

RESEARCH ARTICLE

Open Access



Modelling of viral load dynamics and CD4 cell count progression in an antiretroviral naive cohort: using a joint linear mixed and multistate Markov model

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

Abstract

Background: Patients infected with HIV may experience a succession of clinical stages before the disease diagnosis and their health status may be followed-up by tracking disease biomarkers. In this study, we present a joint multistate model for predicting the clinical progression of HIV infection which takes into account the viral load and CD4 count biomarkers.

Methods: The data is from an ongoing prospective cohort study conducted among antiretroviral treatment (ART) naïve HIV-infected women in the province of KwaZulu-Natal, South Africa. We presented a joint model that consists of two related submodels: a Markov multistate model for CD4 cell count transitions and a linear mixed effect model for longitudinal viral load dynamics.

Results: Viral load dynamics significantly affect the transition intensities of HIV/AIDS disease progression. The analysis also showed that patients with relatively high educational levels ($\beta = -0.004$; 95% confidence interval [CI]: -0.207, -0.064), high RBC indices scores ($\beta = -0.01$; 95%CI: -0.017, -0.002) and high physical health scores ($\beta = -0.001$; 95%CI: -0.026, -0.003) were significantly associated with a lower rate of viral load increase over time. Patients with TB co-infection ($\beta = 0.002$; 95%CI: 0.001, 0.004), having many sex partners ($\beta = 0.007$; 95%CI: 0.003, 0.011), being younger age ($\beta = 0.008$; 95%CI: 0.003, 0.012) and high liver abnormality scores ($\beta = 0.004$; 95%CI: 0.001, 0.01) were associated with a higher rate of viral load increase over time. Moreover, patients with many sex partners ($\beta = -0.61$; 95%CI: -0.94, -0.28) and with a high liver abnormality score ($\beta = -0.17$; 95%CI: -0.30, -0.05) showed significantly reduced intensities of immunological recovery transitions. Furthermore, a high weight, high education levels, high QoL scores, high RBC parameters and being of middle age significantly increased the intensities of immunological recovery transitions.

Conclusion: Overall, from a clinical perspective, QoL measurement items, being of a younger age, clinical attributes, marital status, and educational status are associated with the current state of the patient, and are an important contributing factor to extend survival of the patients and guide clinical interventions. From a methodological perspective, it can be concluded that a joint multistate model approach provides wide-ranging information about the progression and assists to provide specific dynamic predictions and increasingly precise knowledge of diseases.

Keywords: Factor analysis, Latent variables, Mixed effect model, Multistate model, Quality of life

* 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 infection is one of the leading causes of death from infectious diseases globally and remains a serious *global public health issue* [1, 2]. AIDS, the last progress stage of HIV infection, leads to severe damage to the body's immune system [3]. The progression of HIV/AIDS is highly variable between populations and individuals and is determined by immunological, genetic, environmental and virological factors [4]. CD4 cell and viral load counts have remained the two strongest correlates and surrogate markers of HIV disease progression regularly used in the clinical setting to monitor the infection [5]. Of the two, CD4 count is a more accurate biomarker of the stage of HIV and is recommended by all guidelines of HIV management [6]. Some studies also argue that CD4 cell count predicts clinical information (event time data) [7] whereas HIV viral load trajectories largely determine the time from initial infection to AIDS: high initial viral load is a marker for rapid progression [8]. Further, many studies report that there is a relationship between these biomarkers, often explaining the disease progression of one biomarker according to the other [9–11].

HIV disease patients progress through normal, mild, advanced and severe clinical stages, and the disease diagnosis can be considered as one of these states. So, instead of modeling of CD4 count without binning into categories, the CD4 count progression should be modeled as a multi-state process, which takes into account viral load dynamics, to explain the development of such diseases. Simultaneous modelling of these two biomarkers may be a better way to capture the complete disease process and progression of HIV/AIDS than individual modelling. It is also essential to understand and predict accurately the dynamic evolution of the disease, which is of particular relevance to physicians who need to distinguish the different types of intermediate events in order to properly adapt treatment plans. Thus, the research focus of the current study is to model viral load dynamics and CD4 cell count progression in an ART-naive cohort.

Previously, a multi-state Markov process modelling of disease progression of HIV/AIDS has been examined by several authors [12–15]. In particular, Binquet et al. [12] estimated the impact of CD8 cell count, weight loss, drug use, gender, viral load and haemoglobin on the progression of HIV/AIDS. Oliveira et al. [14] studied the degrees of chronicity of HIV/AIDS and went further to examine the impact of covariates; adherence, disease duration, and age on CD4 cell count progression. Recently, Shoko et al. [15] and Chikobvu and Shoko [13], also studied disease progression of HIV/AIDS. However, no previous study has jointly modeled the transition intensity of sequential adversity of the events and longitudinal viral load biomarkers of HIV/AIDS. In addition to that, although the factors related to disease progression of HIV are multiple and

complex, very few studies have directly examined the effects of several clinical attributes (ie: white blood cell parameters, RBC parameters, blood chemistry parameters and QoL domain scores) on both viral load trajectories and transitions intensity of sequential events. This study thus gives an insight on presenting a joint multistate model for predicting the clinical progression of HIV infection which takes into account the viral load biomarker, to study several factors that may affect the transition intensities between sequential events.

During the last few decades, some joint models have been developed to study longitudinal biomarker and competing risks using scleroderma lung study data [16, 17]. Han et al. [18] and Cai et al. [19] also provided a joint analysis of longitudinal markers and recurrent events using Epileptic seizure data and Biocard data, respectively. Chi and Ibrahim [20] extended the joint modeling framework to multivariate repeated measurements of longitudinal biomarkers and multivariate survival data using breast cancer data. In contrast, the present work is focusing on joint modeling of longitudinal biomarker and time to transitions between sequential states with application to HIV/AIDS. This is an important aspect that has not been considered in many medical studies particularly in HIV/AIDS cohort studies. This study contains a multi-state model and a linear mixed model, both linked by shared random effects. The goals of joint modelling are to improve inference for the time to transition between multistate events, whilst taking into account the endogenous nature of a biomarker [21]. It also examines the association between the two correlated outcomes [22]. This will lead to an improvement in estimation for a longitudinal biomarker response variable, subject to an informative dropout mechanism that is not of direct interest [23] and will improve inferences as compared to the separate analysis of the two response variables [24].

The goals of this paper are thus threefold. Firstly, we seek to present a joint model of a multistate and longitudinal process, briefly describing previous approaches. Secondly, we seek to apply this model to describe the trajectory of HIV in ART-naive South African patients, so as to aid a deeper understanding of the HIV disease progression process and discover possible factors that influence disease progression. Finally, we will compare the parametric estimator (estimated from the joint multistate model) to nonparametric methods.

Methods

Data description

The data is from an ongoing prospective cohort study conducted by the Centre for the AIDS Program of Research in South Africa (CAPRISA) among ART naïve HIV-infected women. The original study, which started in 2004, enrolled a cohort of HIV uninfected women

whose age was greater than 18 years with the aim to describe immunologic, clinical and virologic characteristics of HIV-1 disease [25]. In this study, study enrollment was conducted from August 2004 to December 2017. The participant who seroconverted during the HIV uninfected stage of CAPRISA_002 and other CAPRISA prevention and sero-incidences trials (including the CAPRISA_004 trials), were enrolled into the Acute HIV Infection phase, and then followed-up during chronic infection and up to ART initiation. Participants were recruited at two sites in KwaZulu-Natal-South Africa, a rural site in Vulindlela and an urban site in the city of Durban. Women without well documented estimated date of HIV infection, and those who did not have at least two follow-up clinical attribute measurements were excluded in this analysis. Finally, 219 participants were included in the study. Further information about the above mentioned ongoing prospective HIV cohort study (CAPRISA_002), including women eligibility criteria and the enrollment procedures were reported in [25–27].

Variables and measurements

Once HIV diagnosis was confirmed, participants were followed-up for a maximum of 11 years up to cART initiation. CAPRISA initially enrolled HIV-negative (phase I) women into different study cohorts. The women who seroconverted were enrolled into acute infection (i.e. phase II: weekly visits up to 3 months post-infection), then into early infection (i.e. phase III: monthly visits from 3 to 12 months) subsequently into established infection (i.e. phase IV: quarterly visits for more than 12 months) and afterward on cART (phase V). Patients were offered to start cART according to South African (SA) treatment guidelines. SA guidelines for ART eligibility have changed over time during the study period: severe clinical disease or CD4 cell count ≤ 200 cells/mm³ until 2010; expansion to CD4 cell count ≤ 350 cells/mm³ for patients with TB and pregnant women from 2010; further to all SA HIV infected patients with CD4 cell count ≤ 350 cells/mm³ until 2015. In 2013, the WHO recommended treatment at ≤ 500 cells/mm³, which was then expanded to universal

treatment in 2015. South Africa moved to a threshold of ≤ 500 cells/mm³ in 2016. For the purpose of this study, samples of immunological, virological and clinical attributes (such as viral load, WBC parameters, RBC parameters, blood chemistry parameters, CD4 cell count, etc.) were measured at each visit (phase II-IV). These longitudinal immunologic, virologic and clinical measurements, were recorded for several followed-up visits.

There was a total of 8760 follow-up visits recorded from 219 HIV infected black women with a median age of 25 years (Interquartile range, IQR, 22–30). Of these patients, 9.2% of them were co-infected with TB. Over two-thirds (69.9%) reported having completed grades 11/12 of schooling. The follow-up time of the participants ranged from 0-year to 11.10 years with a first quartile of 0.37 years, a median of 1.50 years and a third quartile of 4.04 years. The median baseline CD4 count of the participants included in the analysis was 519.0 cells/mm³ (IQR 419–655.5 cells/mm³) and the median CD4 counts at ART initiation was 408.0 cells/mm³ (IQR 307.0–587.0 cells/mm³). The log₁₀ copies/ml VL count of the participants ranged from 1.47 to 6.81 with a first quartile of 3.84, a median of 4.46 and a third quartile of 5.06.

The main outcome variables in this current paper are the transition times between multiple states and the longitudinal viral load dynamics (baseline VL and long-term viral load change). In line with the World Health Organization (WHO) immunological classification criteria, which is used to assess the degree of severity of the HIV infection of patients, we have categorized CD4 count. Medical practitioners and health workers also use such WHO classifications to monitor HIV infected patients. These HIV infection stages are defined as no adverse events ($CD4 \geq 500$), mild ($350 \leq CD4 < 500$), advanced ($200 \leq CD4 < 350$) and severe ($CD4 < 200$) [6]. (See Fig. 1).

The effect of numerous possible factors on the transition times of sequential adverse events and longitudinal viral load biomarkers were evaluated, including (1) demographic variables, (2) risk variables, (3) past opportunistic infections, (4) health-related quality of life domain scores and (5) clinical attributes (see Fig. 2).

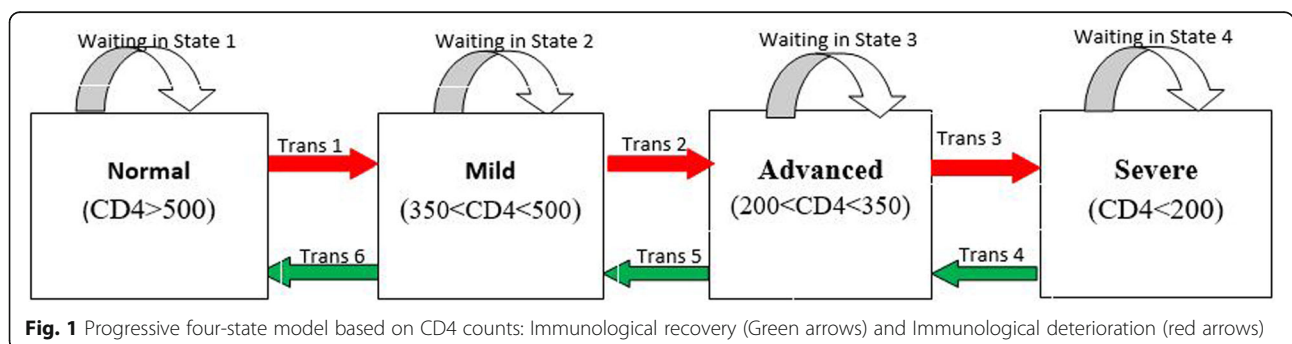


Fig. 1 Progressive four-state model based on CD4 counts: Immunological recovery (Green arrows) and Immunological deterioration (red arrows)

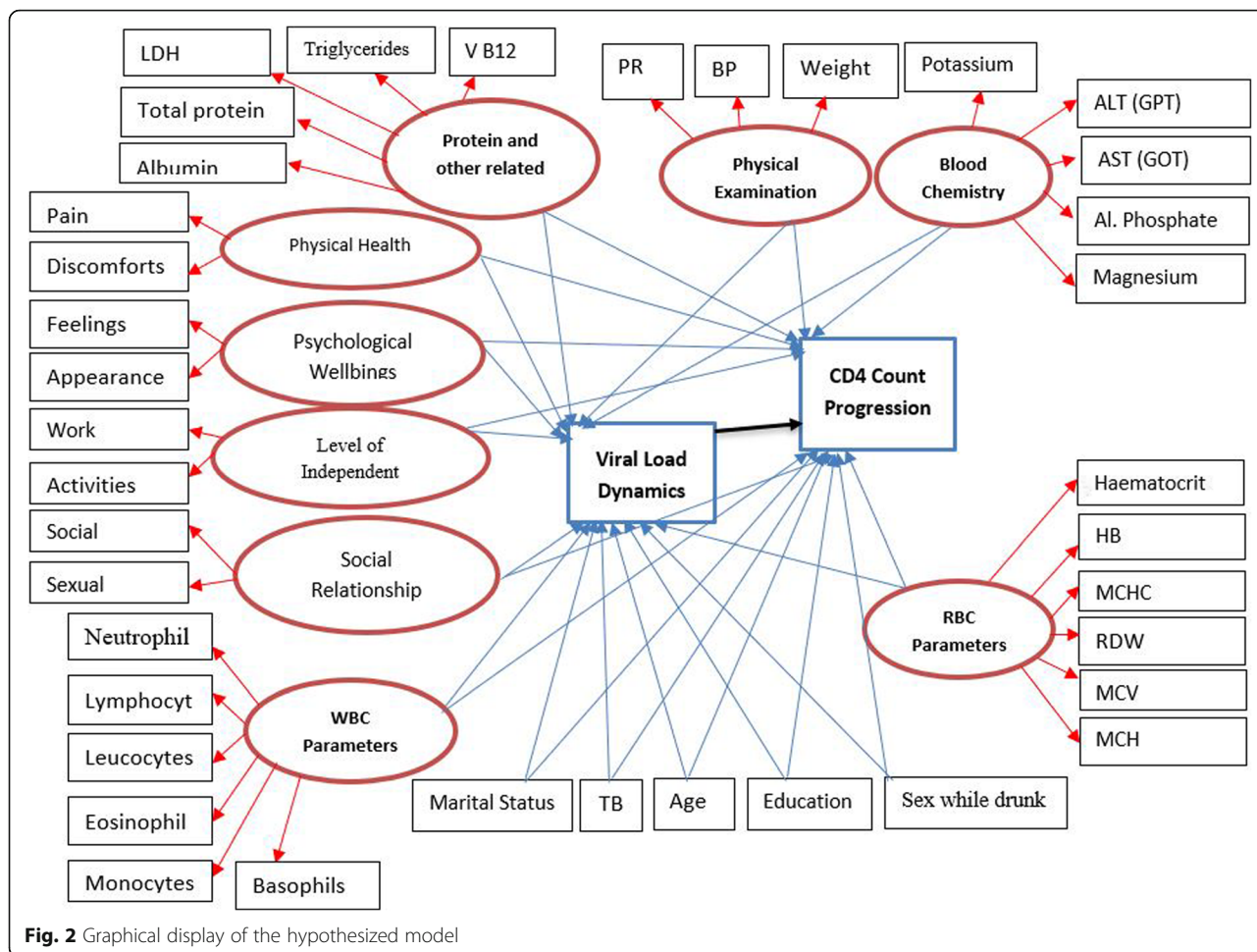


Fig. 2 Graphical display of the hypothesized model

Statistical method

Factor analysis

Since our data have a large number of clinical variables, we used exploratory factor analysis in order to group and minimize the number of variables. Factor analysis 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 [28]. A maximum likelihood extraction method with varimax rotation was used. Factor loadings describe the relationship of each clinical variable with each factor. The Factor loading is considered strong if greater than 0.6, moderate if 0.4–0.6, and weak if less than 0.4 [29]. Each observation was assigned a score for each rotated factor on the basis of the loadings of the subject’s original variable levels. Accordingly, from the 24 clinical variables in the study, we managed to group them in order to create 9 latent variables, defined as granulocytes components, mononuclear components, eosinophils component, RBC component, red blood cell indices, liver abnormality component, electrolyte component, lipid component and protein component. (See Table 1).

Joint multistate model formulation

Let a Markov process $\{S(t), t \in T\}$ that has finite space, denoted by $=\{1, 2, 3, 4\}$, be a representation of the transition process, where for each patient, a multistate process is observed. Here $T = [0, \tau]$ for $\tau < \infty$. This Markov process has an initial probability, denoted by $P(S(0) = m)$, $m \in E$, which evolves over time and with a history (H_E) , which contains the state previously visited, durations and times of transitions [30, 31]. The multi-state process is defined through transition probabilities between two states m and j relative to the given process history, as:

$$P_{mj}(z, t) = P(S(t) = j | S(z) = m, H_E) \text{ for } m, j \in E \text{ and } z, t \in T, z < t.$$

$P_{m,j}(z,t)$ thus denotes the transition probability of the patient being in state j at time t , given that the patient was in state m at time z .

In the current study, we consider a joint model for the long-term trends of viral load biomarker and transition times into the different adverse effect states based on CD4 cell counts. This model consists of two related sub-models: a Markov multistate model for the transition

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

WBC parameters	Rotated Factors Loadings		
	1. Granulocytes component	2. Mononuclear component	3. Eosinophils component
Leucocyte	0.925	0.282	0.146
Neutrophils	0.936	-0.158	0.022
Total lymphocytes	0.226	0.838	-0.109
Monocytes	0.635	0.417	-0.032
Eosinophils	0.085	0.058	0.947
Basophils	-0.035	0.616	0.339
RBC Parameters	Rotated Factors Loadings		
	4. Hb and Haematocrit component	5. Red blood cell indices component	
RBC counts	0.946	-0.130	
Hb	0.886	0.439	
Haematocrit	0.919	0.366	
MCV	0.075	0.953	
MCH	0.024	0.825	
MCHC	0.201	0.521	
RDW	-0.382	-0.592	
Blood Chemistry	Rotated Factors Loadings		
	6. Liver enzyme abnormality component	7. Electrolyte component	
Chloride	-0.023	0.455	
Alkaline	0.174	0.032	
ALT (GPT)	0.829	-0.073	
AST (GOT)	0.967	-0.122	
Sodium	0.103	0.994	
Calcium	-0.020	0.213	
Protein and related	Rotated Factors Loadings		
	8. Lipid component	9. Protein component	
Cholesterol	0.971	0.027	
LDL	0.917	-0.129	
Triglycerides	0.360	0.341	
LDH	0.052	-0.769	
Total protein	-0.009	0.670	

times data and a linear mixed effect model for the longitudinal data (longitudinal measurements of the viral load marker) process, both linked by a function of the shared random effects. HIV-infected ART-naive patients may experience many CD4 cell count fluctuations mostly before ART initiations. This suggests that transition times should be modelled by a multistate process. Figure 1 shows a flow diagram of the multi-state model.

Longitudinal submodel

To model the long-term trend of viral load dynamics, we used a linear mixed effect model. Let y_{it} ($i = 1, 2, \dots, n; t = 1, 2, \dots, n_i$) be the viral load count of HIV infected patient i at follow-up visit time t . Under Gaussian assumptions,

the longitudinal viral load dynamics y_{it} is modelled using the following model which is proposed by [32].

$$y_i(t) = x_i'(t)\beta + z_i'(t)b_i + \varepsilon_i(t) \tag{1}$$

where $x_i(t)$ and $z_i(t)$ represents a vector of potentially time-varying covariates corresponding to the vector of fixed effects β and random effects b_i , respectively. The model assumes that random effects are distributed as multivariate normal with mean $\mathbf{0}$ and covariance matrix \mathbf{D} . We also assume that the errors are independent and follow a normal distribution, $\varepsilon_i(t) \sim N(0, \sigma)$, and that b_i and $\varepsilon_i(t)$ are independent [33]. The normality assumption of longitudinal biomarker measures was checked via

the Q-Q plot using the transformed residuals errors based on the fitted joint model [34]. The random-effects represent the effects of each subject on the longitudinal measures that cannot be explained by the observed covariates.

Multistate submodel

In this sub-section, we model the transition intensities defined by CD4 cell count which takes into account the viral load biomarker. The standard approach of modeling the event history data [35], in which, for patient i , the instantaneous hazard rate of moving to state j conditioned on state m is defined as follows

$$h_{mj}(t) = \lim_{\delta t \rightarrow 0} \frac{P(S(t + \delta t) = j | S(t) = m)}{\delta t} \tag{2}$$

$$h_{mj}(t; \mathbf{X}) = h_{mj,0}(t) \exp(\mathbf{X}'_{mj} \boldsymbol{\alpha}_{mj}) \tag{3}$$

where h_{mj} is the transition intensity for $m = 1, \dots, 4$ and $j = m \pm 1$ sequential states defined by CD4 counts. $h_{mj,0}(t)$ is the baseline intensity from state m to state j and can either be left unspecified or modelled parametrically, \mathbf{X}_{mj} representing a matrix of covariates. $\boldsymbol{\alpha}_{mj}$ is the effect of the covariates on the hazard intensity h_{mj} . For this model, if $m < j$, then a transition from m to j is defined as an immunological deterioration transition and a transition where $m > j$ is defined as immunological recovery. In the extension of joint modelling, we used a Markov multistate model that takes into account the viral load biomarker through the shared subject-specific random effects \mathbf{b}_i . Thus, the instantaneous hazard, for a patient moving from state $m \in E$ to state $j \in E$, at time t will be modeled as:

$$h_{mj}(t; \mathbf{b}) = h_{mj,0}(t) \exp(\mathbf{X}'_{mj} \boldsymbol{\alpha}_{mj} + \mathbf{W}_{mj}(\mathbf{b}, t) \boldsymbol{\gamma}_{mj}). \tag{4}$$

$\mathbf{W}_{mj}(\mathbf{b}, t)$ defines a multivariate function that represents the dependence structure between the longitudinal viral load trajectory and multistate CD4 count transitions. We can choose $\mathbf{W}_{mj}(\mathbf{b}, t)$ as the true baseline and slope of the viral load of HIV-infected patients. $\boldsymbol{\gamma}_{mj}$ represents the impact of the longitudinal viral load biomarker on the CD4 cell count progressive form state m to state j .

Association parameters

As one of our objectives is to examine the relationship between transient states defined by CD4 cell count progression and the longitudinal viral load dynamics, we followed the approach adopted by Ferrer et al. [36], where the two sub-models are linked by $\mathbf{W}_{mj}(\mathbf{b}, t)$. This dependence function was the same for all the CD4 cell count transitions and resulted in combining the true

baseline and rate of change of the viral load fitted to capture the association between viral load dynamics and the instantaneous hazard rate to transitions between sequential events.

Estimation

To estimate the parameters in the shared parameter model, we used maximum likelihood estimation. A joint model can be estimated using the independence between the multi-state and longitudinal processes conditionally on the random effects. So, the likelihood for all the observed longitudinal viral load trajectory (\mathbf{y}_i) and multi-state CD4 count transitions (\mathbf{S}_i) is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{S}) &= \prod_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{S}_i) \\ &= \prod_{i=1}^n \int f_{VL}(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f_{CD4}(\mathbf{S}_i | \mathbf{b}_i, \boldsymbol{\theta}) f_{\mathbf{b}}(\mathbf{b}_i | \boldsymbol{\theta}) d\mathbf{b}_i \end{aligned} \tag{5}$$

$\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma, \mathbf{D})$, representing all the parameters contained in Eqn. (1) and (4); and the corresponding density functions, is then given by

$$\begin{aligned} f_{VL}(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) &= \frac{1}{(2\pi\sigma^2)^{n_i/2}} \exp\left\{-\frac{\|\mathbf{y}_i - \mathbf{x}'_i \boldsymbol{\beta} - \mathbf{z}'_i \boldsymbol{\gamma}\|^2}{2\pi\sigma^2}\right\} \\ f_{CD4}(\mathbf{S}_i | \mathbf{b}_i, \boldsymbol{\theta}) &= \prod_{r=0}^{k-1} \left\{ P_{S(T_r), S(T_{r+1})}(T_r, T_{r+1} | \mathbf{b}) * h_{S(T_r), S(T_{r+1})}(T_{r+1} | \mathbf{b})^{\delta_{r+1}} \right\} \\ &= \prod_{r=0}^{k-1} \left\{ \exp\left(-\int_{T_r}^{T_{r+1}} h_{S(T_r), S(T_{r+1})}(\mathbf{U} | \mathbf{b}) du\right) * h_{S(T_r), S(T_{r+1})}(T_{r+1} | \mathbf{b})^{\delta_{r+1}} \right\} \\ f_{\mathbf{b}}(\mathbf{b}_i | \boldsymbol{\theta}) &= \frac{1}{(2\pi)^{q/2} |\mathbf{D}|^{1/2}} \exp\left\{-\frac{\mathbf{b}'_i \mathbf{D}^{-1} \mathbf{b}_i}{2}\right\} \end{aligned}$$

where $f_{VL}(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta})$ representing the conditional function for the longitudinal viral load trajectory, $\|\mathbf{x}\|$ represents the Euclidean norm of a vector, $f_{CD4}(\mathbf{S}_i | \mathbf{b}_i, \boldsymbol{\theta})$ is the conditional function for the multistate CD4 count transitions, k denotes the number of transitions, $P_{S(T_r), S(T_{r+1})}(T_r, T_{r+1})$ represents the probability of remaining in state $S(T_r)$ between times T_r and T_{r+1} , $h_{S(T_r), S(T_{r+1})}(T_{r+1})$ denotes the transition intensity to state $S(T_{r+1})$, δ denotes transition indicator and $f_{\mathbf{b}}(\mathbf{b}_i | \boldsymbol{\theta})$ the joint distribution of the random effects.

Since the dimension of the random effects \mathbf{b}_i is often high and the density functions of the multi-state Markov process can be highly complicated, evaluation of the above integral (Eqn. 5) can be a major challenge and very intensive. Therefore, for the computations of the above function (eq. 5), a multi-step pseudo-adaptive Gaussian-hermit quadrature rule [24, 36] can be applied to estimate the integrals and to avoid intensive computations of the adaptive quadrature.

The estimation approach was implemented under R (mstate and JM). With JM package initializing the

parameters for the longitudinal and multistate sub-models, then, an extended JM() function, called JMstate-Model(), which was mainly implemented by Ferrer et al. [36], can carry out the estimation process of the joint model for viral load dynamics and multistate CD4 cell count progression.

Results

The complete disease progression situation is visualized and explained by Fig. 1, which shows all possible transitions both immune deterioration (red arrows) and immune recovery (green arrows). In particular, using nonparametric Aalen-Johansen estimation, we calculate the transition probabilities from all starting states to all possible states, between the starting time $t = 0$ and all event times successively. Thus, Fig. 3 (panels A-C) displays the estimated transition probability from higher CD4 count states to lower CD4 count states (immune deterioration) in HIV-infected women. From Fig. 3 (panels A-C), we note that as the years since enrolment increased, the probability of immune deteriorates (particularly from mild to

advanced and advanced to severe stages of the diseases), while the transition probability from lower CD4 count states to higher CD4 count states (immune recovery) not remarkably increasing Fig. 3 (panels D-F). In other words, women who enrolled with a CD4 cell count < 350 (severe and advanced disease stage) had a far smaller chance of immune recovery, and a considerably greater chance of immune deterioration (recurrence) compared to women with higher CD4 cell counts (mild and normal disease stage) at enrollment.

Results of the joint multistate model

The results of the joint multistate model are presented in Table 2. For longitudinal sub-model results, patients with TB-coinfection ($\beta = 0.15$; 95%CI: 0.045, 0.249) were significantly associated with a higher baseline viral load, as compared to those who were not co-infected with TB. Similarly, higher educational levels ($\beta = -0.12$; 95%CI: $-0.207, -0.064$), higher weight ($\beta = -0.02$; 95%CI: $-0.057, -0.018$), higher RBC indices ($\beta = -0.01$; 95%CI: $-0.017, -0.002$), higher physical health score ($\beta = -0.01$;

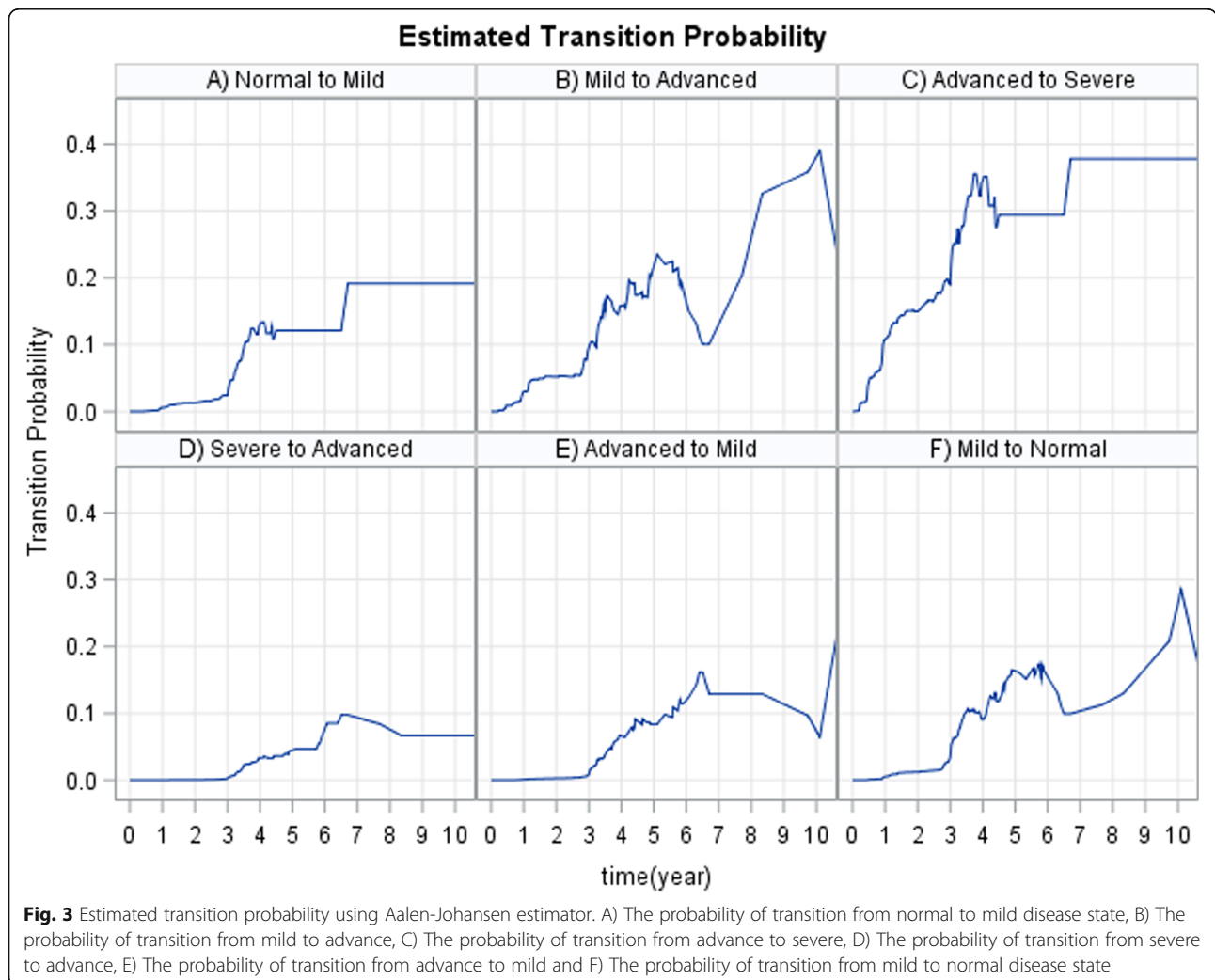


Fig. 3 Estimated transition probability using Aalen-Johansen estimator. A) The probability of transition from normal to mild disease state, B) The probability of transition from mild to advance, C) The probability of transition from advance to severe, D) The probability of transition from severe to advance, E) The probability of transition from advance to mild and F) The probability of transition from mild to normal disease state

Table 2 Estimates and the 95% confidence intervals for parameters of joint multistate and longitudinal model

Variables	Longitudinal Process	Multi-State Process					
	β (95% CI)	Transition 1: Normal to Mild, β (95% CI)	Transition 2: Mild to Advanced, β (95% CI)	Transition 3: Advanced to Severe, β (95% CI)	Transition 4: Severe to Advanced, β(95% CI)	Transition 5: Advanced to Mild, β (95% CI)	Transition 6: Mild to Normal, β (95% CI)
Intercepts	2.197 (2.016, 2.377)***	1.41 (0.53, 2.28)*	1.21 (0.52, 1.91)*	2.08 (0.71, 3.46)*	2.46 (0.13, 4.79)*	-0.01 (-0.84, 0.83)	1.71 (0.79, 2.62)*
TB: Yes	0.147 (0.045, 0.249)**	0.61 (-0.14, 1.37)	0.15 (-0.43, 0.74)	1.50 (-0.39, 3.39)	0.41 (-2.42, 3.23)	-0.47 (-1.06, 0.12)	0.98 (-0.24, 1.72)
Marital status: Married (stable sex partner)	-0.119 (-0.170, -0.068)*	0.44 (-0.15, 1.03)	-0.58 (-1.17, 0.01)	-1.11 (-2.15, -0.06)*	-0.13 (-1.50, 1.24)	-0.09 (-0.64, 0.46)	0.05 (-0.54, 0.64)
Marital status: Many sexual partners	-0.020 (-0.117, 0.078)	-0.07 (-0.40, 0.27)	0.001 (-0.99, 1.00)	-0.50 (-1.23, 0.22)	-0.61 (-0.94, -0.28)*	-0.09 (-0.47, 0.30)	-0.02 (-0.36, 0.32)
Education: 9–10th Grade	-0.053 (-0.132, 0.027)	-0.90 (-1.34, 0.46)	-0.26 (-0.67, 0.15)	-0.32 (-1.13, 0.49)	1.02 (-0.10, 2.15)	0.34 (-0.13, 0.81)	0.28 (-0.21, 0.77)
Education: ≥ 11th Grade	-0.136(-0.207, -0.064)**	-1.20 (-1.66, -0.74)*	0.03 (-0.42, 0.48)	0.80 (-0.06, 1.66)	0.71 (0.50, 1.91)*	0.19 (-0.31, 0.68)	0.59 (0.15, 1.03)*
Age_Cat: < 20 years	0.018(-0.039, 0.075)	0.36 (0.01, 0.70)**	0.34 (-0.03, 0.70)	0.15 (-0.61, 0.92)	0.55 (-0.44, 1.53)	0.22 (-0.18, 0.62)	0.17 (-0.15, 0.48)
Age_Cat: 21–39 years	-0.119(-0.229, -0.009)*	0.08 (-0.47, 0.63)	0.11 (-0.58, 0.79)	-1.36 (-3.17, 0.45)	1.23 (-0.58, 3.04)	1.09 (0.32, 1.85)***	1.01 (0.48, 1.53)**
Sex while drunk: No	-0.060(-0.134, 0.015)	0.10 (-0.31, 0.51)	-0.23 (-0.67, 0.20)	-0.55 (-1.29, 0.18)	1.23 (0.28, 2.17)*	-0.17 (-0.58, 0.23)	0.17 (-0.26, 0.59)
Weight	-0.037(-0.057, -0.018)**	-0.23 (-0.33, -0.12)*	-0.30 (-0.42, -0.17)*	0.03 (-0.27, 0.33)	0.21 (-0.24, 0.65)	-0.03 (-0.19, 0.13)	-0.03 (-0.12, 0.07)
Level of independence score	-0.006(-0.017, 0.005)	-0.48 (-0.60, -0.37)*	-0.11 (-0.23, 0.01)*	-0.37 (-0.65, -0.09)*	-0.05 (-0.35, 0.26)	-0.03 (-0.16, 0.09)	0.14 (0.03, 0.26)*
Social relationship score	-0.015(-0.026, -0.003)*	-0.12 (-0.24, 0.01)	-0.04 (-0.16, 0.08)	-0.25 (-0.49, 0.01)	-0.14 (-0.44, 0.17)	0.001 (-0.13, 0.14)	-0.26 (-0.38, 0.14)
Physical health score	-0.014(-0.027, -0.001)*	-0.45 (-0.64, -0.27)*	-0.36 (-0.54, -0.17)*	-0.25 (-0.61, -0.11)*	-0.49 (-0.93, 0.06)	-0.36 (-0.53, 0.19)	-0.36 (-0.52, 0.20)
Psychological well-being score	0.005(-0.008, 0.018)	0.51 (0.32, 0.70)	0.14 (-0.04, 0.32)	0.13 (-0.19, 0.45)	0.50 (0.11, 0.90)**	0.07 (-0.10, 0.23)	0.37 (0.19, 0.55)*
Liver enzymes abnormality component	0.006(-0.002, 0.013)	-0.08 (-0.18, 0.03)	-0.04 (-0.15, 0.08)	0.04 (-0.15, 0.22)	0.11 (-0.16, 0.38)	-0.17 (-0.30, -0.05)*	0.04 (-0.07, 0.16)
RBC indices	-0.009(-0.017, -0.002)*	-0.03 (-0.13, 0.07)	-0.19 (-0.30, -0.07)*	0.04 (-0.19, 0.26)	-0.07 (-0.35, 0.21)	-0.08 (-0.21, 0.04)	0.16 (0.04, 0.27)*
Granulocytes component	-0.002(-0.009, 0.006)	-0.08 (-0.18, 0.02)	-0.13 (-0.25, -0.02)*	-0.05 (-0.28, 0.18)	0.001 (-0.31, 0.32)	0.12 (0.001, 0.25)*	0.12 (0.01, 0.23)*
Mononuclear component	0.017(-0.009, 0.024)	-0.35 (-0.48, -0.23)*	-0.45 (-0.58, -0.33)*	-0.56 (-0.84, -0.28)*	0.03 (-0.23, 0.28)	0.33 (0.22, 0.45)***	0.32 (0.21, 0.43)*
Time Slope		Associations between Transitions and Longitudinal Outcomes					
Visit	-0.007(-0.015, 0.001)	Transitions based on CD4 count Vs Viral Load	Coefficient	95% CI of β	P-value		
Weight	0.001(-0.001, 0.001)	Trans 1 vs baseline VL	0.51	(-0.35, 1.37)	0.054		

Table 2 Estimates and the 95% confidence intervals for parameters of joint multistate and longitudinal model (Continued)

Variables	Longitudinal Process β (95% CI)	Multi-State Process					
		Transition 1: Normal to Mild, β (95% CI)	Transition 2: Mild to Advanced, β (95% CI)	Transition 3: Advanced to Severe, β (95% CI)	Transition 4: Severe to Advanced, β(95% CI)	Transition 5: Advanced to Mild, β (95% CI)	Transition 6: Mild to Normal, β (95% CI)
Psychological well-being score	0.001(−0.001, 0.002)	Trans 2 vs baseline VL	−0.33	(−0.66, −0.001)	0.043*		
Physical health score	−0.004(−0.008, −0.001)*	Trans 3 vs baseline VL	0.18	(−0.11, 0.47)	0.227		
Level of independence score	0.001(−0.001, 0.001)	Trans 4 vs baseline VL	0.56	(0.24, 0.88)	0.001**		
Social relationship score	0.001(−0.001, 0.002)	Trans 5 vs baseline VL	1.14	(0.75, 1.54)	0.000***		
Marital Status: Married (stable sex partner)	0.005(−0.003, 0.007)	Trans 6 vs baseline VL	2.21	(1.20, 3.21)	0.000**		
Marital Status: many sex partners	0.007 (0.003, 0.011)*	Trans 1 vs Time slope VL	0.06	(−0.93, 1.05)	0.909		
RBC indices	−0.01(−0.09, −0.002)*	Trans 2 vs Time slope VL	−0.32	(−1.30, 0.67)	0.528		
TB: Yes	0.002 (0.001, 0.004)*	Trans 3 vs Time slope VL	0.62	(−1.92, 3.16)	0.632		
Sex while drunk: No	0.0001(−0.003, 0.003)	Trans 4 vs Time slope VL	3.34	(0.26, 6.42)	0.034*		
Age_Cat:< 20	0.008 (0.003, 0.012)*	Trans 5 vs Time slope VL	2.15	(0.69, 3.61)	0.004**		
Age_Cat: 21–39 years	0.001(−0.001, 0.003)	Trans 6 vs Time slope VL	1.22	(0.05, 2.40)	0.042*		
Education: 9–10th Grade	0.001(−0.002, 0.004)						
Education: ≥ 11th Grade	−0.001(−0.004, −0.0003)*						
Liver enzymes abnormality component	0.004 (0.001, 0.01)*						

Keys:- Statistical significance: (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; Reference category: Age [> 40]; Education [≤ 8 grade]; Marital status [single]; TB [No] and Sex while drunk [Yes]

95%CI: $-0.026, -0.003$), higher social relationship score ($\beta = -0.12$; 95%CI: $-0.027, -0.001$) and patients with stable sex partners ($\beta = -0.12$; 95%CI: $-0.207, -0.064$), were significantly associated with lower baseline viral load. When we consider the effects of covariates on the long-term viral load trajectories (the slope), patients with higher educational levels and higher physical health scores were significantly associated with a lower rate of viral load increase over time. Similarly, patients with many sex partners, of younger age (< 20 years) and with higher liver abnormality scores were significantly associated with a bigger increase in viral load in ART-naïve patients.

For the multi-state sub-model results, patients with higher educational levels had statistically significantly increased intensities of transitions from mild to normal ($\beta = 0.59$; 95%CI: $0.15, 1.03$) and severe to advanced ($\beta = 0.71$; 95%CI: $0.50, 1.91$) disease stages, but had reduced the intensities of

transitions from normal to mild ($\beta = -1.20$; 95%CI: $-1.66, -0.74$) disease stages. Patients with many sex partners had significantly decreased the intensities of transitions from severe to advanced ($\beta = -0.61$; 95%CI: $-0.94, -0.28$) disease stages. Middle-aged patients were significantly associated with a higher intensity of transitions from advanced to mild ($\beta = 1.09$; 95%CI: $0.32, 1.85$) and from mild to normal ($\beta = 1.01$; 95%CI: $0.48, 1.53$) disease stages, compared to those patients in the older age group. The psychological wellbeing score had significant effects on severe to advanced ($\beta = 0.50$; 95%CI: $0.11, 0.90$) and mild to normal ($\beta = 0.37$; 95%CI: $0.19, 0.55$) transitions. Patients with high scores of latent variable related RBC indices had significantly increased the intensity of transition from mild to normal ($\beta = 0.16$; 95%CI: $0.04, 0.27$) disease stage, but decreased the intensity of transition from mild to advance ($\beta = -0.19$; 95%CI: $-0.30, -0.07$) disease stage. Moreover, having high weight, high physical

health score and high level of independent score significantly reduced the intensities of transitions from normal to mild, mild to advanced and advanced to severe disease stages. Furthermore, having a high liver abnormality score significantly reduced the intensities of transitions from advanced to mild disease stage.

Regarding the association parameters between the longitudinal viral load biomarker process (baseline and time slope) and disease progressions of HIV/AIDS, viral load dynamics had a significant effect on the intensities of transitions of HIV/AIDS disease progression. As the baseline viral load of women in the study increased, the likelihood that women transitioned from advanced to mild ($aHR = 0.72 = \exp(-0.33)$) stages of the disease decreased. Similarly, a higher baseline viral load significantly increased the intensities of immunological deterioration transitions. Moreover, as the long-term viral load trajectories of women in the study increased, the likelihood that women transitioned from normal to mild, mild to advanced and advanced to severe stages of the disease increased.

Assessment of the fitted model

The estimates of the joint longitudinal viral load biomarker and multistate CD4 cell count transition model were validated by using the graphical methods presented in Figs. 4 and 5. For the mixed effect submodel, the plotted fitted values versus residuals (standardized) of the viral load marker confirmed no heteroscedasticity error (see Fig. 4). For the multistate submodel, the estimates of these multistate models were compared with a non-parametric Aalen-Johansen estimate to assess model fit (as discussed by Ieva et al. [37] and Titman and Sharples

[38]). The summary results are presented in Fig. 5 and the six plots showed overall good performances of the joint multistate model in terms of the fit for the transitions cumulative hazard estimate.

Discussion

In this study, we have presented a joint model for multistate and longitudinal biomarker data. Such data are common for many medical and clinical studies in monitoring chronic diseases. Simultaneous modelling of this multistate and longitudinal biomarker outcome in joint models offers advantages over a separate model of each outcome, including more clinically meaningful adjusted association parameters, improved predictive accuracy, and more efficient parameter estimates. Also, using this joint multistate model not only an improved inference but also the opportunity for dynamic prediction. Król et al. [39] developed dynamic prediction tools for their recurrent events joint model. Others have also presented dynamic prediction in the context of joint models involving multiple event time data [36, 40, 41]. Hence, in this article the joint model for multistate CD4 cell count progression and longitudinal viral load outcomes provides a complete model of HIV/AIDS disease progression in an ART-naive cohort, which takes into account longitudinal viral load dynamics, to study possible factors that affect time to transition between sequential adverse events of HIV/AIDS. We assumed that the dependency between the transitions time for a given patient is entirely explained by the longitudinal viral load biomarker and the prognostic factors. This assumption could be relaxed by including frailty term in the multistate submodel.

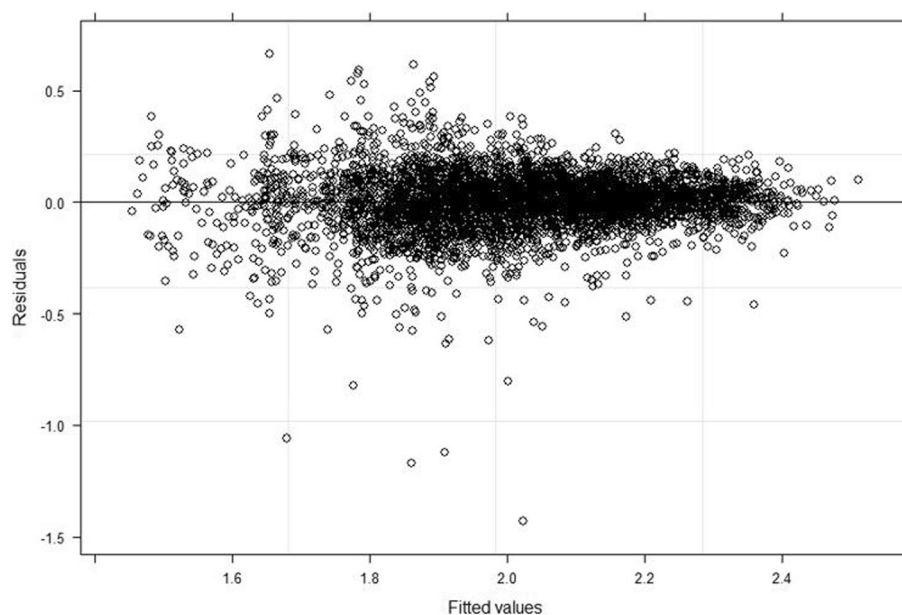
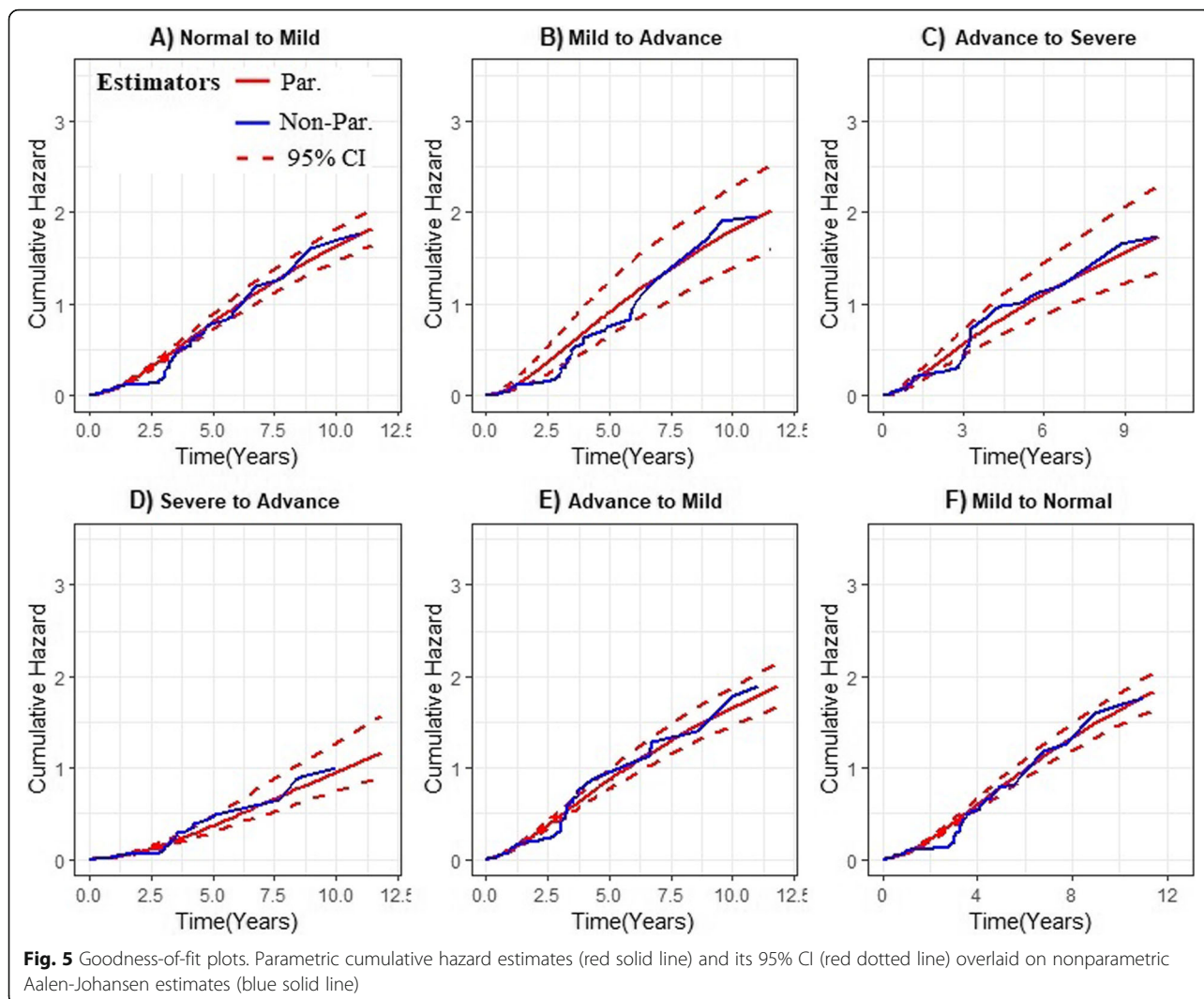


Fig. 4 Goodness-of-fit plots for longitudinal viral load biomarker



Our results showed that older age has been associated with a higher baseline viral load, which is in agreement with the previous studies [42, 43]. We also found that a faster rise in viral load was significantly associated with young adolescents (age < 20 years) compared to those patients in the older age group (age > 40 years). Having multiple sexual partners significantly increased patients' viral load throughout the follow-up time. This might be due to patients with high-risk behaviors, such as an increased number of sexual partners, have been associated with depression [44], and patient with chronic depression was significantly associated with an increase in viral load [45]. In other words, viral load has a mediating effect through depression caused by having multiple sexual partners. Moreover, patients with a higher quality of life domain score and higher educational levels were associated with a lower rate of viral load increase over time.

Patients with opportunistic infections and tuberculosis infection, in particular, are associated with increases in

HIV viral load [46, 47]. These studies have also been interpreted as tuberculosis accelerating the loss of CD4 count and promote progression from HIV infection to AIDS. Our data add to this literature by showing patients with TB co-infection were associated with a bigger increase in viral load in ART-naïve patients. Contrary to our findings, studies from Nigeria [48] showed that there was no significant relationship between TB co-infection and HIV viral load. Possible explanations for this controversial report might be that data for our study was conducted in a cohort of acutely infected patients and followed up repeatedly over an extended ART-free period. Therefore, patients with active TB should thus be prioritized for viral load monitoring.

Latent variable related to aminotransferase were significantly associated with the change in viral load. Faster rises in the rate of change of viral load over time were observed in patients with higher liver enzymes abnormality. This finding concurs with the prior report [49],

which noted that a positive correlation exists between viral load and aminotransferase (ALT and AST). Thus, there is a need to monitor alanine aminotransferase and aspartate aminotransferase levels before initiation of ART, mainly in high-risk patients, to reduce side effect concerns. Furthermore, we found that patients with higher RBC indices scores were associated with a lower rate of viral load increase over time.

The rate of change of immune recovery of a patient with higher education levels increased with greater rates, compared to those with low levels of education. This is supported by the findings of Maurya et al. [50], who found that the level of education significantly affects CD4 cell count and wellness of HIV infected patients. This could be explained as access to education provides a better understanding of hygiene and sanitary practices. Subsequently, proper hygiene and sanitation practices increase the CD4 count. We further observed that patients in the middle-age group experienced higher rates of immunological recovery compared to those patients in the older age group. Consistent with the finding of this study, some studies showed that patients seroconverting at older ages progress more quickly to lower CD4 cell count categories [43, 51]. Furthermore, we found that patients having a high weight significantly increased the intensities of immunological recovery transitions.

The results showed there was a significant positive relationship between QoL domain scores and intensities of recovery of HIV-infected patients. Studies conducted in South Africa [52] and China [53] revealed a similar finding where a better QoL score significantly associated with a higher CD4 count.

Among the different hematological parameters for HIV infected patients, as expected, high scores of latent variable related to total lymphocytes and basophils counts in the blood had significantly reduced intensities of immunological deterioration transitions, a finding that is in accordance with the literature [54], where it was found that *total lymphocytes* and *basophils* are positively and independently associated with CD4 cell count responses. Many studies also suggested that total lymphocytes can adequately serve as a surrogate biomarker for predicting CD4 count progression in resource-limited settings [55–57]. Similar results have been found for latent variable related to neutrophils, monocytes, and leucocytes. Patients having a high score of latent variable related to neutrophils, monocytes and leucocyte count in the blood had significantly reduced intensities of immunological deterioration transitions. This was supported by the previous study [58], where absolute neutrophils and total white cell (leucocyte) counts are independently associated with CD4 cell count responses. Moreover, we found that patients having high RBC parameters score significantly increased the intensities of immunological recovery transitions, but reduced the intensities of immunological deterioration transitions.

Furthermore, patients with a high liver abnormality score and with many sex partners showed significantly increased intensities of immunological deterioration transitions.

Viral load biomarker significantly affects the transition intensities of HIV/AIDS disease progression. Having a high baseline viral load significantly decreased the intensities of immunological recovery transitions, but increased the intensities of immunological deterioration transitions. This was supported by previous studies, Farahani et al. [9] and Martinson et al. [59], where higher baseline viral load in early infection has been associated with faster CD4 count decline. Moreover, patients experiencing a higher rate of viral load increase over time have been associated with increased intensities of immunological deterioration transitions. This was supported by a previous study [60], where viral load increases over time were strongly associated with CD4 cell decline. Furthermore, in agreement with the world health organization treatment guidelines [61], which recommends ART in all PLHIV regardless of CD4 cell count, early identification of patients with poor clinical characteristics and initiation of treatment will improve programmatic success and treatment prognosis.

This study has some limitations, including the missing data, which are expected for a study conducted on data collected from patients' files and when dealing with a long term follow-up period. We did not evaluate model performance with an alternative multistate sub-model, such as a full parametric multistate sub-model (including accelerated failure time models). However, model diagnostics have been performed and the residual and influence diagnostics affirmed no violation of implicit and explicit assumptions in our model. The other limitation is that the study was limited to adult females. Moreover, this joint model did not take into account the simultaneous modeling of many biomarkers and multistate outcomes before and after cART initiation. Bayesian multivariate joint models for analyzing many biomarkers (i.e. viral load and quality of life scores) and multistate outcomes before and after cART initiation, will be our future research work.

Conclusions

Overall, from a methodological perspective, it can be concluded that the joint multistate model approach provides wide-ranging information about the progression and assists to provide specific dynamic predictions and increasingly precise knowledge of diseases. Joint multistate modelling is necessary to explore the impact of the longitudinal biomarker outcome on the transitions between clinical states. Joint models are also an improvement over separate multisite models because they consider all the longitudinal observations that are predictive of the transition event of interest. Though this

research presented the usefulness of the joint multistate model for analyzing the HIV/AIDS cohort data, the approach is applicable to a wide variety of chronic diseases. There is a need for increased research in terms of methods, so hopefully, this article will be helpful applied researchers (for medical research) to familiarize with the method and interpretation of the results therefrom. From a clinical perspective, the findings of this study contribute to extend the survival of the patients and guide clinical interventions.

Abbreviations

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

Acknowledgments

We would like to acknowledge Dr. Nigel Garrett and the entire CAPRISA Acute Infection team including research leaders who formulated the research ideas that generated such rich data, thank you very much for allowing us to use this dataset. We also thank the participants for their involvement in this study. We would also like to thank the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, South Africa for providing their guidance and support.

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. All 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 study data is available upon 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: 18 July 2019 Accepted: 13 March 2020

Published online: 26 March 2020

References

- Moraga P, Collaborators GCoD. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the global burden of disease study 2016. *Lancet*. 2017;390(10100):1151–210.
- World Health Organization. Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000–2016. Geneva. 2018. In: 2018.
- Hare CB. Clinical overview of HIV disease. *HIV InSite Knowledge Base Chapter-January*.2006.
- Haynes BF, Pantaleo G, Fauci AS. Toward an understanding of the correlates of protective immunity to HIV infection. *Science*. 1996;271(5247):324–8.
- Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384(9939):258–71.
- 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.
- 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.
- Korenromp EL, Williams BG, Schmid GP, Dye C. Clinical prognostic value of RNA viral load and CD4 cell counts during untreated HIV-1 infection—a quantitative review. *PLoS One*. 2009;4(6):e5950.
- Farahani M, Novitsky V, Wang R, Bussmann H, Moyo S, Musonda RM, et al. Prognostic value of HIV-1 RNA on CD4 trajectories and disease progression among antiretroviral-naïve HIV-infected adults in Botswana: a joint modeling analysis. *AIDS Res Hum Retrovir*. 2016;32(6):573–8.
- Guedj J, Thiébaud R, Commenges D. Joint modeling of the clinical progression and of the biomarkers' dynamics using a mechanistic model. *Biometrics*. 2011;67(1):59–66.
- Thiébaud R, Jacqmin-Gadda H, Lepout C, Katlama C, Costagliola D, Le Moing V, et al. Bivariate longitudinal model for the analysis of the evolution of HIV RNA and CD4 cell count in HIV infection taking into account left censoring of HIV RNA measures. *J Biopharm Stat*. 2003;13(2):271–82.
- Binquet C, Le Teuff G, Abrahamovic M, Mahboubi A, Yazdanpanah Y, Rey D, et al. Markov modelling of HIV infection evolution in the HAART era. *Epidemiol Infect*. 2009;137(9):1272–82.
- Chikobvu D, Shoko C. A Markov model to estimate mortality due to HIV/AIDS using CD4 cell counts based states and viral load: a principal component analysis approach. *Biomed Res*. 2018;29(15):3090–8.
- RdVCd O, Shimakura SE, Campos DP, Victoriano FP, Ribeiro SR, Veloso 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.
- Shoko C, Chikobvu D, Bessong PO. A Markov model to estimate mortality due to HIV/AIDS using viral load levels-based states and CD4 cell counts: a principal component analysis approach. *Infect Dis Ther*. 2018;7(4):457–71.
- Huang X, Li G, Elashoff RM, Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Anal*. 2011;17(1):80–100.
- Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*. 2008;64(3):762–71.
- Han J, Slate EH, Peña EA. Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in medicine*.2007; 26(29):5285–5302.
- Cai Q, Wang M-C, Chan KCG. Joint modeling of longitudinal, recurrent events and failure time data for survivor's population. *Biometrics*. 2017;73(4):1150.
- Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*. 2006;62(2):432–45.
- Ferrer L, Putter H, Proust-Lima C. Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment. *Stat Methods Med Res*. 2017;0962280218811837.
- Schluchter MD, Piccorelli AV. Shared parameter models for joint analysis of longitudinal and survival data with left truncation due to delayed entry—applications to cystic fibrosis. *Stat Methods Med Res*. 2018;0962280218764193.
- J-f D, Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Anal*. 2002;8(2):99–115.
- Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Comput Stat Data Anal*. 2012;56(3):491–501.

25. van Loggelenberg 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.
26. Garrett N, Norman E, Leask K, Naicker N, Asari V, Majola N, et al. Acceptability of early antiretroviral therapy among south African women. *AIDS Behav*. 2018;22(3):1018–24.
27. Karim QA, Karim SSA, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, Kharsany AB, Sibeko S, Mlisana KP, Omar Z. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010;329(5996):1168–74.
28. Byrne BM. Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *J Pers Assess*. 2005;85(1):17–32.
29. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958;23(3):187–200.
30. 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.
31. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956–71.
32. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963–74.
33. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000;1(4):465–80.
34. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*, vol. 998: John Wiley & Sons; 2012.
35. 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.
36. Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust-Lima C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Stat Med*. 2016;35(22):3933–48.
37. 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.
38. Titman AC, Sharples LD. Model diagnostics for multi-state models. *Stat Methods Med Res*. 2010;19(6):621–51.
39. Król A, Ferrer L, Pignon JP, Proust-Lima C, Ducreux M, Bouché O, et al. Joint model for left-censored longitudinal data, recurrent events and terminal event: predictive abilities of tumor burden for cancer evolution with application to the FFCO 2000–05 trial. *Biometrics*. 2016;72(3):907–16.
40. Rouanet A, Joly P, Dartigues JF, Proust-Lima C, Jacqmin-Gadda H. Joint latent class model for longitudinal data and interval-censored semi-competing events: application to dementia. *Biometrics*. 2016;72(4):1123–35.
41. Dantan E, Joly P, Dartigues J-F, Jacqmin-Gadda H. Joint model with latent state for longitudinal and multistate data. *Biostatistics*. 2011;12(4):723–36.
42. Korenromp EL, Williams BG, Schmid GP, Dye C. Clinical prognostic value of RNA viral load and CD4 cell counts during untreated HIV-1 infection—a quantitative review. *PLoS one*. 2009;4(6).
43. Nakagawa F, Grp N, Ep C. Factors associated with short-term changes in HIV viral load and CD4 R cell count in antiretroviral-naive individuals. *Aids*. 2014; 28(9):1351–6.
44. Nduna M, Jewkes RK, Dunkle KL, Shai NPJ, Colman I. Associations between depressive symptoms, sexual behaviour and relationship characteristics: a prospective cohort study of young women and men in the eastern cape, South Africa. *J Int AIDS Soc*. 2010;13(1):44.
45. Kelson-Chichetto N, Okafor C, Cook R, Abraham AG, Bolan R, Plankey M. Association between depressive symptom patterns and clinical profiles among persons living with HIV. *AIDS Behav*. 2018;22(5):1411–22.
46. Anejo-Okopi J, Abah IO, Barshep Y, Ebonyi AO, Daniyam C, Isa SE, Simji G, Oguche S, Agaba P, Lar P. Demographic and clinical correlates of HIV-1 RNA levels in antiretroviral therapy-naive adults attending a tertiary hospital in Jos, Nigeria. *J Virus Eradication*. 2017;3(1):51.
47. Gautam H, Bhalla P, Saini S, Uppal B, Kaur R, Baveja CP, et al. Epidemiology of opportunistic infections and its correlation with CD4 T-lymphocyte counts and plasma viral load among HIV-positive patients at a tertiary care hospital in India. *J Int Assoc Phys AIDS Care*. 2009;8(6):333–7.
48. Ebonyi A, Meloni ST, Anejo-Okopi J, Akanbi M, Oguche S, Agaba P, Sagay a, Kanki PJ. Factors associated with pulmonary tuberculosis-HIV co-infection in treatment-naive adults in Jos, North Central Nigeria. *J AIDS Clin Res: Okonkwo PI*; 2013.
49. Mata-Marín JA, Gaytán-Martínez J, Grados-Chavarría BH, Fuentes-Allen JL, Arroyo-Anduiza CI, Alfaro-Mejía A. Correlation between HIV viral load and aminotransferases as liver damage markers in HIV infected naive patients: a concordance cross-sectional study. *Virology*. 2009;6(1):181.
50. Maurya S, Singh R, Negi N, Kapil A, Chaudhry R, Das B. The level of education affects CD4 cell count and wellness among HIV infected adult between age group 18 to 60 years. *Int J Infect Dis*. 2016;45:264.
51. Cori A, Pickles M, van Sighem A, Gras L, Bezemer D, Reiss P, Fraser C. CD4+ cell dynamics in untreated HIV-1 infection: overall rates, and effects of age, viral load, sex and calendar time. *AIDS (London, England)*. 2015;29(18):2435.
52. Tomita A, Garrett N, Werner L, Burns JK, Mpanza L, Mlisana K, van Loggelenberg F, Karim SSA. Health-related quality of life dynamics of HIV-positive South African women up to ART initiation: evidence from the CAPRISA 002 acute infection cohort study. *AIDS Behav*. 2014;18(6):1114–23.
53. Jiang T, Zhou X, Wang H, Luo M, Pan X, Ma Q, Chen L. Psychosocial Factors Associated with Quality of Life in Young Men Who Have Sex with Men Living with HIV/AIDS in Zhejiang, China. *Int J Environ Res Public Health*. 2019;16(15):2667.
54. 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.
55. Chen J, Li W, Huang X, Guo C, Zou R, Yang Q, Zhang H, Zhang T, Chen H, Wu H. Evaluating total lymphocyte count as a surrogate marker for CD4 cell count in the management of HIV-infected patients in resource-limited settings: a study from China. *PLoS one*. 2013;8(7):e69704.
56. Obirikorang C, Quaye L, Acheampong I. Total lymphocyte count as a surrogate marker for CD4 count in resource-limited settings. *BMC Infect Dis*. 2012;12(1):128.
57. Wondimeneh Y, Ferede G, Yismaw G, Muluye D. Total lymphocyte count as surrogate marker for CD4 cell count in HIV-infected individuals in Gondar University Hospital, Northwest Ethiopia. *AIDS Res Ther*. 2012;9(1):21.
58. Vanker N, Ipp H. The use of the full blood count and differential parameters to assess immune activation levels in asymptomatic, untreated HIV infection. *S Afr Med J*. 2014;104(1):45–8.
59. Martinson NA, Gupte N, Msandwa R, Moulton LH, Barnes GL, Ram M, Gray G, Hoffmann C, Chaisson RE. CD4 and viral load dynamics in antiretroviral-naive HIV-infected adults from Soweto, South Africa: a prospective cohort. *PLoS one*. 2014;9(5):e96369.
60. Lima VD, Fink V, Yip B, Hogg RS, Harrigan PR, Montaner JS. Association between HIV-1 RNA level and CD4 cell count among untreated HIV-infected individuals. *Am J Public Health*. 2009;99(5):S193–6.
61. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach: World Health Organization; 2016.

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

