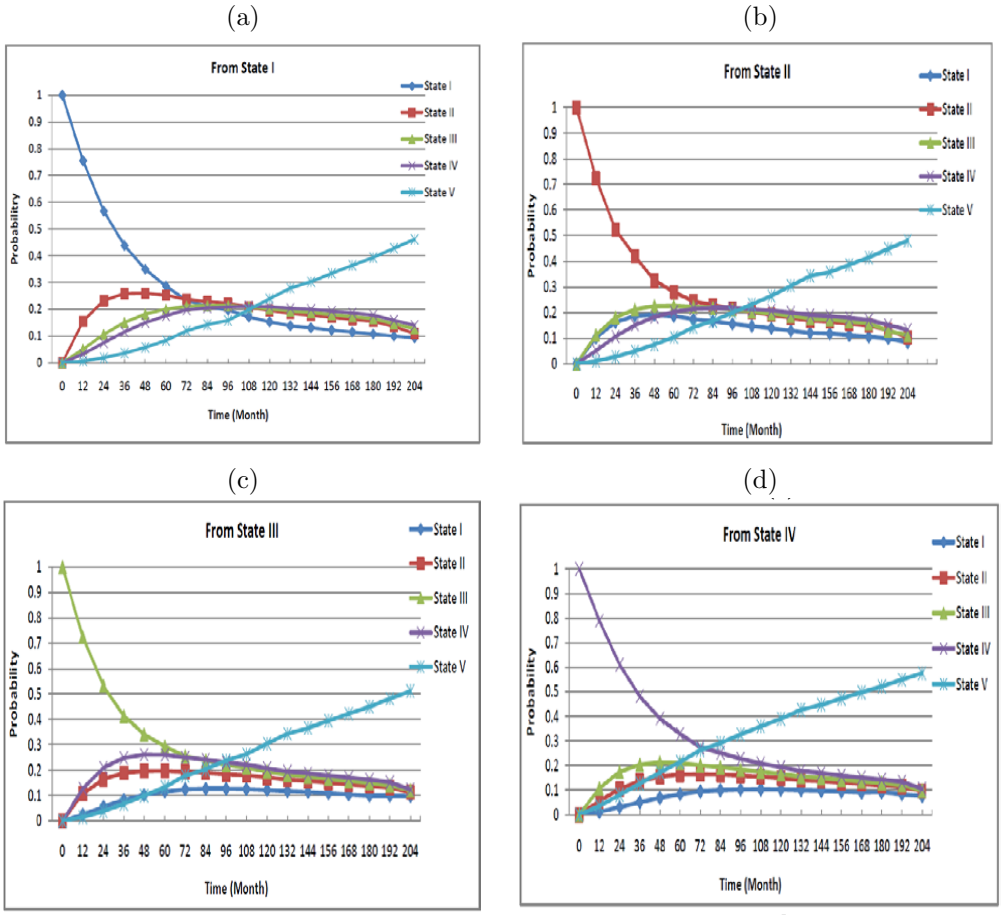

Figure 3: Conditional probabilities of being in next state j after a month t given the starting state i. (a) starting state SI, (b) starting state SII, (c) starting state SIII, (d) starting state SIV

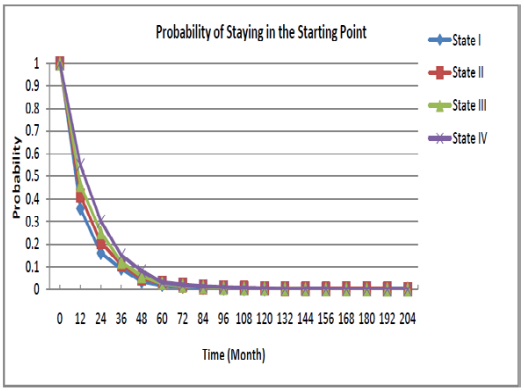

Figure 4: The probability that a patient stays in same state of disease for at least t months

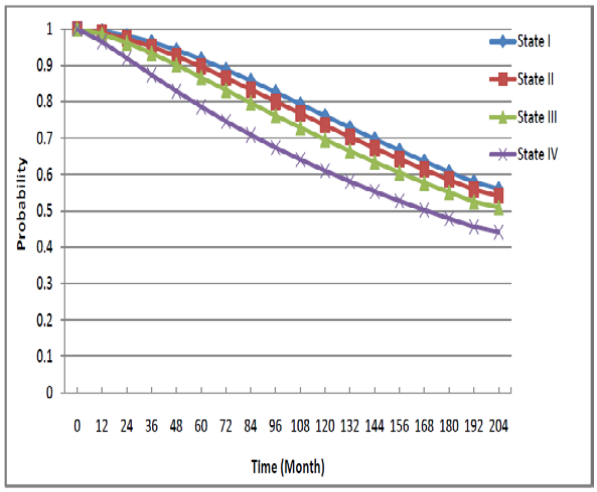

Figure 5: Reliability or survival probability of a patient at t months with known starting state

5. Conclusions and Recommendations

The semi-Markov process model is applied to capture the AIDS dynamic progression of a patient. The model considers the randomness of the time that a patient spends in a given state of the disease. The following can be concluded from this study.

Within the good states, the transition probability from a given state to the next worse state increases with time, gets optimum at a time and then decreases with increasing time. This means that there is some period of time when such probability is highest for a patient to transit to a worse state of the disease. Intervention for patient care may minimize such effect.

The probability that an HIV/AIDS patient with any one of the good states will be in death state is increasing with time, irrespective of the current state of the patient. Being in the fourth state (low CD counts) leads to the highest probability of dying at a specific time as compared to the other states. More generally, the probability of dying decreases with increasing CD4 counts over time. For an HIV/AIDS patient in a specific state of the disease, the probability of being in same state decreases over time. Within the good states, the results show that probability of being in a better state is non-zero, but less than the probability of being in worse state. At any time of the process, there is more likely to be in worse state than to be in better one.

Within the good states, it is more likely for a patient to stay in a worse state than in a better one at any time of the process. The death state is an absorbing state, i.e., once a patient is death state he/she stays in same state forever. The

conditional probability of staying in same state until a given number of month decreases with increasing time. So there is a possibility of changing from one state to another with non-zero probability. The reliability analysis indicates that the survival probabilities are all decreasing with increasing time, implying that patient conditions should be improved to maintain the survival probability as high as possible.

In general, the survival probability of an HIV/AIDS patient depends on his/her current state of the disease in such a way that the lower CD4 counts the higher is the risk to be in worse health state or death state. The dynamic nature of the AIDS progression is confirmed with particular findings that there is more likely to be in worse state than better one unless interventions are made. It is recommendable to keep up the ongoing ART treatment services in most effective ways with the careful considerations of recent disease status of patients.

Appendix: Derivation of the HSMP Model

Giuseppe *et al.* (2007) defines homogenous semi-Markov process (HSMP) model as follows:

Let $X_n : \Omega \to S$ be the stochastic process with state space $S = \{S_1, S_2, \cdots, S_m\}$ and $T_n : \Omega \to \Re$ be the time of the n^{th} transition, with Ω domain of the process and $\Re$ set real numbers. Here the time is a random variable.

The kernel $Q = [Q_{ij}]$ associated with the process and the transition probability P_{ij} of the embedded Markov chain is defined as follows:

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

$$P_{ij} = \lim_{t\to\infty} Q_{ij}(t). \tag{2}$$

Define the probability that the process will leave a state i in a time t as

$$H_i(t) = P[T_{n+1} - T_n \leq t | X_n = i] = \sum_{j=1}^{m} Q_{ij}(t). \tag{3}$$

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

$$G_{ij}(t) = P[T_{n+1} - T_n \leq t \,|\, X_n = i, X_{n+1} = j], \tag{4}$$

which can be computed as:

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

For any homogeneous semi-Markov process$\{X(t),\ t \geq 0\}$, the transition probabilities are given by (6) for which the solutions should be obtained using the progression (7).

$$\phi_{ij}(t) = P[X(t) = j | X(0) = i], \tag{6}$$

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

Here δ_{ij} represents the Kronecker delta δ. An approximate solution of (7) can be obtained using the general numerical integration formula given in Corradi *et al.* (2004). In the same paper, they proved that the numerical solution of the process converges to the discrete time HSMP described as an infinite countable linear system:

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

where h stands for the step measure of the approximation and

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

$$v_{ij}^h(kh) = \begin{cases} 0, & \text{if } k \neq 0, \\ Q_{ij}^h(kh) - Q_{ij}^h((k-1)h), & \text{if } i = j. \end{cases}$$

$$\Rightarrow \boldsymbol{\Phi}^h(kh) - \sum_{\tau=1}^{k} V(\tau h)\boldsymbol{\Phi}^h((k - \tau)h) = \boldsymbol{D}^h(kh). \tag{9}$$

The fact that the matrix $\boldsymbol{\Phi}^h(kh)$ is stochastic is already proved in Corradi *et al.* (2004) and Janssen and Monica (2001).

For solving the progression equation, Corradi *et al.* (2004) proposed the following algorithm with suggested matrix form:

$$\boldsymbol{V}^T\boldsymbol{\Phi}^T = \boldsymbol{D}^T. \tag{10}$$

The variables involved are the following:

m= number of states of HSMP, which 5 in this case.

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

$\boldsymbol{P}$ = matrix of order m of the embedded Markov chain in HSMP.

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

$\boldsymbol{Q}^T$ = kernel of SMP.

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

$\boldsymbol{D}^T$ = block vector of order $T+1$ the block which are the diagonal square matrix of order m.

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

$\boldsymbol{S}^T$ = block vector of order $T+1$ the block which are the diagonal square matrix of order m. The diagonal element of each block t are $s_{ii} = \sum_{j=1}^{m} Q_{ij}(t)$.

Given an epoch T is fixed, matrices $\boldsymbol{G}$ and $\boldsymbol{P}$, the algorithm solves the linear system (10) for the unknown matrix $\boldsymbol{\Phi}^T$ by means of a purely iterative procedure.

The algorithm is:

(i) Read the inputs: m, T, $\boldsymbol{P}$, $\boldsymbol{G}^T$

(ii) Construct: $\boldsymbol{Q}^T$, $\boldsymbol{V}^T$, $\boldsymbol{D}^T$

$$
\begin{aligned}
&V_{(0)} = I;\ \ Q_{(0)} = 0;\ \ S_{(0)} = 0;\ \ D_{(0)} = I\\
&\quad for\ \ t = 1\ \ to\ \ T\\
&\qquad Q_{(t)} = P * G_{(t)}\\
&\qquad for\ \ i = 1\ \ to\ \ m\\
&\qquad\quad s_{ii}(t) = Q_{i^*}(t) \bullet \mathbf{1}\\
&\qquad end\ \ for\\
&\qquad V_{(t)} = Q_{(t)} - Q_{t-1}\\
&\qquad D_{(t)} = D_{(0)} - S_{(t)}\\
&\quad end\ \ for
\end{aligned}
$$

(iii) Given $\Phi_{(0)} = D_{(0)}$, solve for Φ^T

$$
\begin{aligned}
&for\ \ t = 1\ \ to\ \ T\\
&\quad \Phi_{(t)} = D_{(t)}\\
&\quad for\ \ s = 1\ \ to\ \ t\\
&\qquad \Phi_{(t)} = \Phi_{(t)} + V_{(s)} \bullet \Phi_{(t-s)}\\
&\quad end\ \ for\\
&end\ \ for
\end{aligned}
$$

(iv) Print the results, $\boldsymbol{\Phi}^T$, $\boldsymbol{Q}^T$

In the algorithm, the symbol $(\bullet)$ represents row column matrix product while $(*)$ represents element by element product.

Furthermore, for the transition matrix $\boldsymbol{P}$ of the embedded MC in HSMP we used $\hat{P}_{ij} = n_{ij}/n_i$, where n_{ij} is the number of transition in state i given the state j, and n_i is the number of observed elements in state i. It is assumed that process spends some time in a given state and random time has distribution $\boldsymbol{G}$, exponential distribution $G_{ij}(t) = 1 - \exp(-\lambda_{ij}t)$, where λ_{ij} is expected time the process spends in state i before it enters state j from i.

In predicting the survival probability of a patient, let us first group the states of the process into two sets A and B, where A contains all "good" states in which the patient is alive and set B contains all "bad" states in which the patient is not alive. Then survival probability or reliability function R_i of a patient by the time t is given as:

$$R_i(t) = \sum_{j \in A} \hat{\phi}_{ij}(t). \tag{11}$$

This represents the probability that a patient who is currently in any good state $i \in A$ is alive until time t; while the probability that the patient is not alive is $1 - R_i$.

The algorithm is programmed in the R statistical software version 2.6.2.

References

Braitstein, P., Brinkhof, M. W., Dabis, F., Schechter, M., Boulle, A., Miotti, P. *et al.* (2006). Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* **367**, 817-824.

Corradi, G., Janssen, R. and 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. and 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* (Edited by L. Sakalauskas and C. Skiadas), 473-476.

Davidov, O. and Zelen, M. (2000). Designing cancer prevention trials: a stochastic model approach. *Statistics in Medicine* **2**, 1983-1995.

Egger, M. (2007). Outcomes of antiretroviral treatment in resource limited and industrialized countries. In *Proceedings of the 14th Conference on Retroviruses and Opportunistic Infections*, February 25-28. Los Angeles.

Giuseppe, D., D'Amico, G., Girolamo, A., Janssen, J., Iacobelli, S., Tinari, N. and Manca, R. (2007). A stochastic model for the HIV/AIDS dynamic progression. *Mathematical Problems in Engineering*, Volume 2007, Article ID 65636, 14 pages.

Janssen, J. and Monica, N. (2001). Numerical solution of non-homogenous semi-Markov processes in transient case. *Methodology and Computing in Applied Probability* **3**, 271-279.

Levy, P. (1954). Processus semi-Markoviens. In *Proceedings of the International Congress of Mathematicians*, Amsterdam, Volume III, 416-426. N. V. Erven P. Noordhoff, Groningen.

Ross, S. M. (2007). *Introduction to Probability Models*, 9th edition. Wiley, New York.

Satten, G. A. and Sternberg, M. R. (1999). Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics* **55**, 507-513.

Viladent, C. and van Ackere, A. (2007). HIV/AIDS modeling, a two-angle-retrospective. Toward generic deterministic model for pattern II countries? University of Lausanne, Institute of Research in Management, Switzerland.

Received July 21, 2012; accepted November 7, 2012.

Ayele Taye Goshu
School of Mathematical and Statistical Sciences
Hawassa University
P.O. Box 05, Hawassa, Ethiopia
ayele_taye@yahoo.com

Zelalem Getahun Dessie
Department of Statistics
Bahir Dar University
P.O. Box 79. Bahir Dar, Ethiopia
zelalem_getahune@yahoo.com