Joint estimation of CD4$^+$ cell progression and survival in untreated individuals with HIV-1 infection

Tara D. Mangal, on behalf of the UNAIDS Working Group on CD4 Progression and Mortality Amongst HIV Seroconverters including the CASCADE Collaboration in EuroCoord

Objective: We compiled the largest dataset of seroconverter cohorts to date from 25 countries across Africa, North America, Europe, and Southeast/East (SE/E) Asia to simultaneously estimate transition rates between CD4$^+$ cell stages and death, in antiretroviral therapy (ART)-naive HIV-1-infected individuals.

Design: A hidden Markov model incorporating a misclassification matrix was used to represent natural short-term fluctuations and measurement errors in CD4$^+$ cell counts. Covariates were included to estimate the transition rates and survival probabilities for each subgroup.

Results: The median follow-up time for 16 373 eligible individuals was 4.1 years (interquartile range 1.7–7.1), and the mean age at seroconversion was 31.1 years (SD 8.8). A total of 14 525 individuals had recorded CD4$^+$ cell counts pre-ART, 1885 died, and 6947 initiated ART. Median (interquartile range) survival for men aged 20 years at seroconversion was 13.0 (12.4–13.4), 11.6 (10.9–12.3), and 8.3 years (7.9–8.9) in Europe/North America, Africa, and SE/E Asia, respectively. Mortality rates increase with age (hazard ratio 2.22, 95% confidence interval 1.84–2.67 for >45 years compared with <25 years) and vary by region (hazard ratio 2.68, 1.75–4.12 for Africa and 1.88, 1.50–2.35 for Asia compared with Europe/North America). CD4$^+$ cell decline was significantly faster in Asian cohorts compared with Europe/North America (hazard ratio 1.45, 1.36–1.54).

Conclusion: Mortality and CD4$^+$ cell progression rates exhibited regional and age-specific differences, with decreased survival in African and SE/E Asian cohorts compared with Europe/North America and in older age groups. This extensive dataset reveals heterogeneities between regions and ages, which should be incorporated into future HIV models. Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc.

Introduction

A good characterization of HIV natural history is essential to track epidemic trends, estimate incidence, determine treatment needs, and estimate the impact of interventions to reduce AIDS mortality. Current estimates of CD4$^+$ cell progression and survival have been derived from analyses employing different approaches, combined to approximate a Markov transition model [1–3]. CD4$^+$ cell progression estimates tend to come from linear mixed

effects models [4–7], and mortality estimates are generally derived from survival analyses [8,9]. To consistently use both sources of data requires a Markov model for estimation, which simultaneously describes the process of $CD4^+$ cell decline and survival.

There is evidence for increased disease progression by age across all exposure categories, but suggestions of differences between the sexes are inconclusive [4,10–14]. Faster progression in persons from Africa compared with Europe, after adjusting for age, has been found in some studies, whereas others have found no significant differences [6,10,15]. In addition, some studies report faster $CD4^+$ cell decline and shorter survival in Asian cohorts compared with Europe or Africa. For example, Todd et al. [8] report median gross survival times of 11.1 and 11.6 years in East and South Africa, respectively, and 7.5 years for Thai cohorts. Similarly, a comparison between the Concerted Action on Seroconversion to AIDS and Death in Europe (CASCADE) collaboration and the Beijing PRIMO cohort found significantly faster $CD4^+$ cell decline in the Beijing cohort [16].

No previous study has directly compared progression and survival times in multiple regions using empirical data from individuals with well estimated seroconversion dates (i.e. the date at which an infected individual begins to develop antibodies to the HIV virus). $CD4^+$ cell counts are subject to measurement error and natural short timescale variability, producing a noisy observable marker of progression. For that reason, we use a hidden Markov model that allows joint estimation of $CD4^+$ cell progression and survival without prior knowledge of time of infection along with estimation of the impact of time-dependent covariates [17–19]. Informative censoring such as initiation of ART, provided selectively to those with severe disease, can be accommodated by introducing a competing hazards element.

Here we summarize data from seroconverter cohorts from 25 countries across four geographic regions: Africa, North America, Europe, and Southeast/East (SE/E) Asia. We use this extensive dataset to describe the natural history of HIV infection using $CD4^+$ cell counts as a marker, estimating differences among age groups, sex, and regions. We simultaneously estimate transition rates between $CD4^+$ cell stages and time from seroconversion to death and discuss the likely impact of demographic characteristics on the progression of HIV infection.

Methods

Study population
We identified seroconverter cohorts from a systematic review of published articles and conference abstracts,

inviting those with laboratory evidence of recent seroconversion or reported intervals between negative and positive HIV tests results along with $CD4^+$ cell counts to participate. The CASCADE collaboration comprises the majority of seroconverters in our dataset, with 29 separate cohorts in 13 countries. An additional 24 studies were included, resulting in a pooled dataset comprising 30 164 individuals monitored during 1979–2014 across 25 countries (Table S1, http://links.lww. com/QAD/B52). Data were grouped by region (Europe/North America, Africa, and SE/E Asia) for analysis to assess geographical variations in natural history.

Many of the studies contained detailed immunological, virological, and clinical data along with individual characteristics. A limited number contained no immunological information but were included as they furnished information on survival prior to the availability of ART. For 98.5% of individuals, the date of infection (seroconversion) was estimated as the midpoint between the last seronegative and first seropositive results. Seroconversion dates for the remaining individuals were estimated using laboratory evidence of seroconversion, seroconversion illness, or using unspecified methods. Individuals between 15 and 65 years of age at seroconversion with seroconversion intervals of less than 1 year were included in the main analysis. We repeated the analysis using seroconversion intervals of up to 2 years to examine the potential effect of this criterion on initial state probability estimates.

The hidden Markov model
The Markov model represents $CD4^+$ cell decline following seroconversion due to HIV-1 infection as transitions through a series of seven $CD4^+$ cell count categories and from each $CD4^+$ cell stage to death (Fig. 1). Transition rates to ART, included as an absorbing state, were estimated as ancillary parameters to reflect the data [20].

The transition rates (λ_{rs}) describe the instantaneous rate of movement between states r and s and values of the transition matrix (Q) on the diagonal (λ_{rr}) were fixed to minus the sum of the entries in each row (Fig. S1, http://links.lww.com/QAD/B52). We allowed only forward transitions between adjacent $CD4^+$ cell-based states, consistent with the underlying biology [17–19]. A misclassification matrix with entries r,s describes the probability of observing a $CD4^+$ cell count in state r conditionally on the true state s. Thus, any apparent backwards transitions between states with lower $CD4^+$ cell count to states with higher $CD4^+$ cell count were assumed to be explained by short-term reversible fluctuations in combination with measurement error. The two absorbing states corresponding with ART and death were assumed to be recorded error-free.
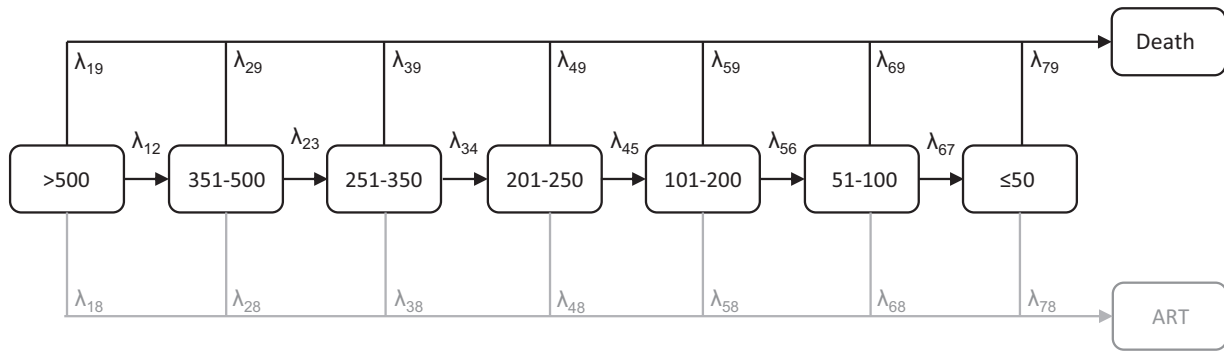
Fig. 1. The staged Markov model of HIV disease progression.

Individuals can enter into any of the transient states at seroconversion, defined by CD4$^+$ cell counts of more than 500, 351–500, 251–350, 201–250, 101–200, 51–100, and 50 CD4$^+$ cells/μl or less. Short time-scale variability in the measured CD4$^+$ cell counts was stabilized by log-transformation.

Individual-level covariates age (time-dependent), sex, and region were included as explanatory variables on the transition rates. The transition rate (λ) for individual i at observation time j is conditional on a vector of covariate values z with a vector of coefficients (β) estimated using multinomial logistic regression:

$$\lambda_{rs}(z_{ij}) = \lambda_{rs}^{(0)}\exp(\beta_{rs}^{T} z_{ij})$$

where r and s represent a pair of states, for each nonzero entry of the transition matrix (Q) [21].

We included a separate absorbing state for transitions to ART to reduce bias caused by informative censoring, that is, individuals with lower CD4$^+$ cell counts are more likely to be censored due to ART initiation than individuals with higher counts. We use the recorded date of ART initiation as the exact transition time to the ART absorbing state and consider only those individuals receiving HAART. These parameters are jointly estimated during the model fitting but discarded for the survival time calculations. That is, all survival estimates are calculated conditionally on not receiving ART.

The transition rates were used to calculate the median survival time from seroconversion to death, using the cumulative transition probability distributions of the time from all CD4$^+$ cell-based states to death, weighted by the initial state probabilities, and accounting for the changing age of the person.

Model fitting and parameter estimation

The Markov model was fit using the msm package in R (R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/) [21]. We fitted the model directly to the data, without smoothing,

incorporating the measurement error structure directly into the likelihood function. To ensure convergence to the true maximum likelihood, we used Latin hypercube sampling to generate 10 000 sets of initial values from a plausible range and selected the 10 parameter combinations with the greatest log-likelihood [22].

We then started optimization algorithms using these 10 parameter sets to find the optimum parameter estimates. Single-covariate models were developed incorporating current age, sex, and region variables on transition intensities and initial state probabilities, and compared with a baseline model containing no covariates. Constrained models that estimated one covariate effect for each transition type (i.e. for transitions among CD4$^+$ cell states, to ART, and to death) were compared with a saturated model allowing covariate effects to be estimated separately for every transition. The likelihood ratio test was used to select covariates that improved the model fit and we selected the most parsimonious model representing the associations between each covariate and the disease process. Confidence limits were obtained by nonparametric bootstrap resampling using 5000 resampled sets.

To examine the effect of clinical AIDS diagnosis as a potential bias in the survival estimates, under the assumption that at this point individuals may be receiving additional medical care prolonging their lives, a sensitivity analysis was performed that excluded all observations recorded after an AIDS diagnosis.

Ethical approval

Separate ethical approval was not required for this study as this was a secondary analysis of fully anonymized existing datasets. Ethical approval was obtained by each individual study and can be accessed from the original publications.

Results

Study characteristics

Of the 30 164 individuals in the 25 included studies, 16 373 were eligible for inclusion in the analysis (Fig. 2);
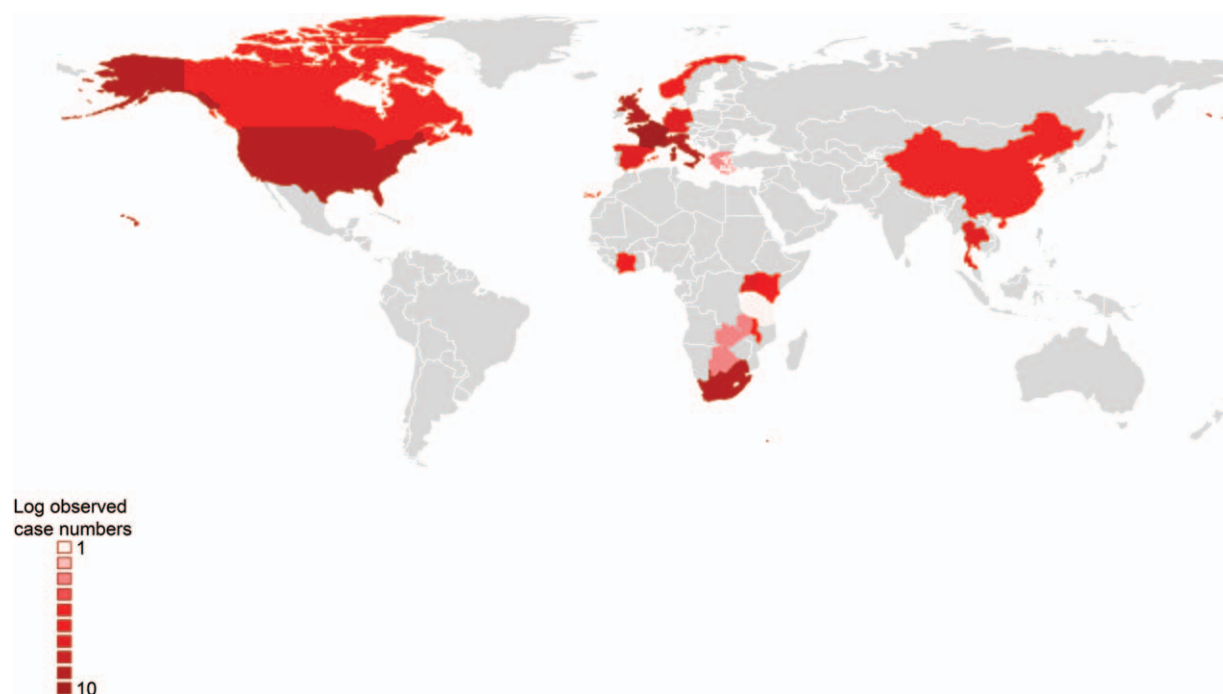
Fig. 2. Geographic distribution of 16 373 HIV-1-infected individuals enrolled throughout 1979–2014, in 25 countries*. *Missing values: 699 cases categorized as 'sub-Saharan Africa' and six cases in 'Europe'.

14 525 had recorded CD4$^+$ cell counts, 6947 initiated ART, and 1885 deaths were observed (Table 1). The mean age at seroconversion was 31.1 years (SD 8.8), 77% were men, and a median of 6 [interquartile range (IQR) 3–13] CD4$^+$ cell measurements had been recorded over 4.1 years (IQR 1.7–7.1).

Model selection
Each covariate (age, sex, and region) significantly improved the model fit (likelihood ratio test, all $P < 0.001$) compared with the baseline model and so were retained in the final model. There was no significant improvement when allowing covariate effects to be estimated independently for each transition rate compared with constraining the covariate effects ($P > 0.05$), and so we opted for the more parsimonious model.

Initial state probabilities
As individuals age, their estimated CD4$^+$ cell count at seroconversion decreases (Fig. 3). The probability of being in the highest state (CD4$^+$ cell count >500 cells/ μl) after seroconversion falls from 0.57 for ages 15–24

years to 0.41 for those aged at least 45 years for European and North American men and from 0.54 to 0.38 for African men. CD4$^+$ cell counts at seroconversion were lower among Asian cohorts; only 0.41 of men aged 15– 24 years in the SE/E Asian cohorts seroconvert into the highest CD4$^+$ cell state. Including individuals with longer seroconversion intervals, that is, up to 2 years between the last negative and the first positive test produced qualitatively similar findings (Table S2, http:// links.lww.com/QAD/B52).

Progression between CD4$^+$ cell states and mortality rates
The progression rates between CD4$^+$ cell states were not strongly influenced by age (Table 2); however, older individuals experience higher mortality rates [hazard ratio 2.22, 95% confidence interval (CI) 1.84–2.67 for ≥45- year-olds compared with <25 years]. Estimates of the transition rates for 15–24-year-olds, by sex and region, are presented in Tables 3 and 4. Women were estimated to have slightly slower progression rates between CD4$^+$ cell states than men (hazard ratio 0.92, 95% CI 0.86–0.99),

Table 1. Selected characteristics of 16 373 HIV-infected individuals, by region.

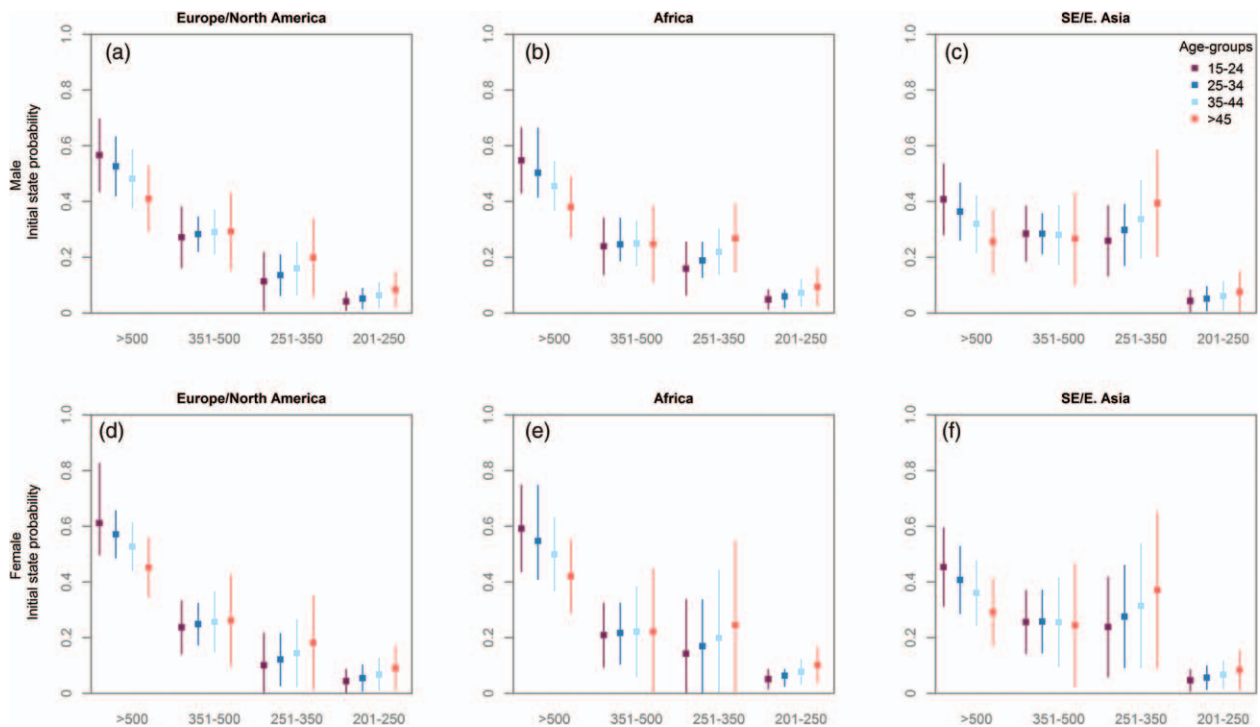|	Eligible seroconverters	Mean age at seroconversion [years (SD)]	Proportion men	Study period	Median follow-up time [years (2.5th and 97.5th percentiles)]	Median CD4$^+$ cell counts (range)	Number of deaths
North America	1492	29.06 (7.99)	0.95	1984–2012	3.41 (0.06–12.42)	3 (1–75)	56
Europe	10 950	31.75 (8.99)	0.79	1979–2012	4.36 (0.61–15.53)	7 (1–199)	1199
Africa	3341	30.42 (8.10)	0.59	1987–2014	4.45 (0.37–11.41)	7 (1–41)	545
SE/E Asia	590	27.34 (7.89)	0.96	1989–2013	3.37 (0.12–7.39)	6 (1–17)	85

SE/E Asia, Southeast/East Asia.

Fig. 3. Initial state probabilities at seroconversion for HIV-1-infected individuals stratified by region and age for 12 567 men (a–c) and 3806 women (d–f). States 1–4 represent CD4$^+$ cell counts of more than 500, 351–500, 251–350, and 201–250 cells/μl, respectively.

but there were no significant differences in overall mortality rates between men and women (hazard ratio 1.04, 95% CI 1.00−1.08). Mortality rates were higher in cohorts from Africa and SE/E Asia (hazard ratio 2.68, 95% CI 1.75−4.12 and hazard ratio 1.88, 95% CI 1.50−2.35, respectively) compared with Europe and North America, but CD4$^+$ cell decline was faster only for Asian cohorts (hazard ratio 1.45, 95% CI 1.36−1.54).

Overall survival times

Survival in SE/E Asia was substantially shorter than in other regions, with median survival estimates of 8.3 years (IQR 7.9−8.9) for men aged 20 years compared with 13.0 and 11.6 years in Europe and Africa, respectively

(Table 5). Removing observations occurring after a clinical AIDS diagnosis results in qualitatively similar findings, although survival estimates are shorter by an average of 10 months (Table S3, http://links.lww.com/QAD/B52).

Model assessment

There is substantial misclassification between observed and underlying states in the intermediate stages of disease as expected, given the narrow ranges of these states and the high variability in individual CD4$^+$ cell dynamics (Table S4, http://links.lww.com/QAD/B52). We see the greatest agreement between observed and underlying states in those with the highest CD4$^+$ cell counts (0.935,

Table 2. Estimated hazard ratios (95% confidence interval) for covariates (current age, sex, and region) on transition rates of 16 373 individuals using a proportional intensities model.

	Progression	ART	Mortality
Age-group (years)			
15–24 (reference)	1.0		1.0
25–34	1.01 (0.94−1.10)		1.25 (1.13−1.40)
35–44	0.94 (0.87−1.03)		1.55 (1.35−1.78)
>45	0.93 (0.86−1.01)		2.22 (1.84−2.67)
Sex			
Male (reference)	1.0		1.0
Female	0.92 (0.86−0.99)		1.04 (1.00−1.08)
Region			
Europe/North America (reference)	1.0	1.0	1.0
Africa	0.97 (0.77−1.21)	0.21 (0.16−0.28)	2.68 (1.75−4.12)
SE/E Asia	1.45 (1.36−1.54)	0.50 (0.33−0.76)	1.88 (1.50−2.35)

ART, antiretroviral therapy; SE/E Asia, Southeast/East Asia.

Table 3. Estimated annual transition rates (95% confidence interval) of the CD4$^+$ cell-based Markov model of HIV natural history by region for 3072 men aged 15–24 years.

	Europe/North America	Africa	SE/E Asia
Progression rates			
λ_{12}	0.205 (0.141–0.299)	0.198 (0.127–0.309)	0.297 (0.203–0.436)
λ_{23}	0.341 (0.180–0.644)	0.329 (0.178–0.610)	0.493 (0.261–0.932)
λ_{34}	0.348 (0.143–0.848)	0.336 (0.123–0.917)	0.504 (0.205–1.242)
λ_{45}	0.527 (0.188–1.478)	0.509 (0.165–1.574)	0.763 (0.266–2.187)
λ_{56}	0.593 (0.229–1.530)	0.573 (0.185–1.771)	0.858 (0.321–2.295)
λ_{67}	0.897 (0.178–4.529)	0.867 (0.169–4.443)	1.300 (0.255–6.616)
Mortality rates			
λ_{19}	0.001 (0–0.004)	0.003 (0.001–0.014)	0.002 (0.001–0.009)
λ_{29}	0.003 (0.001–0.012)	0.009 (0.002–0.039)	0.006 (0.002–0.025)
λ_{39}	0.003 (0.001–0.006)	0.007 (0.003–0.018)	0.005 (0.002–0.012)
λ_{49}	0.005 (0.001–0.023)	0.013 (0.002–0.067)	0.009 (0.002–0.045)
λ_{59}	0.008 (0.003–0.024)	0.022 (0.006–0.073)	0.015 (0.005–0.048)
λ_{69}	0.014 (0.002–0.113)	0.037 (0.004–0.349)	0.026 (0.003–0.237)
λ_{79}	0.315 (0.217–0.459)	0.845 (0.418–1.709)	0.592 (0.359–0.976)

Refer to Fig. 1 for schematic representation of transition rates. SE/E Asia, Southeast/East Asia.

Table 4. Estimated annual transition rates (95% confidence interval) of the CD4$^+$ cell-based Markov model of HIV natural history by region for 1444 women aged 15–24 years.

	Europe/North America	Africa	SE/E Asia
Progression rates			
λ_{12}	0.190 (0.131–0.276)	0.184 (0.121–0.280)	0.275 (0.189–0.400)
λ_{23}	0.315 (0.164–0.606)	0.305 (0.164–0.567)	0.457 (0.239–0.876)
λ_{34}	0.322 (0.136–0.763)	0.311 (0.118–0.820)	0.466 (0.195–1.114)
λ_{45}	0.488 (0.175–1.362)	0.471 (0.154–1.438)	0.707 (0.248–2.014)
λ_{56}	0.548 (0.217–1.385)	0.530 (0.176–1.596)	0.794 (0.304–2.075)
λ_{67}	0.830 (0.162–4.257)	0.802 (0.155–4.152)	1.202 (0.233–6.21)
Mortality rates			
λ_{19}	0.001 (0–0.004)	0.003 (0.001–0.013)	0.002 (0.001–0.008)
λ_{29}	0.004 (0.001–0.014)	0.009 (0.002–0.039)	0.007 (0.002–0.027)
λ_{39}	0.003 (0.001–0.006)	0.008 (0.003–0.020)	0.005 (0.002–0.011)
λ_{49}	0.005 (0.001–0.024)	0.013 (0.002–0.069)	0.009 (0.002–0.046)
λ_{59}	0.008 (0.003–0.024)	0.022 (0.006–0.075)	0.016 (0.005–0.051)
λ_{69}	0.014 (0.002–0.113)	0.039 (0.004–0.364)	0.027 (0.003–0.243)
λ_{79}	0.328 (0.221–0.486)	0.878 (0.424–1.817)	0.615 (0.366–1.033)

Refer to Fig. 1 for schematic representation of transition rates. SE/E Asia, Southeast/East Asia.

95% CI 0.896–0.976 for CD4$^+$ cell count >500 cells/μl) and the lowest counts (0.956, 95% CI 0.882–1.0 for CD4$^+$ cell count ≤50 cells/μl). Using the fitted transition probability matrix and the misclassification matrix to forecast expected prevalences shows a similar pattern (Fig. S2, http://links.lww.com/QAD/B52).

Discussion

We used the largest database to date of seroconverters infected with HIV-1 to jointly estimate the rates of immune function decline and survival across four continents, thus providing the most comprehensive

Table 5. Median survival estimates in years (interquartile range) from 5000 bootstrapped samples by region for 16 373 HIV-1-infected individuals in 25 countries during 1979–2014.

Age at seroconversion (years)	Europe/North America		Africa		SE/E Asia	
	Men	Women	Men	Women	Men	Women
20	13.0 (12.4–13.4) *3072*	13.5 (12.9–14.2) *1444*	11.6 (10.9–12.3) *402*	10.7 (10.0–12.2) *535*	8.3 (7.9–8.9) *302*	8.2 (7.8–9.0) *18*
30	12.1 (11.6–12.6) *5753*	12.5 (11.9–13.4) *1650*	10.8 (9.9–11.5) *979*	9.7 (9.0–11.5) *618*	7.7 (7.3–8.4) *169*	7.6 (7.1–8.4) *4*
40	10.7 (10.1–11.3) *2630*	11.2 (11.1–11.5) *497*	9.0 (8.2–10.2) *430*	8.1 (7.4–10.2) *174*	6.7 (6.3–7.5) *67*	6.6 (6.1–7.5) *2*

Number of individuals in each category is in italics. SE/E Asia, Southeast/East Asia.

estimates of the natural history of HIV. CD4$^+$ cell counts in HIV-negative populations are highly variable, decreasing with age and lower in men compared with women [23–26]. We found that individuals infected at an older age were more likely to have lower CD4$^+$ cell counts at seroconversion and higher mortality rates similar to other published analyses [8,9,27,28]. Progression rates did not vary by current age, although some studies have found differences by age at seroconversion [4,29,30]. Interestingly, the strongest predictor of progression and mortality is the region from which the cohort originated. Various studies have found slower progression in African compared with European cohorts, and some have also documented lower CD4$^+$ cell counts at seroconversion with no increase in mortality [6,23,31,32]. However, these differences could be driven by environmental factors and differential access to healthcare rather than ethnicity as Africans living in London experienced similar progression rates to non-Africans [10,33]. As we do not differentiate between different ethnicities but stratify by geographic region, and aim to remove the effects of access to treatment, it is not surprising that we do not detect significantly slower progression in the African cohorts compared with the European and North American cohorts. Cohorts from SE/E Asia experience faster CD4$^+$ cell decline and higher mortality rates, and this, in combination with lower CD4$^+$ cell counts at the time of seroconversion, leads to significantly shorter survival times compared with the other regions. A study across 17 clinical sites in Asia showed the total lymphocyte count and CD4$^+$ cell count in untreated HIV-infected persons to be lower compared with a clinical study conducted in France [34]. The authors cite several potential explanations for the finding including variations in immune function or genetic factors. Differences in virus subtype (subtype CRF01_AE is the predominant form in Southeast Asia) could also influence progression rates [35]. In addition, lower CD4$^+$ cell counts at seroconversion [16], faster CD4$^+$ cell decline [16,36], and higher mortality rates in Asian populations compared with Europeans have been observed. Other studies have also noted lower CD4$^+$ cell counts in HIV-negative individuals in Asian cohorts compared with Europeans. After controlling for differences in transition rates by age and region, we find no clear differences in mortality rates between men and women and only marginal differences in progression rates in line with previous studies [7,27].

Our survival estimates agree well with earlier studies, although our estimates in older age groups tend to be slightly longer. Estimates of median survival times in European cohorts in the CASCADE collaboration for ages 15–24, 25–34, and 35–45 years were 12.5 (95% CI 12.1–12.9), 10.9 (10.9–11.3) and 9.1 years (8.7–9.5), respectively [9]. For the same age groups, we predict mean survival times of 12.3–13.2, 10.9–12.3, and 10.7–10.9 years. Published estimates for mean survival in men

aged 25–29 years in South Africa are 11.6 years (95% CI 9.8–13.7), East Africa 11.1 years (95% CI 8.7–14.2), and Thailand 7.5 years (95% CI 5.4–10.4) [8]. Again, we find similar values of 10.3–10.4 years for African men and 7.2–7.7 years in men for Thailand, Singapore, and China. Although we generalize our findings to Europe/North America and Africa, the European data originate from Western Europe, and the African data are clustered in East and South Africa. SE/E Asia region includes data from Thailand, Singapore, and China. The expected survival times estimated by the Markov model are extrapolated under the assumption that the transition probabilities are constant and will remain constant in the future, meaning that the transitions to ART, estimated as a means of reducing bias, are averaged across the study period. However, as expected, the rates of initiating ART are lower in Africa and Asia compared with Europe, consistent with the later availability in these regions along with lower uptake. Applying a Kaplan–Meier analysis that ignores this bias results in implausibly long survival times (Fig. S3, http://links.lww.com/QAD/B52).

One assumption of the staged Markov model is that the observations occur at random intervals (i.e. not following acute events or due to patient self-selection), which could lead to sampling bias. There is also a bias towards longer survivors present in all studies of this type. Individuals are both right-censored (removed from the study due to various circumstances) and left-censored (they enter the study at some unknown stage of the disease). Although some evidence suggests longer survival times in later stages of the epidemic, this is likely to be confounded with access to ART or improved healthcare [37]. Data on mode of acquisition are available for 70% of individuals in this database, and the majority of our data are from MSM (53% of known risk groups) followed by men who have sex with women (28%). We do not distinguish between risk groups here, as additional stratification in our model would result in very low numbers in each subgroup, preventing us from obtaining reliable estimates. There is mounting evidence suggesting that progression rates can be influenced by host characteristics such as human leukocyte antigen, virus subtype, and set-point viral load [38–43]. A natural next step would be to include this information in our model and investigate whether the regional differences found here can be explained by variations in viral dynamics [44–46].

Our results are based on the largest ever compiled dataset of seroconverters and give important insights into the influence of demographic factors on CD4$^+$ cell progression and survival, highlighting a need for HIV care guidelines to be tailored to specific subgroups. The faster progression and shorter survival times in the cohorts from SE/E Asia are of particular importance, and further investigation is necessary to understand the biological mechanisms of these differences.

Acknowledgements

Members of the UNAIDS Working Group on CD4 Progression and Mortality Amongst HIV Seroconverters including the CASCADE Collaboration in EuroCoord: Nikos Pantazis[a], Brian Agan[b,c], Connie Celum[d], Tsungai Chipato[e], Myron Cohen[f], Amelia C. Crampin[g], Marcel E. Curlin[h], Christine Danel[i], Delphine Gabillard[j], Judith Glynn[g], Simon Gregson[k], Marguerite Guiguet[l], Kobus Herbst[m], Timothy H. Holtz[n], Xiaojie Huang[o], Manjeetha Jaggernath[p], Etienne Karita[q], William Kilembe[q], Cheng C. Lee[r], Wanna Leelawiwat[n], Yee-Sin Leo[r], Jairam R. Lingappa[d], Tom Lutalo[s], Grace Macalino[b,c], Billy Mayanja[t], Denna Michael[u], Jessica N. Miiro[t], Albert Minga[v], Charles Morrison[w], Sikhulile Moyo[x,y], Jill Murray[y], Rosemary Musonda[x], Emanuel M. Mwendo[u,z], Thumbi Ndung'u[aa], Kenrad Nelson[bb], OonTek Ng[r], Vladimir Novitsky[cc], Rebecca N. Nsubuga[t], Constance Nyamukapa[k], Julien Nyombayire[q], Deenan Pillay[m], Alison Price[g], Ram Rangsin[dd], Steven J. Reynolds[ee], Robert Salata[ff], Emma Slaymaker[g], Kelly Soderberg[gg], Pam Sonnenberg[hh], Albert Takaruza[ii], Jim Todd[g], Mark Urassa[ii], Hao Wu[kk], Tong Zhang[kk], Giota Touloumi[a], Peter D. Ghys[ll], Kholoud Porter[hh], Timothy B. Hallett[k], Jeffrey W. Eaton[k]

[a]Department of Hygiene, Epidemiology and Medical Statistics, Athens University Medical School, Athens, Greece; [b]Uniformed Services University of the Health Sciences; [c]Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, Maryland; [d]University of Washington, Seattle, Washington, USA; [e]University of Zimbabwe-UCSF Collaborative Research Program, Harare, Zimbabwe; [f]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; [g]London School of Hygiene and Tropical Medicine, London, UK; [h]Division of Infectious Diseases, Department of Medicine, Oregon Health and Sciences University, Portland, Oregon, USA; [i]PACCI Program, ANRS Research Site of Côte d'Ivoire, INSERM U897, University of Bordeaux; [j]INSERM, Université Bordeaux Segalen, Bordeaux, France; [k]Imperial College London, London, UK; [l]INSERM, France; [m]Africa Centre for

Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa; ^{n}Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ^{o}Capital Medical University, Beijing, China; ^{p}HIV Pathogenesis Programme, Durban, South Africa; ^{q}Projet San Francisco, Kigali, Rwanda; ^{r}Tan Tock Seng Hospital, Singapore, Singapore; ^{s}Rakai Health Sciences Program (RHSP); ^{t}Medical Research Council, Entebbe, Uganda; ^{u}National Institute for Medical Research (NIMR), Mwanza, Tanzania; ^{v}Programme PAC-CI, Abidjan, Côte d'Ivoire; ^{w}FHI 360, Durham, North Carolina, USA; ^{x}Botswana Harvard AIDS Institute, Botswana; ^{y}University of Stellenbosch, Stellenbosch, South Africa; ^{y}National Institute for Occupational Health, National Health Laboratory Service and School of Public Health, University of the Witwatersrand, Johannesburg, South Africa; ^{z}Kilimanjaro Christian Medical University College, Moshi, Tanzania; ^{aa}University of Kwazulu-Natal, Mtubatuba, South Africa; ^{bb}Johns Hopkins University, Baltimore, Maryland; ^{cc}Harvard School of Public Health, Boston, Massachusetts, USA; ^{dd}Department of Military and Community Medicine, Phramongkutklao College of Medicine, Bangkok, Thailand; ^{ee}NIAID, Bethesda, Maryland; ^{ff}Case Western Reserve University, Cleveland, Ohio; ^{gg}Duke University, Durham, North Carolina, USA; ^{hh}University College London, London, UK; ^{ii}Manicaland Centre for Public Health, Rusape, Zimbabwe; National Institute for Medical Research-Mwanza Research Centre, Mwanza, Tanzania; ^{kk}Beijing You'an Hospital, Capital Medical University, Beijing, China; ^{ll}UNAIDS, Geneva, Switzerland.

Author contributions: T.D.M. contributed in literature search, figures, study design, data analysis, data interpretation, and writing; Nikos Pantazis and Timothy B. Hallett contributed in study design, data collection, data interpretation, and writing; Brian Agan, Connie Celum, Tsungai Chipato, Myron Cohen, Amelia C. Crampin, Marcel E. Curlin, Christine Danel, Delphine Gabillard, Judith Glynn, Simon Gregson, Marguerite Guiguet, Kobus Herbst, Timothy H. Holtz, Xiaojie Huang, Manjeetha Jaggernath, Etienne Karita, William Kilembe, Cheng Chuan Lee, Wanna Leelawiwat, Yee-Sin Leo, Jairam R. Lingappa, Tom Lutalo, Grace Macalino, Billy Mayanja, Denna Michael, Jessica N. Miiro, Albert Minga, Charles Morrison, Sikhulile Moyo, Jill Murray, Rosemary Musonda, Emanuel M. Mwendo, Thumbi Ndung'u, Kenrad Nelson, OonTek Ng, Vladimir Novitsky, Rebecca N. Nsubuga, Constance Nyamukapa, Julien Nyonbayire, Deenan Pillay, Alison Price, Ram Rangsin, Steven J. Reynolds, Robert Salata, Emma Slaymaker, Kelly Soderberg, Pam Sonnenberg, Albert Takaruza, Jim Todd, Mark Urassa, Hao Wu, Tong Zhang, Giota Touloumi, Peter D. Ghys, and Kholoud Porter contributed in data collection and writing; Jeffrey W. Eaton contributed in study design, data collection, data analysis, data interpretation, and writing.

Conflicts of interest
There are no conflicts of interest.

References

1. Stover J, Andreev K, Slaymaker E, Gopalappa C, Sabin K, Velasquez C, et al. **Updates to the Spectrum model to estimate key HIV indicators for adults and children.** *AIDS* 2014; **28**:S427–S434.
2. Cori A, Ayles H, Beyers N, Schaap A, Floyd S, Sabapathy K, et al. **HPTN 071 (PopART): a cluster-randomized trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: mathematical model.** *PLoS One* 2014; **9**:e84511.
3. Birrell PJ, Gill ON, Delpech VC, Brown AE, Desai S, Chadborn TR, et al. **HIV incidence in men who have sex with men in England and Wales 2001–10: a nationwide population study.** *Lancet Infect Dis* 2013; **13**:313–318.
4. Lodi S, Phillips A, Touloumi G, Geskus R, Meyer L, Thiébaut R, et al. **Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 cells/mm³: assessment of need following changes in treatment guidelines.** *Clin Infect Dis* 2011; **53**:817–825.
5. Touloumi G, Pantazis N, Pillay D, Paraskevis D, Chaix M-L, Bucher HC, et al. **Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline, and viral load set point in European seroconverter cohorts.** *Clin Infect Dis* 2013; **56**:888–897.
6. Pantazis N, Morrison C, Amornkul PN, Lewden C, Salata RA, Minga A, et al. **Differences in HIV natural history among African and Non-African seroconverters in Europe and seroconverters in sub-Saharan Africa.** *PLoS One* 2012; **7**:e32369.
7. CASCADE Collaboration. **Differences in CD4 cell counts at seroconversion and decline among 5739 HIV-1-infected individuals with well estimated dates of seroconversion.** *J Acquir Immune Defic Syndr* 2003; **34**:76–83.
8. Todd J, Glynn JR, Marston M, Lutalo T, Biraro S, Mwita W, et al. **Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy.** *AIDS* 2007; **21**:S55–S63.
9. Babiker A, Darby S, De Angelis D, Kwart D, Porter K, Beral V, et al. **Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis.** *Lancet* 2000; **355**:1131–1137.
10. Morgan D, Mahe C, Mayanja B, Okongo JM, Lubega R, Whitworth JA. **HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries?** *AIDS* 2002; **16**:597–603.
11. CASCADE Collaboration. **Changes in the uptake of antiretroviral therapy and survival in people with known duration of HIV infection in Europe: results from CASCADE.** *HIV Med* 2000; **1**:224–231.
12. Touloumi G, Pantazis N, Babiker AG, Walker SA, Katsarou O, Karafoulidou A, et al. **Differences in HIV RNA levels before the initiation of antiretroviral therapy among 1864 individuals with known HIV-1 seroconversion dates.** *AIDS* 2004; **18**:1697–1705.
13. Donnelly C, Bartley L, Ghani A, Le Fevre A, Kwong G, Cowling B, et al. **Gender difference in HIV-1 RNA viral loads.** *HIV Med* 2005; **6**:170–178.
14. Sterling TR, Chaisson RE, Moore RD. **HIV-1 RNA, CD4 T-lymphocytes, and clinical response to highly active antiretroviral therapy.** *AIDS* 2001; **15**:2251–2257.
15. Stephenson J, Griffioen A, Woronowski H, Phillips A, Petruckevitch A, Keenlyside R, et al. **Survival and progression of HIV disease in women attending GUM/HIV clinics in Britain and Ireland.** *Sex Transm Infect* 1999; **75**:247–252.
16. Huang X, Lodi S, Fox Z, Li W, Phillips A, Porter K, et al. **Rate of CD4 decline and HIV-RNA change following HIV seroconversion in men who have sex with men: a comparison between the Beijing PRIMO and CASCADE cohorts.** *J Acquir Immune Defic Syndr* 2013; **62**:441–446.

17. Guihenneuc-Jouyaux C, Richardson S, Longini IM. **Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline.** *Biometrics* 2000; **56**:733–741.

18. Satten GA, Longini IM Jr. **Markov chains with measurement error: estimating the true course of a marker of the progression of human immunodeficiency virus disease.** *Appl Stat* 1996; **45**:275–309.

19. Hendriks JC, Satten GA, Longini IM, van Druten HA, Schellekens PTA, Coutinho RA, et al. **Use of immunological markers and continuous-time Markov models to estimate progression of HIV infection in homosexual men.** *AIDS* 1996; **10**:649–656.

20. Cappe' O, Moulines E, Ryde'n T. **Inference in hidden Markov models**. *Proceedings of EUSFLAT Conference*, pp. 14–16. 2006.

21. Jackson CH. **Multistate models for panel data: the msm package for R.** *J Stat Softw* 2011; **38**:1–29.

22. Iman R. **Latin hypercube sampling.** *Encyclopedia of quantitative risk analysis and assessment*. Chichester, UK: John Wiley & Sons, Ltd; 2008.

23. Williams BG, Korenromp EL, Gouws E, Schmid GP, Auvert B, Dye C. **HIV infection, antiretroviral therapy, and CD4+ cell count distributions in African populations.** *J Infect Dis* 2006; **194**:1450–1458.

24. Lugada ES, Mermin J, Kaharuza F, Ulvestad E, Were W, Langeland N, et al. **Population-based hematologic and immunologic reference values for a healthy Ugandan population.** *Clin Diagn Lab Immunol* 2004; **11**:29–34.

25. Kassu A, Tsegaye A, Petros B, Wolday D, Hailu E, Tilahun T, et al. **Distribution of lymphocyte subsets in healthy human immunodeficiency virus-negative adult Ethiopians from two geographic locales.** *Clin Diagn Lab Immunol* 2001; **8**:1171–1176.

26. Prins M, Robertson JR, Brettle RP, Aguado IH, Broers B, Boufassa F, et al. **Do gender differences in CD4 cell counts matter?** *AIDS* 1999; **13**:2361–2364.

27. Lutalo T, Gray RH, Wawer M, Sewankambo N, Serwadda D, Laeyendecker O, et al. **Survival of HIV-infected treatment-naive individuals with documented dates of seroconversion in Rakai, Uganda.** *AIDS* 2007; **21**:S15–S19.

28. Darby SC, Ewart DW. **Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population.** *Lancet* 1996; **347**:1573–1579.

29. Operskalski EA, Stram DO, Lee H, Zhou Y, Donegan E, Busch MP, et al. **Human immunodeficiency virus type 1 infection: relationship of risk group and age to rate of progression to AIDS.** *J Infect Dis* 1995; **172**:648–655.

30. Touloumi G, Hatzakis A, Rosenberg PS, O'Brien TR, Goedert JJ. **Effects of age at seroconversion and baseline HIV RNA level on the loss of CD4+ cells among persons with hemophilia.** *AIDS* 1998; **12**:1691–1697.

31. Mekonnen Y, Geskus RB, Hendriks JCM, Messele T, Borghans J, Miedema F, et al. **Low CD4 T cell counts before HIV-1 seroconversion do not affect disease progression in Ethiopian factory workers.** *J Infect Dis* 2005; **192**:739–748.

32. May M, Wood R, Myer L, Taffé P, Rauch A, Battegay M, et al. **CD4+ T cell count decreases by ethnicity among untreated patients with HIV infection in South Africa and Switzerland.** *J Infect Dis* 2009; **200**:1729–1735.

33. Del Amo J, Petruckevitch A, Phillips A, Johnson AM, Stephenson J, Desmond N, et al. **Disease progression and survival in HIV-1-infected Africans in London.** *AIDS* 1998; **12**:1203–1209.

34. Achhra AC, Zhou J, Dabis F, Pujari S, Thiebaut R, Law MG, et al. **Difference in absolute CD4+ count according to CD4 percentage between Asian and Caucasian HIV-infected patients.** *J AIDS Clin Res* 2010; **1**:1–4.

35. Buonaguro L, Tornesello M, Buonaguro F. **Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications.** *J Virol* 2007; **81**:10209–10219.

36. Rangsin R, Chiu J, Khamboonruang C, Sirisopana N, Eiumtrakul S, Brown AE, et al. **The natural history of HIV-1 infection in young Thai men after seroconversion.** *J Acquir Immune Defic Syndr* 2004; **36**:622–629.

37. CASCADE Collaboration. **Effect of ignoring the time of HIV seroconversion in estimating changes in survival over calendar time in observational studies: results from CASCADE.** *AIDS* 2000; **14**:1899–1906.

38. Deeks SG, Kitchen CMR, Liu L, Guo H, Gascon R, Narváez AB, et al. **Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load.** *Blood* 2004; **104**:942–947.

39. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. **Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis.** *Proc Natl Acad Sci U S A* 2007; **104**:17441–17446.

40. Lavreys L, Baeten JM, Chohan V, McClelland RS, Hassan WM, Richardson BA, et al. **Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women.** *Clin Infect Dis* 2006; **42**:1333–1339.

41. Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, et al. **HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads.** *J Infect Dis* 2007; **195**:1177–1180.

42. Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, et al. **Advantage of rare HLA supertype in HIV disease progression.** *Nat Med* 2003; **9**:928–935.

43. Kaleebu P, French N, Mahe C, Yirrell D, Watera C, Lyagoba F, et al. **Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda.** *J Infect Dis* 2002; **185**:1244–1250.

44. Brown AE, Malone JD, Zhou SYJ, Lane JR, Hawkes CA. **Human immunodeficiency virus RNA levels in US adults: a comparison based upon race and ethnicity.** *J Infect Dis* 1997; **176**:794–797.

45. Anastos K, Gange SJ, Lau B, Weiser B, Detels R, Giorgi JV, et al. **Association of race and gender with HIV-1 RNA levels and immunologic progression.** *J Acquir Immune Defic Syndr* 2000; **24**:218–226.

46. Saul J, Erwin J, Sabin CA, Kulasegaram R, Peters BS. **The relationships between ethnicity, sex, risk group, and virus load in human immunodeficiency virus type 1 antiretroviral-naive patients.** *J Infect Dis* 2001; **183**:1518–1521.