STATISTICS IN MEDICINE

Statist. Med. 2006; **00**:1–6 *Prepared using* simauth.cls *[Version: 2002/09/18 v1.11]*

A Comparison of Methods for Determining HIV Viral Set Point

Y. Mei[1*†], L. Wang[2], S. E. Holte[2]

[1] *School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30329, U.S.A.*

[2] *Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, U.S.A.*

SUMMARY

During a course of HIV-1 infection, the viral load usually increases sharply to a peak following infection then drops rapidly to a steady state where it remains until progression to AIDS. This steady state is often referred to as the viral set point. It is believed that the HIV viral set point results from an equilibrium between the HIV virus and immune response and is an important indicator of AIDS disease progression. In this paper, we analyze a real data set of viral loads measured before antiretroviral therapy is initiated, and propose two-phase regression models to utilize all available data to estimate the viral set point. The advantage of the proposed methods are illustrated by comparing them with two empirical methods, and the reason behind the improvement is also studied. Our results illustrate that for our data set, the viral load data is highly correlated and it is cost-effective to estimate the viral set point based on one or two measurements obtained between five and twelve months after HIV infection. The utility and limitations of this recommendation will be discussed. Copyright © 2006

*Correspondence to: Yajun Mei, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30329, U.S.A.

†E-mail: ymei@isye.gatech.edu

Contract/grant sponsor: NIAID/NIH; contract/grant number: R01 AI055343

Received 27 July 2006

Revised 25 June 2007

John Wiley & Sons, Ltd.

KEY WORDS: cost effective, experimental design, longitudinal data, random change-point model, set point, repeated observations

1. INTRODUCTION

In a representative course of human immunodeficiency virus 1 (HIV-1) infection the number of HIV-RNA copies per milliliter (mL) of plasma, referred to as viral load, increases sharply to a peak during the first few weeks post infection then drops rapidly to a steady state where it remains relatively stable until progression to AIDS (e.g., [1, 2, 3, 4]). This steady state, called the "viral set point," results from an equilibrium between HIV virus and immune responses, see [5]. It is well-known that the viral set point is an important predictor of outcome in HIV patients in terms of progression to AIDS and survival: the higher the viral set point, the faster the progression to AIDS and death (e.g., [6, 7]). In a prospective cohort study, individuals with viral load in the highest quartile progressed to disease three times more quickly than individuals whose viral load was within the lowest quartile, see [6, 8].

In the past, empirical methods have been proposed in the literature to estimate the viral set point, see, for example, [9, 10, 11]. There are two widely used empirical methods to define or estimate the viral set point. The first method is based on the first measurement taken after a pre-specified time, and the second method uses all measurements taken after the pre-specified time by applying the linear mixed-effects model of [12]. Here the pre-specified time is chosen subjectively by the researchers, and one popular choice is four months since infection,

see [3, 4, 10, 13]. Both empirical methods assume that all observations taken after the pre-specified time are in the steady state (i.e., at the viral set point). However, the empirical methods are used more out of convenience rather than on the basis of scientific arguments.

The primary goal of this paper is to compare these two empirical methods with two different methods, based on the two-phase regression model developed in [14] and [15], which utilize all available data collected during primary and early infection. The two-phase regression model and its extension to the population level allow us to use the data themselves to estimate both the value of the viral set point and the time at which the viral set point is reached. We will apply these four different methods to a data set collected at the University of Washington Primary Infection Clinic (PIC) as well as simulated data.

From the methodology point of view, the population version of the two-phase regression model we use is a special case of the so-called "random change-point models," which have been used recently in many other applications, see, for example, [16] on the onset of a specific disease, and [17] on the CD4 count of HIV patients with treatments. However, our application to the viral load data for untreated patients is new. In the literature, the statistical analysis of viral load data, especially that for untreated patients, is rather limited, partly because of the lack of availability of large data sets. For a recent review on statistical methods and the challenges for viral load data, we refer readers to [18] and the references therein.

A closely related goal of this paper is to consider study design issues when the viral set point is an outcome of interest. For example, the viral set point is used as a co-primary endpoint in a landmark clinical trial, see [19]. It is critical to design a study which will estimate viral set points accurately and effectively while controlling the total cost of the study. Because the importance of sample size (the number of patients) and its calculation have been well-

established in the longitudinal data analysis literature, see, for example, [20, 21, 22], we will focus on the optimal choice of the number of repeated measures per patient. In particular, we will study this from two different viewpoints: one from the cost-effective approach proposed in [23], and the other from the relative precision approach, which has been used in other contexts, see page 24 of [24] and [25]. Our results will provide guidelines for when and how often viral load measurements should be collected in order to estimate the viral set point, subject to a cost or precision constraint.

It is worth pointing out that the problem of the optimal choice for the number of repeated measures per patient has been neglected in the repeated measurements or longitudinal experiments literature. For instance, on page 28 of [24], a widely cited book on longitudinal data, the authors state that "for a given total cost, the investigator may be free to choose between a small value of n (number of repeated measures per patient) and a large sample size (number of patients), or vice versa." However, designs with different combinations will have different statistical properties and the design which leads to the best statistical properties should be chosen. Fortunately, this issue has recently been recognized, see, for example, [26] and [27]. A closely related issue has also been addressed in [28], which proposed a so-called "intentionally incomplete longitudinal design" to overcome the drawbacks of complete longitudinal design.

The remainder of this paper is organized as follows. In Section 2 we describe the PIC data. Section 3 presents two empirical methods and two proposed methods for estimating the viral set point. These four methods are applied to the PIC data in Section 4 and to the simulated data in Section 5. In Section 6, we consider the design issues, and derive the optimal numbers of measurements per patient for determining the viral set point. Technical details are given in

the Appendices.

2. The Data Set

Since 1992, the University of Washington Primary Infection Clinic (PIC) has enrolled individuals with acute or early HIV infection into an observational cohort. Eligible patients had detectable HIV RNA (defined as HIV RNA > 2000 copies per ml) and negative enzyme immunoassay (EIA) (acute HIV infection) or a positive EIA with either negative HIV test within 240 days or negative "detuned" antibody test (less-sensitive EIA) (early HIV infection). The estimated date of HIV infection (Day 0) is the first day of symptoms as reported by the patient or, for asymptomatic individuals, the midpoint between last negative HIV test and first positive HIV test. Patients are enrolled as soon as they are verified to be eligible. Following study enrollment, if possible, patients provided blood samples every week for the first 4 weeks, then at weeks 8, 12, 16, 24, 32, 48, and every 6 months thereafter. Note that while the sampling schedule is based on the time after enrollment, all analyses in this paper use time since HIV infection as defined above. This cohort study was approved by the University of Washington Institutional Review Board, and all patients gave written consent.

Here we consider only the viral load data taken prior to treatment and during the period of $15 - 360$ days following the estimated date of infection. That is, we only consider those data when patients are untreated. The techniques of measuring the viral load in the PIC study are standard, and a detailed explanation can be found in [29]. Because we included only untreated patients from the cohort, there are very few measurements below the standard levels of detection, and for those few that were, we used the value reported from the ultra-sensitive assay described in [29]. Although this approach for censored data seems to be appropriate in

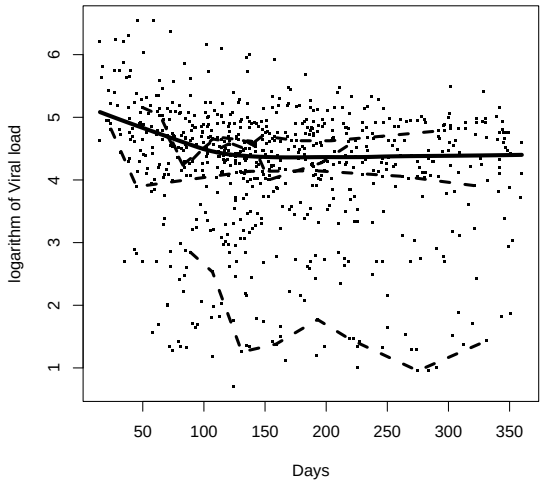

Figure 1. $\log_{10}$ Viral Load versus estimated Days after infection based on 938 observations from 229 patients. The dashed curves are sequences of measurements for selected patients. The solid curve is a lowess smooth fit to the data.

our context, it may cause significant bias for other data sets, particularly those from treated patients, see, for example, [30].

Figure 1 displays 938 viral load measurements plotted against estimated days after infection (verified by clinicians and described above) for 229 infected patients enrolled in the PIC study. Repeated measurements for some patients are connected by dashed curves to illustrate the longitudinal nature of the data. Also shown with a solid curve is a lowess smooth fit ([31]) which ignores the correlation of the repeated measurements on the same patient and bias caused by censoring. Among these 229 patients in this dataset, 57 patients have only 1 measurement, 47 patients have 2 measurements, 2 patients have 12 measurements, the maximum number observed, and the remaining 123 patients have between 3 and 11 measurements. More specifically, $24, 12, 15, 10, 21, 22, 9, 6$ and 4 patients have $3, 4, 5, 6, 7, 8, 9, 10$

and 11 measurements, respectively.

Of the 229 patients in the cohort, 110 (48%) had only untreated viral load measurements observed prior to four months after infection, and these 110 patients have 215 observed viral load measurements in the first four months after infection. In other words, only 119 (52%) of 229 patients have at least one (untreated) measurement during the period between four and twelve months since infection. It turns out this fact will play an important role, and will be explained in more detail in our data analysis section.

3. Methods

3.1. Notation and Model

Suppose there are m untreated HIV infected patients, each with n_i viral load measurements with $i = 1, \ldots, m$. Denote by Y_{ij} and t_{ij}, respectively, the logarithm (base 10) of the viral load and the days since infection, for the jth examination of the ith patient, $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. For the PIC data, or more generally in practice, it is rare to observe the increase to peak viral load, occurring in the first several days after initial infection. Thus we assume that observed viral loads are decreasing from the peak or have already reached steady state. Specifically, we consider viral load collected between $T_{\min} = 15$ and $T_{\max} = 365$ days post estimated date of infection.

As in the longitudinal data analysis literature, the general model for the viral load takes the form

$$Y_{ij} = f(t_{ij}, \phi_i) + \epsilon_{ij},$$

for $i = 1, \cdots, m$ and $j = 1, \cdots, n_i$, where ϕ_i denotes patient-specific parameters, and ϵ_{ij} are

independent error terms. Since we assume that viral load decreases to a stationary level, the function $f(t, \phi_i)$ has the form

$$f(t, \phi_i) = \begin{cases} g(t, \psi_i) & \text{if } t < \nu_i \\ \theta_i & \text{if } t \geq \nu_i \end{cases}, \tag{1}$$

where θ_i is the individual viral set point, the change-point ν_i is the time at which the ith patient reaches the viral set point, $g(t, \psi_i)$ is a decreasing function of t, and $\phi = (\theta, \nu, \psi)$ denotes the vector of parameters. Combining this with the empirical plot of the viral load and the corresponding lowess smooth fit shown in Fig. 1, it seems reasonable to model $g(t, \psi_i)$ as a linear function, and thus $f(t, \phi_i)$ can be modelled parametrically as

$$f(t, \phi_i) = \begin{cases} \alpha_i - \beta_i t, & \text{if } t < \nu_i; \\ \theta_i, & \text{if } t \geq \nu_i. \end{cases} \tag{2}$$

Since the behavior of the viral load is expected to be smooth, we assume that $f(t, \phi_i)$ is a continuous function of t, which requires that $\alpha_i - \beta_i \nu_i = \theta_i$, i.e., $\alpha_i = \theta_i + \beta_i \nu_i$. Plugging this into (2), we get $f(t, \phi_i) = \theta_i - \beta_i \min(t - \nu_i, 0)$, and the parametric model we are interested in can be rewritten as

$$Y_{ij} = \theta_i - \beta_i \min\big(t - \nu_i, 0\big) + \epsilon_{ij}, \tag{3}$$

where $(\theta_i, \beta_i, \nu_i)$ are patient-specific parameters and ϵ_{ij} are independent error terms. Note that model (3) is a special case of the "two-phase" regression models, developed in [14] and [15].

In a random-effects model, some of the patient-specific parameters $(\theta_i, \beta_i, \nu_i)$ may be modelled as independent deviations from the population average value (θ, β, ν). Then θ can be thought of as the population level viral set point, and ν as the population level time at which the viral set point is reached.

Now we will describe the two empirical approaches as well as two proposed approaches to estimate θ, the viral set point at the population level in model (3). Both empirical approaches assume that all patients reach the viral set point after a pre-specified time ν_0, e.g., $\nu_0 = 4$ months post-infection, while our proposed approaches do not make such an assumption and estimate the time ν_i at which the viral set point is reached for the ith patient from the data.

3.2. Method A: Single-measurement

The first empirical approach, the single-measurement method, utilizes a subset of data which contains a single measurement for each patient. Specifically, assume all patients reach the viral set point at $\nu_0 = 4$ months since infection, where the value $\nu_0 = 4$ is chosen in advance, not based on data analysis. Let $\hat{Y}_i$ be the first of the measurements Y_{ij} taken on the i-th patient during the period of $[\nu_0, T_{\max}]$, i.e., between 4 and 12 months post estimated infection. Then $\hat{Y}_i$ is an estimate of the $\log_{10}$ viral set point for the i-th patient, and thus the $\log_{10}$ viral set point θ at the population level can be estimated by

$$\bar{\theta}_1 = \sum_{i=1}^{m} \frac{\hat{Y}_i}{m}. \tag{4}$$

3.3. Method B: Linear Mixed-Effects

The second empirical approach, the linear mixed-effects (LME) method, uses all measurements in $[\nu_0, T_{\max}]$ to estimate the viral set point θ, The assumption is that all patients have reached the set point when $t_{ij} \geq \nu_0 = 4$ months post-infection, the same as Method A. By (1), for $i = 1, \ldots, m$ and $t_{ij} \geq \nu_0$, we have $Y_{ij} = \theta_i + \epsilon_{ij}$. Furthermore, if we assume θ_i's are patient-specific random effects, i.e., $\theta_i = \theta + b_i$, then

$$Y_{ij} = \theta + b_i + \epsilon_{ij}, \tag{5}$$

where θ is the desired viral set point at the population level, b_i's are independent with distribution $N(0, \sigma_b^2)$, and ϵ_{ij}'s are random errors with distribution $N(0, \sigma_w^2)$.

Equation (5) is a standard form of the LME model. In the literature, the parameters in the LME model can be estimated by the maximum likelihood (ML) or restricted maximum likelihood (REML) methods, both of which are available in many standard computer packages, e.g, the "lme" function in R. In this paper we will use REML methods for parameter estimation in the LME models.

3.4. Method C: Two-Phase LME

The first of our proposed methods is the two-phase LME method, which extends the individual-level two-phase model (3) to the population level. The idea is to use model (3) to estimate the time when the viral set point is reached for each patient, and then to derive the viral set point at the population level based on all observations after the patient-specific estimated time when set point is reached. Specifically, we can first fit model (3) for each of the m patients and obtain the corresponding estimators $(\widetilde{\theta}_i, \widetilde{\nu}_i, \widetilde{\beta}_i)$ for $i = 1, \ldots, m$. Since model (3) has three unknown parameters, there is no unique fit if a patient has less than 3 observations. In that scenario, this method is not applicable, and as in the empirical methods, we will define $\widetilde{\nu}_i = 4$ months post-infection. After we obtain the estimated time $\widetilde{\nu}_i$ for each patient, we use the LME method (5) to estimate the viral set point of the population based on all observations $\{Y_{ij}\}$ with $t_{ij} \geq \widetilde{\nu}_i$.

From the statistical point of view, the two-phase LME method may not be appropriate to estimate the viral set point at the population level, as it does not take into account the random nature of the estimated change-points $\widetilde{\nu}_i$. Nevertheless, it can still be applied to estimate the

viral set point at the population level. Moreover, the two-phase LME method allows us to look at the empirical distribution of $\widetilde{\nu}_i$, which is very valuable when building more appropriate models at the population level. For example, estimates of viral change-points obtained from the two-phase LME method can be used the check the normality assumption for $\log(\nu_i)$.

The two-phase LME method can be easily implemented. To see this, it suffices to demonstrate how to fit model (3) to the data $\{Y_{ij}\}_{j=1}^{n_i}$ for each i. An efficient procedure, proposed in [32], can be summarized as follows.

In order to calculate the likelihood ratio of $\{Y_{ij}\}_{j=1}^{n_i}$ under model (3), let us first define τ as the largest integer such that $t_{i\tau} \leq \nu_i$. That is, $t_{i,\tau} \leq \nu_i < t_{i,\tau+1}$. Then, following the notation in (2), maximizing the likelihood ratio of $\{Y_{ij}\}_{j=1}^{n_i}$ under model (3) is equivalent to minimizing

$$L(\alpha_i, \beta_i, \theta_i, \nu_i, \tau_i) = \sum_{j=1}^{\tau}(Y_{ij} - (\alpha_i - \beta_i t_{ij}))^2 + \sum_{j=\tau+1}^{n_i}(Y_{ij} - \theta_i)^2,$$

subject to the constraints $\alpha_i - \beta_i\nu_i = \theta_i$ and $t_{i,\tau} \leq \nu_i < t_{i,\tau+1}$. It was shown in [32] that L is minimized at either (1) $\nu_i = t_{i,\tau}$ for some τ, or (2) $\nu_i = (\hat{\alpha}_\tau - \hat{\theta}_\tau)/\hat{\beta}_\tau \in [t_{i,\tau}, t_{i,\tau+1})$ for some τ, where $(\hat{\alpha}_\tau, \hat{\beta}_\tau)$ and $\hat{\theta}_\tau$ minimize the two sums in L separately.

Let us first discuss finding the solution in the second case. For a given τ, we can separate the data into two parts depending on whether $j \leq \tau$ or $j \geq \tau + 1$, fit simple linear regression lines separately to obtain the estimate $(\hat{\alpha}_\tau, \hat{\beta}_\tau)$ and $\hat{\theta}_\tau$, and calculate $\hat{\nu}_\tau = (\hat{\alpha}_\nu - \hat{\theta}_\tau)/\hat{\beta}_\tau$. We obtain a local maximum likelihood estimator (MLE), a candidate for the global MLE, if and only if $\hat{\nu}_\tau \in [t_{i,\tau}, t_{i,\tau+1})$,

Now let us go back to the first case. At a given τ, we have $\alpha_i = \beta_i t_{i,\tau} + \theta_i$ and thus

$$L = \sum_{j=1}^{\tau}(Y_{ij} - (\theta_i + \beta_i(t_{ij} - t_{i,\tau})))^2 + \sum_{j=\tau+1}^{n_i}(Y_{ij} - \theta_i)^2.$$

Treating $Y_{ij} - \theta_i$ as $Y_{ij} - (\theta_i + \beta_i \times 0)$ will reduce the problem to a simple linear regression

problem, and we can easily obtain another local MLE. It is worth emphasizing that the first case deals with the boundary points in each subinterval and is often neglected.

We repeat the above calculation for both cases and record L for each local MLE; then the local MLE for which L is minimized is just the global MLE $(\widetilde{\theta}_i, \widetilde{\nu}_i, \widetilde{\beta}_i)$ in model (3). Finally, once the estimated individual change-points, $\widetilde{\nu}_i$'s, have been obtained, the LME model can be used on all observations $\{Y_{ij}\}$ with $t_{ij} \geq \widetilde{\nu}_i$ to derive the estimate of the viral set point.

3.5. Method D: Two-Phase NLME

The second of our proposed methods is the two-phase NLME method, which incorporates random effects when extending model (3) to the population level. As described earlier, one or more of the patient-specific parameters $(\theta_i, \beta_i, \nu_i)$ can be modelled as random effects. When treating all parameters as random effects, we have difficulties in obtaining numerical results for the PIC data due to the high degree of correlation between θ_i, β_i and ν_i. For that reason, we consider the following two-phase NLME model with θ_i and ν_i being random effects:

$$Y_{ij} = \theta_i - \beta \min\left(t_{ij} - \nu_i, 0\right) + \epsilon_{ij}, \tag{6}$$

assuming the patient-specific effects $(\theta_i, \log(\nu_i))$ are mutually independent with a common underlying bivariate normal distribution with mean $(\theta, \log(\nu))$, and ϵ_{ij} are independent error terms. The reason we assume the normality of $\log(\nu_i)$ instead of ν_i is that ν_i is always positive. Further details can be found in Section 4.

To obtain the estimates in the two-phase NLME model (6), it is useful to approximate the function $\min(u, 0)$ by a smooth function $u/(\exp(\delta u) + 1)$, where δ is a (large) positive smoothing parameter chosen by the statistician. This approximation will allow us to use standard computer packages to fit this model, for example, the function "nlme" in R. For

Table I. Estimated $\log_{10}$ Viral Set Point θ ($\log 10$ copies/mL) for the PIC data

Method	Estimates	SE
A: Single-measurement	4.163	0.087
B: LME	4.174	0.087
C: Two-Phase LME	4.246	0.078
D: Two-Phase NLME	4.227	0.065

details regarding numerical implementations and smoothing, see [16] and [33].

4. Example: PIC data

4.1. Results

We now apply the four methods mentioned in the previous section to the PIC data shown in Figure 1: (A) the single-measurement method; (B) the LME method; (C) the two-phase LME method; and (D) the two-phase NLME method. We used of the statistical software R (version 2.3.1) in our numerical computation.

For each of these four methods, the estimate and standard error (SE) for the $\log_{10}$ viral set point θ are summarized in Table I. On the one hand, from the bias point of view, these four methods produce similar point estimates of the viral set points in the sense that the estimates are within the range of standard error. On the other hand, from the efficiency point of view, the proposed methods, Methods C and D, have significantly smaller standard errors of the estimate than the empirical methods, Methods A and B. In fact, Methods C and D reduced the standard errors of the empirical methods by 10% and 25%, respectively.

4.2. Results after taking into account the effects of sample sizes

From the above results, we see that the proposed methods are better than the empirical methods (in terms of smaller variance of the estimate) for the viral load data set. In this subsection we take a further step to understand the reason behind the improvement.

A more detailed analysis shows that the sample size (number of patients) in the empirical methods are 119, i.e., only 119 patients have any viral load measurements after the viral change point, whereas the sample sizes for Methods C and D are 143 and 229, respectively. That is, the empirical methods cannot be applied to patients who only have observations during the first four months since infection, and the proposed methods, particularly Method D (the two-phase NLME model) can be applied to data from more patients so that the patient sample size for these two methods is larger. Therefore, it is not surprising that the reduced sample size would result in the empirical methods being less efficient in estimating the viral set point at the population level than our proposed two-phase methods.

As discussed in the Introduction, the effect of sample sizes is well-known in the literature, and a closely related goal of this paper is to understand how often viral load measurements should be collected per patient in order to estimate the viral set point. For that purpose, in the following we exclude the patients who only have observations during the first four months post-infection, and apply all four methods to this reduced data set. This will allow us to see whether or not the empirical methods are still less efficient than the proposed methods after taking into account the effects of the patient sample size.

After excluding the patients who only have observations during the first four months post-infection, the estimate and standard error (SE) for the $\log_{10}$ viral set point θ for each of these four methods are summarized in Table II. From Table II, there is little difference in the point

Table II. Estimated log 10 Viral Set Point θ (log 10 copies/mL) for the PIC data after taking in account the effects of sample size

Method	Estimates	SE
A: Single-measurement	4.163	0.087
B: LME	4.174	0.087
C: Two-Phase LME	4.197	0.088
D: Two-Phase NLME	4.172	0.085

estimates of θ by these four methods in light of the standard errors. Moreover, the standard errors of the estimates for these four methods are almost the same. Thus, all four methods yield similar results for the estimated viral set point when applied to the same patient population. These results indicate that for the PIC data, the empirical methods, including both the single-measurement and LME methods, seem to be efficient, when applied to datasets where each patient has at least one measurement obtained after reaching the viral set point (in this case, between four to twelve months after estimated infection).

4.3. Estimation of Change-Point and Normality Assumption

In this subsection let us focus on the reduced PIC data after excluding the patients who only have observations during the first four months post-infection. Note that both Method C (the two-phase LME model) and Method D (the two-phase NLME model) can estimate the change-point ν at which the set point is reached from the data, and the respective results are summarized as follows.

Method C provides an estimate, $\widetilde{\nu}_i$, for each individual with at least three observations. For

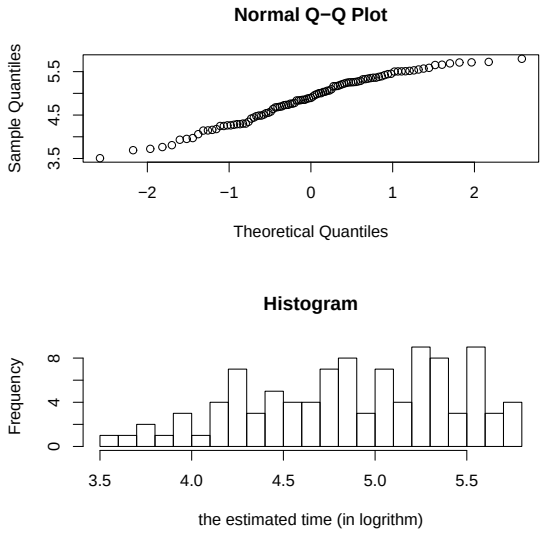

Figure 2. The normal QQ Plot and Histogram of the logarithm of the estimate time ν_i by Method C (the two-phase LME model)

the PIC data, Method C produced 101 such $\widetilde{\nu}_i$'s, whose median is 130.7. The 95% confidence interval of the change-point ν at the population level is $[117, 146]$.

The normal QQ plot and histogram of the $\log(\widetilde{\nu}_i)$ by Method C are shown in Fig. 2. Note that although the histogram may not support the normality assumption, the normal QQ plot seems to suggest that it is reasonable to assume the $\log(\nu_i)$ are normally distributed, an assumption which was used in Method D: the two-phase NLME model.

Applying Method D, $\log \nu$ (in terms of natural logarithm) was estimated to be normally distributed with mean 3.9943 (approximately 54 days) and variance 0.0621. In addition, the 95% confidence interval of the change-point ν at the population level is $[48, 62]$.

Therefore, Methods C and D provide different 95% confidence intervals for the change-point ν at the population level: one is [117, 146] and the other is $[48, 62]$. Note that estimation

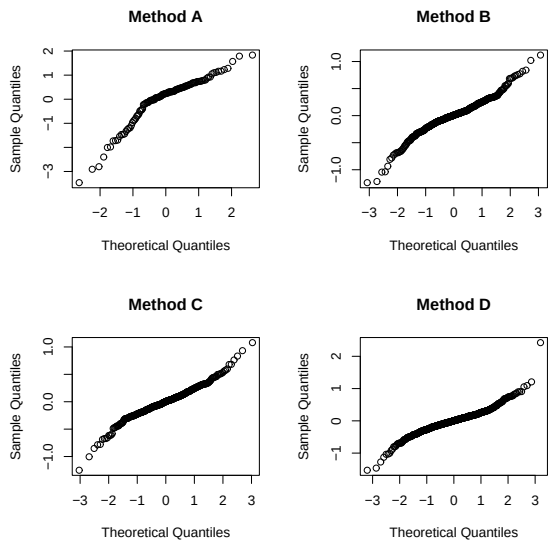

Figure 3. The normal QQ Plot of the residuals for the four methods: Method A (single-measurement); Method B (LME); Method C (two-phase LME); and Method D (two-phase NLME). Note that these four methods are working on different subsets of the observations.

of (random) change points is generally a very challenging problem, especially with possible misspecification of the underlying model. After considering the HIV dynamics and the lowess smooth fit in Fig. 1, we believe the estimate of ν from Method D (the two-phase NLME model) is likely too early, i.e., Method D underestimates the true ν. See Section 5 for a more detailed discussion.

Fig. 3 shows the normal QQ plots of the residuals of the four methods applied to the reduced data set. Although these four methods have the same sample size (number of patients $n = 119$), the number of observations per patient is different for these four methods. Thus the total numbers of observations in each analysis is different (119, 472, 413, 723 for Methods A, B, C, D, respectively), implying that we should not compare the plots with each other. It is interesting to see that the QQ plot of Method A seems to suggest that the first measurements after the

four months post-infection in the PIC study are from a mixture of normal distributions. Fig. 3 also suggests a non-linearity of the QQ plots of Methods B-D, in which the patient-specific viral set point is a random effect.

5. Simulations

Since the effect of sample sizes is well-known in the literature, our simulations will focus on comparing the four methods we have discussed after taking into account the effects of the sample size. To compare these four methods, we use the two-phase NLME model (3) with all of the parameters θ_i, β_i and ν_i as random effects to generate the simulation data and then apply all four methods in Section 3 to the simulated data. Specific attention is focused on how well each method performed in terms of its accuracy and efficiency in determining the viral load set point at the population level, when the sample sizes are the same for these four methods, that is when all patients have at least one viral load measurement obtained four months or more post infection.

In our simulation, we performed Monte Carlo analysis with 1000 runs. In each run, data were simulated for 120 individuals at $t = 2, 4, 8, 12, 16, 20, 24$, and 32 weeks. The parameters used in generating our simulated data were motivated by the PIC data. When fitting all four methods to the simulated data, only the two-phase NLME method failed to converge in 35 out of 1000 replicates (3.5%) with "nlme" in R.

Table III shows the estimated parameters for the four methods when using R and the convergence status of each. From Table III, the Monte Carlo means of the viral set point θ are very close to the true value for all four methods, except possibly Method C, the two-phase LME method. As previously mentioned, Method C may have some bias due to the randomness

Table III. Comparison of Four Methods by Simulation. The true values in the simulation for the viral set point and the time at which the viral set point is reached are $\theta = 4.2$ and $\nu = 54.6$, respectively.

	θ	ν	Non-convergence
A: Single-measurement	4.198 ± 0.085	—	None
B: LME	4.197 ± 0.082	—	None
C: Two-Phase LME	4.163 ± 0.081	66 ± 3	None
D: Two-Phase NLME	4.190 ± 0.081	55 ± 2	35/1000 (=3.5%)

of the estimated change-point, but the difference between its estimate and the true value is very small (0.037) compared to the standard error (0.081).

Table III also suggests that Method C, the two-phase LME method, overestimates the time ν at which the viral set point is reached, whereas Method D, the two-phase NLME method, provides an unbiased estimate of ν. It is worth mentioning that our simulations were carried out in an ideal setting with the correct underlying model specified, and so could be misleading if there are any misspecifications of the underlying model.

From Table III, these four methods have similar performances (in terms of both bias and precision of the estimated set point) in our simulations. It is useful to point out that it is mainly because the within-patient correlation we used is high ($\rho = 0.86$ based on the PIC data), and the patient sample sizes for these four methods are the same.

Our simulations were carried out using a balanced design with complete data for all patients, and from a practical perspective, one may want to see what happens in an unbalanced design with a sparser and less uniformly complete data environment. A subtle point here is whether or not the unbalanced design will lead to different numbers of patients (sample size) available

for analysis by the different methods. If the sample sizes are significantly different in the unbalanced design, then it is not surprising that the estimated set point obtained from Methods C and D will have significantly smaller standard errors than the estimated set point obtained from the empirical methods, Methods A and B, as shown in Table 1 of Section 4.1. On the other hand, if the sample sizes are the same for all methods in the unbalanced design, then the conclusion will be similar to our simulation results using the balanced design.

To see this, note that Method A (single measurement method) only uses one measurement for each patient, whereas Methods C and D use all available measurements for each patient. If we treat the unbalanced design as the balanced design with missing data, the unbalanced design will not change the performance of Method A, but will make the performance of Methods C and D worse (as long as the sample sizes are the same). Therefore, if these four methods perform similarly using a balanced design (as in our simulations), then they will also agree using an unbalanced design with a less uniformly complete data environment (as long as the sample sizes are the same).

6. Design Issues

In this section, we will assume that the true model is the general model defined in (1), and the patient-specific parameter $(\theta_i, \nu_i, \psi_i)$ is a random deviation from the population parameter (θ, ν, ψ). Denote by θ^* the restricted MLE estimate of θ when fitting this random-effects model to the data. In designing a study with viral set point as an outcome, two design issues must be considered: when and how often should viral load measurements be obtained so that the variance of the estimate θ^* is small? In order to estimate the viral set point, intuitively it would be best to take all measurements after the viral set point has been reached. Unfortunately, it

is unclear when the viral set point will be reached for each patient. Nevertheless, we can first study the properties of good designs, and then consider the practical issues of such designs. For that purpose, we will assume all observations are taken after the viral set point is reached in subsections 6.1 and 6.2, and then provide more practical guidelines for a cost-effective design in subsection 6.3.

6.1. Number of repeated Measurements: Cost Consideration

As unbalanced designs can be thought of as balanced designs with missing data, in this and next subsection, we will only consider the balanced design where each of m patients has the same number of repeated observations (n). Since we can assume that all observations are taken during the stationary state (i.e., at the viral set point), the model for the viral load becomes the linear mixed model in (5). To determine how often viral load should be measured on each patient, it is sufficient to find the optimal choice of n so the estimate $\hat{\theta}_n$ of the viral set point has desired properties, where $\hat{\theta}_n$ denotes the estimated viral load based on n measurements per patient.

Viral load tests are expensive (compared to other tests such as CD4) and therefore it is reasonable to introduce a cost function. As in [23], suppose it costs c_1 to recruit a patient, and costs c_2 to take a measurement. Then the "total cost" of the design is

$$C = c_1 m + c_2 mn = m(c_1 + c_2 n),$$

where m is the number of patients. Now the cost-effective choice of the number of repeated observations per person is equivalent to finding the number of observations per patient, n, that achieves a balance between the variance of $\hat{\theta}_n$ and the total cost $C = c_1 m + c_2 mn$. In this paper, the balance we are seeking is based on minimizing the product of $Var(\hat{\theta}_n)$ and the total

cost $C = c_1 m + c_2 mn$. As shown in the Appendix, under our formulation, the cost-effective choice of the number of repeated observations per patient is given by

$$n_{opt} = [\sqrt{\frac{1-\rho}{\rho}\frac{c_1}{c_2} + \frac{1}{4}} + \frac{1}{2}], \tag{7}$$

where $[x]$ is the integer part of x, and ρ is the so-called intra-class correlation (ICC) or within-patient correlation

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} = Corr(Y_{ij}, Y_{ik}) \quad \text{for all } j \neq k, \tag{8}$$

where σ_b^2 and σ_w^2 are the variances of b_i and ϵ_{ij} in model (5), respectively.

There are many other mathematical formulations of the balance between the variance of $\hat{\theta}_n$ and the total cost of the study $C = c_1 m + c_2 mn$. For example, one intuitive formulation is to find the n that minimizes $Var(\hat{\theta}_n)$ subject to the constraint that the total cost $C = c_1 m + c_2 mn \leq C_0$, where $C_0 > 0$ is a pre-specified total cost. Another intuitive formulation is to find the n that minimizes the total cost $C = c_1 m + c_2 mn$, subject to the constraint the variance $Var(\hat{\theta}_n) \leq V_0$, where $V_0 > 0$ is a pre-specified desired variance. By some elementary arguments, it is easy to show that these two intuitive formulations are equivalent to our formulation of minimizing the product of $Var(\hat{\theta}_n)$ and the total cost $C = c_1 m + c_2 mn$, if we treat the parameter m (the number of patients) as a variable and optimize over all possible (not necessarily an integer) values of m. This equivalence demonstrates that our formulation is indeed useful and reasonable. Moreover, our formulation is simpler as it does not introduce any new constraints (C_0 or V_0).

It is interesting to point out that the value of n_{opt} given in (7) does not depend on the number of patients, and thus is optimal at either the individual or population level. From (7), it is straightforward to see that the single-measurement design is cost-effective, i.e., $n_{opt} = 1$,

if and only if

$$\frac{1-\rho}{\rho}\frac{c_1}{c_2} < 2.$$

Now we can apply these results to answer how often one should take viral load tests in order to efficiently estimate the viral set point. Based on the PIC data, the estimate of the within-patient correlation is $\rho \approx 0.86$. In practice, it currently takes $c_2 \approx 100$ U.S. dollars to perform each viral load test, and c_2 can be more expensive in the developing countries, where the blood samples are often sent to other countries for measuring viral load. Thus, if it takes $c_1 \leq 1228$ U.S. dollars to recruit a patient, then equation (7) suggests that $n_{opt} = 1$, implying that the cost-effective design will take 1 measurement per patient after the viral set point is reached. Of course if the cost to recruit a patient $c_1 \geq 1229$ U.S. dollars, then the investigators may want to take at least 2 measurements per patient, as suggested by equation (7).

6.2. Number of Repeated Measurements: Relative Precision

An alternative approach to determine how often viral load should be measured per patient is to require that the estimate should attain a certain accuracy compared to a benchmark. For comparison purposes, we assume that m, the number of patients, is fixed. The benchmark we used is an unrealistic design which assumes that each of m patients provides infinitely many observations. In the benchmark, the variance of the estimate of the viral set point is the smallest over all possible designs, and we denote this smallest value by V_{inf}. Other benchmarks could be used depending on the specific study. As in [24, p. 24], [25], define the relative precision (RP) of $\hat{\theta}_n$ with respect to the benchmark by

$$RP(n) = \frac{V_{\text{inf}}}{Var(\hat{\theta}_n)}. \tag{9}$$

Because $V_{\inf}$ is the lower bound, we have $0 < RP(n) \leq 1$. One natural formulation of determining the number of measurements per patient is to determine n so that

$$RP(n) \geq r_0$$

where $0 < r_0 < 1$ is a pre-determined constant, e.g., $r_0 = 0.9$. This formulation guarantees that the design with n measurements at the viral set point for each patient will have a relative precision of at least r_0 as compared to the benchmark design.

In Appendix B, it is shown that the smallest n satisfying $RP(n) \geq r_0$ is given by

$$n_0 = \left\lceil \frac{1-\rho}{\rho} \frac{r_0}{1-r_0} \right\rceil, \tag{10}$$

where ρ is the intra-class correlation defined in (8) and $\lceil x \rceil$ is the smallest integer not smaller than x. In particular, if the required relative precision $r_0 \leq \rho$, then $n_0 = 1$. Otherwise, $n_0 \geq 2$.

As an example, suppose we want to design an experiment with $r_0 = 0.9$ based on the PIC data where $\rho \approx 0.86$. Then, $n_0 = 2$, i.e., we need to take 2 measurements per patient after the viral set point is reached.

6.3. Practical Considerations

From the above arguments, the design we recommend for estimating the viral set point in future studies is based on one or two measurements obtained after the viral set point is reached.

It is worth emphasizing that we made a major assumption in this recommendation based on our analysis of the PIC data: for the viral loads taken at the viral set point from untreated patients, the observed within-patient correlation ρ is very large ($= 0.86$). This is consistent with our intuition: the viral loads at the stationary state (viral set point) should be highly correlated, and probably only a few measurements will be sufficient to estimate the viral set point. If the value of ρ is smaller, then the corresponding optimal choice of the number of

Table IV. Relationship between the optimal value and within-patient correlation ρ when $c_1/c_2 = 5$.

ρ	[0.71,1]	[0.45,0.71)	[0.29, 0.45)	[0.20, 0.29)	[0.14, 0.20)	[0.11, 0.14)
n_{opt}	1	2	3	4	5	6

repeated observations per patient can still be derived from (7) or (10), although the actual values can be larger. For instance, if $c_1/c_2 = 5$, then the relationship between the optimal value n_{opt} and the within-patient correlation ρ is illustrated in Table IV.

Unfortunately, this recommendation is not applicable, since it will be difficult, even if possible, to know the exact time at which the viral set point is reached for each patient. As in the empirical methods, a natural idea is to derive an upper bound on the times at which the viral set points are reached, and then recommend that all viral loads be taken beyond this upper bound. On the one hand, such an upper bound should be conservative, since otherwise the observed viral loads may not be in the stationary state (i.e., at the viral set point). On the other hand, such an upper bound is desired to be as small as possible so as to allow early medical decision making and to maximize sample sizes (due to treatment and dropout). Thus, we need to derive a conservative yet reasonably small upper bound on the times at which the viral set points are reached.

The empirical methods claim that such an upper bound can be $\nu_0 = 120$ days, or 4 months post-infection. For the PIC data, our analysis in Section 4 shows the 95% confidence interval of the change-point ν at the population level is $[117, 146]$ for Method C (the two-phase LME model), and $[48, 62]$ for Method D (the two-phase NLME model). In addition, our simulation in Section 5 suggests that when the two-phase NLME model is the true underlying model, Method D (the two-phase NLME model) will give an unbiased estimate for the change-point

ν and Method C (the two-phase LME model) may overestimate ν. Unfortunately, we cannot determine whether the two-phase NLME model is the true underlying model. Moreover, some estimated individual change-points ν_i's in Method C are larger than 120 days post-infection. This suggests adopting the conservative estimate provided by Method C (the two-phase LME model), and a more appropriate upper bound on the change-point ν seems to be 150 days, or five months, post infection.

Therefore, a more practical guideline for effectively estimating viral set point in future designs may be based on one or two measurements obtained between five and twelve months since infection. This is consistent with the empirical methods for estimating the viral set points, see [3, 4, 10, 13]. The only difference is that we are more conservative on the time of viral load tests to make sure all or most observed viral loads are after the viral set point has been reached. However, our conclusion is based on a single data set and needs further verification.

It is well-known in the experimental design literature that an optimal design depends heavily on the assumptions used and ignores the robustness of the assumptions. This does not mean that the optimal design is useless, but emphasizes that caution should be used when considering an optimal design. In our context, from the practical point of view, if financially feasible, one may want to take as many observations as possible in future studies, especially before the viral set point is reached, so that new studies are able to estimate the viral set point efficiently as well as verify the validity of the assumptions. Moreover, it is also important to emphasize that while viral load measurements obtained after the viral set point has been reached are suitable for estimating the viral set point, they are not appropriate for other purposes, such as understanding HIV dynamics, especially in the early stage.

7. Discussion

In this paper, we have shown that when observations within patients are highly correlated, which is often the case with repeated viral load measurements obtained after the viral set point has been reached, it is accurate and efficient to estimate the viral set point based on 1 or 2 viral load measurements taken after the viral set point is reached. This observation is very useful in practice and supports the widely-used single-measurement approach currently used by many investigators. More importantly, it also suggests that if an investigator wants to compare HIV viral set points between two or several groups, then it is sufficient to take 1 or 2 viral load measurements per patient (after the viral set point is reached). In other words, while adding more observations per patient from the viral set point will improve the power to detect a difference in the viral set point, this improvement is not significant. The improvements we observed in precision of the estimated viral set point using the entire PIC data set, including patients who had no measurements four or more months post infection, were due to larger patient sample size rather than increased additional observations per patient. Thus the investigator should focus on recruiting more patients rather than taking more observations per patient.

We proposed two different methods to use all viral load data to estimate the viral set point when viral load data are available prior to reaching the viral set point, and compared them with the empirical methods. On the one hand, if there are patients who only have viral load measurements prior to the pre-specified time, then our proposed methods will include those patients and their measurements, whereas the empirical methods will not. Thus, it is not surprising that our proposed methods will have better performance in this scenario due to larger patient sample sizes. In other words, adding pre-set point viral load data increases

precision in the estimation of viral set point only when additional patients can be included.

On the other hand, as shown by theoretical arguments as well as by simulation, if each patient has at least one observation collected after the viral set point has been reached, then our proposed methods have the same performance as the empirical methods when estimating the viral set point. In other words, once a patient has viral load measurements available after a pre-specified time at which viral set point is assured to have occurred, adding additional measurements prior to that time via a model and alternate estimation procedures does little to affect the precision of the estimated viral set point on a population level.

Unlike the empirical methods, our proposed methods can be used to estimate the time at which viral set point occurs at the individual or population level. Our simulations suggest that the two-phase LME method can provide an upper bound on the time to viral set point, and the two-phase NLME method provides an unbiased estimate of time to viral set point. Unfortunately, the simulations were carried out in an ideal setting with the correct model specified, and the results could be misleading if there are any departures from the correct model.

In general it is difficult to estimate change-points accurately, particularly in the random change-point models when the number of observations per patient is limited. To avoid such difficulties, we focus on the upper bound on change-points, and the two-phase LME method seems to be able to provide such a bound. Our results based on analysis of the PIC data suggests that a *conservative* estimate of time to viral set point is five months post infection. We have shown that relatively few measures of viral load per patient are necessary for estimating the viral set point, so that using a conservative estimate of time when the viral set point is reached in a population should not adversely affect most studies.

It would also be interesting to study other important aspects of a design, such as the appropriateness of the models, serial correlation, missing data, non-normal distribution, and truncated data. Results of these investigations will further demonstrate whether the simple, widely-used single-measurement method is indeed sufficient to estimate the HIV viral set point.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the University of Washington Primary Infection Clinic for permission to use the viral load data. The authors would like to thank Dr. Joanne Stekler for clarifying some clinical issues, and Ms. Valerie Dodgin for proofreading the earlier draft of this paper. The authors also would like to thank two reviewers for helpful suggestions, which led to significant improvements in organization and presentation.

APPENDIX

Recall that $\hat{\theta}_n$ is the estimate of the viral set point in the LME model (5) because all observations are taken at the viral set point. By (5),

$$Var(\hat{\theta}_n) = \frac{\sigma_b^2}{m} + \frac{\sigma_w^2}{mn} = \frac{\sigma_{tot}^2}{m}(\rho + \frac{1-\rho}{n}), \tag{11}$$

where ρ is the intra-class correlation (ICC) defined in (8), and $\sigma_{tot}^2 = \sigma_b^2 + \sigma_w^2$. To achieve a balance between the variance of $\hat{\theta}_n$ and the total cost $C = c_1 m + c_2 mn = m(c_1 + c_2 n)$, observe that

$$Var(\hat{\theta}_n)C = \sigma_{tot}^2(\rho + \frac{1-\rho}{n})(c_1 + c_2 n),$$

does not depend on m, the number of patients. Thus, to minimize the product of the variance of $\hat{\theta}_n$ and the total cost C, an optimal choice of n, the number of repeated measures per patient, will minimize

$$h_n = (\rho + \frac{1-\rho}{n})(c_1 + c_2 n).$$

Note that

$$h_n - h_{n-1} = \rho c_2 - \frac{(1-\rho)c_1}{n(n-1)},$$

so $h_n \leq h_{n-1}$ if and only if

$$n(n-1) \leq \frac{(1-\rho)c_1}{\rho c_2}. \tag{12}$$

A simple calculation shows that n_{opt} in (7) is the largest integer n satisfying (12). Thus h_n is decreasing if $n \leq n_{opt}$, and increasing if $n \geq n_{opt}$. Hence, h_n is minimized at $n = n_{opt}$, or equivalently, the optimal value of the number of repeated measures per patient is given by $n = n_{opt}$.

APPENDIX

By the results in Sections 6.1, 6.2 and equation (11),

$$V_{\text{inf}} = \lim_{n\to\infty} Var(\hat{\theta}_n) = \rho\frac{\sigma_{tot}^2}{m}.$$

Using the definition of the relative precision in (9), we have

$$RP(n) = \frac{V_{\text{inf}}}{Var(\hat{\theta}_n)} = \frac{\rho}{\rho + (1-\rho)/n},$$

Thus, $RP(n) \geq r_0$ is equivalent to requiring that

$$n \geq \frac{1-\rho}{\rho}\frac{r_0}{1-r_0},$$

and so the smallest n satisfying $RP(n) \geq r_0$ is n_0 defined in (10).

REFERENCES

1. Coffin JM, Hughes SH, Varmus HE. *Retroviruses.*Cold Spring Harbor Laboratory Press, 1997.
2. Daar ES, Moudgil T, Meyer RD, Ho DD. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. *New England Journal of Medicine* 1991; **324**:961-964.
3. Schacker T, Collier AC, Hughes J, Shea T, Corey L. Clinical and Epidemiologic Features of Primary HIV Infection. *Annals of Internal Medicine* 1996; **125**(4):257-264.

4. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC. Initial Plasma HIV-1 RNA Levels and Progression to AIDS in Women and Men. *New England Journal of Medicine* 2001; **344**:720-725.
5. Henrard DR, Phillips JF, Muenz LR, Blattner WA, Wiesner D, Eyster ME, Goedert JJ. Natural history of HIV-1 cell-free viremia. *Journal of the AMerican Medical Association* 1995; **274**(7): 554-558.
6. Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P, Kingsley LA, Todd JA, Saah AJ, Detels R, Phair JP, Rinaldo CR Jr. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of Internal Medicine* 1997; **126**(12):946-54.
7. 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**(5265):1167-70. Erratum in: *Science.* 1997; **275**(5296):14. DOI: 10.1126/science.272.5265.1167
8. de Wolf F, Spijkerman I, Schellekens PT, Langendam M, Kuiken C, Bakker M, Roos M, Coutinho R, Miedema F, Goudsmit J. 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**(15):1799-1806.
9. Garcia F, Vidal C, Gatell JM, Miro JM, Soriano A, Pumarola T. Viral load in asymptomatic patients with CD4+ lymphocyte counts above 500 x 10(6)/l. *AIDS.* 1997;**11**(1):53-57.
10. Richardson BA, Mbori-Ngacha D, Lavreys L, John-Stewart GC, Nduati R, Panteleeff DD, Emery S, Kreiss JK, Overbaugh J. Comparison of human immunodeficiency virus type 1 viral loads in Kenyan women, men, and infants during primary and early infection. *Journal of Virology* 2003; **77**(12):7120-7123. DOI: 10.1128/JVI.77.12.7120-7123.2003
11. Vidal C, Garcia F, Romeu J, Ruiz L, Miro JM, Cruceta A, Soriano A, Pumarola T, Clotet B, Gatell JM. Lack of evidence of a stable viral load set-point in early stage asymptomatic patients with chronic HIV-1 infection. *AIDS.* 1998; **12**(11):1285-1289.
12. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963-974. DOI:10.2307/2529876
13. Kaufmann GR, Cunningham P, Kelleher AD, Zaunders J, Carr A, Vizzard J, Law M, Cooper DA.Patterns of viral dynamics during primary human immunodeficiency virus type 1 infection. The Sydney Primary HIV Infection Study Group. *The Journal of infectious diseases* 1998; **178**(6):1812-1815. DOI: 10.1086/314480
14. Hinkley DV. Inference about the intersection in two-phase regression. *Biometrika* 1969; **56**: 495-504. DOI:10.2307/2334655
15. Hinkley DV. Inference about the change-point in a sequence of random variables. *Biometrika* 1970; **57**:

1-17. DOI:10.2307/2334932

16. Slate EH, Turnbull BW. Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine* 2000; **19**: 617-637.
17. Chu H, Gange SJ, Yamashita TE, Hoover DR, Chmiel JS, Margolick JB, Jacobson LP. Individual variation in CD4 cell count trajectory among human immunodeficiency virus-infected men and women on long-term highly active antiretroviral therapy: an application using a Bayesian random change-point model. *American Journal of Epidemiology* 2005; **162**(8):787-97. DOI:10.1093/aje/kwi268
18. Wu H. Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research* 2005; **14**(2):171-192. DOI: 10.1191/0962280205sm390oa
19. Mehrotra DV, Li X, Gilbert PB. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial. *Biometrics* 2006; **62**(3): 893-900. DOI: 10.1111/j.1541-0420.2005.00516.x
20. Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics* 1997; **53**: 937-947. DOI:10.2307/2533554
21. Overall J, Doyle S. Estimating sample size for repeated measurement designs. *Controlled Clinical Trials* 1994; **15**: 100-123.
22. Rochon J. Sample size calculations for two-group repeated-measurements experiments. *Biometrics* 1991; **47**: 1383-1398. DOI:10.2307/2532393
23. Cochran WG. *Sampling Techniques.* Wiley, New York, 1977.
24. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data.* Oxford University Press, New York, 2002.
25. Wang YG, Chen Z, Liu J. General ranked set sampling with cost consideration. *Biometrics* 2004; **60**(2): 556-561. DOI:10.1111/j.0006-341X.2004.00204.x
26. Carroll RJ, Freedman L, Pee D. Design aspects of calibration studies in nutrition, with analysis of missing data in linear measurement error models. *Biometrics* 1997; **53**: 1440-1457. DOI:10.2307/2533510
27. Lindsey, JK. *Models for Repeated Measurements.* Oxford University Press, New York, 1999.
28. Helms RW. Intentionally incomplete longitudinal designs: methodology and comparison of some full span designs. *Statistics in Medicine* 1992; **11**(14-15):1889-1913.
29. Holte SE, Melvin AJ, Mullins JI, Tobin NH, Frenkel LM. Density-dependent decay in HIV-1 dynamics. *Journal of Acquired Immune Deficiency Syndromes* 2006; **41**(3):266-276.
30. Hughes MD. Analysis and design issues for studies using censored biomarker measurements with an

example of viral load measurements in HIV clinical trials. *Statistics in Medicine* 2000; **19**(23):3171-3191.

31. Cleveland WS, Devlin SJ. Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**: 596-610. DOI:10.2307/2289282
32. Hudson DJ. Fitting segmented curves whose joint points have to be estimated. *Journal of the American Statistical Association* 1966; **61**: 1097-1129. DOI:10.2307/2283203
33. Seber GAF, Wild CJ. *Nonlinear Regression.* Wiley, New York, 1989.

