

RESEARCH

Open Access



Modelling HIV disease process and progression in seroconversion among South Africa women: using transition-specific parametric multi-state model

Zelalem G. Dessie^{1,2*}, Temesgen Zewotir¹, Henry Mwambi¹ and Delia North¹

Abstract

Background: HIV infected patients may experience many intermediate events including between-event transition throughout their follow up. Through modelling these transitions, we can gain a deeper understanding of HIV disease process and progression and of factors that influence the disease process and progression pathway. In this work, we present transition-specific parametric multi-state models to describe HIV disease process and progression.

Methods: The data is from an ongoing prospective cohort study conducted amongst adult women who were HIV-infected in KwaZulu-Natal, South Africa. Participants were enrolled during the acute HIV infection phase and then followed up during chronic infection, up to ART initiation.

Results: Transition specific distributions for multi-state models, including a variety of accelerated failure time (AFT) models and proportional hazards (PH) models, were presented and compared in this study. The analysis revealed that women enrolling with a CD4 count less than 350 cells/mm³ (severe and advanced disease stages) had a far lower chance of immune recovery, and a considerably higher chance of immune deterioration, compared to women enrolling with a CD4 count of 350 cells/mm³ or more (normal and mild disease stages). Our analyses also showed that older age, higher educational levels, higher scores for red blood cell counts, higher mononuclear scores, higher granulocytes scores, and higher physical health scores, all had a significant effect on a shortened time to immunological recovery, while women with many sex partners, higher viral load and larger family size had a significant effect on accelerating time to immune deterioration.

Conclusion: Multi-state modelling of transition-specific distributions offers a flexible tool for the study of demographic and clinical characteristics' effects on the entire disease progression pathway. It is hoped that the article will help applied researchers to familiarize themselves with the models, including interpretation of results.

Keywords: AFT models, Markov model, Latent variables, Factor analysis, Waiting probabilities, Transitions and quality of life domain scores

* Correspondence: zelalem_getahune@yahoo.com

¹School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

²College of Science, Bahir Dar University, Bahir Dar, Ethiopia



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

HIV is one of the leading causes of mortality amongst infectious diseases globally and is recognized as a main public health problem [1]. AIDS, the last disease progress stage of HIV infection, leads to severe damages to the body's resistance system [1, 2]. Progression of HIV/AIDS is highly variable between individuals and populations and is determined by immunologic, environmental, genetic, and virology factors [3]. Improved awareness of the HIV disease process and progression pathway, including influencing can add great value to understanding HIV pathogenesis and developing treatment strategies [4].

CD4 cell count is the surrogate marker of HIV disease progression regularly used in the clinical setting to monitor the infection [5]. It is also an accurate marker of the stage of HIV disease and is recommended by all guidelines of HIV management [6]. Researchers also argue that the HIV viral load appears to be the best predictor of long term clinical outcomes, whereas CD4 cell count predicts clinical information (event time data) [7]. CD4 cell count has informed us when to start and stop opportunistic infection management or prophylaxis when to start antiretroviral therapy, OI risk stratification, as well as in monitoring response to treatment [8, 9]. Thus, CD4 cell count has been an important factor in the clinical investigation of HIV as well as used as a prognostic marker for assessing HIV disease process and progression.

HIV patients go through severe, advanced, mild, and normal clinical stages and *CD4 cell count* provides a biomarker for characterizing these *clinical stages* [6]. Also, the disease diagnosis can be considered as one of these clinical stages. Modelling of these sequential clinical stages may be a better capture of the complete HIV/AIDS disease progression. It is also essential to understand and predict accurately the course of the disease evolution, and it is of particular relevance for the clinicians that need to distinguish the different types of events in order to properly adapt the treatment.

For modelling the disease progression of HIV/AIDS, there are some specific challenges: there may be several intermediate events for the patients, from HIV infection to death; and some clinical variables may be measured over time and are therefore often not available for some cohort members. These challenges require a specific arrangement and preparation of the data for analysis as well as the specification of appropriate models, in order to allow for the cyclic patient longitudinal data. To address the above-mentioned challenges, we have presented multi-state transition-specific parametric models. This is an important point that has not been considered in HIV/AIDS cohort studies particularly in Sub-Saharan Africa. A multi-state transition-specific parametric model allows rich approaching into complex disease

processes and progression pathways, where the patients may experience some intermediate endpoints, and in addition, the model permits the analysis to examine the possible covariate effects on each specific transition [10–12]. Multi-state models are very useful for describing event-history data, giving a deeper understanding of disease process and progression and how other patient's demographic and clinical characteristics affect the entire disease progression pathway.

Most methodological developments have usually focused on semi-parametric methods, by using the Cox regression method as the basic framework of multi-state modelling [11, 13–16]. The fully parametric multi-state approach is less well known and somewhat of new development in medical data, where the same distribution is assumed for all transitions [2, 17]. Through a fully parametric survival model approaching, there are many advantages: predicting, modelling time-varying variable factors, gaining a better understanding of how factors change with time, etc. as compared to a semi-parametric method [18]. The basis of this paper is to present a multi-state model considering the requirement of distribution for each transition, approaching a variety of fully parametric models, and allowing for structures to be shared across the particular transitions. In some cases, there will be limited data for particular transitions and we may have an inclination to assume simple parametric distributions, for example, the exponential distribution; and while we may have more information for a specific transition, we may use a more complex parametric distribution [19]. Furthermore, to our knowledge, little research has emerged that examines the consequences of fitting transition specific distributions to multi-state models for medical data with time to sequential adverse events.

The purposes of this study are thus three-fold. Firstly, we seek to introduce the multi-state Markov modelling framework, briefly describing previous approaches. Secondly, we use a non-parametric method (Aalen-Johansen estimator) to estimate transition probabilities and length of stay in a particular state. Finally, we seek to develop multi-state models, which allow for transition-specific distributions and to apply this model to describe the complete evolution process of HIV in ART-naive individuals so as to aid in obtaining a deeper understanding of HIV disease progression and to discover possible factors that influence immune deterioration.

Methods

Study population

The data is from an ongoing prospective cohort study from the CAPRISA. CAPRISA initially enrolled HIV-negative (phase I) women into different study cohorts. The main study was a prospective cohort study (the

CAPRISA 002), aimed at documenting acute infection with an extensive follow up to determine the natural history of the HIV-1 subtype C infection. The establishment of the CAPRISA 002 acute infection study was between August 2004 and May 2005 [20]. It was conducted at the Doris Duke Medical Institute (DDMRI) situated at the Nelson R. Mandela School of Medicine of the University of KwaZulu Natal in Durban, South Africa. Participants were recruited at two sites in KwaZulu-Natal: an urban site in Durban and a rural site in Vulindlela. Participants without well documented estimated date of HIV infection and those who were lost to follow-up during the observation period, were excluded in this analysis. Further information about these ongoing prospective HIV cohort studies, including women's eligibility criteria, were reported in [20–22]. Finally, two hundred and nineteen (219) participants were included in the study.

Variables and measurements

Once HIV diagnosis was confirmed, participants were enrolled into the acute HIV infection cohort and were then followed up for a maximum of 13 years at the time of this analysis. The follow-up assessment phase was classified as follows: acute infection (i.e. phase II: this was taken as follow up time within 0–3 months after infection), early infection (i.e. phase III: the time period during which the patients were followed up between 3 to-12 months' post-infection), established infection (i.e. phase IV: It started from 12 months' post-infection and ended once the patient has initiated antiretroviral therapy) and on cART (phase V: the patient was on ART- in this study, it was initiated when the CD4 cell count was below 500 cells/mm³). Samples for immunologic, virologic, and clinical parameters (such as viral load, CD4 counts, etc.) were measured at each visit [23]. There was a total of 8760 followed-up visits recorded from 219 HIV infected women.

The main outcome variable in this current paper is time to sequential adverse events. World Health Organization immunological classifications were used to assess the degree of severity of HIV infection of patients in the study. These HIV infection stages are defined as no adverse events ($CD4 \geq 500$), mild ($350 \leq CD4 \leq 499$), advanced ($200 \leq CD4 \leq 349$) and severe ($CD4 < 200$) [6].

The effect of several factors on time to sequential adverse events was evaluated including: (1) demographics: age, marital status, educational level, and sex under the influence of alcohol; (2) OI: hypertension and tuberculosis; (3) risk variables: substance use, contraceptive use, and sex under the influence of alcohol; (4) clinical parameters: WBC components (neutrophils, lymphocyte count, monocytes, eosinophils count, and leucocyte count), Blood chemistry (sodium, chloride, calcium, ALT, AST, total

protein and LDH), RBC parameters (Hb, RDW, MCH, MCV, MCHC, and hematocrit) and (5) QoL domain scores. The QoL questionnaire [24], was used to measure QoL of HIV infected patients. Therefore, the QoL scales contain the following domain. The first is physical health scores, that assess perceived working capacity, lack of energy and initiative, fatigue, the presence of pain, dependence on therapeutic substances, and the impact of the disease on the activities of daily living. The second is the psychological-wellbeing score domain, which measures the patient's thoughts about body appearance, negative and positive suicide, anxiety, higher cognitive functions, self-esteem and personal beliefs, feelings, depression, and spirituality. The third domain is social relationships, which measures sexual activity, social support, social contacts, and personal relationships. The fourth domain is devoted to the level of independence and measures areas such as work capacity, dependence on treatments, activities of daily living, and mobility. Further information about the above-mentioned factors was reported in [25, 26]. (See Fig. 1).

Statistical method

Factor analysis

Since the dataset has numerous clinical parameters, we used the factor analysis (FA) methods to minimize and group the parameters. Exploratory FA was done by creating the principal components of the original variables and then creating the eigenvectors. By using the Kaiser-criterions, eigenvectors with eigenvalues greater than 1 were kept [27]. A maximum likelihood extraction method with varimax rotation was used. Each observation was assigned a score for each rotated factor, based on the loading of the subject's original variable levels. Accordingly, we managed to group the 24 clinical variables in the study, to create 9 latent variables, defined as protein component, lipid component, electrolyte component, liver abnormality component, red blood cell indices, Hb and haematocrit component, eosinophils component, mononuclear component, and granulocytes component. (See Table 1).

Multi-state model formulation

HIV infected, ART-naive patients may experience many CD4 cell count fluctuations, mostly before ART initiations. This indicates that the HIV disease process and progression should be modelled by a multi-state process. Figure 2 shows a flow diagram of the multi-state model. In this study, we considered six transitions: (1) Normal \rightarrow Mild, (2) Mild \rightarrow Advance, (3) Advance \rightarrow Severe, (4) Severe \rightarrow Advance, (5) Advance \rightarrow Mild and (6) Mild \rightarrow Normal. These mentioned transitions were modeled by a four-state multi-state transition-specific model. Attention was on predicting the probability of transition and the probability of staying in the same

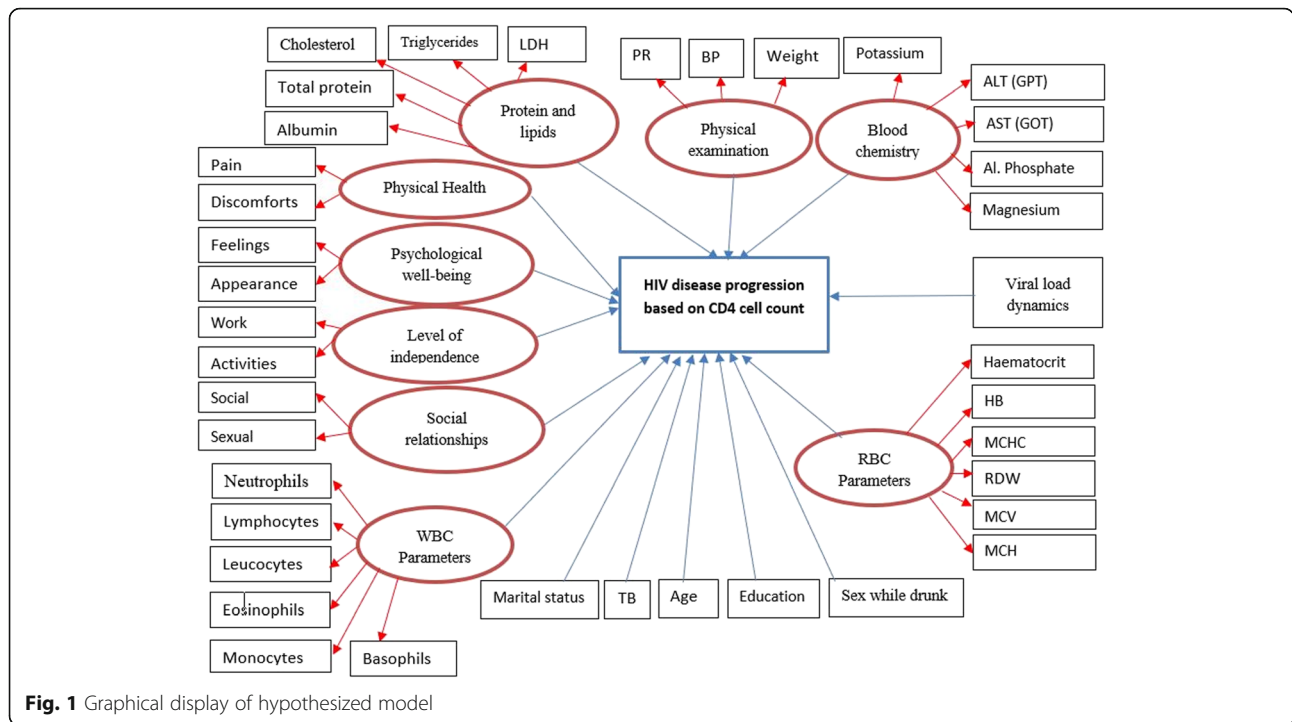


Fig. 1 Graphical display of hypothesized model

Table 1 Clinical parameters and corresponding factor loadings from the rotated factors

Clinical parameters	Principal Components	Variables	Factor loadings	Commutative variations
White blood cell parameters	1. Granulocytes component	Leucocyte	0.925	77%
		Neutrophils	0.936	
		Monocytes	0.635	
	2. Mononuclear component	Lymphocytes	0.838	
		Basophils	0.616	
		3. Eosinophils component	Eosinophils	
Red blood cell parameters	4. Hb and haematocrit component	RBC counts	0.946	81%
		Hb	0.886	
		Haematocrit	0.919	
	5. RBC indices component	MCV	0.953	
		MCH	0.825	
		MCHC	0.521	
Blood chemistry	6. Liver enzyme abnormality component	RDW	-0.592	72%
		ALT (GPT)	0.829	
	7. Electrolyte component	AST (GOT)	0.967	
		Chloride	0.455	
		Sodium	0.994	
		Calcium	0.213	
Protein and lipids	8. Lipid component	Cholesterol	0.971	65%
		LDL	0.917	
		Triglycerides	0.360	
	9. Protein Comp	LDH	-0.769	
		Total protein	0.670	

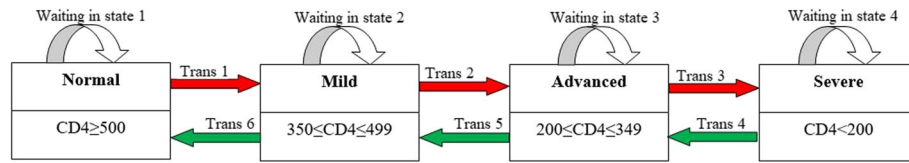


Fig. 2 Progressive four-state model based on CD4 counts. Immunological recovery (green arrows), Immunological deterioration (red arrows) and waiting time (black and white arrows)

disease stage as well as examining the effects of possible factors on the transition intensities. We employed a multi-state model based on multivariate parametric transition specific distribution.

A multi-state process is stated as using a Markov chain of $\{X(t), t \in T\}$ that has finite space, denoted by $S = \{1, 2, 3, 4, \dots, M\}$. Here $T = [0, \tau]$ for $\tau < \infty$. This Markov chains process has an initial probability, denoted by $P(X(0) = m)$, $m \in S$, evolves over time and with a history (H_s) which is containing the stage previously visited, and times of transitions [11, 28]. The multi-state process is differentiated through transition probabilities between two states m and j relative to the given process history is defined as

$$P_{mj}(s, t) = P(X(t) = j | X(s) = m, H_s) \text{ for } m, j \in S \text{ and } s, t \in T, s < t \quad (1)$$

where $P_{mj}(s, t)$ denotes the transition probability of the individual being in state j at time t , given that the individual was in state m at time s and $\sum_{j \in S} P_{mj} = 1$. If a transition probability not only depends on the current state $X(t)$, but also on the entry time of the current state then the process is said to be a semi-Markov process [18]. Thus, the corresponding transition intensities (instantaneous hazard rate) is defined as follows

$$q_{mj}(t) = \lim_{\delta t \rightarrow 0} \frac{P(X(t + \delta t) = j | X(t) = m)}{\delta t} \quad (2)$$

Consequently, in our application, the 4×4 transition intensity matrix $Q(t)$ is defined as

$$Q(t) = \begin{bmatrix} -q_{12}(t) & q_{12}(t) & 0 & 0 \\ q_{21}(t) & -(q_{21}(t) + q_{23}(t)) & q_{23}(t) & 0 \\ 0 & q_{32}(t) & -(q_{32}(t) + q_{34}(t)) & q_{34}(t) \\ 0 & 0 & q_{43}(t) & -q_{43}(t) \end{bmatrix}$$

Note that the rows sum to zero since $\sum_{j \in S} P_{mj} = 1$. The off-diagonal entries are the immunological deterioration and immunological recovery transition intensities). The diagonal entries are defined by $q_{mm}(t) = -\sum_{m \neq j} q_{mj}(t)$. The average length of stay in a single state before making any transitions to either a state associated with higher or lower CD4 count, is estimated by the negative inverse of the m^{th} diagonal entry of $Q(t)$, that is $\frac{-1}{q_{mm}}$. The amount of time spent (L_j), in-state j during the time

period from s to τ , conditional on the patient being in state m at time s is defined as

$$L_j = \int_s^\tau P_{mj}(t) dt$$

In a multi-state process, the transition intensities of a patient moving to state j conditioned on state m , describes and characterizes the multi-state model [28]. A multi-state framework is stated as a combination of parametric transition-specific distributions [18]. The suitable way to do this is, using a “stacked data” representation [12]. Therefore, to examine the effect of covariates on such transitions, we defined a hazard function of the particular transition $m \rightarrow j$.

$$q_{mj}^i(t) = q_0(t) \exp\left(\sum_{k=1}^n X_k^i \beta_{kmj}\right) \quad (3)$$

Where q_{mj}^i represent the transition intensity of patient i from state m to state j , after adjusting a set of covariates. $q_0(t)$ is the baseline intensity and can be modelled parametrically and β_{kmj} is the log-linear effect of the k th covariate (X_k^i) on the transition intensity q_{mj}^i . For this model transition where $m > j$ is defined as immune recovery, then if $m < j$, it is defined as immune deterioration. Any standard survival model fitting software can be used to predict the transition hazard rate, once the cohort data is properly arranged. However, this implies that all transitions used in the same parametric survival model applies. In the next section, relaxing the above-mentioned assumption, allows each transition to use different potential parametric survival models, generating considerable flexibility in model building.

Multi-state transition-specific parametric model

Since it is common to use the semi-parametric multi-state model in modelling transitions intensity, there is a risk that if the PH assumption of some specific transition is not fulfilled, the results derived from this model will have a bias and will be flawed [29, 30]. Although some previous studies tend to turn a blind eye to this defect in their findings due to the ease of a semi-parametric multistate model application and its interpretation, it is essential to use alternative transition-specific parametric

models with a higher degree of reliability for more precise investigations in such cases. Thus, in this paper, we present and compare different multi-state parametric transition-specific models.

Models can be fitted separately to each of the transitions. These can be more efficient computationally and allow the use of appropriate parametric models for each specific transition. This is also important as usually there would be sparse data for some specific transitions, and hence we could fit a simple parametric model (such as an exponential model) for a limited data transition, and a flexible parametric model (such as generalized Gamma, etc) for more information data transitions. This would allow an efficient approach to still make use of all medical cohort datasets [18]. Therefore, when allowing for arbitrary baseline hazard rates between each state, the above model (Eq. 3) generalizes to:

$$q_{mj}^i(t) = q_{mj,0}(t) \exp\left(\sum_{k=1}^n Z_k^i \beta_{kmj}\right) \quad (4)$$

Where $q_0(t) \rightarrow q_{mj,0}(t)$, represents the baseline hazard function for a particular transition $m \rightarrow j$, which allows taking appropriate parametric distribution approaches. In Eq. 3, we assumed a similar parametric model for all transitions. To keep flexibility, we have a vector of patient-level variables included in the particular transition $m \rightarrow j$, Z_{mj}^i , where $Z_{mj}^i \in Z$ (a set of available covariates). This also allows different variables to be included in different transitions. We present different distributional models including Exponential distribution, Weibull distribution, Log-logistic distribution, Log-normal distribution, and Generalized Gamma distribution. This allows for a variety of accelerated failure time models and proportional hazards models, emphasizing the flexibility of this structure.

The presence of time-varying effects is common in many health studies. For instance, in our HIV cohort data, where follow-up is usually over a long period, it is very important to examine the occurrence of time-varying effects. Including time-varying factors within a multi-state structure has received very little attention. In this current multi-state transition-specific parametric modelling method, we extend the structure to an accelerated failure time model which is used for allowing time-varying factors, in either separate modelling approaches or combined modelling approaches fitted to a “stacked data”. Therefore, one of the main advantage of the transition-specific parametric approach is an ease of incorporating time-dependent effects. Thus, the extended transition specific hazard function of the transition $m \rightarrow j$ is given by

$$q_{mj}^i(t) = q_{mj,0}(t) \exp\left(\sum_{k=1}^n Z_k^i \beta_{kmj}(t)\right) \quad (5)$$

Where $\beta_{kmj}(t)$ now allows to vary over time through some standard parametric distribution, relaxing the assumption of proportionality. For example, the baseline hazard function for a Weibull distribution is given by: $q_{mj,0}(t) = \lambda \gamma t^{\gamma-1}$. We can allow the independent variables in the linear predictor of the shape parameter (γ) on the log scale. Similarly, for the generalized gamma AFT model we can allow the three parameters to vary depending on the independent variables. So, with this flexibility, we can apply a parametric survival distribution with any specific transition containing time-varying variables and then, using one of the estimation techniques, we can estimate the parameters.

In this paper, in addition to comparing different parametric multi-state models in modelling transition intensity among different states, the estimates of the selected parametric multistate models were compared with the non-parametric estimate to assess model fit (as discussed by Ieva et al. [31] and Titman and Sharples [32]). The parameters are estimated by maximum likelihood.

Prediction from non-parametric and parametric multistate models

To predict the probability of transitions and probability of staying in each disease state at a fixed time in the future, we calculated the probability of transition matrix $P_{mj}(s, t)$, where $P_{mj}(s, t)$ denotes the transition probability of the individual being in state j at time t , given that the individual was in state m at time s ($s < t$). Under all models, this is calculated by simulating a large number of patients disease states histories from non-parametric and the transition-specific multistate models given the cumulative hazards or covariate-specific hazards for each transition. The implementation was carried out using Stata package (streg and multistate command) and R package (mstate and flexsurvreg codes).

Results

Baseline characteristics of the study population

Table 2 showed the characteristics of the demographic and clinical variables at the baseline of patients followed-up and the observed transitions. All participants were black women ($n = 219$), with a mean age of 26.67 years (standard deviation of 6.9 years). The majority of participants were overweight or obese 137 (62.8%), not with anaemia 208 (95.0%), not co-infected with TB 201 (91.8%), and married or with a stable partner 174 (79.5%). Over half 153 (69.9%) reported having completed Grades 11 of schooling. The maximum transition count was at a mild disease stage, recorded from normal disease stage at 447

Table 2 Baseline Socio-demographic and clinical characteristics in the CAPRISA 002 trials

Variables	Count/Mean (Percentage/SD)
Marital Status, n(%)	
Single/no partner	34(15.5)
Married/stable partner	174 (79.5)
Many partners	11 (5.0)
Educational Status, n(%)	
≤ Grade 8	16 (7.31)
Grade 9–10	50 (22.8)
≥ Grade 11	153 (69.9)
Age Categories, n(%)	
18–20 years	29 (13.2)
21–39 years	178 (81.3)
40–59 years	12 (5.5)
BMI Categories, n(%)	
Underweight	5 (2.3)
Healthy weight	76 (34.9)
Overweight/Obese	137 (62.8)
Baseline immunologic state, n(%)	
< 200 CD4 cells/mm ³	9 (4.1)
200–349 CD4 cells/mm ³	41 (18.7)
350–499 CD4 cells/mm ³	89 (40.6)
≥ 500 CD4 cells/mm ³	80 (36.5)
Contraceptive use, n(%)	
No	40 (18.3)
Yes	179 (81.7)
Anaemia, n(%)	
No	208 (95.0)
Yes	11 (5.0)
TB Co-infection, n(%)	
No	201 (91.8)
Yes	18 (9.2)
Sex acts under the influence of alcohol, n(%)	
No	197 (90.0)
Yes	22 (10.0)
Viral load log₁₀ copies/ml, Mean (SD)	
	4.46 (0.85)
Number of transitions between stage, n(%)	
Transition 1: Normal to Mild	447 (26.1)
Transition 2: Mild to Advance	374 (21.9)
Transition 3: Advanced to Severe	103 (6.03)
Transition 4: Severe to Advance	83 (4.9)
Transition 5: Advance to Mild	318 (18.9)
Transition 6: Mild to Normal	386 (22.6)

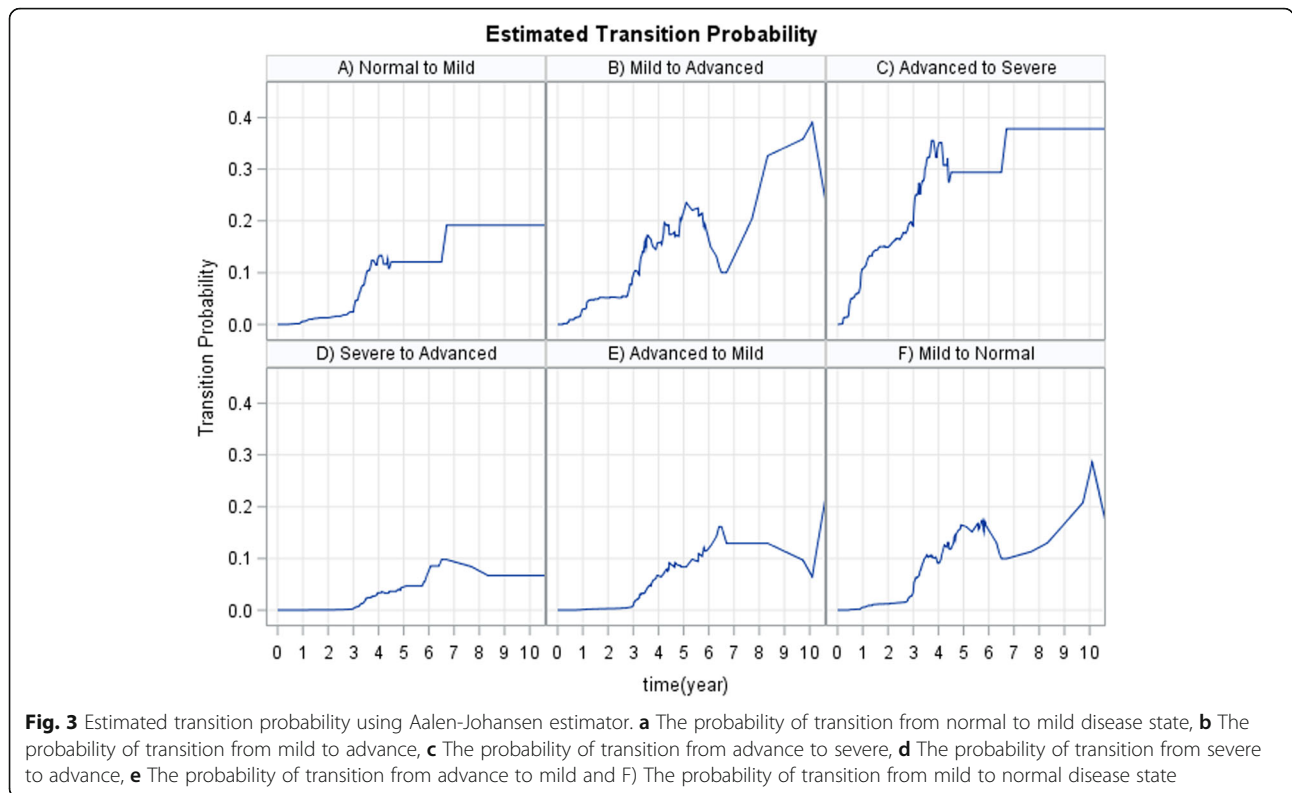
(26.1%). The viral load of the participants ranged from 1.47 log₁₀ copies/ml to 6.81 log₁₀ copies/ml with the first quartile of 3.56 log₁₀ copies/ml, a median of 4.23 log₁₀ copies/ml and the third quartile of 4.79 log₁₀ copies/ml.

Estimated probability of transitions and state-specific duration

The plot in Fig. 3(D-F) displays the probability of transitions from a state of lower CD4 cell count to a state of higher CD4 cell count (immune recovery) in HIV infected women. From this plot, we note that the probability of immune recovery (i.e. from advanced to mild, and severe to advanced stages) did not increase much, whilst the transition probability from a state of higher CD4 cell count, to a state of lower CD4 cell count (immune deterioration), increased with increasing years since entrance into a particular state (Fig. 3a-c). In other words, women who had enrolled with a CD4 cell count of less than 350 cells/mm³ (severe and advanced disease stage) had a far smaller chance on immune recovery, and a considerably greater chance of immune deterioration compared to women with a CD4 cell count of 350 cells/mm³ and more (mild and normal disease stage). The probability of staying in the same disease stage was also computed and represented graphically (see Fig. 4). From these plots, we further note that the probability of staying in the same disease state, decreased with increasing years following entrance into a particular state. The plot also showed that the probability of remaining in severe disease state over time was higher than remaining in other disease states.

Results of transition-specific parametric multi-state model

We applied multi-state transition specific parametric distribution, such as the generalized Gamma, Exponential, Log-logistic, Weibull and Log-Normal models. In Table 3, we present the model selection criteria for each transition and model fitted. We found that the best parametric fitting distribution for transition 4 was the log-normal distribution, based on model selection criteria. The best parametric fitting distribution for transition 1, 2, 3 and 5 were found to be the Log-logistic model, based on model selection criteria. Similarly, the best parametric fitting distribution for transition 6 was found to be the Weibull model, based on the BIC and AIC. Furthermore, we included all possible demographic and clinical variables and assessed the assumptions. For instance, we fitted a Cox regression to each specific transition and then used the interaction between logarithm transformation of time and each explanatory variable to test for proportionality. For each transition, there is a violation of the proportional hazards assumption. Thus, to incorporate time-dependent effects, we used accelerated failure time models.



Results of the multi-state model (see Table 4) showed that hemoglobin abnormality had a significant effect on the time lag for improving from severe to advanced stages of the disease. The interpretation is that the recovery from severe to advanced disease was delayed to a relative rate of 0.71 in those who were anaemic compared with those with normal hemoglobin (aTR = 0.71, 95% CI: 0.55–0.92). Having a high VL, significantly accelerated the

deterioration from normal to mild (aTR = 1.26, 95% CI: 1.19–1.33), mild to advanced (aTR = 1.27, 95% CI: 1.20–1.35) and advanced to severe disease states (aTR = 1.32, 95% CI: 1.18–1.48). Similarly, as the viral load of women increased, the recovery from advanced to mild (aTR = 0.95, 95% CI: 0.89–0.98) and mild to normal disease stage (aTR = 0.94, 95% CI: 0.89–0.99) was delayed to a relative rate of 0.95 and 0.94, respectively. Patients in the middle-

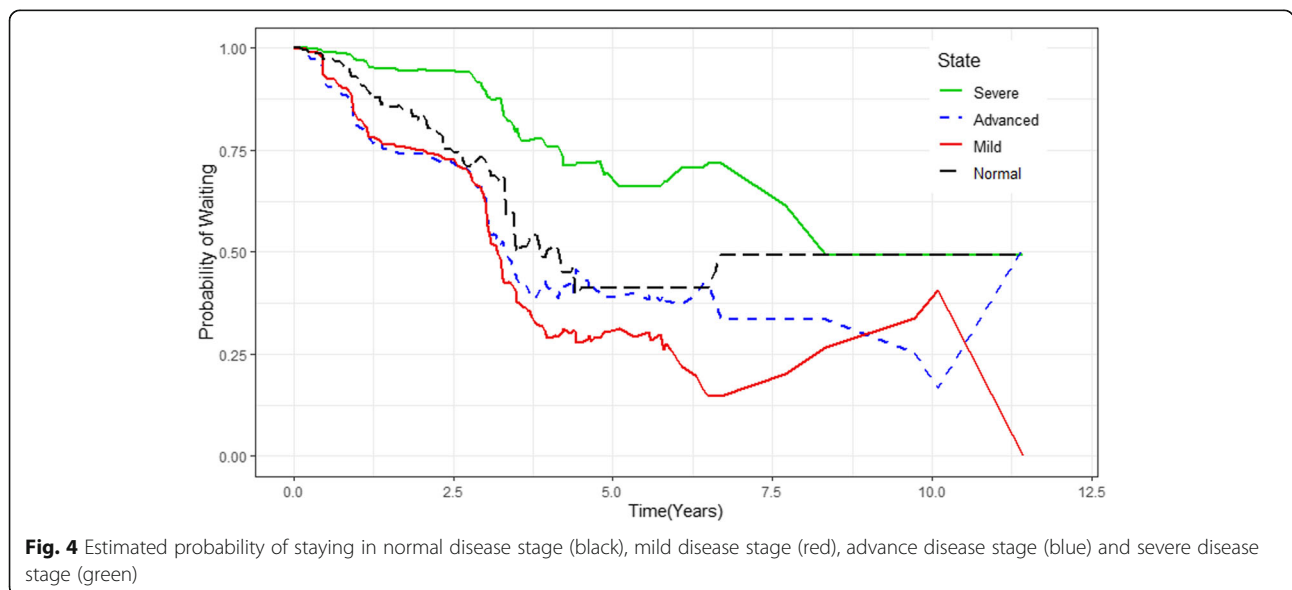


Table 3 Model selection criteria for each parametric model to each transition separately

Transition	Criteria	Exp.	Weibull	Log-logistic	Log-Normal	Gen-Gamma
1	AIC	2021.85	1880.14	1867.61	2114.48	1868.15
	BIC	2168.97	2032.71	2020.17	2270.56	2026.17
2	AIC	1992.44	1841.26	1822.15	1887.68	1824.26
	BIC	2143.07	1997.47	1978.35	2040.24	1986.04
3	AIC	690.28	638.59	633.21	636.67	637.12
	BIC	825.39	778.69	773.32	776.78	782.23
4	AIC	341.37	322.52	375.61	321.32	323.16
	BIC	426.98	411.29	470.90	410.09	415.09
5	AIC	1606.12	1538.31	1524.84	1528.19	1527.79
	BIC	1741.73	1678.95	1665.48	1668.82	1673.46
6	AIC	2185.01	2109.77	2117.94	2114.48	2117.64
	BIC	2335.51	2265.84	2265.01	2270.55	2289.44

aged group were significantly decelerating the recovery from severe to advanced (aTR = 0.90, 95% CI:0.83–0.98) disease stage compared with those patients in the older-aged group. Patients who reported many sex partners were significantly decelerating the recovery from severe to advanced stages (aTR = 0.71, 95% CI:0.65–0.78) compared with those patients with no sexual partner (single). Moreover, patients with high liver enzyme abnormalities significantly decelerated the recovery from advance to mild stage of the disease (aTR = 0.93, 95% CI:0.87–0.98) for the HIV infected patients in the study.

With regard to factors that accelerate recovering or decelerate deterioration, patients without TB co-infection were significantly associated with a shortened time to immunological recovery, compared to those with TB co-infection (particularly from severe to advanced (aTR = 1.51, 95% CI:1.09–1.22) and normal to mild disease stages (aTR = 1.50, 95% CI:1.13–2.01)). Having a high physical health score, significantly accelerated the recovery from severe to advance (aTR = 1.09, 95% CI:1.02–1.16) and from mild to normal disease stage (aTR = 1.07, 95% CI: 1.04–1.10). Having a high level of independence score, significantly decelerated the deterioration from normal to mild (aTR = 0.95, 95% CI:0.94–0.96), and also from advanced to severe disease stage (aTR = 0.97, 95% CI:0.94–0.99). Similarly, having a high social relationship score significantly decelerated the deterioration from normal to mild and from advanced to severe disease stages. As the psychological wellbeing score increased, the deterioration from advanced to severe, from mild to advanced, and from normal to mild stages, was decelerated to a relative rate of 0.95, 0.95, and 0.90, respectively.

Patients with a stable sexual partner, were found to be associated with decelerating immune deterioration from mild to advanced disease stages compared with those with no sex partner. Patients with higher educational levels, were associated with a longer time to immunological

deterioration (particularly from normal to mild disease stages), compared with those women with lower educational levels. Moreover, having a high weight significantly decelerated the deterioration from mild to advanced disease stage (aTR = 0.98, 95% CI:0.97–0.99) and from normal to mild disease stage (aTR = 0.98, 95% CI:0.97–0.99). The time for transition from advance to severe disease stages (aTR = 0.76, 95% CI:0.60–0.98), of women in the study who did not have sex under the influence of alcohol, was decelerated by a factor of 0.76, as compared to those who had sex under the influence of alcohol.

Having a high RBC indices score significantly decelerated the deterioration from mild to advance disease stage (aTR = 0.95, 95% CI:0.91–0.99). Similarly, we noted that having a high mononuclear score significantly accelerated the recovery time from advanced to mild and from mild to normal disease stages, but significantly decelerated the deterioration from normal to mild and from advanced to severe stages of the disease. Finally, we noted that as the latent variable related to granulocytes increased, the time for transition from advanced to mild stages of the disease was accelerated by a factor of 1.08 (aTR = 1.08, 95% CI:1.03–1.14).

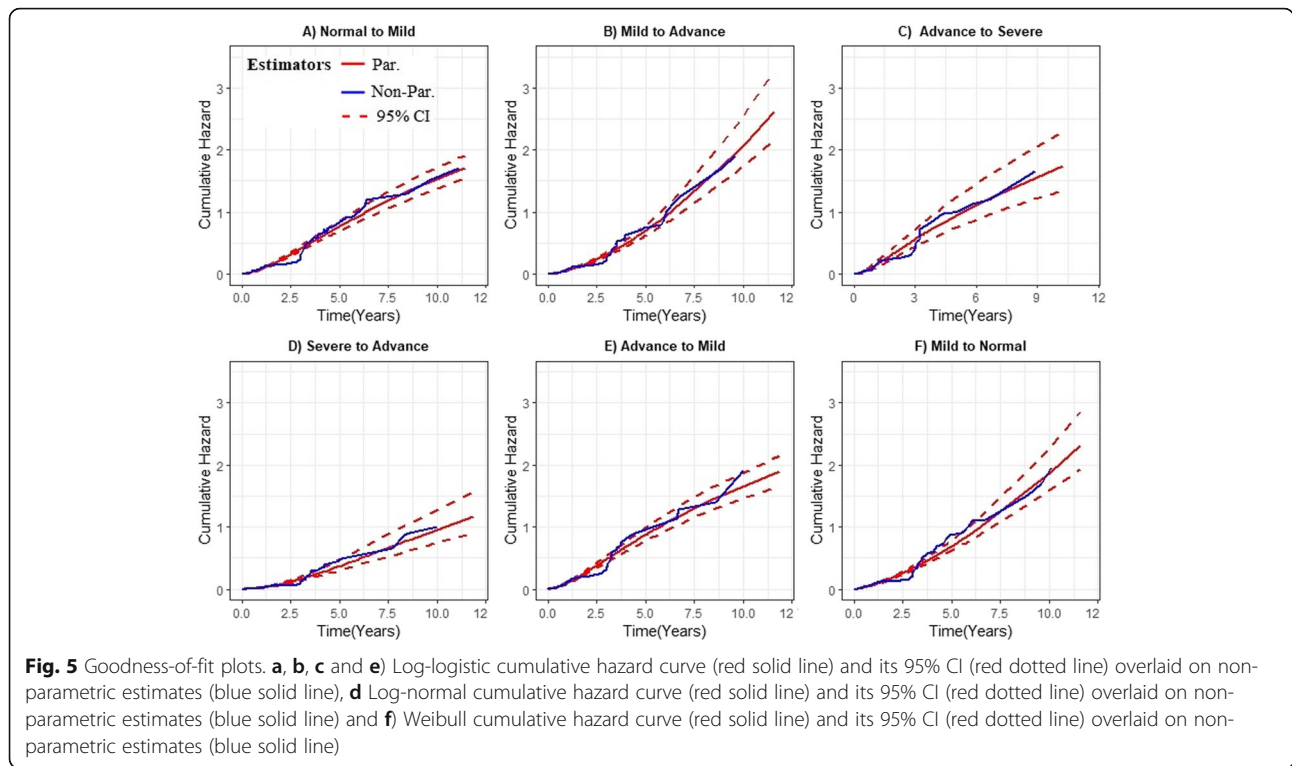
Assessment of goodness of fit of the model

The estimates of the transition specific parametric multi-state model were validated by using graphical methods presented in Fig. 5. The estimates of these multistate models were compared with a non-parametric estimate to assess model fit. The fitted cumulative hazard functions for each of the transitions (obtained from the transition specific parametric multistate model), are shown in Fig. 5(A-F), overlaid on the nonparametric estimator of the transitions (Aalen-Johansen estimator). These plots showed the overall good performances of our transition-specific multi-state models (ie. log-normal model for transition 4, log-logistic model for transition

Table 4 Estimates and the 95% confidence intervals for parameters of multistate transition-specific parametric models

Variables	Transition 1: Normal to Mild, TR(95% CI)	Transition 2: Mild to Advanced, TR (95% CI)	Transition 3: Advanced to Severe, TR (95% CI)	Transition 4: Severe to Advanced, TR (95% CI)	Transition 5: Advanced to Mild, TR (95% CI)	Transition 6: Mild to Normal, TR (95% CI)
Viral Load	1.26 (1.19, 1.33)^a	1.27 (1.20, 1.35)^a	1.32 (1.18, 1.48)^a	1.16 (0.98, 1.37)	0.95 (0.89, 0.98)^a	0.94 (0.89, 0.99)^b
Anaemia: Yes	1.01 (0.91, 1.13)	1.09 (0.99, 1.20)	1.23 (1.03, 1.48)	0.71 (0.55, 0.92)^a	0.94 (0.82, 1.06)	0.92 (0.82, 1.03)
TB: No	1.11 (0.83, 1.50)	1.19 (0.98, 1.44)	1.07 (0.69, 1.68)	1.15 (1.09, 1.22)^a	0.92 (0.73, 1.16)	1.50 (1.13, 2.01)^a
Marital Status: Married	1.01 (0.88, 1.16)	0.80 (0.70, 0.90)^a	0.96 (0.75, 1.23)	0.91 (0.66, 1.27)	0.92 (0.80, 1.07)	0.96 (0.84, 1.10)
Marital Status: Many Partners	1.12 (0.89, 1.42)	0.74 (0.59, 1.09)	0.78 (0.53, 1.14)	0.71 (0.65, 0.78)^a	0.92 (0.72, 1.17)	0.98 (0.75, 1.26)
Education: Grade 9–10	0.74 (0.61, 0.89)^a	0.95 (0.80, 1.12)	1.15 (0.87, 1.51)	1.09 (0.69, 1.73)	0.97 (0.79, 1.18)	1.06 (0.87, 1.29)
Education: Grade ≥ 11	0.82 (0.69, 0.99)^a	0.84 (0.72, 1.04)	0.81 (0.63, 1.04)	1.02 (0.68, 1.51)	0.96 (0.81, 1.15)	1.10 (0.93, 1.32)
Contraceptive use: No	1.03 (0.91, 1.16)	0.99 (0.89, 1.11)	1.02 (0.83, 1.25)	0.99 (0.73, 1.34)	0.85 (0.74, 0.97)^a	0.89 (0.78, 1.01)
Age_Cat: ≤20 years	1.08 (0.95, 1.22)	1.04 (0.92, 1.18)	0.99 (0.78, 1.24)	1.12 (0.78, 1.60)	1.15 (0.99, 1.34)	1.03 (0.91, 1.17)
Age_Cat: 21–39 years	1.11 (0.89, 1.38)	0.87 (0.68, 1.11)	0.74 (0.44, 1.24)	0.90 (0.83, 0.98)^b	1.42 (0.96, 1.94)	1.33 (0.94, 1.65)
Sex while Drunk: No	1.04 (0.89, 1.21)	0.88 (0.75, 1.04)	0.76 (0.60, 0.98)^a	1.07 (1.01, 1.12)^a	0.88 (0.74, 1.04)	1.01 (0.85, 1.21)
Weight	0.99 (0.98, 1.01)^a	0.99 (0.97, 1.02)^a	1.01 (0.99, 1.01)	1.01 (1.00, 1.02)^a	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Independence Score	0.95 (0.94, 0.96)^a	0.99 (0.97, 1.01)	0.97 (0.94, 0.99)^a	0.99 (0.94, 1.03)	1.00 (0.99, 1.02)	1.01 (0.99, 1.02)
Social relationship score	0.98 (0.96, 0.99)^a	0.99 (0.97, 1.01)	0.96 (0.93, 0.99)^a	0.99 (0.94, 1.07)	1.00 (0.98, 1.02)	0.96 (0.95, 1.02)
Physical Health Score	1.13 (0.92, 1.16)	1.03 (1.00, 1.06)	1.01 (0.97, 1.07)	1.09 (1.02, 1.16)^a	1.01 (0.97, 1.03)	1.07 (1.04, 1.10)^a
Psychological Score	0.90 (0.88, 0.93)^a	0.95 (0.93, 0.97)^a	0.95 (0.90, 0.99)^a	0.94 (0.88, 1.01)	0.95 (0.92, 1.03)	0.94 (0.92, 1.04)
Liver Abn Comp	0.97 (0.94, 1.02)	0.99 (0.95, 1.04)	1.01 (0.94, 1.08)	1.00 (0.89, 1.12)	0.93 (0.87, 0.98)^a	1.03 (0.99, 1.08)
RBC indices	0.98 (0.94, 1.03)	0.95 (0.91, 0.99)^b	0.99 (0.91, 1.08)	0.96 (0.85, 1.07)	0.94 (0.89, 1.00)	1.04 (1.00, 1.10)
Hb and Haematocrit Comp	0.97 (0.92, 1.01)	0.99 (0.95, 1.04)	1.01 (0.94, 1.10)	1.04 (0.93, 1.16)	1.01 (0.96, 1.07)	1.04 (0.99, 1.09)
Granulocytes comp	0.95 (0.92, 1.00)	0.95 (0.90, 1.01)	0.99 (0.91, 1.08)	1.01 (0.88, 1.15)	1.08 (1.03, 1.14)^a	1.05 (1.00, 1.10)
Mononuclear comp	0.87 (0.83, 1.09)	0.82 (0.78, 0.87)^a	0.82 (0.74, 0.91)^a	1.12 (0.98, 1.27)	1.14 (1.08, 1.18)^a	1.18 (1.12, 1.23)^a
Eosinophils comp	1.01 (0.97, 1.05)	0.97 (0.93, 1.01)	1.01 (0.92, 1.11)	0.97 (0.89, 1.05)	0.98 (0.93, 1.03)	1.00 (0.96, 1.04)
Electrolyte Comp	0.98 (0.94, 1.03)	1.01 (0.97, 1.06)	1.05 (0.97, 1.14)	0.94 (0.85, 1.05)	1.04 (0.99, 1.10)	1.04 (0.99, 1.08)
Protein Comp	0.96 (0.91, 1.01)	1.03 (0.98, 1.08)	0.99 (0.90, 1.09)	0.93 (0.83, 1.05)	1.03 (0.97, 1.09)	1.00 (0.96, 1.04)
Intercept	0.40 (0.22, 0.71)^a	0.43 (0.27, 0.70)^a	0.23 (0.08, 0.60)^a	0.21 (0.05, 0.92)^a	1.01 (0.57, 1.76)	0.35 (0.20, 0.61)^a

Keys:- ^(a) the coefficient is significant at $\alpha = 0.05$, ^(b) the coefficient is significant at $\alpha = 0.01$, Reference category: Age [> 40 years], education [\leq Grade 8], marital status [Single], Sex while Drunk [Yes], Contraceptive use [Yes], Anemia [No], and TB [Yes]



1, 2, 3 and 5 and Weibull model for transition 6), in terms of fit for the transitions cumulative hazard estimate.

Discussion and conclusion

Intermediate endpoints play a significant role in HIV disease process and progression in many survival analysis studies. Using separate statistical analyses for every single event is the usual approach in many medical studies [33], but does not give the possibility of detecting associations between the endpoints [34, 35]. Utilizing the multi-state statistical model improves understanding of the variation in the possible factors related to HIV disease process and progression [11]. With the increasing use of electronic medical cohort data, allowing the integration of clinical registers and organizational records, there will be considerable opportunities to use multi-state models for expanded modelling of patient profiles across disease progression histories [36]. This has an advantage not only for clinical efficiency modelling, but also for the efficiency of cost modelling due to the parametric methods [37].

Within a multi-state structure, parametric modelling approaches have usually considered distributions that assume the same model for all transitions. In this current work, we have presented a relatively new and under-utilized method for analyzing time to sequential ordinal response variables. We presented the use of the multi-state parametric transition-specific model, allowing a combination of parametric models (such as proportional

hazards and accelerated failure time). We then used an Aalen-Johansen estimator to estimate the probabilities of transitions and probability of staying in the same disease stage. Furthermore, we introduced a structure to allow the flexibility of multi-state transition-specific parametric AFT models, but still enabling the sharing of parameters across the transitions.

The results also showed that patients with higher educational levels were significantly associated with decelerating immune deterioration compared to those patients with lower educational levels. Our finding is concurrent with those from prior reports [38, 39], which noted that patients having a higher educational level significantly associated with a better rate of change of immunological recovery. This might be due to literate patients having work capacity, financial resources, and access to quality health care. Patients in the middle-aged group were significantly decelerating time to immunological deterioration compared with those older patients, a finding that is in accordance with the literature [40, 41], which noted that middle-aged adults experienced higher rates of CD4 recovery than the elderly. Furthermore, patients with many sex partners were significantly decelerating the recovery from severe to advanced stages compared with those with no sexual partner. As has been previously shown [42], patients with higher incidences of sexual risk-taking behavior (such as many sex partners) were significantly associated with low QoL and chronic depression of HIV patients. Chronic depression and low QoL scores are significantly linked

to lower CD4 cell count [43, 44], showing that the effect of many sex partners on incomplete immune recovery, is mediated through depression and QoL.

We have also found that patients with high scores in *quality of life* significantly accelerated the immune recovery time, but significantly decelerated the immune deterioration time of HIV infected patients. This was supported by studies in South Africa: Venter et al. [45] and Ingumbor et al. [46] have found a significant positive association between trends of CD4 cell count recovery and health-related QoL scores of HIV infected patients.

Among the different clinical attributes of patients, patients having higher mononuclear scores significantly accelerated the recovery from advanced to mild and from mild to normal disease stages but significantly decelerated the deterioration from normal to mild and from advanced to severe stages of the disease. Our finding is concurrent with those from study [47], which observed that the increase in the *basophils counts and total lymphocytes counts* corresponded to an increase in the CD4 cell count. We also observed that patients with high liver enzyme abnormalities significantly decelerated the recovery from advance to mild stage of the disease. A previous study also reported a similar finding [48], which observed that lower CD4 count was found to be associated with elevated ALT and AST. Thus, there is a need to monitor ALT and AST levels of the patients before the initiation of cART to reduce side effect concerns. Moreover, having high scores of red blood cell latent and higher granulocytes scores had a significant effect on a shortened time to immunological recovery.

Finally, it can be concluded that transition-specific distributions for multi-state modelling offer a flexible tool for the study of covariate effects on the various transition rates. These models may reveal important biological insights that could otherwise be overlooked when using a model for the marginal survival distribution. The tools are available in terms of methods and software, so hopefully, this paper has helped researchers familiarize themselves with some of these model approaches as well as with the interpretation of multi-state model results (particularly for medical research). In future work, we plan to develop a joint model for multivariate longitudinal biomarkers and multi-state processes in HIV/AIDS diseases on some relevant factors, covariates and a set of latent variables.

Abbreviations

AIDS: Acquired Immune Deficiency Syndrome; *ALT*: Alanine aminotransferase; *ART*: Antiretroviral therapy; *ARV*: Antiretroviral drug; *AST*: Aspartate aminotransferase; *BP*: Blood pressure; *CAPRISA*: Center of the AIDS program of research in South Africa; *HB*: Hemoglobin; *HIV*: Human Immunodeficiency Virus; *LDH*: Lactate dehydrogenase; *MCH*: Mean corpuscular hemoglobin; *MCHC*: Mean corpuscular hemoglobin concentration; *MCV*: Mean corpuscular volume; *OI*: *Opportunistic infections*; *PLHIV*: people living with HIV; *PR*: Pulse rate; *QoL*: Quality of life; *RBC*: Red blood cells; *RDW*: Red cell distribution width; *V B12*: Vitamin B12

Acknowledgments

The Centre for the AIDS Programme of Research in South Africa (CAPRISA team (Dr. Nonhlanhla Yende-Zuma and Dr. Nigel J. Garrett), your kindness in facilitating the data availability and explaining the technicalities in preparing manuscripts for publication is greatly appreciated.

Authors' contributions

ZGD designed the study, collected the data, analyzed the data and wrote the article. TZ designed the study, advised on analysis and edited the manuscript. HM and DN reviewed the study designed and critically edited the manuscript. The authors read and approved the final manuscript.

Authors' information

ZGD is a Ph.D. student and TZ, HM, and DN are senior professors at the University of KwaZulu-Natal.

Funding

This study was supported through the DELTAS Africa Initiative SSACAB (Grant No. 107754/Z/15/Z). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) and the UK Government. DELTAS Africa Initiative SSACAB did not have any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The dataset used and analyzed during the current study is available from the corresponding author on reasonable request.

Ethics approval and consent to participate

All procedures performed in this study were approved by the Research Ethics Committee of the University of KwaZulu-Natal and Centre for the AIDS Programme of Research in South Africa (CAPRISA). Written informed consent was obtained from all participants, and ethical approval for the original study was granted by the University of KwaZulu-Natal (E013/04), the University of Cape Town (025/2004), and the University of the Witwatersrand (M040202).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 1 May 2019 Accepted: 21 May 2020

Published online: 23 June 2020

References

1. Mirzaei M, Poorolajal J, Khazaei S, Saatchi M. Survival rate of AIDS disease and mortality in HIV-infected patients in Hamadan, Iran: a registry-based retrospective cohort study (1997–2011). *Int J STD AIDS*. 2013;24(11):859–66.
2. Hamidi O, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, et al. A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theor Appl Climatol*. 2015; 119(3–4):723–31.
3. Haynes BF, Pantaleo G, Fauci AS. Toward an understanding of the correlates of protective immunity to HIV infection. *Science*. 1996;271(5247):324–8.
4. Sabin CA, Mocroft A, Cozzi Lepri A, Phillips AN. Cofactors and markers of disease progression in human immunodeficiency virus infection. *J R Stat Soc A Stat Soc*. 1998;161(2):177–89.
5. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384(9939):258–71.
6. Organization WH. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children; 2007.
7. Erb P, Battegay M, Zimmerli W, Rickenbach M, Egger M. Effect of antiretroviral therapy on viral load, CD4 cell count, and progression to acquired immunodeficiency syndrome in a community human immunodeficiency virus-infected cohort. *Arch Intern Med*. 2000;160(8): 1134–40.

8. Fauci AS, Bartlett J, Goosby E, Smith M, Kaiser H, Chang S, et al. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. *Ann Intern Med*. 1998;128(12 PART 2):1079–100.
9. Organization WH. Scaling up antiretroviral therapy in resource-limited settings: guidelines for a public health approach: executive summary. Geneva: World Health Organization; 2002.
10. Hamidi O, Poorolajal J, Tapak L. Identifying predictors of progression to AIDS and mortality post-HIV infection using parametric multistate model. *Epidemiol Biostat Public Health*. 2017;14(2):1–9.
11. Hamidi O, Tapak L, Poorolajal J, Amini P. Identifying risk factors for progression to AIDS and mortality post-HIV infection using illness-death multistate model. *Clin Epidemiol Global Health*. 2017;5(4):163–8.
12. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389–430.
13. Schmoor C, Schumacher M, Finke J, Beyersmann J. Competing risks and multistate models. *Clin Cancer Res*. 2013;19(1):12–21.
14. Oliveira RVC, Shimakura SE, Campos DP, Victoriano FP, Ribeiro SR, Velloso VG, et al. Multi-state models for defining degrees of chronicity related to HIV-infected patient therapy adherence. *Cad Saude Publica*. 2013;29:801–11.
15. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res*. 2009;18(2):195–222.
16. Andersen PK. Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Stat Med*. 1988;7(6):661–70.
17. Jackson CH. Flexsurv: a platform for parametric survival modeling in R. *J Stat Softw*. 2016;70(8):1–33.
18. Crowther MJ, Lambert PC. Parametric multistate survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. *Stat Med*. 2017;36(29):4719–42.
19. Cox C, Chu H, Schneider MF, Munoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*. 2007;26(23):4352–74.
20. van Loggenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, et al. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS One*. 2008;3(4):e1954.
21. Dessie ZG, Zewotir T, Mwambi H, North D. Modelling of viral load dynamics and CD4 cell count progression in an antiretroviral naive cohort: using a joint linear mixed and multistate Markov model. *BMC Infect Dis*. 2020;20(1):1–14.
22. Dessie ZG, Zewotir T, Mwambi H, North D. Modeling viral suppression, viral rebound and state-specific duration of HIV patients with CD4 count adjustment: parametric multistate frailty model approach. *Infect Dis Ther*. 2020;9(2):1–22.
23. Mlisana K, Naicker N, Werner L, Roberts L, Van Loggenberg F, Baxter C, et al. Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *J Infect Dis*. 2012;206(1):6–14.
24. Group W. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998;28(3):551–8.
25. Dessie ZG, Zewotir T, Mwambi H, North D. Multivariate multilevel modeling of quality of life dynamics of HIV infected patients. *Health Qual Life Outcomes*. 2020;18(1):1–14.
26. Dessie ZG, Zewotir T, Mwambi H, North D. Modelling immune deterioration, immune recovery and state-specific duration of HIV-infected women with viral load adjustment: using parametric multistate model. *BMC Public Health*. 2020;20(1):1–13.
27. Byrne BM. Factor analytic models: viewing the structure of an assessment instrument from three perspectives. *J Pers Assess*. 2005;85(1):17–32.
28. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956–71.
29. Collett D. Modelling survival data in medical research. London:Chapman and Hall; 2015.
30. Hosmer DW Jr, Lemeshow S, May S. Applied survival analysis: regression modeling of time-to-event data. New York: Wiley; 1999.
31. Ieva F, Jackson CH, Sharples LD. Multi-state modelling of repeated hospitalisation and death in patients with heart failure: the use of large administrative databases in clinical epidemiology. *Stat Methods Med Res*. 2017;26(3):1350–72.
32. Titman AC, Sharples LD. Model diagnostics for multi-state models. *Stat Methods Med Res*. 2010;19(6):621–51.
33. Andersen PK, Esbjerg S, Sørensen TI. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Stat Med*. 2000;19(4):587–99.
34. Eulenburg C, Schroeder J, Obi N, Heinz J, Seibold P, Rudolph A, et al. A comprehensive multistate model analyzing associations of various risk factors with the course of breast cancer in a population-based cohort of breast cancer cases. *Am J Epidemiol*. 2016;183(4):325–34.
35. Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res*. 2002;11(2):91–115.
36. Asaria M, Walker S, Palmer S, Gale CP, Shah AD, Abrams KR, et al. Using electronic health records to predict costs and outcomes in stable coronary artery disease. *Heart*. 2016;heartjnl-2015-308850.
37. Crowther MJ, Lambert PC. stgenreg: a Stata package for general parametric survival analysis; 2013;53(12):1–17.
38. Jiang H, Xie N, Cao B, Tan L, Fan Y, Zhang F, et al. Determinants of progression to AIDS and death following HIV diagnosis: a retrospective cohort study in Wuhan, China. *PLoS One*. 2013;8(12):e83078.
39. Seyoum A, Temesgen Z. Joint longitudinal data analysis in detecting determinants of CD4 cell count change and adherence to highly active antiretroviral therapy at Felege Hiwot teaching and specialized hospital, North-West Ethiopia (Amhara region). *AIDS Res Ther*. 2017;14(1):14.
40. Saracino A, Zaccarelli M, Lorenzini P, Bandera A, Marchetti G, Castelli F, et al. Impact of social determinants on antiretroviral therapy access and outcomes entering the era of universal treatment for people living with HIV in Italy. *BMC Public Health*. 2018;18(1):870.
41. Wong NS, Chan KCW, Cheung EKH, Wong KH, Lee SS. Immune recovery of middle-aged HIV patients following antiretroviral therapy: an observational cohort study. *Medicine*. 2017;96(28):e7493.
42. Vu T, Boggiano V, Tran B, Nguyen L, Tran T, Latkin C, et al. Sexual risk behaviors of patients with HIV/AIDS over the course of antiretroviral treatment in northern Vietnam. *Int J Environ Res Public Health*. 2018;15(6):1106.
43. Ickovics JR, Hamburger ME, Vlahov D, Schoenbaum EE, Schuman P, Boland RJ, et al. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV epidemiology research study. *Jama*. 2001;285(11):1466–74.
44. Rivera-Rivera Y, Vázquez-Santiago FJ, Albino E, MdC S, Rivera-Amill V. Impact of depression and inflammation on the progression of HIV disease. *J Clin Cellular Immunol*. 2016;7(3):423.
45. Venter E, Gericke GJ, Bekker P. Nutritional status, quality of life and CD4 cell count of adults living with HIV/AIDS in the Ga-Rankuwa area (South Africa). *S Afr J Clin Nutr*. 2009;22(3):124–9.
46. Ingumbor J, Steward A, Holzemer W. Comparison of the health related quality of life, CD4 count and viral load of AIDS patients with HIV who have been on treatment for 12 months in rural South Africa; 2013.
47. Tinarwo P, Zewotir T, Yende-Zuma N, Garrett NJ, North D. An evaluation to determine the strongest CD4 count covariates during HIV disease progression in women in South Africa. *Infect Dis Ther*. 2019;8(2):269–84.
48. Shiferaw MB, Tulu KT, Zegeye AM, Wubante AA. Liver enzymes abnormalities among highly active antiretroviral therapy experienced and HAART naive HIV-1 infected patients at Debre Tabor hospital, North West Ethiopia: a comparative cross-sectional study. *AIDS Res Treatment*. 2016; 2016:1–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

