

Fluctuations in HIV-1 Viral Load Are Correlated to CD4⁺ T-Lymphocyte Count During the Natural Course of Infection

*Joanna Masel, *Ramy A. Arnaout, †Thomas R. O'Brien, †James J. Goedert, and *Alun L. Lloyd

*Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, Oxford, U.K.; †Viral Epidemiology Branch, National Cancer Institute, Public Health Service, U.S.A., and ‡Department of Health and Human Services, Rockville, Maryland, U.S.A.

Summary: Viral load fluctuates during the natural course of asymptomatic HIV-1 infection. It is often assumed that these fluctuations are random around a set point or underlying growth trend. Using longitudinal data, we tested whether fluctuations in viral load can be better explained by changes in CD4⁺ T-cell count than by a set point or trend of exponential growth. The correspondence between viral load and CD4⁺ T-cell count could be described by a simple mathematical relation. Using a bootstrapping approach, the hypothesis that viral load fluctuations are random around a set point was rejected with $p < .00005$. The hypothesis that viral load fluctuations are random around a trend of exponential growth was rejected with $p < .005$. Viral load data was explained better by changes in CD4⁺ T-cell counts than by a set point or by a trend of exponential growth. The implications of this finding for improved prognostication are discussed. **Key Words:** CD4⁺ cell count—Prognosis—Predictive value of tests—Longitudinal HIV—Viral load.

Polymerase chain reaction (PCR) has proved to be a powerful technique for the quantitation of plasma viral load in HIV-1 infection (1,2). It was instrumental in showing that infection is active—and not latent, as was long thought—during the asymptomatic period, and today this is the principal method for quantitating plasma viremia. Measurements are commonly used for diagnostic assessments of disease progression and responsiveness to antiviral therapy. As a result, much is known about both the sensitivity and limitations of PCR assays. The standard deviation (SD) due to measurement error is $0.18 \log_{10}$ units for the Amplicor HIV Monitor assay (Roche Molecular Systems, Branchburg, NJ, U.S.A.) (3). Fluctuations of variance $0.4 \log_{10}$ units (standard deviation

[SD], 0.63) are observed over the course of 24 hours (4), and of variance $0.5 \log_{10}$ units (SD 0.7) over the course of 8 weeks (5). It is plausible from these figures that observed fluctuations in HIV-1 RNA reflect more than just measurement error.

To date, most longitudinal studies of viral load have focused on long-term trends, with the goal of explaining pathogenesis or defining prognostic indicators (6–9); medium-term fluctuations (i.e., those on the order of a few years) have received considerably less attention. The prevailing view is that these fluctuations center around a set point and contain no information relevant to prognosis (6,10,11). Other groups have reported that a set point is not always a sufficient description, and that fluctuations center around an underlying trend of increasing viremia (12–14). In either case, if fluctuations in viral load contain no information, we would expect no relation between the fluctuations in viral load and the fluctuations in CD4⁺ cell counts. Here, we test this assumption on longitudinal data using simple mathematical models. We find that the CD4⁺ counts are better explained by fluctuations in viral load than by random error around a set

Address correspondence and reprint requests to Alun Lloyd, Program in Theoretical Biology, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, U.S.A.; email: alun@ias.edu.

R. A. Arnaout is currently affiliated with Harvard Medical School, Boston, Massachusetts, U.S.A.

A. L. Lloyd is currently affiliated with Institute for Advanced Study, Princeton, New Jersey, U.S.A.

Manuscript received October 8, 1999; accepted February 16, 2000.

point or exponential growth of viral load. Fluctuations in viral load are not independent of the CD4⁺ cell count.

METHODS

Patient Data

Patient data were drawn from a 15-year-long U.S. National Cancer Institute study of 134 HIV-1-seropositive men as previously reported (6,12). Serum HIV-1 RNA level, CD4⁺ T-cell count, and clinical and treatment status were recorded on average once yearly for each patient as described, except for 1983 when the Washington study subjects were not observed because of limited resources (12). Serum HIV-1 RNA was measured using the Amplicor HIV Monitor assay (Roche Molecular Systems). Because we were concerned with the natural course of infection, measurements following the start of antiviral therapy were excluded. Moreover, because we fitted parameters to each patient individually, we included only those patients with five or more measurements for each of viral load and CD4⁺ T-cell count. This reduced the number of patients to 29, with a bias against rapid progressors. Viral load measurement and CD4⁺ T-cell count data sets for a given patient were designated $\{v_1, v_2, v_3, \dots, v_m\}$ and $\{x_1, x_2, x_3, \dots, x_n\}$, respectively.

Statistical Analysis

The first null hypothesis is that viral load reaches a set point, and that fluctuations around this set point are random and independent of the CD4⁺ cell count. We fit a function $f(v)$ in which viral load was used to predict CD4⁺ cell count. Many functions f were tried (see Results). Where no viral load measurement coincided with a CD4⁺ T-cell measurement, viral load was linearly interpolated over time. This interpolation allowed the generation of a set of viral load estimates $\{v_1', v_2', v_3', \dots, v_n'\}$ coincident with CD4⁺ T-cell measurements. We optimized arbitrary parameters in f by nonlinear least squares (Marquand) fitting and calculated the goodness of fit using the least squares criterion

$$g = \sum_{i=1}^n \frac{(x_i - f(v_i'))^2}{\sigma_i^2} \quad (1)$$

where σ_i^2 is the variance for the measurement of viral load for each time point. We assume that σ_i^2 is constant for all time points.

Under the first null hypothesis, observed viral load measurements follow some distribution around a set point. A bootstrapping approach avoids making assumptions about the nature of the distribution because it reuses the measurements themselves. If fluctuations in viral load are random and independent of the CD4⁺ T-cell count, then changing the order of the measurements should not make any difference to the goodness of fit. We simulated the viral load measurements by resampling with replacements from the observed data set. If the null hypothesis is true, then these simulated data sets will fit the function f just as well as the original data set. We produced 10,000 simulations and calculated what proportion of these gave a better fit than the original data set. This gives the probability p that as a good a fit as the real data would be generated by chance alone, given the null hypothesis.

The second null hypothesis is that viral load grows exponentially, and that fluctuations around this underlying trend are random and independent of the CD4⁺ T-cell count. The same procedure was used as for the first null hypothesis, except that simulated data sets were generated differently. Viral load measurements were transformed onto a

log scale, and a straight line $\log v = at + b$ was fitted. The residual of each measurement was calculated as $r_i = \log v_i - (at_i + b)$. We assume that the distribution of residuals remains constant on a log scale, that is, that if the error on a measurement of 10^3 is $\pm 0.5 \log_{10}$, then the error on a measurement of 10^8 will also be $\pm 0.5 \log_{10}$. This assumption is supported by studies on variation in viral load (3-5). We then simulated data sets by resampling the residuals r_j with replacement to get simulated measurements $v_j' = 10^{at_j + b + r_j}$.

Following this method, a separate p value was generated for each patient. Those p values generated for a particular null hypothesis can be combined by calculating the quantity

$$-2 \sum_{i=1}^{29} \ln p_i.$$

Given the null hypothesis, this quantity should be distributed as χ^2 with 58 degrees of freedom (15). This allowed us to combine statistical information from our 29 patients.

RESULTS

Many functions f were tried, starting with linear functions, progressing to those increasingly nonlinear. Functions using the logarithm of viral load were tried, as well as functions using untransformed viral load measurements. The best function for two arbitrary parameters was found to be

$$f(v) = 1/(\gamma v + \alpha) \quad [2]$$

Here, v denotes viral load, whereas $f(v)$ denotes CD4⁺ cell count; γ and α are arbitrary constants specific to a particular patient. This function gave a good estimate of the CD4⁺ T-cell count from viral load measurements for many, although not all, of these 29 patients. For 7 of 29 patients, the optimum value of γ was 0, that is, the CD4⁺ T-cell count was predicted to be constant. For these patients, $p = 1.0$, providing no evidence against the set point hypothesis. Other patients showed a good fit; an example with $p = .02$ for the first null hypothesis and $p = .08$ for the second is shown in Figure 1. Values for α and γ for all patients are shown in Figure 2.

Results for individual patients remain inconclusive, but when the results from all 29 patients were combined with a χ^2 test, the null hypothesis of constant viral load gave $p < .00005$, whereas the null hypothesis of exponential growth in viral load gave $p < .005$. In either case, the null hypothesis is rejected, and the data show that viral load and CD4⁺ cell counts are negatively correlated. In neither case did the low p value of a single patient bias the results.

We also examined how well Equation 2 fitted data points taken after treatment and observed that the fit was maintained in almost all cases, and improved in most.

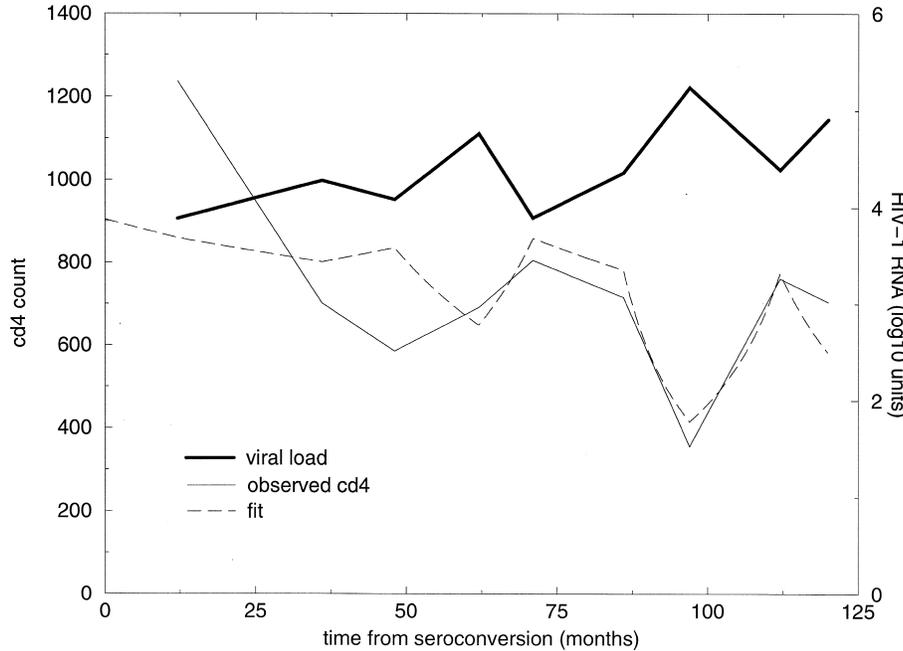


FIG. 1. Longitudinal data for a single patient. Viral load is plotted on a log scale. CD4+ count is predicted from viral load according to Equation 2 and provides a good estimate for the observed CD4+ count. For this patient, the null hypothesis of constant viral load gives $p = .02$ and the null hypothesis of exponential growth in viral load gives $p = .08$. The result for this single patient is inconclusive, but when the results from all 29 patients are combined, the first null hypothesis gives $p < .00005$, whereas the second null hypothesis gives $p < .005$.

The improvement was probably due to the larger number of data points and the magnitude of the drop in viral load on treatment; the greater the biologic fluctuation, the more easily it can be seen above random noise. A correlation between the extent of viral suppression and the extent of CD4+ T-cell count recovery following treatment has been noted elsewhere (16).

DISCUSSION

Using a statistical approach, we have shown that viral load measurements over a period of years are better explained by confluctuation with the CD4+ T-cell count than by a set point or trend of exponential growth. This means that despite the considerable measurement error involved

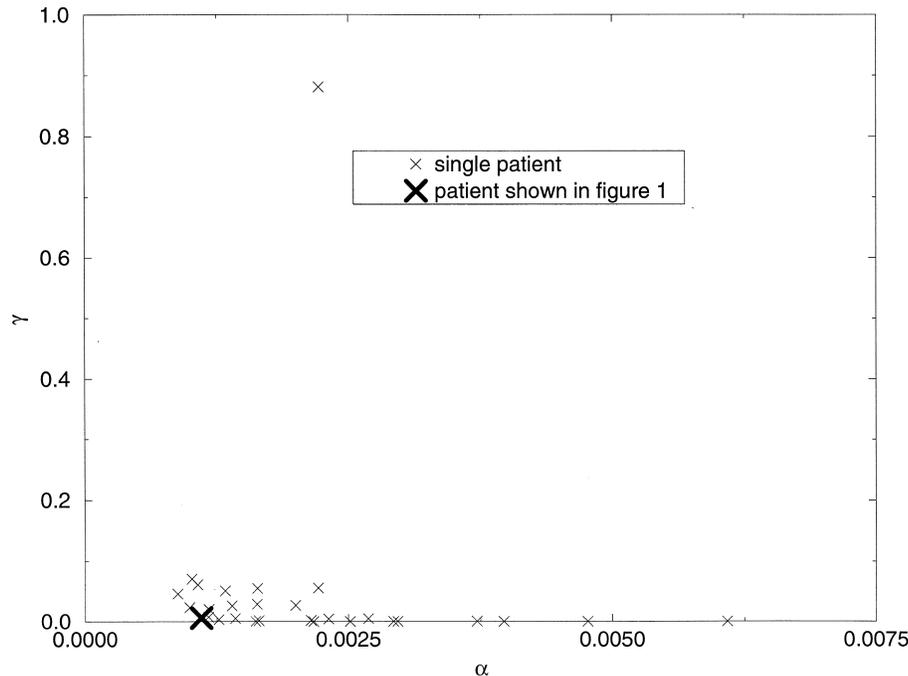


FIG. 2. Best fit values for the parameters α and γ in Equation 2 for all 29 patients. For 7 patients, $\gamma = 0$, meaning that there is no correlation between fluctuations in CD4+ count and viral load for those patients. The single outlying point with a high γ value corresponds to a patient whose longitudinal data fit poorly with Equation 2.

in quantifying viral load, some observed fluctuations may be biologically significant. The interaction between CD4⁺ count and viral load seems to be highly dynamic during the natural course of infection and does not reach a single set point.

Although both null hypotheses were rejected for the patients as a group, they were not rejected for every patient. This could be entirely due the small number of data points for each patient, making it easy for random error to obscure the trend. Alternatively, there may be a distinct subset of patients whose condition is more stable. For this subset, measurement errors may be significantly greater than biologic fluctuations.

The cofluctuation of viral load and CD4⁺ cell count may be superimposed on other fluctuations. For example, CD4⁺ cell counts have been shown to fluctuate diurnally whereas viral RNA levels do not (4).

A weak inverse correlation between viral load and CD4⁺ T-cell count is well-known (8,17). In principle, the relation between viral load and CD4⁺ T-cell count can be described by various mathematical functions. The one that gave the best fit to the data was Equation 2, which was also one of the simplest. This equation has an interesting interpretation. In the standard model of viral dynamics (18,19), uninfected cells and viral load are related by the following differential equation:

$$dx/dt = \lambda - dx - \beta xv \quad [3]$$

Variables x and v denote uninfected (CD4⁺ T) cells and viral load, respectively; the constants λ , d , and β are rates of uninfected cell production, death, and infection, respectively. The short-term equilibrium of uninfected cells is given by $x^* = \lambda/(\beta v^* + d)$, where v^* is the short-term equilibrium viral load. We can divide the numerator and denominator by λ to obtain a relation of the form $x^* = 1/(\gamma v + \alpha)$ where $\gamma = \beta/\lambda$ and $\alpha = d/\lambda$. Because this expression is identical to our best-fit function, our results may support the applicability of Equation 3. Alternatively, the relation between Equations 2 and 3 may simply be coincidental. Equation 3 describes change in the number of uninfected, activated cells, whereas Equation 2 models the total CD4⁺ T-cell count. For Equation 2 to support Equation 3, the number of uninfected, activated cells divided by the total CD4⁺ T-cell count must be fairly constant. It is not clear whether this is the case.

In the standard model, viral load also follows a given differential equation. Here, we treat v as an independent variable the behavior of which is not predicted, and for which no steady state is assumed. It seems that v does indeed fluctuate significantly, and that these fluctuations are mirrored in the CD4⁺ cell count. This correlation

could reflect a number of processes, such as increased death of CD4⁺ T cells or increased trapping in the lymphoid compartment in response to a rise in viral load. The relation between CD4⁺ count and viral load underscores the importance of viral dynamics in the persistence of HIV-1 infection, although the etiology and biologic significance of the fluctuations remain to be determined.

Our finding may be used to provide improved prognostication. Previous reports have shown that although both viral load and CD4⁺ T-cell count have predictive value for development of disease, viral load was the more powerful (7–9,20), although CD4⁺ counts gain value later in the short-term prediction of AIDS or death (21). Predictive powers of the two indicators have been shown to be at least partially independent of the other, that is, after prognosis has been predicted from one indicator, additional information can be gained from the other (6–8,17). A regression tree using both indicators to group patients into four or five categories gives a more accurate prognosis than either indicator used on its own (9,20). For the best prognostication, we must maximize the information from both viral load and CD4⁺ T-cell count. In a regression tree, arbitrary cutoffs divide patients into groups, and information is lost in the process. A weighted mean uses all available information and may potentially provide better prognostication than a regression tree.

Viral load and CD4⁺ T-cell count cannot be combined in a weighted mean unless they can be transformed onto comparable scales. The quantitative relation we have found between viral load and CD4⁺ T-cell count (Equation 2) provides a possible way of doing so. A continuous function combining viral load and CD4⁺ T-cell count may form the basis of a new prognostic measure for patients using longitudinal data.

Acknowledgments: We thank M. Nowak and H. Korthals Altes for helpful discussion. J. Masel was supported by the Rhodes Trust; R. A. Arnaout was supported by the Marshall Aid Commemoration Commission (U.K.); and A. L. Lloyd was supported by the Medical Research Council (U.K.).

REFERENCES

1. Piatak M Jr, Saag MS, Yang LC, et al. Determination of plasma viral load in HIV-1 infection by quantitative competitive polymerase chain reaction. *AIDS* 1993;7(Suppl 2):S65–71.
2. Piatak M Jr, Saag MS, Yang LC, et al. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 1993;259:1749–54.
3. Schuurman R, Descamps D, Weverling GJ, et al. Multicenter comparison of three commercial methods for quantification of human immunodeficiency virus type 1 RNA in plasma. *J Clin Microbiol* 1996;34:3016–22.
4. Deeks SG, Coleman RL, White R, et al. Variance of plasma human

- immunodeficiency virus type 1 RNA levels measured by branched DNA within and between days. *J Infect Dis* 1997;176:514–17.
5. Raboud JM, Montaner JS, Conway B, et al. Variation in plasma RNA levels, CD4 cell counts, and p24 antigen levels in clinically stable men with human immunodeficiency virus infection. *J Infect Dis* 1996;174:191–4.
 6. Henrard DR, Phillips JF, Muenz LR, et al. Natural history of HIV-1 cell-free viremia. *JAMA* 1995;274:554–8.
 7. O'Brien TR, Blattner WA, Waters D, et al. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. *JAMA* 1996;276:105–10.
 8. Mellors JW, Rinaldo CR Jr, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996;272:1167–70.
 9. Mellors JW, Munoz A, Giorgi JV, et al. Plasma viral load and CD4⁺ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* 1997;126:946–54.
 10. Garcia F, Vidal C, Gatell JM, Miro JM, Soriano A, Pumarola T. Viral load in asymptomatic patients with CD4⁺ lymphocyte counts above $500 \times 10^6/l$. *AIDS* 1997;11:53–7.
 11. Saag MS. Use of virological markers in clinical practice. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997;16:S3–13.
 12. O'Brien TR, Rosenberg PS, Yellin F, Goedert JJ. Longitudinal HIV-1 RNA levels in a cohort of homosexual men. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998;18:155–61.
 13. Vidal C, Garcia F, Romeu J, et al. Lack of evidence of a stable viral load set-point in early stage asymptomatic patients with chronic HIV-1 infection. *AIDS* 1998;12:1285–9.
 14. Rinaldo CR Jr, Gupta P, Huang SL, et al. Anti-HIV type 1 memory cytotoxic T lymphocyte responses associated with changes in CD4⁺ T cell numbers in progression of HIV type 1 infection. *AIDS Res Hum Retroviruses* 1998;14:1423–33.
 15. Sokal RR, Rohlf FJ. *Biometry*, 3rd ed. New York: W.H. Freeman and Company, 1995.
 16. Staszewski S, Miller V, Sabin C, et al. Determinants of sustainable CD4 lymphocyte count increases in response to antiretroviral therapy. *AIDS* 1999;13:951–6.
 17. Hughes MD, Johnson VA, Hirsch MS, et al. Monitoring plasma HIV-1 RNA levels in addition to CD4⁺ lymphocyte count improves assessment of antiretroviral therapeutic response. ACTG 241 Protocol Virology Substudy Team. *Ann Intern Med* 1997;126:929–38.
 18. De Boer RJ, Perelson AS. Target cell limited and immune control models of HIV infection: a comparison. *J Theor Biol* 1998;190:201–14.
 19. Nowak MA, Bangham CR. Population dynamics of immune responses to persistent viruses. *Science* 1996;272:74–9.
 20. O'Brien WA, Hartigan PM, Martin D, et al. Changes in plasma HIV-1 RNA and CD4⁺ lymphocyte counts and the risk of progression to AIDS. Veterans Affairs Cooperative Study Group on AIDS. *N Engl J Med* 1996;334:426–31.
 21. de Wolf F, Spijkerman I, Schellekens PT, et al. AIDS prognosis based on HIV-1 RNA, CD4⁺ T-cell count and function: markers with reciprocal predictive value over time after seroconversion. *AIDS* 1997;11:1799–1806.