Neural, Parallel, and Scientific Computations 25 (2017) 181-194

TIME-INHOMOGENEOUS MARKOV MODELLING OF HIV/AIDS PROGRESSION IN SOUTH AFRICA

CLARIS SHOKO*, DELSON CHIKOBVU*, AND WINSTON GARIRA

*Department of Mathematical Statistics and Actuarial Sciences, University of the Free State, Bloemfontein (9300), Republic of South Africa
Department of Mathematics, University of Venda, Thohoyandou (0920), South Africa

ABSTRACT. This paper explores the concept of time-inhomogeneous Markov modelling and its application on the surveillance data for HIV-infected patients on anti-retroviral therapy (ART) from one of the Wellness clinics in Bela Bela, South Africa. Time-homogeneous and time-inhomogeneous models are fitted to analyse the progression of HIV/AIDS for these individuals. A variety of diagnostic methods, formal and informal, are employed to assess the fitted models. The results from the assessment of the fitted models showed an improvement in the use of the time-inhomogeneous model compared to the time-homogeneous model for the progression of HIV/AIDS. Transition intensities from the fitted models were used to analyse the effectiveness of treatment. The estimated transition rates showed that rates of immune recovery were generally higher than the rates of immune deterioration for both models. The fitted time-inhomogeneous model shows that continuous uptake of treatment reduces transitions to death from most of the HIV defined states.

Key Words : Non-homogeneous Markov models, Likelihood ratio test, Longitudinal data.

1. INTRODUCTION

Human immuno-virus (HIV), the virus that causes acquired immunodeficiency syndrome (AIDS) disease, is a continuum of progressive damage to the immune system from the time of infection to the manifestation of severe immunologic damage by opportunistic infections, neoplasm, wasting or low CD4+ cell count that define AIDS [6]. Some clinical markers such as the CD4+ cell count and the RNA viral load help a lot in providing information on the progress of the disease [13, 2]. The normal CD4+ cell count varies from individual to individual, and it is usually between 800 and 1600 cells per mm^3 [5]. CD4+ cell count values below 500 are usually an indication of immune suppression and vulnerability to opportunistic infections [10].

Stochastic processes are helpful when subjects in a population can spend time in different states. This is supported by [9] when she says that Multi-state stochastic models are useful tools for studying complex dynamics such as chronic diseases. HIV/AIDS is also an example of a chronic disease in which patients can be in different

Received February 7, 2017

states and the changes among states that occur together with the time spent in these states are of interest. For HIV patients these states are normally defined according to their CD4 cell count levels.

The Markov model is an appropriate stochastic approach when the present state of the disease summarises all the previous information because of its memoryless property. Homogeneous models have been widely used in the modelling of different disease progressions such as; cancer [14, 3] , stroke [17] and diabetic retinopathy [8]. However the hypothesis of homogeneity is unrealistic in the sense that as times goes on the disease evolve. Time homogeneity models put severe limitations on disease history behaviour. In particular when dealing with HIV, it is more realistic to assume that science and medicine evolve, hence the rate at which people change their opinions is likely to change as people are likely to believe that newer medications will improve the quality of their life [7]. This justifies the need for time inhomogeneous models in analysing disease progressions. This can simply be addressed by using piecewise Markov models that preserve the tractability of constant intensities [11].

The next section in this study explores literature of time inhomogeneous Markov models and explains its importance when investigating treatment effects. It also looks at the diagnostic techniques for Markov models. Section 3 explains the methods used to formulate the model. In Section 4, the time inhomogeneous model is fitted and is compared to the homogeneity assumption using likelihood ratio test and survival functions from fitted models. Finally, section 5 concludes the findings from the analysis.

2. LITERATURE

There are various approaches that can be employed for fitting the inhomogeneous models. These approaches include;

- Allowing transition intensities to be piece-wise constant (PWC). This is the most commonly used method;
- Allowing transitions to have smooth parameters forms, for example the Weibull hazard functions. For Markov models this requires solving the Kolmogorov forward equations which are set to be non-linear ordinary differential equations for time homogeneous model; and
- Non-parametric or semi-parametric techniques

This research is going to focus on the PWC model. Define $Q(t)$ as a non-homogeneous vector of transition intensities, $Q = Q(t) : 0 < t < T$ to be estimated, where T is the maximum observed time. By using the HIV states defined for this research,

$$Q(t) = [q_{12}(t), q_{16}(t), q_{17}(t), q_{21}(t), q_{23}(t), q_{26}(t), q_{27}(t), \ldots, q_{54}(t), q_{56}(t), q_{57}(t)],$$

with each entry a function of time were $q_{uv}(t)$ is the transition rate from state u to state v, see section 3.

2.1. **Piece-wise constant model (PWC).** The PWC model partitions the entire time interval into r continuous, disjoint intervals, $\tau_1, \ldots, \tau_r$ where τ_r is the interval from time a_{r-1} to a_r. This approach to non-homogeneity in a Markov process is a step-wise method that assumes constant transition intensities in different time intervals. This method leads to a non-homogeneous Markov model in which the transition intensities are step-functions of time defined as follows:

$$q_{uv}(t) = \begin{cases} q_{uv,0} & \tau_1 = a_0 \leq t < a_1, \\ q_{uv,1} = q_{uv,0} \exp\{\beta^*_{uv,1}\} & \tau_2 = a_1 \leq t < a_2, \\ \vdots & \\ q_{uv,r-1} = q_{uv,0} \exp\{\beta^*_{uv,1} + \beta^*_{uv,}\} + \ldots \beta^*_{uv,r} & \tau_r = a_{r-1} \leq t < a_r. \end{cases} \tag{2.1}$$

The parameters $q_{ij0}, q_{ij1}, \ldots, q_{ijr-1}$ are the transition intensities for intervals $[a_0, a_1)$, $[a_1, a_2)$ up to the interval $[a_{(r-1)}, a_r)$ respectively and $\beta^*_{uv,r}$ is the vector of regression coefficients associated with the artificial time-dependent covariates. The baseline intensities are represented by the parameter $q_{ij,0}$. Computing $P(0, t_i)$ for a t_i in segment τ_r entails multiplying all the transition matrices across the various intervals as shown below;

$$P(0, t_i) = [\Pi_{b=1}^{r-1} P^{(b)}(\pi_b)] P^{(r)}(t_{(r-1)}, t_j)$$

where $P^{(b)}$ is the transition probability matrix obtained using q_{uvb} for the b^{th} segment denoted by τ_b. If subjects are observed on an equal spaced grid and segments are divided up along these time points, then $P_{uv}(0, t_i)$ would simply be the $(uv)^{th}$ element of the matrix in the above equation. When data is not equally spaced, then observations would be considered missing at the breakpoints. To resolve this, a model that accounts for all possible pathways between the last observed state in the segment b_{i-1} and the first observation in segment b_i was suggested. For example, if a breakpoint t' is created between two points t_j and t_k, then via Chapman-Kolmogorov equations the likelihood contribution from interval (t_j, t_k) for individual i can be found as;

$$L_i = \sum_{l=1}^{k} P^{(1)}_{ul}(t_j, t') P^{(2)}_{lv}(t', t_k)$$

for states u, v. A likelihood ratio test can then be used to determine whether a PWC model is a better fit than the constant model.

2.2. **Diagnostic methods for Markov Models.** Titman suggested a number of methods that can be used as diagnostic tools for the fitted Markov models [1]. He classified the methods into two groups; the formal methods and the informal methods. The formal methods included the use of the log-likelihood ratio tests (LRT) and the

Pearson's chi-squared test. The LRT is used to compare nested models, a class of models were one has got more covariates than the other. The Pearson's chi-squared test is used to compare prevalence counts by the use of observed and expected frequencies. The informal methods use graphs to compare the fitted model with the observed data and do not involve calculation of likelihood ratios.

2.2.1. *The Likelihood ratio test (LRT) for homogeneity.* The homogeneous assumption assumes that transition intensities are constant throughout time, that is, $q_{uv}(t) = q_{uv}$. This assumption is tested using a formal likelihood ratio test for independence of the piecewise model and the time-homogeneous model. Under H_0 (homogeneity assumption) the test statistic has approximately a χ^2_{k-q} where q is the number of parameters under H_0 and k is the number of parameters under H_1 (the inhomogeneity assumption) as follows.

Suppose

$$LRT = -2\log_e\left(\frac{L_0(\hat{\theta})}{L_1(\hat{\theta})}\right)$$

is the ratio of two likelihood functions, for a simpler (homogeneous) model with fewer parameters and the alternative (inhomogeneous) model. The test statistic is asymptotically distributed as a chi-square random variable, with degrees of freedom equal to the difference between the number of parameters for the two models. The likelihood ratio test can be performed provided the simpler model is a special case of the complex model.

The LRT can also be presented in terms of deviance, that is;

$$\begin{aligned} LRT &= -2[\log_e(L_0(\hat{\theta})) - \log_e(L_1(\hat{\theta}))] \\ &= -2\log_e(L_0(\hat{\theta})) + 2\log_e(L_1(\hat{\theta})) \\ &= deviance_0 - deviance_1. \end{aligned}$$

Thus, the LRT can be computed as a difference between the deviances for the two fitted models.

2.2.2. *Kaplan-Meier survival curves.* Kaplan-Meier product limit estimates of the survival function can be compared to survival estimates from the fitted Markov models. This can be done in cases where a model has an absorbing state for which time of entry is precisely known. Kaplan-Meier estimates are only valid for the assumption of homogeneous subjects.

2.2.3. *Contingency table based methods.* This method provides an assessment of the overall fit of the assumed model. Kalbfleisch and Lawless dealt with balanced observation with categorical covariates [1]. They fitted the model by considering observed

and expected transition frequencies either through likelihood ratio test or asymptotically equivalent Pearson chi-squared statistic. However Titman argues that Pearson chi-square has low power particularly when the degrees of freedom are very large and that the asymptotic null distribution cannot be applied when counts in table are small [1].

In the next section, time inhomogeneous models are fitted using the methods discussed above. The observed and expected prevalence are used to diagnose the fitness of the model for both the time-homogeneous model and the time-inhomogeneous model.

3. METHODS

The model was applied on 318 HIV patients under ART from a Wellness clinic in Bela Bela, South Africa. Some of the patients were enrolled at the clinic with TB being the initial marker of HIV. All of the patients in the study were monitored after every 6 months from 2005 to 2009. For every visit the CD4 count, BMI, viral load, any adverse reaction to treatment and development of TB were noted. The time homogeneous models and the time inhomogeneous models were built. These models were used to assess effectiveness of the treatment by comparing the forward transition and the reverse transitions for different time intervals. This then leads to building of models that allow transitions in both directions.

In this section, a piece-wise constant model is fitted for the HIV data to examine the possibility of intensities to change over time. The 2.5 and 4 year cut points are used implying the use of 3 segments. The use of three segments necessitates the computation of all possible pathways from the last observation time in segment 1 to the first observed time in segment 2 and from the last observed time in segment 2 to the first observed time in segment 3. The estimates of $Q(t)$, the transition intensity matrix, are broken into three segments as shown below;

$$Q(t) = \begin{cases} Q_1; & 0 \leq t < 2.5, \\ Q_2; & 2.5 \leq t < 4, \\ Q_3; & 4 \leq t < \infty \end{cases} \tag{3.1}$$

At any time, $t+\Delta t$, where $\Delta t = 0.5$ years the state of an HIV-infected individual is defined basing on the CD4 cell count level or based on whether the individual is dead or has withdrawn as follows:

$$\begin{array}{llll} 1- & CD4 \geq 750; & 2- & 500 \leq CD4 < 750; \\ 3- & 350 \leq CD4 < 500; & 4- & 200 \leq CD4 < 350; \\ 5- & CD4 < 200; & 6- & Dead; \\ 7- & Withdrawn. & & \end{array} \tag{3.2}$$

where sates 1 to 5 are transient and state 6 and 7 are absorbing that is, once entered they cannot be exited. State 5 is normally defined as the AIDS defining stage. The good state which is normally defined as $CD4 \geq 500$ was further split into two states in order to observe if there are any possible backward transition once an individual is in that state.

The decision for taking CD4+ cell count levels to define the transient states was based on the fact that when HIV enters the human body its main target is the white blood cell called the CD4+ cells. When the virus enters the CD4+ cell it destroys the cell as it replicates after which more CD4+ cells are targeted. Continued destruction of these cells result in immune deterioration since the CD4+ cells are the markers of the immune system.

Due to this HIV progression trend of individuals on treatment, we decided to use a bi-directional Markov model to analyse the data. The bidirectional model allows for backward and forward movements between the five transient states, that is, if an individual is in state i at time t, at time $t + \Delta t$ that same individual can either be in state $i - 1$ or $i + 1$, or can remain in the same state i, for $i = 1; 2; \ldots; 5$ or can transition to the absorbing state 6 or state 7. According to our model, transition to state $i - 1$ is an indication of immune recovery and transition to state $i + 1$ is an indication of immune deterioration.

Based on this structure, two-way transition intensity matrices are computed for the constant model as well as for the piece wise constant model and comparisons are made. The Multi-State Model (MSM) package for R developed by Jackson is used for all the analysis in this study [4].

4. RESULTS AND ANALYSIS

The matrix below shows the estimated transition rates for the interval $0 \leq t < 2.5$ which represents the baseline intensity matrix Q_1;

$$Q_1 = \left(\begin{array}{cc|ccccccc} & To & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline From & 1 & -0.696 & 0.667 & 0 & 0 & 0 & 0.020 & 0.009 \\ & 2 & 0.494 & -1.261 & 0.358 & 0 & 0 & 0.076 & 0.334 \\ & 3 & 0 & 0.379 & -0.772 & 0.277 & 0 & 0.059 & 0.056 \\ & 4 & 0 & 0 & 0.535 & -1.023 & 0.224 & 0.008 & 0.256 \\ & 5 & 0 & 0 & 0 & 0.640 & -2.016 & 1.086 & 0.290 \\ & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

The results from the baseline intensity matrix (Q_1) show higher rates of immune recovery compared to immune deterioration. That is, for all the transition intensities $q_{i,i-1} > q_{i,i+1}$ for $i = 2, 3, 4, 5$.

Denoting the transition intensities by $q_{ij}^{(1)}$, $q_{ij}^{(2)}$ and $q_{ij}^{(3)}$for the periods $r = 1; 2; 3$, representing the intervals $0 \leq t < 2.5$, $2.5 \leq t < 4$ and $4 \leq t < \infty$ respectively. A model allowing different effects for each of the transition intensity for the time period 2.5 and 4 years is fitted. Table 1 below shows the effects of the two time periods, $2.5 \leq t < 4$ and $4 \leq t < \infty$, on the baseline transition intensities represented by Q_1.

TABLE 1. Linear Effects of different time periods on the baseline transition intensities

Parameter	$Q_2 : 2.5 \leq t < 4$	$Q_3 : 4 \leq t < \infty$
β_{12}	-0.006	0.316
β_{16}	-13.56	1.197
β_{17}	12.46	-6.419
β_{21}	0.258	0.618
β_{23}	0.491	0.364
β_{26}	-0.864	-13.969
β_{27}	-0.163	-0.251
β_{32}	0.149	-0.197
β_{34}	-0.204	0.258
β_{36}	5.24	6.674
β_{37}	2.112	-7.854
β_{43}	0.0134	0.159
β_{45}	0.141	0.568
β_{46}	-17.06	-19.978
β_{47}	-0.041	-0.733
β_{54}	0.240	0.877
β_{56}	-0.186	-0.169
β_{57}	2.0218	2.930

The results from Table 1 show an estimated reduction on transitions rates to death, from most of the states except state 3 for time periods 2.5 and 4 years post commencement of treatment, the values of β_{36} are 5.24 and 6.74 respectively. This is shows that most deaths for individuals who start treatment with CD4 cell count between 350 and 500 increase with time. This calls for further investigations by future researchers to find out what could be cause of this anomaly. The rates of reduction of transitions to death from state 4 are tremendously high compared to the other states, $\beta_{46} = 17.06$. The rate of progression from state 3 to state 2 is expected to increase after 2.5 years (5 half-years), Q_2, $\beta_{32} = 0.149$, but after 4 years (8 half-years), Q_3, it is expected to decline, $\beta_{32} = -0.197$ and from state 3 to state 4 it is the opposite. For these two intervals, transition rates from state 4 to the AIDS defining state, state 5, are higher than transitions to a better state. The same period contributed positively

to transition to withdrawal from most of the states. For the time interval represented by Q_3, there is generally an estimated reduction on the rate of withdrawal from most of the states. The interval also has lower transitions to death compared to the time interval $2.5 \leq t < 4$.

Table 2 shows the hazard rates for the two time periods, $2.5 \leq t < 4$ and $4 \leq t < \infty$, relative to the time interval $0 \leq t < 2.5$. Hazard rates for the constant model are also shown for comparison with the piecewise model.

TABLE 2. Hazard ratios for the time periods $0 \leq t < 2.5$ (Q_1), $2.5 \leq t < 4$ (Q_2) and $4 \leq t < \infty$ (Q_3) years

Parameters	Q_{homo}	Q_1	Q_2	Q_3
q_{12}	0.887	0.667	3.656	1.372
q_{16}	0.102	0.020	0.003	3.311
q_{17}	0.142	0.009	2.577	0.081
q_{21}	0.518	0.494	1.294	1.856
q_{23}	0.496	0.358	1.634	1.439
q_{26}	0.081	0.076	1.549	0.008
q_{27}	0.038	0.334	3.124	2.863
q_{32}	0.423	0.379	1.160	3.021
q_{34}	0.332	0.277	3.000	1.295
q_{36}	0.015	0.059	13.94	58.56
q_{37}	0.031	0.056	8.266	0.071
q_{43}	0.518	0.535	1.013	1.172
q_{45}	0.252	0.224	1.151	1.764
q_{46}	0.003	0.008	0.001	0.000
q_{47}	0.054	0.256	3.531	1.766
q_{54}	0.516	0.640	1.271	2.404
q_{56}	0.091	1.086	8.301	3.106
q_{57}	0.029	0.290	7.552	5.094

The results from Table 2 show that there is much variation between transition intensities in the time homogeneous model (Q_{homo}) and transition intensities in the piece wise constant model represented by Q_1, Q_2 and Q_3. Regardless of the originating state, transitions to a better HIV state represented by ($q_{21}; q_{32}; q_{43}$ *and* q_{54}) are increasing with time. The highest transitions to a better state (recovery) are recorded from the period of 4 years onwards. The individuals who originated from state 1 and state 3 had the highest transition intensities to worst states for the period 2.5 to 4 years. From 4 years onwards these transition intensities dropped significantly and

transitions to better states increased significantly which is an indication of effectiveness of treatment with time. Unlike the individuals who started in state 1 and 3, those who started in sate 2 and 4 showed a decreasing trend on the rates of transition to worst states and an increasing trend on the rates of transition to better states. Transitions to death (state 6) from state 1 were relatively low for the periods 0 to 2.5 years and 2.5 to 4 years and then increased significantly after 4 years. Deaths from state 2 and state 5 were very high during the period 2.5 to 4 years and then drop significantly thereafter. Transitions to death for individuals who originated in state 3 increased tremendously with time as shown by the values $(q_{36}^{(1)}; q_{36}^{(2)}; q_{36}^{(3)}) = (0.059; 13.94; 58.56)$. However, for those who started in state 4, transitions to the death state decreased with time and these transitions were relatively low.

The probability matrix (P) for the fitted piece-wise constant model is as follows;

$P =$

	To						
From	1	2	3	4	5	6	7
1	0.593	0.339	0.049	0.0045	0.00022	0.0097	0.00571
2	0.220	0.569	0.156	0.0212	0.00140	0.0122	0.0204
3	0.041	0.2003	0.584	0.150	0.0148	0.00394	0.00662
4	0.0068	0.0496	0.273	0.533	0.1036	0.0173	0.0167
5	0.000897	0.0087	0.0711	0.274	0.567	0.0689	0.00921
6	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1

Matrix P shows that the probabilities of moving from a lower CD4 count level to a higher CD4 count level are higher than the probability of immune deterioration. This confirms the result obtained from the transition intensities which is an indication of effectiveness of treatment. The probabilities of maintaining the same state are almost the same, all between 0.5 and 0.6. These individuals tend to take more time in the same state before they transition to another state. However there is a highest probability of maintaining state 1 followed by state 3. This indicates that patients on antiretroviral therapy spend more time in state 1 or state 3 than all the other states. The probabilities of transitions to death increase as HIV/AIDS progresses to a worst state. Patients who are in the AIDS defining state, 5, have the highest probability of dying, $P_{56} = 0.0689$, followed by patients in state 4 compared to death from all the other states. Although transitions to death for individuals who originated from state 3 increases with time as indicated by the piecewise transition intensities, these individuals tend to have relatively low transitions to death compared to all the other states.

4.1. **Comparison of Time Homogeneous and Inhomogeneous models.** This subsection applies the diagnostic techniques discussed in part of section 2 to justify

the importance of time inhomogeneity in describing the HIV/AIDS progression of patients on treatment.

4.2. **Likelihood ratio test for time homogeneity.** To determine whether the piece-wise constant model is a better fit to the HIV data than the straight forward constant approach, the likelihood ratio test was performed in R using the MSM package. The null hypothesis, $q_{ij,r} = q_{ij}$ for all intervals $\tau_r, r = 1; 2; 3$ is used. The alternative hypothesis is that at least two of the $q_{ij,r}$ are not equal. Under the null hypothesis we have 18 parameters to estimate and the alternative we have 54 parameters to estimate, implying that we have 36 degrees of freedom. From the fitted homogeneous and inhomogeneous models we have; $-2L_0 = 3941.971$ and $-2L_1 = 3804.12$, respectively. This leads to:

$$\chi^2 = -2(L_0 - L_1) = 137.850$$

which has a chi-square distribution on 36 degrees of freedom. This can also be shown by using the MSM command, lrtest.msm(proj.msm,proj.pci.msm) where proj.msm is the fitted homogeneous model and proj.pci.msm is the fitted inhomogeneous model. The outcome is shown below;

	$-2LogLR$	df	p
$Proj.pci.msm$	137.850	36	0.00024

Therefore we reject $H_0, (p = 0.00024)$ at 5% level of significance and conclude that there is an improvement in using the piece-wise constant model.

4.2.1. *Survival probability.* By defining survival as "not dying from HIV/AIDS", the estimated survival functions were plotted for each of the 5 transient stages of infection. For each of the states the probability of survival was initially set to be 1. Survival probabilities were computed as follows;

$$S_{T_i}(t) = 1 - \hat{p}_{i6}(t), \quad i = 1; 2; \ldots; 5$$

where $\hat{p}_{i6}(t)$ is the estimate of the probability of dying from each of the transient states i. Survival functions were fitted for the time homogeneous model with no covariates and time inhomogeneous (piecewise constant) model. The results are shown in Figure 1 below.

The graphs in Figure 1 show that probabilities of survival decrease with time as expected but not at an alarming rate. By the end of the period of study survival probabilities from each of the five transient states were all above 0.8 except for both the time-homogeneous model and time-inhomogeneous model. The results show that the chances of survival are highest for patients whose previous CD4 count was below 200 (state 5) since they have the highest prevalence, and the chances of survival are lowest for individuals with initial CD4 count between 500 and 750 (state 2). The fitted

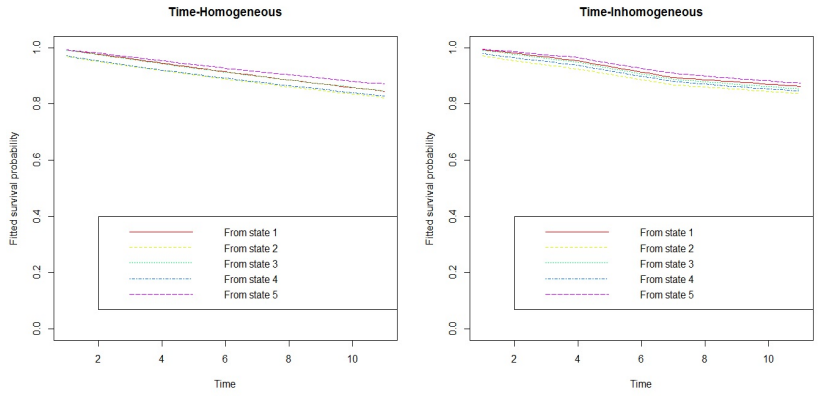

FIGURE 1. *Comparison of survival functions from each of the fitted models*

time-inhomogeneous model chances of survival from all states differ by a very small margin particularly from 8 half-years (4 years) onwards. For the time-homogeneous model chances of survival are very high from state 5 and differ with a very high from the other states.

4.2.2. *Diagnostic plots.* One simple diagnostic compares model predictions of the entry time into a particular state with nonparametric estimates, for example Kaplan-Meier curves. If the entry time is not observed exactly, then the nonparametric estimate is an approximation [4]. The fitted time-homogeneous and time-inhomogeneous models are assessed in Figure 2. The results from the diagnostic plots in Figure 2 show

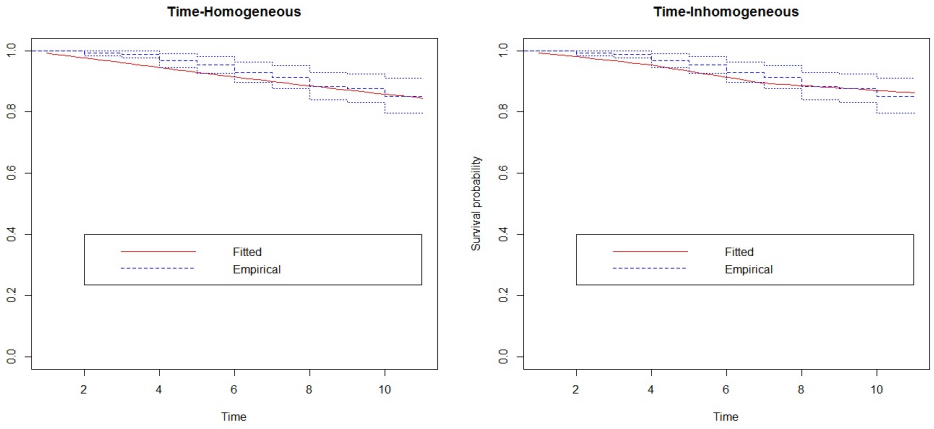

FIGURE 2. *Comparison of observed and fitted survival for four multi-state models*

that both the fitted time-homogeneous model and the time-inhomogeneous model underestimate survival of the observed individuals up to 8 half-years (4 years). From 8 half-years onwards, although the fitted homogeneous model still underestimates survival of the observed individuals the inhomogeneous model displays a perfect fit.

Although none of the models fit the data adequately, the time-inhomogeneous model fits better than the time-homogeneous model.

4.3. **Prevalence for the time-inhomogeneous model.** Examination of the expected frequencies versus the predicted frequencies for each state was done in order to identify areas of poor fit of the model. Figure 3 shows that the expected counts are relatively close to the observed counts in most of the states. However, for the death state which is the absorbing state, the expected model underestimates the fitted model from 8 half-years (4 years) to 11 half-years (5.5 years) by a large margin. The expected model underestimates the fitted model for the patients with CD4 cell counts between 0 and 200 (state 5). There is a sharp decrease on the prevalence for this state.

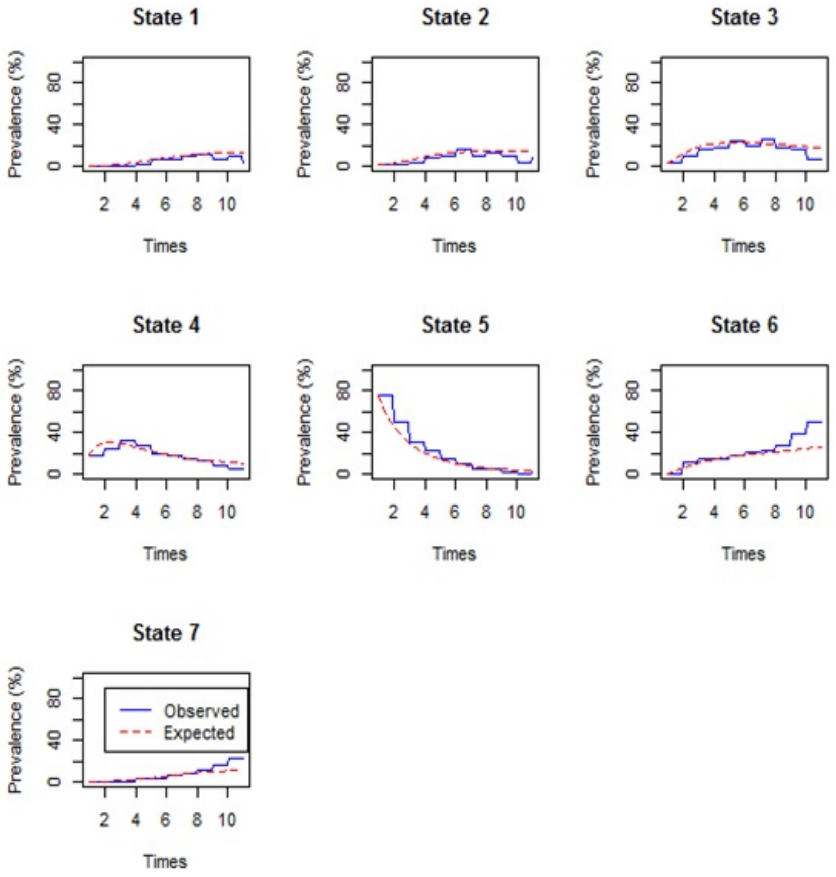

FIGURE 3. *Comparison of observed and expected prevalence from the time-inhomogeneous model 5 and 8 half-year change points*

5. CONCLUSION

This paper explains the theory of non-homogeneous Markov models with particular reference to the piece-wise constant model. It also discusses the methods that can be used in model diagnostics. The methods include the formal likelihood ratio test

and the informal diagnostic plots for model comparison. The 3-segment piece-wise constant model to the HIV data was compared with the time homogeneous model using the likelihood ratio test. The test showed that the piecewise constant model fits the data best. Diagnostic plots were also used to compare the models and the results still confirm that the inhomogeneous model is the best model for the data. The fitted piece-wise constant model confirms that rates of immune recovery are generally higher than the rates of immune deterioration confirming the effectiveness of treatment. These rates of immune recovery increase as individuals continue to take their medication. The rates of immune deterioration showed a generally decreasing trend with time although for individuals who originated from state 1 and 3 it showed an increase followed by a decrease.

Deaths from all the other states decreased significantly with time except for those individuals who originated in state 3 which is defined by a CD4 cell count between 350 and 500 cells per micro litre. For these individuals the transitions to death rose from 13.94 between 2.5 and 4 years to 58.56 there after. This calls for a need by further researchers to investigate the causes of these deaths.

ACKNOWLEDGEMENTS

This study would not have been a success without the assistance of the University of Venda's Department of Microbiology which provided the data through Professor Pascal O. Bessong.

REFERENCES

[1] A. C. Titman, Model diagnostics in multi-state models of biological systems, *Department of Pure Mathematics and Mathematical Statistics*, University of Cambridge, 2007.

[2] AIDSinfo, Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents, http://aidsinfo.nih.gov/guidelines, Last accessed: 9/15/2016

[3] A. M. F. Yen, T. H. H. Chen, Mixture Multi-State Markov Regression Model, *J Appl Stat*, 34: 11–21,2007.

[4] C. H. Jackson, Multi-state models for Panel Data: The msm package for R, *Medical Research Council Biostatistics Unit*, 38(8), 2011.

[5] Fact Sheet 124, CD4 Cell Tests, *aidsinfonet.org*, May 2014

[6] G. Biase, G. D'Amico, A. Girolamo, J. Janssen, S. Iacobelli, N. Tinari and R. Manca, A Stochastic Model for the HIV/AIDS Dynamic Evolution, *Mathematics Problems in Engineering*, Vol.65636, 14 pages 2007.

[7] G. Lebovic, Estimating Non-Homogeneous Intensity Matrices in Continuous Time Multi-State Markov Models, A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Dalla Lana School of Public Health, University of Toronto, 2011.

[8] G. Marshall, and R. H. Jones, Multi-state Markov models and diabetic retinopathy. *Statist. Med.*, 14:1975–1983, 1995.

[9] G. T. Zelalem, Multi-state models of HIV/AIDS by homogeneous semi-Markov processes, *American Journal of Biostatistics*, 4 (2): 21–28, 2014.

[10] Health 24 2000-2013, The interaction between viral load and CD4 cell count and disease progression, 2003.

[11] J. C. Lindsay and L. M. Ryan, A three-state multiplicative model for rodent tumorigenicity experiments, *Applied Statistics*, 42: 283–300, 1993.

[12] J. D. Kalbfleisch, J. F. Lawless, The analysis of panel data under a Markov assumption, *J Am Stat Assoc*, 80:863–871, 1985.

[13] M. Carter and C. Hughson, Cd4 cell count, aidsmap, 14 March 2016

[14] M. J. Duffy, J. Blaser, C. Duggan, E. McDermott, N. O'Higgens, J. J. Fennelly and H. Tschesche, Assay of metalloproteases types 8 and 9 by ELISA in human breast cancer. *Br. J. Cancer*, 71:1025–1028, 1995.

[15] P. Saint-Pierre, C. Combescure, J. P. Daures and P. Godard, The analysis of asthma control under the Markov assumption with use of covariates, *Internatinal Society of Clinical Biostatistics*, Dijon, France, September 2002.

[16] R. Perez-Ocon, J. E. Ruiz-Castro and M. L. Garmiz-Perez, A piecewise Markov process for analysing survival from breast cancer in different risk groups, *Statist. Med.*, 20:109–122, 2001

[17] W. Pan, and A. J. Kastin, Tumor necrosis factor and stroke: role of the blood-brain barrier, *Progress in neurobiology* 83:363–374, 2007.