CD4\(^+\) cell dynamics in untreated HIV-1 infection: overall rates, and effects of age, viral load, sex and calendar time

Anne Cori\(^a,*\), Michael Pickles\(^a,*\), Ard van Sighem\(^b\), Luuk Gras\(^b\), Daniela Bezemer\(^b\), Peter Reiss\(^b\) and Christophe Fraser\(^a\)

**Background:** CD4\(^+\) cell count is a key measure of HIV disease progression, and the basis of successive international guidelines for treatment initiation. CD4\(^+\) cell dynamics are used in mathematical and econometric models for evaluating public health need and interventions. Here, we estimate rates of CD4\(^+\) decline, stratified by relevant covariates, in a form that is clinically transparent and can be directly used in such models.

**Methods:** We analyse the AIDS Therapy Evaluation in the Netherlands cohort, including individuals with date of seroconversion estimated to be within 1 year and with intensive clinical follow-up prior to treatment initiation. Owing to the fact that CD4\(^+\) cell counts are intrinsically noisy, we separate the analysis into long-term trends of smoothed CD4\(^+\) cell counts and an observation model relating actual CD4\(^+\) measurements to the underlying smoothed counts. We use a monotonic spline smoothing model to describe the decline of smoothed CD4\(^+\) cell counts through categories CD4\(^+\) above 500, 350–500, 200–350 and 200 cells/\(\mu L\) or less. We estimate the proportion of individuals starting in each category after seroconversion and the average time spent in each category. We examine individual-level cofactors which influence these parameters.

**Results:** Among untreated individuals, the time spent in each compartment was 3.32, 2.70, 5.50 and 5.06 years. Only 76% started in the CD4\(^+\) cell count above 500 cells/\(\mu L\) compartment after seroconversion. Set-point viral load (SPVL) was an important factor: individuals with at least 5 log\(_{10}\) copies/ml took 5.37 years to reach CD4\(^+\) cell count less than 200 cells/\(\mu L\) compared with 15.76 years for SPVL less than 4 log\(_{10}\) copies/ml.

**Conclusion:** Many individuals already have CD4\(^+\) cell count below 500 cells/\(\mu L\) after seroconversion. SPVL strongly influences the rate of CD4\(^+\) decline. Treatment guidelines should consider measuring SPVL, whereas mathematical models should incorporate SPVL stratification.

**Keywords:** CD4\(^+\) cell count, HIV, mathematical model, set-point viral load, viral load

Introduction

Initiation of HAART in HIV-positive individuals is currently recommended to occur once their CD4\(^+\) cell count decreases below 500 cells/\(\mu L\), according to WHO guidelines [1], although in many settings national guidelines are still based on the earlier recommendation [2] of CD4\(^+\) 350 cells/\(\mu L\) or less.

Apart from being the main clinical marker for HAART initiation, CD4\(^+\) cell count is also a measure of HIV disease progression. Most mathematical models of HIV...
transmission, including those used in economic evaluation, include a representation of CD4⁺ decline over time, usually via compartments: above 500, 350–500, 200–350 and below 200 cells/μl. This structure is based on the historical evolution of the WHO guidelines for treatment initiation.

Models are increasingly playing key roles in policy decision-making in HIV healthcare. Modelling has been influential in the adoption of current WHO treatment guidelines [3] and was used in designing the recent PEPFAR combination prevention trials [4]. Accurately modelling the progression of HIV-infected individuals through these different stages is crucial to provide reliable estimates of incidence, prevalence and number in need of treatment, and hence for robust policy recommendations and assessments.

Although the mean rate of CD4⁺ cell count decline has been measured in many studies, the times to reach each of the aforementioned CD4⁺ cell count thresholds have, to our knowledge, only been previously analysed in Lodi et al. [5]. That study did not explicitly model censoring because of treatment initiation, death or loss to follow-up, and its results had to be processed to be incorporated in mathematical models [6]. Here, we propose to analyse CD4⁺ cell count data in a form that can be directly integrated into mathematical and econometric models, and include stratifications that would a priori be thought to affect epidemiological dynamics.

Age and sex may influence the CD4⁺ cell count at seroconversion [7,8]. Similarly, the rate of decline in CD4⁺ cell count depends on a number of factors, including age at seroconversion [7–10] and viral load [10,11]. Set-point viral load (SPVL) is known to also affect the time to AIDS [12–14]. However, to date, no work has quantified how these factors affect the time to reach these CD4⁺ thresholds, and therefore mathematical models of HIV transmission seldom stratify disease progression by any of them.

In this study, we use data from a cohort of individuals with date of seroconversion known to be within 1 year to estimate the CD4⁺ cell count category at seroconversion and time taken to progress from one CD4⁺ category to the next. We explore how these vary according to age at seroconversion, SPVL, sex and HIV transmission route in a treatment-naive population. These stratifications are to our knowledge novel to this analysis.

Material and methods

Data used

We used data from the AIDS Therapy Evaluation in the Netherlands (ATHENA) observational cohort [15] including HIV-infected patients followed in one of the 27 HIV treatment centres in the Netherlands since 1996. At entry into the cohort, as well as during follow-up visits, clinical, virological and immunological data were collected. The ATHENA cohort includes all patients registered for HIV care in Netherlands, provided they were in clinical care in or after 1996. Data collection was initiated in 1998. All data prior to 1998 were included retrospectively. Data from some patients who died prior to 1996 has also been included retrospectively.

Data selection

The ATHENA cohort comprises 21,999 individuals in total, of whom we selected the 1,039 patients with a seroconversion window (defined by a negative HIV test and a positive HIV test) no longer than a year, at least 6 months between first positive HIV test and antiretroviral therapy (ART) initiation, and at least six CD4⁺ cell counts prior to HAART initiation.

Sensitivity analyses were performed for patients who had between three and seven CD4⁺ measurements (n = 1,571 and n = 903, respectively); the results as well as further detail on patient selection are shown in the Supplementary Material, http://links.lww.com/QAD/A766.

Ethical approval

Ethical approval in the ATHENA cohort was not obtained as the data used are collected from patients as part of their routine HIV care. Patients can opt out from data collection after being informed by their treating physician of the purpose of collection.

Progression model

We modelled HIV progression as shown in Fig. 1a. Upon seroconversion, individuals enter one of four categories, defined by their CD4⁺ cell count, measured in cells/μl: CD4⁺ > 500, 350 < CD4⁺ ≤ 500, 200 < CD4⁺ ≤ 350 and CD4⁺ ≤ 200. They then successively progress through decreasing CD4⁺ categories, at a rate qₖ (k = 1, . . ., 4) depending on their current CD4⁺ category (k). At any stage, individuals may be lost to follow-up or die, from HIV-related or unrelated causes. They can also initiate treatment if eligible according to local guidelines.

Statistical analysis

In the following, the CD4⁺ categories referred to are those defined in the previous subsection and illustrated in Fig. 1a.

Smoothed and observed CD4⁺ cell counts

Owing to the fact that CD4⁺ cell counts are highly variable [16], we hypothesize that the model shown in Fig. 1a best describes the underlying dynamics of smoothed CD4⁺ cell counts, and that actual CD4⁺ cell counts are subject to additional random variability. We thus focus our survival analysis on smoothed CD4⁺ cell counts and estimate additional parameters associated with
random variation of actual CD4$^+$ cell counts relative to their underlying trend.

**Estimation of smoothed CD4$^+$ cell counts**

We used a smoothing model to describe the dynamics of CD4$^+$ cell count between seroconversion and initiating HAART. More specifically, for each individual, we fitted a monotonic decreasing cubic regression spline to describe the decline of CD4$^+$ cell count as a function of time since seroconversion [17] (see Supplementary Material for detail, http://links.lww.com/QAD/A766).

We used these predicted trajectories of smoothed CD4$^+$ cell counts to define, for each individual, their CD4$^+$ category at the estimated date of seroconversion as well as the times of transition to lower CD4$^+$ categories. The projected smoothed CD4$^+$ cell counts after acute infection were used to estimate $f_k$ as the empirical proportion of individuals entering CD4$^+$ cell count category $k$ upon seroconversion. We performed maximum likelihood parametric exponential survival analyses to estimate $1/q_k$, that is the average time for all individuals to transition from one CD4$^+$ category to the next. A progression endpoint was defined if the smoothed CD4$^+$ cell count crossed a boundary between CD4$^+$ cell count categories, whereas a censoring event was defined if the patient started treatment, died or reached their final CD4$^+$ measurement in the database. Confidence intervals (CIs) were obtained for all parameters by bootstrapping patients. Results were visually compared with nonparametric Kaplan–Meier survival estimates.

**Estimation of set-point viral load**

For each patient, we defined SPVL as the geometric mean of all viral load measurements taken during the set-point window (defined as 6–24 months following the first positive HIV test) and prior to HAART initiation. SPVL was available for 873 of the 1039 individuals considered in the main analysis. Viral load units are per millilitre of
peripheral blood plasma (see Supplementary Material for details, http://links.lww.com/QAD/A766).

**Subgroup analyses**

The analysis was carried out both unstratified and stratified by each of the following characteristics, in order to assess the influence of these factors on individual clinical progression: SPVL, age at seroconversion, sex, calendar time and HIV transmission route. The categories used for log_{10} SPVL (<4, 4–4.5, 4.5–5, ≥5) and age at seroconversion (<30, 30–35, 35–40, ≥40) corresponded approximately to the quartiles of SPVL and age at seroconversion.

**Sensitivity analyses**

Extensive sensitivity analyses were performed to assess the robustness of the method. Other regression methods were considered to assess the influence of the regression choice on the results. These included an unconstrained cubic spline and a monotonic linear regression, respectively less and more constrained than the monotonic spline, as well as a model with no smoothing. To analyse the robustness of the results to the minimum number of CD4+ measurements, the analysis with monotonic cubic splines was repeated with a minimum of five to seven CD4+ measurements, whereas for the linear regression, we further varied the minimum number of CD4+ measurements from three to seven. To assess the effects of mono and dual ART, we repeated these analyses for individuals using only CD4+ measurements prior to the initiation of mono or dual therapy, respectively. We also examined the influence of retrospective inclusion of patients diagnosed before 1996 by comparing individuals recruited after 1996 to the whole study population (see Supplementary Material for more individuals, http://links.lww.com/QAD/A766).

**Set-point viral load as a continuous variable**

Additional analysis of progression to the next CD4+ category was carried out by applying a Cox proportional hazards model to estimate the progression rates q_k as linear functions of SPVL.

**Validation**

We used as a validation set the 514 patients with seroconversion window between 1 and 2 years, and at least six CD4+ cell counts prior to HAART initiation. For each of these patients, we predicted, given the first observed CD4+ category, the times at which the smoothed CD4+ cell count would transition from one category to the next. Our predictive model accounted for the potential mismatch between the first observed and smoothed CD4+ categories as well as the smoothed CD4+ cell count progression over time. The predicted times of transition were then compared with those directly derived by monotonic spline smoothing of the observed CD4+ cell counts for each patient (see Supplementary material for detail, http://links.lww.com/QAD/A766).

The analysis was carried out using R version 3.1.0. Unconstrained and monotonic splines were created using the smooth.spline function in the stats package and the gam and mono.con functions in the mgcv package, respectively. Survival analyses were performed using the survival package.

**Results**

**Study population**

The median age at seroconversion of the 1039 individuals included in the main analysis was 35.4 years (interquartile range 28.8–42.2 years); 66 (6.4%) were women, whereas 867 (83.4%) of the study population were MSM. Seventy-four patients had died by the end of the study period, although only two were still HAART naïve. The median initial CD4+ cell count was 570 (426–720) cells/μl and a median of 11 CD4+ measurements were taken over 2.9 years. Only 433 patients had a recorded HIV subtype, and of these, 397 (91.7%) were infected with subtype B.

SPVL was available for 873 (84.0%) patients and based on 1–21 measurements. The median SPVL was 4.6 (4.0–4.9) log_{10} copies/ml. Across the cohort, SPVL varied over time (Figure S2, http://links.lww.com/QAD/A766) [18]. Mean SPVL in patients who seroconverted during 1995–2000 was 3.9 log_{10} copies/ml, significantly lower than before 1995 (4.3 log_{10} copies/ml, T-test P value <0.01) and after 2000 (4.5 log_{10} copies/ml, T-test P value <0.001).

Figure S3, http://links.lww.com/QAD/A766, shows the distribution of all individual viral load measurements throughout the set-point window, stratified by SPVL and CD4+ cell count. Most viral load measurements of a patient lie in the range of their assigned SPVL category, even when the patient progresses to lower CD4+ categories, showing that SPVL is a good predictor of viral load throughout untreated chronic infection. Only in patients with SPVL below 4 log_{10} copies/ml, do individuals with low CD4+ cell counts appear to have an average lower viral load measurements.

**Individual CD4+ cell dynamics**

Figure 1b–f shows, for five selected individuals, their observed and smoothed CD4+ cell counts (see Supplementary Material for more individuals, http://links.lww.com/QAD/A766). Coloured rectangles show the CD4+ category each individual is inferred to be in over time, from seroconversion onwards, by the smoothed model. The figure illustrates how variable the CD4+ dynamics are from one individual to the next, both in
Table 1. Observed versus smoothed CD4⁺ category (defined as pᵢ in the text).

<table>
<thead>
<tr>
<th>Smoothed CD4⁺ category</th>
<th>Observed CD4⁺&gt;500</th>
<th>Observed CD4⁺ 350–500</th>
<th>Observed CD4⁺ 200–350</th>
<th>Observed CD4⁺ ≤200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed CD4⁺&gt;500</td>
<td>4853 (83.98%)</td>
<td>910 (18.69%)</td>
<td>24 (0.85%)</td>
<td>1 (0.16%)</td>
</tr>
<tr>
<td>Observed CD4⁺ 350–500</td>
<td>855 (14.79%)</td>
<td>3122 (64.12%)</td>
<td>566 (20.11%)</td>
<td>8 (1.29%)</td>
</tr>
<tr>
<td>Observed CD4⁺ 200–350</td>
<td>65 (1.21%)</td>
<td>815 (16.74%)</td>
<td>1961 (69.66%)</td>
<td>66 (10.61%)</td>
</tr>
<tr>
<td>Observed CD4⁺≤200</td>
<td>6 (0.10%)</td>
<td>22 (0.45%)</td>
<td>264 (9.38%)</td>
<td>547 (87.94%)</td>
</tr>
<tr>
<td>Total</td>
<td>5779 (100%)</td>
<td>4869 (100%)</td>
<td>2815 (100%)</td>
<td>622 (100%)</td>
</tr>
</tbody>
</table>

Each column in the table shows the distribution of observed CD4⁺ cell count for a given category of the smoothed CD4⁺.

Table 2. Estimates of mean time 1/qᵢ (in years) spent in CD4⁺ compartment (>500, 350–500, 200–350, ≤200 cells/μL), fraction of individuals fᵢ starting in each CD4⁺ compartment and expected time (in years) to reaching CD4⁺ cell count of 200 cells/μL given current stage of infection.

<table>
<thead>
<tr>
<th>Average time spent in each CD4⁺ category</th>
<th>1/q₁</th>
<th>1/q₂</th>
<th>1/q₃</th>
<th>1/q₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstratified</td>
<td>3.32 (3.07–3.58)</td>
<td>2.70 (2.48–2.94)</td>
<td>5.50 (4.69–6.54)</td>
<td>5.06 (3.61–7.29)</td>
</tr>
<tr>
<td>Stratified by SPVL log₁₀ SPVL &lt; 4.0</td>
<td>3.12 (2.68–3.64)</td>
<td>2.35 (2.08–2.64)</td>
<td>3.09 (2.65–3.64)</td>
<td>8.39 (5.46–15.55)</td>
</tr>
<tr>
<td>stratified by SPVL log₁₀ SPVL 4.5–5.0</td>
<td>1.51 (1.28–1.76)</td>
<td>1.44 (1.22–1.69)</td>
<td>1.93 (1.2–1.49)</td>
<td>2.14 (1.32–3.59)</td>
</tr>
<tr>
<td>Stratified by SC age SC age ≤30</td>
<td>2.67 (2.34–3.02)</td>
<td>2.17 (1.89–2.48)</td>
<td>4.75 (3.64–6.62)</td>
<td>2.57 (1.69–4.02)</td>
</tr>
<tr>
<td>SC age &gt;30</td>
<td>0.76 (0.73–0.78)</td>
<td>0.19 (0.17–0.22)</td>
<td>0.05 (0.03–0.06)</td>
<td>0.00 (0.00–0.01)</td>
</tr>
<tr>
<td>SC age 30–35</td>
<td>0.86 (0.81–0.91)</td>
<td>0.11 (0.07–0.16)</td>
<td>0.02 (0.00–0.04)</td>
<td>0.00 (0.00–0.00)</td>
</tr>
<tr>
<td>SC age 40–50</td>
<td>0.78 (0.72–0.83)</td>
<td>0.19 (0.14–0.25)</td>
<td>0.03 (0.01–0.05)</td>
<td>0.00 (0.00–0.00)</td>
</tr>
<tr>
<td>SC age ≥60</td>
<td>0.74 (0.69–0.79)</td>
<td>0.21 (0.16–0.26)</td>
<td>0.05 (0.03–0.08)</td>
<td>0.00 (0.00–0.01)</td>
</tr>
<tr>
<td>Proportion in each CD4⁺ category after seroconversion</td>
<td>f₁</td>
<td>f₂</td>
<td>f₃</td>
<td>f₄</td>
</tr>
<tr>
<td>Unstratified</td>
<td>0.72 (0.67–0.77)</td>
<td>0.21 (0.17–0.26)</td>
<td>0.06 (0.03–0.09)</td>
<td>0.01 (0.00–0.02)</td>
</tr>
<tr>
<td>Stratified by SC age SC age ≤30</td>
<td>0.77 (0.71–0.83)</td>
<td>0.17 (0.12–0.23)</td>
<td>0.05 (0.02–0.09)</td>
<td>0.01 (0.00–0.02)</td>
</tr>
<tr>
<td>SC age &gt;30</td>
<td>0.77 (0.73–0.82)</td>
<td>0.20 (0.15–0.24)</td>
<td>0.03 (0.01–0.05)</td>
<td>0.00 (0.00–0.01)</td>
</tr>
<tr>
<td>SC age 30–35</td>
<td>10.54 (9.68–11.60)</td>
<td>11.51 (10.59–12.58)</td>
<td>8.19 (7.32–9.26)</td>
<td>5.00 (4.69–6.54)</td>
</tr>
<tr>
<td>SC age 50–55</td>
<td>13.82 (10.75–21.03)</td>
<td>14.60 (11.46–21.76)</td>
<td>11.48 (8.43–18.73)</td>
<td>8.39 (5.46–15.55)</td>
</tr>
<tr>
<td>Average time to reaching CD4⁺ 200</td>
<td>3.32 (3.07–3.58)</td>
<td>2.70 (2.48–2.94)</td>
<td>5.50 (4.69–6.54)</td>
<td>5.06 (3.61–7.29)</td>
</tr>
</tbody>
</table>

Mean estimates (95% confidence interval) are shown unstratified and stratified by set-point viral load and seroconversion age. SC, seroconversion.
proportion of individuals entering each of the CD4⁺ categories upon entering chronic infection \( \left( f_{jk} \right) \) for the whole population, and then stratified by SPVL and by age at seroconversion. We use these results to derive the estimated average time to reach a CD4⁺ cell count of 200 cells/µl, a threshold which is used in many settings to define AIDS, from different time points in the course of infection.

These findings were relatively insensitive to the choice of smoothing method used (see Supplementary Material, http://links.lww.com/QAD/A766). They differed from estimates obtained when actual CD4⁺ measurements rather than smoothed CD4⁺ cell counts were used (i.e. we considered the earliest time when CD4⁺ measurements fell below each CD4⁺ threshold) which, while giving a comparable time to death, consistently underestimated both the time spent with CD4⁺ cell count above 350 cells/µl and the proportion of individuals with CD4⁺ above 350 cells/µl after seroconversion, while overestimating the time spent with CD4⁺ 350 cells/µl or less.

Patients were only included in the analysis if they had a minimum number (six) of CD4⁺ cell counts so that smoothing could be reliably performed; the choice of minimum number of individual CD4⁺ measurements used to estimate the smoothing model was also seen not to have an effect on the results (Supplementary Material, http://links.lww.com/QAD/A766). Similarly, results were similar when considering only CD4⁺ measurements prior to mono or dual ART initiation, or when considering only patients recruited after 1996 (Supplementary Material, http://links.lww.com/QAD/A766).

SPVL was found to be strongly associated with CD4⁺ progression, with individuals with higher SPVL progressing much faster through all CD4⁺ stages than those with lower SPVL (Table 2). Overall, individuals with \( \log_{10} \) SPVL of at least 5.0 progress from seroconversion to CD4⁺ cell count of 200 cells/µl approximately twice as fast as those with \( \log_{10} \) SPVL below 4.0 (see Table 2).

In the Cox proportional hazards model analysis, we analysed the data under the assumption that the rate (hazard) of progressing from one CD4⁺ category to the next was a linear model of SPVL, an assumption we assessed through a Kaplan–Meier analysis stratified into narrow bands of SPVL (Supplementary Material, http://links.lww.com/QAD/A766). Using this model, the relative hazard of progressing to the next CD4⁺ category was 2.17 (95% CI 1.89–2.49), 1.88 (95% CI 1.61–2.19), 1.96 (95% CI 1.41–2.73) and 1.63 (95% CI 0.77–3.44) for each \( \log_{10} \) increase in SPVL.

Substantial changes have been seen in population-level SPVL over time (Figure S2, http://links.lww.com/QAD/A766), and similarly large differences can be observed in the unstratified CD4⁺ progression rates between different time periods. However, when stratified by SPVL, there was little difference remaining in the rate of CD4⁺ progression between 1995 and 2000 compared with before or after (Supplementary Material, http://links.lww.com/QAD/A766).

The rate of CD4⁺ progression was not found to differ substantially by age at seroconversion, although individuals seroconverting at older ages were found to have a slightly faster progression (Table 2). Comparing MSM/bisexual men to other HIV transmission routes (79% via heterosexual sex, but also including people who inject drugs and one haemophiliac), there was little difference in the rate of CD4⁺ progression, although results indicate that MSM/bisexual men may progress more quickly to death once they reach a CD4⁺ cell count of 200 cells/µl (Supplementary Material, http://links.lww.com/QAD/A766). No significant difference was found in CD4⁺ progression between men and women, but there were very few women in the sample and hence only unstratified analysis was possible (Supplementary Material, http://links.lww.com/QAD/A766). However, median SPVL was lower in women (4.00 \( \log_{10} \) copies/ml) compared with men (4.57 \( \log_{10} \) copies/ml).

Figure 2 shows, for each CD4⁺ category, the survival functions until transition to the next CD4⁺ category, for all patients as well as stratified by SPVL.

The nonparametric Kaplan–Meier estimated survival curves are shown alongside the fitted exponential survival curve (i.e. \( e^{-\lambda t}, k=1,\ldots,4 \)). This suggests that the exponential survival function provides a very good fit to the nonparametric curve, indicating that the hazard of progression between categories is approximately constant. From this observation, we conclude that the compartmental model shown in Fig. 1a, with constant rates of progression between successive CD4⁺ stages, is well supported by the data. In particular, it is well matched to the mean rates of progression, but also provides a good description of interperson variability in progression times.

Figure 3 presents the results of our predictive model, which combines the true CD4⁺ progression model and the misclassification model shown in Table 1. The figure shows the predicted future distribution of true CD4⁺ categories given current observed CD4⁺ category, for all patients and stratified by SPVL. It highlights the large influence of SPVL on patients’ future clinical progression.

This predictive model was able to broadly reproduce the true CD4⁺ dynamics of the patients selected for validation, given their first observed CD4⁺ category (see Figure S6 and Supplementary Material for detail, http://links.lww.com/QAD/A766).
Discussion

In this study, we used data from the ATHENA cohort to provide a comprehensive quantitative analysis of CD4+ cell progression in HIV-infected individuals and the factors affecting the rate of decline. These estimates allow for more accurate long-term projections in mathematical and econometric models. Furthermore, we have found that the structural assumptions in a simple compartmental model of CD4+ progression (Fig. 1a) provides a very good statistical description of interperson variability in progression rates (Fig. 2) increasing the confidence in such parsimonious modelling approaches.

We showed that a constrained monotonic spline regression model is suitable for smoothing the individual decline of CD4+ cell counts between the four commonly used CD4+ categories: CD4+ > 500, 500 < CD4+ ≤ 500, 200 < CD4+ ≤ 350 and CD4+ ≤ 200 cells/μl. We provided estimates of the proportion of individuals who are in each of those categories following seroconversion, as well as estimates of the rates at which individuals progress through the decreasing CD4+ categories. We found that nearly a quarter of HIV-infected individuals have a CD4+ cell count below 500 cells/μl immediately following seroconversion. Many mathematical models assume that all individuals start with a CD4+ cell count above...

Fig. 2. Survival figures, overall and stratified by set-point viral load, showing decline to CD4+ 500 cells/μl, from 500 to 350, from 350 to 200 and from 200 to death, respectively (darker shades indicate higher SPVL). Crosses and shaded areas indicate nonparametric Kaplan–Meier survival estimates with 95% confidence intervals. Dotted lines show the fitted exponential survival functions.
500 cells/\mu l, and thus they may potentially underrepresent the heterogeneity in CD4\(^+\) progression between individuals, in particular the time to AIDS or death.

In an analysis of data from the CASCADE collaboration, Lodi \textit{et al.} [5] estimated the median time from seroconversion to a CD4\(^+\) cell count below 500, below 350 and below 200 cells/\mu l as 1.19, 4.19 and 7.93 years, respectively. Gras \textit{et al.} [19] found a similar estimate of over 4 years from seroconversion to a CD4\(^+\) cell count 350 cells/\mu l for individuals seroconverting before 1996, but a shorter estimate of 2.7 years from seroconversion to CD4\(^+\) cell count 350 cells/\mu l for those seroconverting after 2002. A number of other studies have reported the median time from estimated date of seroconversion to a CD4\(^+\) cell count cell count below 200 cells/\mu l. Kilmarx \textit{et al.} [20] reported 6.9 years among Thai female sex workers, whereas Wandel \textit{et al.} [21] used cohorts from Uganda, Thailand and Côte d’Ivoire and estimated a median of 6.1 years to CD4\(^+\) cell count below 200 cells/\mu l based on a Weibull survival model, with a shorter duration among the Thai cohorts than the Africans. The difference between our estimates and those of these other studies may be partly explained by their use of the median time. Motivated by the fact that the median time of an exponential process is lower than the mean time, we conducted numerical experiments, and we found a median time of 9.5 years given our model of disease progression.

Cori \textit{et al.} [6] used the data from [5] to parameterize a mathematical model of CD4\(^+\) progression similar to that
presented in this study, finding the total mean time from seroconversion to a CD4⁺ cell count of 200 cells/μL was 9.98 years, similar to our estimate of 10.54 years in this analysis. They estimated the proportion of individuals in each CD4⁺ category after infection [6] to be \( f_1 = 0.58, f_2 = 0.23, f_3 = 0.16 \) and \( f_4 = 0.03 \). Our analysis led to a slightly higher proportion starting at CD4⁺ above 500 cells/μL (\( f_1 = 0.76 \)), and a lower proportion portion starting at 200 < CD4⁺ ≤ 350 cells/μL (\( f_3 = 0.05 \)) but similar values for the two remaining proportions (Table 2). The average times to progression to the next CD4⁺ category estimated in [6] were \( 1/q_1 = 6.37, 1/q_2 = 2.86, 1/q_3 = 3.54 \) and \( 1/q_4 = 2.30 \) years, respectively. These estimates are quite different from ours \( (1/q_1 = 3.32, 1/q_2 = 2.70, 1/q_3 = 5.50 \) and \( 1/q_4 = 3.54 \) years, see Table 2), although in the absence of CI in [6], it is difficult to assess how significant these differences are. However, Cori et al. [6] based their analysis on a 3-year seroconversion window, much longer than the 1-year window used in this study, and this may partly account for the lower fraction of individuals starting at high CD4⁺ cell counts found previously.

Importantly, although we have found SPVL to be highly influential in terms of CD4⁺ decline, none of these previous studies have reported the population distribution of SPVL, nor have they stratified their analyses by SPVL. Two studies looking at the average rate of CD4⁺ decline in South Africa [22] and in Europe [10] similarly found that SPVL was the most influential factor with sex and age playing smaller roles.

It is well known that SPVL and CD4⁺ progression are tightly associated [10,11,23]. Here, we provide, to our knowledge, the first analysis which quantifies the association by CD4⁺ stage. Our results have multiple direct implications. Clinically, this highlights the value of monitoring viral load. Current WHO guidelines for patient monitoring in pre-ART care do not take SPVL into account. However, patients with higher SPVL prior to initiating treatment will progress faster, and so may need to be monitored more regularly than those with low SPVL (see Fig. 3). For mathematical modelling, it may be important to incorporate SPVL stratification, especially as mean population SPVL has changed over time [18,24] and thus the average rate of progression between CD4⁺ stages in the population may have varied over time, consistent with findings from Gras et al. [19].

We found some suggestion that individuals seroconverting at older ages may progress more quickly to lower CD4⁺ categories, in agreement with the previous studies [7–10], but the effect was small, consistent with [19]. The ATHENA cohort contains mostly MSM, as this is the group with the highest burden of HIV in the Netherlands, and thus we had limited ability to distinguish differences by sex and transmission route. Most patients with recorded subtype were of subtype B, and we were unable to the examine differences between viral subtypes which have been seen in the previous studies in CD4⁺ dynamics [25–28] and SPVL [28]. Ethnicity was not routinely recorded in the ATHENA cohort, and thus we could not test differences between ethnic groups [8,9,29,30], and in particular assess the finding by Muller et al. that the rate of CD4⁺ decline depended strongly on SPVL in patients of European descent, but not in patients of African descent [30]. Extending the study performed here to a wider range of similarly well characterized cohorts will enable a more robust assessment of the extent of variation between populations.

A strength of our study lies in that we have challenged our model through extensive sensitivity analyses. Although all sensitivity analyses gave broadly consistent results, our unstratified estimates were slightly sensitive to the minimal requested number of CD4⁺ cell counts, with faster progression inferred when including individuals with few CD4⁺ measurements (see Table S4, http://links.lww.com/QAD/A766). However, the estimates stratified by SPVL were insensitive to the number of CD4⁺ measurements. These results suggest that individuals with very few CD4⁺ cell counts may have higher SPVL, leading them to progress faster, hence their small number of CD4⁺ measurements. Therefore, although our results stratified by SPVL appear very robust, our unstratified population estimates might be slightly biased towards slow progressors because of our inclusion criteria. Results were very similar whether we considered patients before initiation of HAART, or before initiation of either mono or dual therapy. Similarly, results obtained using the entire dataset and using only individuals diagnosed after 1996 were similar. It is worth noting that our estimates of the average time from a CD4⁺ cell count of 200 cells/μL to death are based on very few observations, in particular when stratified by SPVL, as reflected by the wide CIs. Indeed diagnosed HIV-positive individuals with low CD4⁺ cell counts are almost always initiated on ART before dying, precluding precise assessment of late-stage progression. Future models using our estimates should account for this uncertainty. In particular, modellers could consider using alternative approaches to modelling the influence of SPVL on late-stage progression, for example, by using the results of our Cox proportional hazard analysis which suggest that, for individuals with CD4⁺ less than 200 cells/μL, the death hazard increases by 63% for each \( \log_{10} \) increase in SPVL.

Another novel aspect of our approach is the separation of long-term underlying trends (represented by smoothed CD4⁺ cell counts) and random variation (represented by differences between smoothed and observed counts). We hypothesize that, as a result, our findings are more robust to setting and time-specific variation in frequency of CD4⁺ testing and can be more readily extrapolated to different settings. Our validation analysis confirmed the validity of our model and our CD4⁺ progression
estimates, and indicates that these could be used to predict, at the population level, future true CD4\(^+\) trajectories given current observed CD4\(^+\) cell count distributions.

Our results suggest that SPVL could also be considered in clinical guidelines (see Table 2 and Fig. 3). Current guidelines for pre-ART monitoring recommend a CD4\(^+\) test every 6–12 months, irrespective of SPVL. We estimated that 28% (respectively 38%) of individuals with observed CD4\(^+\) cell count above 500 cells/\(\mu\)l would progress to a true CD4\(^+\) cell count below 500 cells/\(\mu\)l after 6 months (respectively 12 months). However, 43% (respectively 59%) of those with SPVL of at least 5.0 \(\log_{10}\) copies/\(\mu\)l would have already progressed by then. Our results suggest that if the patient’s SPVL is known, the frequency of CD4\(^+\) monitoring could be adapted to account for that, with a CD4\(^+\) cell count test every 13–23, 6–12, 3–8 and 2–4 months for individuals with SPVL below 4.0, 4.0–4.5, 4.5–5.0 and at least 5.0 \(\log_{10}\) copies/\(\mu\)l, respectively (see Supplementary Material for detail, http://links.lww.com/QAD/A766). These proposed delays are only indicative and should be tailored to each patient in particular given how close their observed CD4\(^+\) cell count is to the threshold 500 cells/\(\mu\)l. One should also keep in mind that 16% of individuals with observed CD4\(^+\) cell count above 500 cells/\(\mu\)l may in fact already have their true CD4\(^+\) cell count below 500 cells/\(\mu\)l, and therefore are only considered ineligible for ART initiation because of stochastic fluctuations in CD4\(^+\) cell counts. Further studies should examine, in light of these results, the costs and benefits of pre-ART monitoring strategies including viral load testing in different settings, to assess the potential cost-effectiveness of such approaches at a population level.

Our analysis has highlighted the fundamental and well known role of viral load in affecting CD4\(^+\) progression dynamics. It has also confirmed that viral load is very stable during the set-point window, suggesting that viral load itself is a good predictor of CD4\(^+\) progression. We have provided estimates of dependence of CD4\(^+\) progression rates on SPVL, both in a subgroup analysis and in a continuous Cox proportional hazards analysis. We suggest that the effect sizes are sufficiently large that models and analyses should be stratified by SPVL wherever possible, if dynamical considerations are of importance. Clinically, these results are relevant for patient monitoring, which should involve both CD4\(^+\) cell count and viral load testing wherever possible.

**Acknowledgements**

A.C., M.P. and C.F. thank the NIH for funding through the NIAID cooperative agreement UM1AI068619. C.F. thanks the European Research Council (PBDR-339251). The authors thank Chris Wymant, Tara Mangal, Jeff Eaton and Tim Hallett for useful discussions.

The ATHENA database is supported by a grant from the Dutch Ministry of Health, Welfare and Sport and was set up and is maintained by Stichting HIV Monitoring. The ATHENA observational cohort has been made possible by the participating patients and through the collaborative efforts of the following physicians working at Netherlands HIV Treatment Centers:

**Clinical centres**

‘*’ denotes site coordinating physician.

References


