Estimation of Cure Fraction and Misclassification Probabilities for HIV/AIDS Patients Under ART Using Continuous Time Hidden Markov Model

Gurprit Grover[1], Arpan Kumar Thakur[1,*], Sejong Bae [2]

[1]Department of Statistics, University of Delhi, Delhi- 110007, India
[2]Division of Preventive Medicine, University of Alabama, Birmingham.

Abstract

The central thrust of this paper is to accentuate the impact of Anti-Retroviral Therapy (ART) on cure rate of HIV/AIDS patients and on the transition intensities between the stages of disease using cure rate model and Hidden Markov model (HMM) respectively. Hidden Markov Model (HMM) is a captivating algorithm for temporal pattern recognition like automated speech, handwriting and gesture recognition in the signal processing field. Although it is based on Markov processes which are more widely used in estimating the transition rates between the different stages of a disease, but HMM is hardly being used in survival data modeling.

Keywords: AIDS, CD4, Cure Rate Model, Hidden Markov Model.

---

*Corresponding author: Arpan Kumar Thakur

Abstract

The central thrust of this paper is to accentuate the impact of Anti-Retroviral therapy (ART) on cure rate of HIV/AIDS patients and on the transition intensities between the stages of disease using cure rate model and Hidden Markov model (HMM) respectively. Hidden Markov Model (HMM) is a captivating algorithm for temporal pattern recognition like automated speech, handwriting and gesture recognition in the signal processing field. Although it is based on Markov processes which are more widely used in estimating the transition rates between the different stages of a disease, but HMM is hardly being used in survival data modeling.

Keywords: AIDS, CD4, Cure Rate Model, Hidden Markov Model.

---

1. Introduction

Human Immunodeficiency Virus (HIV) is a kind of virus that ushers and leads to Acquired Immune Deficiency Syndrome (AIDS). HIV taints a particular type of white blood cells, known as T- cells (or CD4+ T-cells), that helps in fighting diseases. As time passes, HIV kills CD4+ T- cells and multiplying itself, that leads to weakening of the immune system. In due course of time, the infected person's immune system can no longer fight off diseases. So, proper measurement of CD4+ T cell count may be viewed as the snapshot of how good a patient's immune system is functioning.

Till date, there is no vaccine that can claim of curing HIV/AIDS. Although, a medication called antiretroviral (ARV) drug can steady the deteriorating immune system. The initiation of ARV drug is generally based on two clinical observations, one is CD4+ T cell count and another is viral load (HIV RNA concentrations) that measures HIV in the blood, lower is better. The purpose of the ARV drug is to make viral load undetectable and if it is able to do so, then

infected person can't transmit HIV to partner [Veterans' Health Administration]. According to WHO guidelines also, initiation of ARV drug and for measuring disease progression, viral load should be preferred over the CD4+ cell count.

But, in India due to scanty of resources, the decision about the commencement of treatment and disease progression is taken merely based on CD+ cell count. In spite of the fact that, national AIDS control organization (NACO) issued new guidelines that mandated to "*treat all persons living with HIV (PLHIV) with antiretroviral therapy regardless of CD4+ cell count, clinical stage, age or population*" [NACO on May, 2017], CD4+ cell count play an indispensable role in entire treatment protocol.

To study the transmission of the virus to next-generation Bature et al. (2010) used a Markov chain model. The same model has been used for observing disease progression in liver cancer Kay et al. (1986), for Hepatitis C disease progression Sweeting et al. (2010), for tuberculosis (TB) progression Debanne et al. (2000), Alzheimer's disease Commenge et al. (2004), liver-cirrhosis progression Grover et al. (2014). Discretized Markov model has been developed and employed to AIDS prediction in England and Wales, Aalen et al. (2018), Grover et al. (2013) used Markov model to study disease progression among HIV/AIDS patients.

New and ameliorated statistical methods are always entailed for making decisions about initiation and switching treatment protocols. Nevertheless, antecedent studies have appropriately modeled disease progression using multistate Markov processes, very few have explored the aptness of the hidden Markov model.

The aftermath of lung transplantation is studied by Jackson and Sharples (2002), Guihenneuc-Jouyaux et al. (2000) used a Bayesian hierarchical model for hidden Markov processes by exemplifying HIV infected patient's data. On the contrary to the simple Markov model, where

the state is directly observable, in HMM the true state is not directly visible (that's what name hidden symbolizes). The HMM canvasses to recuperate the true sequence of states from the visible (observed) sequence of states It has a plethora of applications in speech recognition, in part of speech tagging, in object tracking, in computational molecular biology. HMM in one sense may be treated as an artefact in the sense that it has developed way back in late 1960's by Baum & Petrie (1966) but it's use is now ubiquitous in science including survival analysis.

In India, ART centers are compelled to use CD4+ T cell count instead of the viral load while staging the HIV patients. This may lead to a mismatch in staging additionally measurement of CD4+ cell count itself is prone to error mainly due to intraindividual variability and to some extent due to measurement error. In this paper an attempt has been made to underline the mismatch using HMM.

The paper is organized as follows: in next section 2, a short explanation of material and method to be used is given. In section 3, results are provided followed by section 4 where discussions, limitations, future ambits and pipelined research is presented.

2. Material and Methods

2.1 Materials

It is a longitudinal retrospective follow-up study of HIV/AIDS patients undergoing treatment at ART center of Dr. Ram Manohar Lohia hospital in New Delhi, during the period April 2004 to December 2014. Exclusion criteria were the age at enrollment should be >= 18 years, should have baseline CD4+ cell count available, periodic CD4+ cell count available for at least two visits. By filtering using complete case analysis on variables like sex, smoking and alcohol consumption status, treatment (virocomb-N combination and others), we are left with only 1063 observations.

2.2 Methods

2.2.1 Cure Fraction Model

Assume that C be the probability of an HIV patient being a long-term survivor and $(1 - C)$ be the probability of a patient being susceptible to death (Stage 5 of the disease). Then, Berkson et al. (1952) defined the survival function at any time t as:

$$S(t) = C + (1-C) * S_u(t) \qquad (1)$$

where, $Su(t)$ is the survival function of the susceptible population which may be assumed to follow some life time distribution. Probability density function $f(t)$ of the overall population is written as

$$f(t) = (1 - C) * fu(t) \qquad (2)$$

where $fu(t)$ is the probability density function of susceptible population.

Now let (t_i, δ_i) be the observed data of size n , where t_i is the survival time of the i^{th} patient and δ_i is censoring indicator variable which is defined as follows: $\delta_i = 0$ for right-censored observation and $\delta_i = 1$ for uncensored observation (i = 1, 2, . . , n).

Accordingly, the individual patient's contribution to the likelihood function can be written as

$$L_i = [f(t_i)]^{\delta i} \ [S(t_i)]^{(1-\delta i)}$$

$$= [(1 - C) f_u(t_i)]^{\delta i} \ [C + (1 - C) S_u(t_i)]^{(1-\delta i)} \qquad (3)$$

So, complete likelihood is given by

$$L = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} [(1-C) f_u(t_i)]^{\delta_i} [C + (1-C) S_u(t_i)]^{(1-\delta_i)} \quad (4)$$

Parameters are estimated by maximizing the complete data likelihood in equation (4) using WinBUGS software package using Gibbs sampling approach. Here we have used various lifetime distributions like exponential, Weibull, gamma, exponentiated Weibull etc., based on least deviance information criteria (DIC) value we found exponentiated Weibull distribution to the best model. For detailed review of the foregoing model one may refers to Farewell (1982), Yamaguchi (1992), Maller and Zhou (1995), Chen et al. (1999), Peng and Dear (2000), and Sy and Taylor (2000), Kannan et al. (2010), Achcar et al. (2012), Swain et al. (2016).

2.2.2 Hidden Markov Model

Before applying HMM, we have used a time-homogenous multistate Markov model to study disease progression among HIV/AIDS patients. For this purpose, stages of HIV/AIDS patients have been defined in terms of CD4+ cell count as:

Stage/State	1	2	3	4	5
CD4+ cell count range	>500	351-500	200-350	<200	Death

It is well established that ARV drugs improve the CD4+ cell count in most of the cases, but unfortunately for some patients, it might not do so, that results in deterioration of health. That is, the patients may move from a lower stage to higher stages of the disease, a significant proportion of patients move to end-stage, i.e. death stage too. So, backward progression/transition is also a possibility. Consequently we used reversible transition model that is depicted in **Figure 1**. Except for stage 5, which is absorbing stage all other stages are transient in nature.

With the passage of time, a patient may move in possible state space S={1,2,3,4,5}. Let X(t) = r be the current state of the patient, then the transition intensity λ_{rs} of advancing to state s in infinitesimal time δt is given by

$$\lambda_{rs} = \lim_{\delta t \to 0} \frac{P(X(t+\delta t) = s \mid X(t) = r)}{\delta t}.$$

Then the transition intensity matrix Q can be written as $Q = [\lambda_{rs}]_{r,s \in S}$ and possess the following two properties (a) $\sum_{s \in S} \lambda_{rs} = 0$ for all r and (b) $\lambda_{rs} = -\sum_{r \neq s} \lambda_{rs}$.

The maximum likelihood estimation technique developed by Kalbfleish and Lawless (1986) can be used to estimate the transition intensities λ_{rs}. Estimated transition intensities in turn can be used to find the transition probability matrix $P(t) = [P_{rs}(t)]_{r,s \in S}$ and $P_{rs}(t)$ is defined as:

$$P_{rs}(t) = P(X(t+v) = s / X(t) = r)$$

Also, Cox and Miller (1965) defined transition probability matrix with the help of the intensity matrix as a Kolmogorov equation $P(t) = e^{tQ}$. Similarly, mean sojourn time, that is the time of stay in any transient state, is given by - $1/\lambda_{rr}$. Let us denote covariates vector as **Z**, then the effect of covariates on transition intensity can be modeled by $q_{ij}(t)$, and defined in terms of Cox- proportional hazard regression as suggested by Marshall and Jones (1995):

$$q_{ij}(t) = q_{ij}(0) e^{\beta_{ij}^T Z}$$

Here $q_{ij}(0)$, is the baseline intensity, β_{ij} is the coefficient of regression. Here it is assumed that covariates are time independent. Estimates can be obtained using the maximum likelihood procedure suggested by Kalbfleish and Lawless (1986).

A hidden Markov model is generally used for defining a probability distribution over a sequence of observations. For brief elucidation, consider the observation at time t by the variable X_{it}. It is presumed that t is an integer-valued index. Additionally, it is based on two assumptions: (i) the observations at time t is fostered by some process that is hidden from the observer and generated by misclassification matrix, (ii) it is also assumed that hidden state follows the Markov property with transition matrix Q, put in another way current state envelopes all information that is required to know about the historicity of the process to predict the subsequent future of the process, Ghahramani (2001), this intricate relationship for HMM is given in **Figure 2**. Generalized regressions can be used to model the covariates effect on transition intensity and misclassification probabilities.

For mathematical formulation of the HMM, let $X_{iT} = [X_{i1},...,X_{iT_i}]$ denotes the observed state that triggered by the hidden state S_{it}. The observed states X_{it} are assumed to be conditionally independent of true hidden states. The likelihood contribution for patient i is given by

$$L_i = f(X_{i1},...,X_{iT_i})$$

$$= \sum_{S_i} f(X_{i1},...,X_{iT_i} \mid S_{i1,}.....S_{iT_i}) f(S_{i1,}.....S_{iT_i})$$

Given the values of the underlying Hidden state, observed states are conditionally independent, using, Markovian property of Hidden states

$$P(S_{ij} / S_{i,j-1},....,S_{i1}) = P(S_{ij} / S_{i,j-1})$$

The resulting likelihood can be rewritten as,

$$= \sum_{S_i} \prod_{t_i=t_1}^{T_i} f(X_{it_i} \mid S_{it_i)}) \left\{ f(X_{i1}) \prod_{t_i=t_2}^{T_i} f(X_{it_i} \mid S_{it_i-1}) \right\}$$

In HMM, for the observable state X_{t_i} are conditionally emitted by hidden states S_{t_i} through misclassification matrix $M = [e_{rs}]_{r,s \in S}$, whose elements are defined by

$$e_{rs} = P\{X_{t_i} = s / S_{t_i} = r\}, \; r, s \in S$$

An assumption about disease stages is that a stage can be misclassified only to the adjacent disease stage. By employing the Viterbi algorithm technique, we can recreate the optimal sequence in HMM using dynamic programming algorithm. It was disseminated by Viterbi (1967), but more elaborate elucidation was given by Bellman (1957).

3. Results and Discussions

The progression of disease stages in HIV/AIDS patients are given in **Table 1**. Diagonal entries in the table is the number of times a patient remains in the same stage. The number 19 signify that number of occasions where patient of stage 1 moves to stage 2. Likewise, there are 5,12, 22 and 35 number of cases of reaching end stage 5 from stage1, stage 2, stage 3 and stage 4 respectively.

Table 1: Number of state transitions

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	130	19	7	1	5
Stage 2	131	128	65	7	12
Stage 3	75	251	314	64	22
Stage 4	28	133	484	363	35

The estimated parameters of cure rate model have been presented in **Table 2**. Here stages are observed after one year of initiation of ARV drug. Following table shows that patients who are in stage 1 have **86%** chance of being long-term survivors, and chances are shrinking with severity of the disease. Patients who are in stage 4 even after one year of treatment have comparatively less chance (only **58%**) of being long-term survivors.

Table 2: Estimated cure rate model parameters

		Mean	S.D.	MC- error
Stage 1	C	0.862	0.0587	0.05011
	α	4.85E-03	0.003741	6.57E-04
	β	0.06538	0.0995	0.00113
	γ	1.547	0.1095	0.0221
Stage 2	C	0.724	0.0418	0.00735
	α	5.74E-03	0.00411	2.51E-04
	β	0.00856	0.0997	0.001306
	γ	1.632	0.1014	0.0113
Stage 3	C	0.657	0.0156	0.00815
	α	6.85E-03	0.00412	5.27E-04
	β	0.006449	0.01317	0.001614
	γ	1.0546	0.2514	0.01822
Stage 4	C	0.587	0.0248	0.00139
	α	7.54E-03	0.00417	4.28E-04
	β	0.009324	0.0243	0.000908
	γ	0.693	0.168	0.099

Table 3: Estimated transition intensities with 95% confidence interval

From	To	Intensity	C.I.
Stage 1	Stage 1	-0.5306	(-0.759,-0.371)
Stage 1	Stage 2	0.303	(0.249, 0.730)
Stage 1	Stage 3	0.14	(0.023, 0.3621)
Stage 1	Stage 4	0.09	(0.01, 0.1625)
Stage 1	Stage 5	1.32E-06	(0, 8.035e+39)
Stage 2	Stage 1	0.983	(0.734, 1.158)
Stage 2	Stage 2	-1.94	(-2.240,-1.371)
Stage 2	Stage 3	0.553	(0.335, 1.210)
Stage 2	Stage 4	0.405	(0.272, 1.116)
Stage 2	Stage 5	1.98E-05	(0, 2.920e+16)
Stage 3	Stage 1	0.33	(0.234, 0.621)
Stage 3	Stage 2	0.841	(.603, 1.331)
Stage 3	Stage 3	-1.64	(-1.837,-1.456)

Stage 3	Stage 4	0.462	(0.3604, 0.6001)
Stage 3	Stage 5	8.68E-03	(0.00067,0. 1115)
Stage 4	Stage 1	0.27	(0.13502, 0.3402)
Stage 4	Stage 2	0.7504	(0.613, 1.712)
Stage 4	Stage 3	0.716	(0.571, 1.966)
Stage 4	Stage 4	-1.76	(-1.966,-1.571)
Stage 4	Stage 5	2.63E-02	(0.0059, 0.118)

Table 3 presents the intensity of disease progression in the absence of prognostic factors. Patients of stage 3 are 1.82 times (0.841/0.462) more likely to move to less severe disease stage 1 than moving to severe stage 4. Similarly, the patients of stage 4 are 27.2 times (0.716/0.0263) more likely to move to stage 3 than moving to death stage 5.

Table 4: Mean Sojourn Times at Different Stages

	Estimates (Std. error)	**95 % C.I.**
Stage 1	1.884 (0.343)	(1.318,2.694)
Stage 2	0.517 (0.038)	(0.446,0.598)
Stage 3	0.812 (0.036)	(0.544,0.987)
Stage 4	0.769 (0.032)	(0.508,0.963)

From **Table 4** it can be observed that on an average a patient elapsed 1.88 years in stage 1, and 0.517 years, 0.812 years, 0.769 years in stage 2, stage 3 and stage 4 respectively.

Table 5: Estimated transition intensities and misclassification probabilities for misclassification model

From	**To**	**Intensity**	**Probability**	
Stage 1	Stage 1	-0.517	e_{11}	0.894
Stage 1	Stage 2	0.233	e_{12}	0.106
Stage 1	Stage 3	0.15		
Stage 1	Stage 4	0.09		
Stage 1	Stage 5	0.046		
Stage 2	Stage 1	0.933	e_{21}	0.106

Stage 2	Stage 2	-1.845	e_{22}	0.834
Stage 2	Stage 3	0.514	e_{23}	0.06
Stage 2	Stage 4	0.382		
Stage 2	Stage 5	8.28E-03		
Stage 3	Stage 1	0.232		
Stage 3	Stage 2	0.625	e_{32}	0.152
Stage 3	Stage 3	-1.223	e_{33}	0.743
Stage 3	Stage 4	0.366	e_{34}	0.105
Stage 3	Stage 5	1.98E-05		
Stage 4	Stage 1	0.24		
Stage 4	Stage 2	0.783		
Stage 4	Stage 3	0.267	e_{43}	0.063
Stage 4	Stage 4	-1.29	e_{44}	0.937
Stage 4	Stage 5	1.27E-06		

Table 6: Mean sojourn times for misclassification model

	Estimates (Std. error)	**95 % C.I.**
Stage 1	1.934 (0.215)	(1.734,2.159)
Stage 2	0.542 (0.093)	(0.345.747)
Stage 3	0.817 (0.082)	(0.651,0.892)
Stage 4	0.775 (0.136)	(0.650,0.893)

Table 5 presents the estimated transition intensities for misclassification model along with misclassification probabilities. Therein e_{rs}, r denotes true stage and s denotes observed stage. So, e_{12} signify that for true stage 1 misclassifying it to stage 2 has probability 0.106, in other words there is 10% chance that patient of stage 1 will be mistakenly treated as stage2, similarly there, is 0.06 probability of treating stage 2 patients as stage 3. Mean sojourn time for misclassification model is given in **Table 6**. Even though prognostic factors effect have not been presented for simple Markov model, it is used for Hidden Markov model in **Table 7**. With sex (female) as reference, overestimation (e_{12}, e_{23}, e_{34}) of misclassification probability has odds ratio 1.46, 1.81 and 2.08 over male patients. Odds ratio for misclassification probability

for age (>35) is 2.412, 1.477, 0.906 for overestimation (e_{12}, e_{23}, e_{34}) with respect to age (<=35).

Table 7: Odds ratios for misclassification probabilities for prognostic factors

	Misclassification					
	e_{12}	e_{21}	e_{23}	e_{32}	e_{34}	e_{43}
Sex	1.466	0.651	1.814	0.578	2.08	0.722
Age	2.412	0.855	1.477	0.881	0.906	0.763
Smoking	1.524	0.743	1.745	0.578	1.79	0.62
Alcohol	2.438	0.835	2.216	0.771	1.823	0.697
CD4 count	1.245	0.529	1.329	0.742	1.074	0.092
Treatment	1.586	0.784	1.157	0.635	5.428	0.083

To decrypt the states that could have most pertinently generated the sequence of stages observed, we employ a Viterbi algorithm **Table 8**. We have randomly taken two patients data to visualize the mismatch between true and observed stage of the patient. We found that for one patient, two times stage have been underestimated, and for another patient, two times stage have been overestimated.

Table 8: Viterbi sequence

Patient	Time	Observed	Actual
870	0	4	4
870	1.542466	3	3
870	2.169863	1	2
870	3.027397	1	3
391	0	4	4
391	0.753425	4	4
391	1.334247	4	4
391	1.632877	4	4
391	1.778082	4	3
391	2.265753	4	3
391	2.671233	5	5

4. Conclusion

The study shows that current ART treatment is successful and effective in making HIV/AIDS patients long-term survivors. Although, sticking to the treatment (adherence) is highly suggested but that isn't easy to comply. Sometimes antiretroviral drugs could cause such side

effects that is severe enough to make patient stop taking them. Unfortunately, if a patient skips drugs the virus may start multiplying itself. This results in HIV to get resistant to drugs, the scenario relatively more prevalent in developing countries including India. That may be the reason of partially high morbidity and mortality due to HIV in India. This also showed by our *cure rate model* where stage 4 patients have less long-term survivors than the lower stages. We have demonstrated the alluring algorithm of pattern recognition, HMM in modeling the survival time data. This paper ventured to decipher the hidden Markov model in HIV/AIDS setup, where simple Markov model is effectively and predominantly being used to study disease progression. We obtained transition intensity for misclassification model and also the misclassification probabilities. Despite the fact that prognostic factor's effects were not considered in simple Markov model it is contemplated whilst studying hidden Markov model. Notwithstanding the evidence that sex of the patient have no significant effect on the disease progression Jackson (2011), when it comes to misclassification of stages it do have effect on odds of misclassification probability. It can be observed that males have more odds of misclassification probability than the females (reference group) patients. In other words males are more vulnerable to exaggeration of stages of disease than the females, it may be distantly attributable to the prejudices towards males with respect to debauchery in general and promiscuity in particular. This finding may be re-verified through large scale meta- analysis of HIV/AIDS data.

Patients with age more than 35 years at enrolment may be subject to overestimation of stages, which is partially understandable as older age is closely related with rapid progression of disease, Ghate et al. (2011), Touloumi et al. (1998). Thus our study solidify the point that person with relatively higher age with even higher CD4+ count should initiate ART. Likewise, smoking and alcohol consumption is associated with overestimation of stages of the disease.

Most significant and compelling finding is related with CD4+ count, whenever CD+ count is below 200 cells/μL, then odds of misclassification (overestimation) probability have increased. We have to further study the subjectivity involved in this result. As we have filtered the data set, therefore out of 1063 patients, majority of patients (694) are those on whom virocomb-N treatment combination were administered and remaining were given Tenolam+ Efravinez-600 etc.,. Hence we classify the treatment protocol as "*virocomb-N*" (reference group) and "*Tenolam+ Efravinez-600*" as target group. With virocomb-N in reference, the others treatment have more odds of misclassification (overestimation), i.e. if treatment combination administered is "*others*" then there is more chance that they will be misclassified to higher stages of the disease. At last, we have randomly taken any two patients data to see the most probable sequence of disease progression stages that may have given rise to the stages that we perceive as observed stage. By employing the Viterbi algorithm, at one go we can get rid of glut of errors committed during staging of the disease.

References

1. Aalen, O. O., Farewell, V. T., De Angelis, D., Day, N. E., & Nöel Gill, O. (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Statistics in medicine*, *16*(19), 2191-2210.

2. Achcar, J. A., Coelho-Barros, E. A., & Mazucheli, J. (2012). Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications*, *51*(1), 1-9.
3. Bature, R. S., Obiniyi, A. A., Absalom, E. E. S., & Sule, O. O. (2010). Markov chain simulation of HIV/AIDS movement pattern. *International Journal of Computer Science and Information Security*, *8*(2), 156-167.
4. Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, *37*(6), 1554-1563.
5. Bellmann, R. (1957). Dynamic programming princeton university press. *Princeton, NJ*.

6. Berkson, J., & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, *47*(259), 501-515.

7. Chen, M. H., Ibrahim, J. G., & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, *94*(447), 909-919.

8. Commenges, D., Joly, P., Letenneur, L., & Dartigues, J. F. (2004). Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in medicine*, *23*(2), 199-210.

9. Cox, D. R. and Miller, H. D. (1965) *The Theory of Stochastic Processes*. London:Chapman and Hall.

10. Debanne, S. M., Bielefeld, R. A., Cauthen, G. M., Daniel, T. M., & Rowland, D. Y. (2000). Multivariate Markovian modeling of tuberculosis: forecast for the United States. *Emerging Infectious Diseases*, *6*(2), 148.

11. Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 1041-1046.

12. Ghate, M., Deshpande, S., Tripathy, S., Godbole, S., Nene, M., Thakar, M., ... & Mehendale, S. (2011). Mortality in HIV infected individuals in Pune, India. *The Indian journal of medical research*, *133*(4), 414.

13. Grover, G., Gadpayle, A. K., & Swain, P. K. (2013). A multistate Markov model based on CD4 cell count for HIV/AIDS patients on antiretroviral therapy (ART). *International Journal of Statistics in Medical Research*, *2*(2), 144.

14. Grover, G., Sreenivas, V., Khanna, S., & Seth, D. (2014). Multi-State Markov Model: An Application to Liver Cirrhosis. *Statistics in Transition new series*, *15*(3), 429-442.

15. Ghahramani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. *International journal of pattern recognition and artificial intelligence*, *15*(01), 9-42.

16. Guihenneuc-Jouyaux, C., Richardson, S. and Longini, I. M. (2000) Modelling markers of disease progression by a hidden Markov process:application to characterising CD4 cell decline. *Biometrics*, **56**, 733–741.

17. Jackson, C. H. (2011). Multi-state models for panel data: the msm package for R. *Journal of statistical software*, *38*(8), 1-29.

18. Jackson, C. H. and Sharples, L. D. (2002) Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statist. Med.*, **21**, 113–128.

19. Kannan, N., Kundu, D., Nair, P., & Tripathi, R. C. (2010). The generalized exponential cure rate model with covariates. *Journal of Applied Statistics*, *37*(10), 1625-1636.

20. Kay, R. (1986) A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*,**42**, 855–865.

21. Kuk, A. Y., & Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 531-541.
22. Lee, S., Ko, J., Tan, X., Patel, I., Balkrishnan, R., & Chang, J. (2014). Markov chain modelling analysis of HIV/AIDS progression: A race-based forecast in the United States. *Indian journal of pharmaceutical sciences*, *76*(2), 107.
23. Maller, R. A., & Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, 1197-1205.

24. Marshall, G. and Jones, R. H. (1995) Multi-state Markov models and diabetic retinopathy. *Statist. Med.*, **14**,1975–1983.

25. National AIDS Control Organization | MoHFW | GoI. *(2018).* Naco.gov.in. *Retrieved 13 August 2018, from* http://naco.gov.in/

26. Peng, Y., & Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, *56*(1), 237-243.
27. Rabiner LR, Fellow. A tutorial on hidden Markov models and selected applications in speech recognition. 1989;257-286.

28. Ross, J. M., Ying, R., Celum, C. L., Baeten, J. M., Thomas, K. K., Murnane, P. M., ... & Barnabas, R. V. (2018). Modeling HIV disease progression and transmission at population-level: The potential impact of modifying disease progression in HIV treatment programs. *Epidemics*, *23*, 34-41.
29. Swain, P. K., Grover, G., & Goel, K. (2016). Mixture and Non-Mixture Cure fraction Models based on Generalized Gompertz Distribution under Bayesian approach. Tatra Mountains Mathematical Publications, Slovakia; *66*.
30. Sweeting, M. J., Farewell, V. T., & De Angelis, D. (2010). Multi-state Markov models for disease progression in the presence of informative examination times: An application to hepatitis C. *Statistics in medicine*, *29*(11), 1161-1174.
31. Sy, J. P., & Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, *56*(1), 227-236.
32. Touloumi, G., Hatzakis, A., Rosenberg, P. S., O'brien, T. R., & Goedert, J. J. (1998). Effects of age at seroconversion and baseline HIV RNA level on the loss of CD4+ cells among persons with hemophilia. *Aids*, *12*(13), 1691-1697.
33. Tsodikov, A. D., Ibrahim, J. G., & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models. *Journal of the American Statistical Association*, *98*(464), 1063-1078.

34. *Veterans Health Administration, a. (2018).* HIV/AIDS Home. Hiv.va.gov. *Retrieved 2 August 2018, from https://www.hiv.va.gov/index.asp*

35. Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE transactions on Information Theory, 13(2), 260-269.

36. WHO guidelines on the use of CD4, Viral load and EID test for initiation and monitoring of ART *WHO | World Health Organization* . (2018). *Who.int*. Retrieved 13 August 2018, from http://www.who.int/

37. Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association*, *87*(418), 284-292.