Published in final edited form as:

Stat Med. 2010 August 15; 29(18): 1932–1946. doi:10.1002/sim.3955.

Assessing agreement in the timing of treatment initiation determined by repeated measurements of novel versus gold standard technologies with application to the monitoring of CD4 counts in HIV-infected patients

Farzad Noubary^{a,*} and Michael D Hughes^a

^a Department of Biostatistics, Harvard School of Public Health

Abstract

Repeated biomarker measurements are often taken over time to help assess risk of disease progression and guide clinical decision-making, such as whether to start treatment. Unfortunately, gold standard methodologies for measuring biomarkers are often prohibitively expensive or unavailable in resource-limited settings. For example, the costs of monitoring HIV-infected subjects to decide when to start or change treatments are a significant burden for many countries, often exceeding the costs of treatments. A major issue concerns how to evaluate changes in timing of key clinical decisions if a new, simpler or less expensive technology were used instead of the gold standard. We develop a framework for addressing this problem and apply it to the case of monitoring CD4 counts in HIV-infected patients. We focus on the practically-important situation in which longitudinal natural history data are available for the gold standard (flow cytometry for CD4 counts), but where the first data expected for a new technology will come from a cross-sectional method comparison study, allowing for estimation of variability and systematic differences (bias) between the two technologies. In a case study, we illustrate how a combination of statistical modeling and simulation study might be used to evaluate the potential impact of using a new technology on treatment starting times in a population of HIV-infected subjects. This gives developers of new CD4 measurement technologies insight into what might constitute acceptable increases in variability and/or bias for novel methods. We finish with a discussion of our findings and some statistical problems that need further work.

Keywords

time to threshold; error; bias; confirmatory measurement; diagnostic testing

1. Introduction

Repeated measurements of a biomarker are often taken over time to help assess risk of disease progression and guide clinical decision-making, such as whether to start treatment. For example, the World Health Organization (WHO) recommends that CD4⁺ T-lymphocyte count be measured every six months [1] in patients infected with the Human Immunodeficiency Virus (HIV) and that antiretroviral therapy (ART) be initiated in patients with CD4 counts below 350 cells/mm³ because of an elevated risk of progression to the Acquired Immunodeficiency Syndrome (AIDS) [2]. Guidelines in the United States recommend that patients initiate therapy when they have a count below 500 cells/mm³,

* Correspondence to: 655 Huntington Avenue, SPH2, 4th Floor, Boston, MA 02115. fnoubary@hsph.harvard.edu.

though the recommendation was stronger for patients with counts below 350 cells/mm³ than between 350 and 500 cells/mm³ [3].

Unfortunately, gold standard methodologies for measuring biomarkers of interest are often prohibitively expensive or not available in resource-limited settings. Flow cytometry, the gold standard CD4 measurement methodology, is expensive and requires trained personnel, reliable electricity, and expensive reagents [4]. Indeed, in the context of HIV, the cost of laboratory monitoring has surpassed the price of ART [5]. To address this issue, the WHO has urged the development of inexpensive methodologies for measuring CD4 counts, while encouraging national HIV programs to increase access to CD4 measurement technologies because of their prognostic value. This recommendation, in turn, has stimulated research and development into new technologies, and some new methods have been developed [6]. These proposed methods vary in many respects: laboratory methods used, equipment costs, assay costs, location of use (e.g., central reference laboratory, district/regional facility, point of care), time to result, and scope (e.g., adult or pediatric) [4]. However, as noted in a recent review, only modest progress has been made toward the development of an affordable technology [6], and no method has been shown to have acceptable performance for widespread use.

A major issue in evaluating the performance of new methods is that agreement between the gold standard and a novel method has been evaluated in cross-sectional laboratory-based studies [4,7,8]. Such studies often use concepts in diagnostic testing such as sensitivity and specificity applied to whether the gold standard measurement of CD4 count is above or below a given threshold, or graphical techniques such as Bland-Altman plots for measuring the agreement between measurements from two different methods [9]. However, these methods do not provide insight into the longitudinal question of how timing of treatment initiation would differ based on the use of a novel method for monitoring biomarker levels versus the use of the gold standard. In Section 2, we provide a framework for investigating this issue. Then in Section 3, through a case study, we show how increased variability or bias in a new method for measuring CD4 counts might impact timing of treatment initiation relative to that based on the gold standard. To do this, we first develop a model for CD4 decline using preexisting longitudinal data for the gold standard. We then use the model to simulate how changes in variability or bias, characteristics that we can study cross-sectionally, would impact timing of treatment initiation. This allows for an assessment of how a new technology might perform in the practically-important situation in which cross-sectional data for a new technology versus the gold standard become available. We also evaluate the potential value of biomarker monitoring algorithms that require an additional “confirmatory” measurement following an initial value below the threshold of interest. Finally, we close with a discussion in Section 4.

2. Framework for Comparing Longitudinal Agreement Between Marker Measurement Methods

Suppose that an underlying continuous time biological stochastic process of interest is characterized by a latent variable, η_{it} , where i denotes an individual who is a part of a population being monitored for progression of a specific disease and t is the time since disease acquisition or some other relevant origin. Suppose also that we can measure η_{it} using a gold standard biomarker measurement methodology, such that

$$Y_{1i}(t) = \eta_{it} + \varepsilon_{1it}(\eta_{it}),$$

where $Y_{1i}(t)$ denotes the gold standard value of the biomarker for the i th individual when measured at time t , and $\varepsilon_{1it}(\eta_{it})$ is the within-subject error which we assume to have mean zero and variance $\sigma_{\varepsilon_1}^2(\eta_{it})$ that may depend on the value of the underlying process.

Next, suppose there exists an established threshold value for gold standard biomarker measurements, c , below which clinical action is taken (the ideas extend easily if clinical action occurs when the biomarker is above c). The rationale for the choice of threshold, though application-specific, will likely be based on some assessment of the balance of risks (and perhaps financial costs) and benefits of clinical intervention at the threshold. Let the times of biomarker measurement for the i th individual be t_{ij} for $j = 1, 2, \dots$. Then, for example, clinical action might be initiated when Y_1 is first observed to be below c , and the time of such action, T_{1i} , can be defined as

$$T_{1i} = \min_j (t_{ij}) \text{ s.t. } Y_{1i}(t_{ij}) < c.$$

Motivated by our case study, we will henceforth refer to T_{1i} as the treatment start time (based on the gold standard technology). The probability of crossing the threshold by time t and, hence, of starting treatment, $P(T_{1i} \leq t)$, is impacted not only by the latent value of the biomarker and the error, but also by the frequency of measurement. This is because more frequent measurements yield more realizations of the error, and therefore more opportunities for the threshold to be crossed. More generally, T_{1i} could be defined in terms of multiple values of Y_{1i} , e.g. the time at which the mean of two successive measurements is first less than c or the time at which both an initial measurement and a confirmatory measurement are less than c .

When a novel measurement method is introduced, we can specify a model for each subject's measurements from this second technology,

$$Y_{2i}(t) = \eta_{it} + B_{2it}(\eta_{it}) + \varepsilon_{2it}(\eta_{it}),$$

where $B_{2it}(\eta_{it})$ allows for the possibility that $Y_{2i}(t)$ is a biased assessment of η_{it} with magnitude that could depend on the value of the underlying process, and $\varepsilon_{2it}(\eta_{it})$ is the within-subject error which we again assume to have mean zero and variance $\sigma_{\varepsilon_2}^2(\eta_{it})$ that may also depend on the value of the underlying process. Correlation between $\varepsilon_{1it}(\eta_{it})$ and $\varepsilon_{2it}(\eta_{it})$ can be accommodated through specification of a bivariate model. We defer discussion of this to the case study. As for the gold standard, we can define the treatment start time based on the novel method, T_{2i} as

$$T_{2i} = \min_{j'} (t_{ij'}) \text{ s.t. } Y_{2i}(t_{ij'}) < c$$

where we allow for the possibility that the sequence of measurement occasions for the novel method ($j' = 1, 2, \dots$) differs from that of the gold standard (which might arise if, for example, the novel method can be administered more frequently at a rural clinic whereas the gold standard requires a visit to a hospital-based clinic). A comparison of the performance of the two methods might then proceed by comparing the marginal distributions of T_1 and T_2 in the population or by evaluating some aspect of the distribution of gap times for individuals given by $T_{2i} - T_{1i}$. We illustrate how this might be approached in the case study.

As defined here, T_1 and T_2 represent first passage times (FPT) of two stochastic processes to a threshold of interest. Despite the large literature in this area, closed-form expressions for the probability density function (pdf) of the FPT for a discretely-sampled stochastic process are only available for limited special cases. These special cases do not include non-homogeneous stochastic processes (those whose probability distributions change when shifted in time) or dynamic measurement schedules which we study here [10]. Because of this, we proceed by conducting a simulation study which generates marker trajectories for individual subjects based on a longitudinal model for Y_1 and Y_2 , and study the corresponding treatment start times based on the gold standard and novel methods.

3. Case Study: Comparing Treatment Start Times Based on Novel and Gold Standard CD4 Enumeration Methods

Guidelines in the United States describe the CD4+ T-lymphocyte count as “the major clinical indicator of immune function in patients who have HIV infection” and “one of the most important factors in the decision to initiate antiretroviral therapy and/or prophylaxis for opportunistic infections” [3]. Though different thresholds have been used, guidelines have typically included a lower limit for CD4 counts below which it is recommended that patients begin therapy, regardless of clinical factors. The thresholds have been defined in terms of measurements from the gold standard technology, using flow cytometry. Different thresholds represent different compromises between benefits and risks of ART. Benefits include decreased risk of many undesirable outcomes including progression to AIDS or death, other potentially HIV-associated complications (e.g., cardiovascular, renal, and liver disease), and possibly HIV transmission; drawbacks include treatment-related side effects and toxicities, development of drug resistance and treatment fatigue, and transmission of drug-resistant virus [3]. Though research is ongoing on the choice of an optimal threshold, that problem is not the focus of this paper.

Instead, given a threshold, we aim to illustrate how treatment start times based on measurements taken on discrete occasions using a novel technology (hereafter referred to as “SimpleCD4”) might differ from those based on the gold standard. Ideally, this would be done using longitudinal sequences of measurements obtained for each of the two technologies over several years. However, such data are unlikely to be available for a new technology for some time. This is due in part to the fact that novel technologies need to be run on freshly drawn blood samples and cannot be run on stored samples; the same is true for flow cytometry. As cross-sectional data comparing SimpleCD4 measurements to gold standard measurements would then be the first type of data available, our approach here is to use existing longitudinal gold standard CD4 data from a study in which subjects were followed from HIV seroconversion to model gold standard CD4 trajectories over time. We then evaluate the impact on the distribution of treatment start times of increased variability (i.e., $\sigma_{\varepsilon_2}^2 > \sigma_{\varepsilon_1}^2$) and bias (i.e., $B_{2it} \neq 0$) of SimpleCD4 measurements compared to gold standard measurements. Based on these distributions, we then show the impact on the expected number of pre-treatment CD4 tests as a measure of monitoring costs and on the expected person-years on treatment as a measure of treatment costs. In doing so, we aim to help inform public health policy by illustrating the potential impact of utilizing a SimpleCD4 technology versus a gold standard technology for CD4 count monitoring.

3.1. Longitudinal Model for CD4 Count Decline

In order to compare treatment start times based on a SimpleCD4 technology to that of the gold standard, we require a longitudinal model for the natural history of CD4 count decline using available data from the gold standard. In this context, a relevant time origin ($t = 0$) is that of HIV seroconversion, when patients change from HIV antibody negative to HIV

antibody positive. To our knowledge, longitudinal CD4 counts are not available for a large cohort of seroconverters from a resource-limited country, and so we use, as an illustrative example, public-use data collected on HIV seroconverters who were a part of a study undertaken in the United States, the Multicenter AIDS Cohort Study (MACS), at sites in Baltimore, Chicago, Pittsburgh, and Los Angeles [11]. The MACS is an ongoing prospective study of the natural and treated history of HIV infection in homosexual and bisexual men. It included 3,384 men who were HIV seronegative at study entry and who were evaluated for HIV infection and, if subsequently found to be infected, for CD4 counts at semiannual visits beginning in 1984 [12]. As our focus is on HIV seroconverters, we use the date of the first seropositive HIV test result as the time origin ($t = 0$) in our model. We exclude those patients whose last negative and first positive test results were more than two scheduled visits (i.e., more than 1 year) apart. We also exclude follow-up after January 1990, the date when the drug zidovudine was introduced into clinical use in the populations included in the study, because we are interested in the natural (untreated) course of decline in CD4 count. The resulting analysis includes 330 HIV seroconverters followed for a maximum of 5 years.

There is a rich statistical literature on modeling CD4 counts over time. For a review, see Boscardin *et al.* [13]. Many of the models used can be expressed as stochastic mixed effects models,

$$Y_{1i}(t_{ij}) = \eta_{it_{ij}} + \varepsilon_{1it_{ij}} = X_i(t_{ij})\beta + Z_i(t_{ij})b_i + W_i(t_{ij}) + \varepsilon_{1it_{ij}}$$

where $Y_{1i}(t_{ij})$ is the measurement (or some transformation of it) made on the i th person at time t_{ij} , and, for our purposes, X_i and Z_i represent design vectors whose elements are restricted to functions of measurement time. $X_i(t_{ij})\beta$ then describes the population mean CD4 count (or mean transformed CD4 count) over time, $Z_i(t_{ij})b_i$ is an individual-level deterministic model of change in CD4 count over time, $W_i(t_{ij})$ is a mean zero stochastic process about this deterministic path which allows for correlation between measurements on subject i , and $\varepsilon_{1it_{ij}}$ is an error term that acts as an aggregate of measurement error and very short-term biological variation. The MACS public-use datasets do not provide exact dates of CD4 counts, so variability in timing of counts contributes to this error term. However, such variability in timing of measurements would also occur in any CD4 monitoring program. It is typically assumed that the b_i , $W_i(t_{ij})$ and $\varepsilon_{1it_{ij}}$ are mutually independent across values of i and times, t_{ij} , and that $b_i \sim N(0, R)$ where R is the covariance matrix of the random effects and $\varepsilon_{1it_{ij}} \sim N(0, \sigma_{\varepsilon_1}^2)$.

The inclusion of a stochastic process, $W_i(t_{ij})$, in the model is important in describing short-term variability in an individual's latent CD4 count, $\eta_{it_{ij}}$, about their long-term trend given by $X_i(t_{ij})\beta + Z_i(t_{ij})b_i$, reflecting the degree of "derivative tracking". If it is hypothesized that each individual's latent CD4 count, $\eta_{it_{ij}}$, follows a deterministic path over time, a model that assumes perfect derivative tracking (i.e., a standard linear mixed effects model in which $W_i(t) = 0$ for all t) would be appropriate. However, this would correspond to a situation in which immunologically weak (strong) patients continue on their initial fast (slow) rates of decline over long periods; for CD4 count trajectories, such an assumption seems untenable [14]. Taylor *et al.* [14] used an integrated Ornstein-Uhlenbeck (IOU) stochastic term, a mean-zero nonstationary Gaussian process with covariance function

$$\text{cov}(W(s), W(t)) = \frac{\sigma_w^2}{2\lambda^3} [2\lambda \min(s, t) + e^{-\lambda t} + e^{-\lambda s} - 1 - e^{-\lambda(t+s)}]$$

where σ_w^2 and λ determine the degree of derivative tracking. The IOU process is attractive because of its flexibility in accommodating a wide range of derivative tracking: small λ and small σ_w^2 result in nearly deterministic CD4 trajectories, while large values for both parameters give a process which is essentially memoryless with regard to the past rate of change [14]. The versatility of this model is further bolstered by the fact that scaled Brownian motion ($\text{cov}(W(s), W(t)) = \phi s$) is a special case of the IOU process achieved when $\lambda \rightarrow \infty$ while $\sigma_w^2/\lambda^2 = \phi$, a constant [13]. We used this model because of its flexibility and because Taylor *et al.* [14] had previously found it to be useful for modeling CD4 count trajectories using data from the MACS. Specifically, we considered Y_1 as the square root of CD4 count and, using a macro developed by Zhang *et al.* [15], found that the following model fit the data well:

$$Y_{1i}(t_{ij}) = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + W_i(t_{ij}) + \varepsilon_{1it_{ij}}, \quad (1)$$

where $W_i(t)$ is a scaled Brownian motion process such that $\text{var}[W(t)] = \sigma_w^2 t$. The parameter estimates from the fitted model are displayed in Table 1. Standard residual analyses and empirical semi-variogram analyses for longitudinal models (e.g., [16]) were used to assess the assumption of homogeneity of within-subject variance and the overall model for the covariance; they confirmed that this model fit the data well. In addition, in the next section, we show that the estimated marginal distribution of time to crossing the threshold of interest was similar to that observed in the MACS. Furthermore, we evaluated other transformations of CD4 counts which have been used in models of CD4 count trajectories in the literature, notably log [17] and fourth root [14], but found that these did not provide as good a fit. Nevertheless, it is useful to note here that when we conducted the simulation study described in the next section using either of these transformations, we found that the basic conclusions of our study were not sensitive to the choice of transformation for the response. Inclusion of a quadratic term in time also did not improve the fit of the model.

3.2. Simulation Study

We use a simulation study to investigate how timing of CD4 threshold crossing and, hence, treatment initiation, in an HIV monitoring/treatment program would change if a SimpleCD4 technology were used for monitoring compared with using the gold standard technology. We consider the setting in which monitoring begins shortly after seroconversion as would be the case if there were ongoing monitoring in the population for HIV infection. As well as studying the timing of treatment initiation, we consider how other parameters that affect the costs of programs, such as the average number of pre-treatment CD4 measurements and person-years on treatment, are impacted by the choice of measurement technology. In our discussion, we consider issues that might arise in extensions of this case study, for example to consider monitoring of a population in which some people may have been infected for some time, or to evaluate the incidence of clinical events such as HIV-related opportunistic infections that characterize AIDS or mortality.

Our approach is to simulate sequences of CD4 measurements for the gold standard and SimpleCD4 technologies and apply specific algorithms to identify the times at which treatment would be started. For the gold standard, we simulate sequences using model (1)

assuming that the true values of the parameters are the same as the estimated values shown in Table 1. We consider the WHO-recommended approach to monitoring involving two semiannual CD4 measurements per year. Thus, for semiannual measurement times $t = 0, 0.5, 1, \dots, 5$ years spanning a five-year period from the time of the first seropositive HIV test, we obtain values for $W_i(t)$ for subject i by generating sample paths of a (Gaussian) scaled Brownian motion using the estimated parameter value $\widehat{\sigma_w^2}$. Then, by generating b_{i0} and b_{i1} from a $N(0, R)$ distribution using the estimated parameter value, $\widehat{R}$, we obtain $\eta_{i0}, \dots, \eta_{i5}$ where $\eta_{it} = \beta_0 + b_{i0} + \beta_1 t + b_{i1} t + W_i(t)$. Finally, we generate independent values for ε_{1itij} from a $N(0, \sigma_{\varepsilon_1}^2)$ distribution and use $Y_{1i}(t_{ij}) = \eta_{itij} + \varepsilon_{1itij}$ to give the five-year sequence of semiannual gold standard CD4 measurements from first seropositive HIV test for subject i , $Y_{1i}(0), \dots, Y_{1i}(5)$. We repeated this process to give simulated gold standard CD4 trajectories for 50,000 subjects. The choice of a five-year period was made to avoid extrapolation of the simulated sequences beyond the range of the data supporting the parameter values, as the maximal follow-up available in the MACS data was 5 years after first seropositive HIV test.

As noted earlier, longitudinal data from seroconversion for a SimpleCD4 technology are not available; we therefore generate SimpleCD4 measurements under two specific scenarios in which model (1) also applies for SimpleCD4 with the only changes being (a) that $\sigma_{\varepsilon_2}^2$ is inflated by a factor of δ relative to $\sigma_{\varepsilon_1}^2$; or (b) that bias $B(\eta_{it})$ is added. We return to (b) later and focus first on (a). To generate a SimpleCD4 sequence for each of the 50,000 subjects for whom we have generated the gold standard sequence $Y_{1i}(0), \dots, Y_{1i}(5)$, we use the same $\eta_{i0}, \dots, \eta_{i5}$ as described above. To generate the ε_{2it} we need to consider whether ε_{1it} and ε_{2it} might be correlated. As previously mentioned, these “error” terms reflect very short-term biological variability and variability in such factors as sample shipping conditions, laboratories, laboratory environment and personnel, and technical measurement error. Indeed, as alternative CD4 enumeration methods can impact when and where a blood sample is obtained from a patient (e.g., at a patient's home vs. at a specialized clinic), it is not unreasonable to assume that all of these factors could change for SimpleCD4. In practice, we can imagine two possible extremes: (i) where a patient's blood sample is shipped to a specialized laboratory and either technology could be applied by the same staff at any given time so that any difference between ε_{1it} and ε_{2it} might be dominated by differences in measurement error of the equipment used for each technology, and (ii) where the gold standard is as in (i) but SimpleCD4 is a “point of care” technology administered by a nurse at a patient's home so that there is no shipping and the time of measurement during the day might also differ from that for a gold standard measurement. For (i), the correlation of ε_{1it} and ε_{2it} might be very high, whereas, for (ii), the correlation might be zero or very close to zero. We consider the case of zero correlation as this would be the situation in which treatment start times would differ most between the gold standard and SimpleCD4 within each individual patient. Note, however, that the choice of correlation does not affect the marginal distributions of outcomes such as number of pre-treatment CD4 measurements or person-years on treatment and, hence, any comparison of the technologies based on these marginal distributions.

Therefore, corresponding to each $Y_{1i}(t_{ij})$, we generate a value $Y_{2i}(t_{ij}) = \eta_{itij} + \varepsilon_{2itij}$ where $\varepsilon_{2itij} \sim N(0, \sigma_{\varepsilon_1}^2 \times \delta)$, so that δ measures the proportionate increase in the error variance, assuming that ε_{1itij} and ε_{2itij} are independent. We consider values for δ of 1, 1.1, 1.25, 1.5, 2, and 5. To define the time of treatment initiation, we have chosen a CD4 threshold of 500 cells/mm³. This threshold was chosen for two reasons. First and most important, the current trend is toward treatment initiation at higher CD4 counts as evidenced by recent changes in treatment guidelines [2,3] and recent evidence that starting treatment using a threshold of

500 cells/mm³ would lead to increased life expectancy and quality-adjusted life years [18,19]. Second, for the purposes of an illustrative case study, a higher threshold leads to fewer censored values for T_1 and T_2 during the 5 year period after first seropositive HIV test which we simulated. This helps clarify some of the issues that we present though results using lower thresholds showed similar patterns, albeit in the presence of more censoring.

3.3. Results

In developing a model for CD4 count trajectories, we are interested in checking not only that the assumptions of the stochastic mixed effects model are reasonably satisfied (see Section 3.1), but also that the selected model provides a good fit to the distribution of threshold crossing times observed in the MACS data. Figure 1 shows a plot of the estimated cumulative probability of threshold crossing until time t based upon the simulated data in comparison to the corresponding Kaplan-Meier estimate based on the data from the MACS and indicates a good fit. Note that in obtaining the Kaplan-Meier estimate, we censored follow-up of a subject prior to their first missing scheduled CD4 measurement. Doing so provides a valid estimate if missingness is completely at random. We evaluated this following the methods described in Diggle [20] and Ridout [21], specifically evaluating whether there was evidence of an association between CD4 count at any visit prior to threshold crossing and subsequent presence or absence of the next scheduled measurement. We did not find a significant association despite the fact that nearly 40% of subjects had a missing measurement before a count less than 500 was obtained. The lack of an association is perhaps not surprising as nearly all subjects with counts above 500 would have asymptomatic HIV infection.

To illustrate how treatment starting times based on a SimpleCD4 technology might compare to those of the gold standard for individual patients, Figure 2 shows the simulated CD4 trajectories for three patients. For each, the “latent CD4” line represents an approximation to the patient's latent CD4 trajectory obtained by extending the simulation to generate 12 monthly values of η_{it} each year. The simulated 6-monthly gold standard CD4 measurements are shown through to the time that the first value was below the threshold of 500 cells/mm³. The simulated SimpleCD4 measurements in the figure were generated for the situations in which $\delta = 2$. The line joining the SimpleCD4 measurements is also discontinued at the time of the first SimpleCD4 measurement below 500 cells/mm³, but is followed by a dashed line which represents a “dual confirmatory” treatment initiation protocol which required a second measurement to be taken one month after a SimpleCD4 measurement is identified as being below the threshold. Under such a strategy, treatment is initiated if the confirmatory measurement is also below the threshold. Otherwise, the schedule of two measurements per year is restarted, with a confirmatory value taken after any subsequent measurement below the threshold. Finally, the asterisk represents the treatment start time under a “mean confirmatory” protocol, whereby an initial measurement below the threshold also leads to a confirmatory measurement being obtained and the mean of the initial and associated confirmatory measurements needs to be below the threshold in order to start treatment.

Figure 2(a) shows a case where treatment would be initiated at the same time irrespective of whether monitoring was based on the gold standard or SimpleCD4 with no confirmatory value. In contrast, requiring a confirmatory value below the threshold would have substantially delayed treatment. In this particular example, treatment would have been initiated 1 year 7 months later using the dual confirmatory protocol compared with using the gold standard, by which time the patient would have had a latent CD4 count of about 300 cells/mm³. The mean confirmatory protocol does not share this problem in this case, as it only delays treatment for the additional month required for the second test. Figure 2(b), on the other hand, illustrates a case where treatment would be initiated too early using SimpleCD4 with no confirmatory value (2 years after first seropositive HIV test using

SimpleCD4 compared to 3 years using the gold standard). Requiring a confirmatory value would have reduced the difference in timing of treatment initiation, as it would have delayed treatment until one month after the gold standard starting time if either the mean or dual confirmatory protocol had been used. Figure 2(c) illustrates the case of a slower progressor whose latent CD4 count stays above the threshold for the initial five years after first seropositive HIV test. This example shows that even the gold standard measures the outcome with error, as treatment would be initiated at the time of first seropositive HIV test using the best available technology. Treatment initiation would be delayed until 2 years 6 months and 2 years 7 months based on the SimpleCD4 no confirmatory and mean confirmatory protocols, respectively. The dual confirmatory protocol would further delay treatment until 5 years 1 month after first seropositive HIV test.

3.3.1. The Effect of Increased Variance on Treatment Starting Times—We now turn to examining the effect of increased variance for SimpleCD4 on treatment starting times at the population level. Figure 3 shows the cumulative simulated probabilities of meeting the criteria to start treatment at the CD4 measurement occasions after first seropositive HIV test for the gold standard, and for the SimpleCD4 technology with no confirmatory value when $\delta = 2$ or 5. The cumulative probability of starting treatment at any given time t increases with δ . For example, 19% of patients would initiate treatment at the time of their first seropositive HIV test using the gold standard (equivalent to SimpleCD4 with $\delta = 1$) compared to 22% for SimpleCD4 when $\delta = 2$ and 29% when $\delta = 5$. The median treatment starting time also reflects this trend: it is 18 months for the gold standard, but 12 months for SimpleCD4 when $\delta = 2$ or 5. By 5 years after first seropositive HIV test, 13% of patients would not have started treatment based on the gold standard while 9% and 4% would not have started based on SimpleCD4 when $\delta = 2$ and $\delta = 5$, respectively.

Table 2 provides summary statistics for the marginal distribution of treatment starting times when no confirmatory measurement is required for $\delta = 2$ and $\delta = 5$, as well as for smaller values of δ . To illustrate how use of a SimpleCD4 technology might affect monitoring and treatment costs of a public health program, the table also shows the mean number of pre-treatment CD4 measurements during the first 5 years after first seropositive HIV test as a measure of monitoring costs, and the mean person-years on treatment during the same period as a measure of treatment costs. These measures are restricted to what happens during the first 5 years after first seropositive HIV test to avoid bias that might arise if the model based on data for 5 years of follow-up is not valid over longer periods of time. Although this is a limitation of the simulation study, it does not affect the basic conclusions that might be drawn from the case study, as the large majority (87%) of patients reach the threshold within 5 years when the gold standard is used. As δ increases, the mean number of pre-treatment CD4 measurements decreases from 5.1 per patient during the first 5 years after first seropositive HIV test for the gold standard to 4.7, 4.4 and 3.4 when $\delta = 1.5, 2$ and 5, respectively. Conversely, as more subjects start treatment earlier with increasing δ , the mean time on treatment during the 5 years increases from 3.1 person-years for the gold standard to 3.3, 3.4 and 3.9 person-years, respectively. Because risk of imminent clinical manifestations of HIV infection would be associated with the underlying latent CD4 count, it is useful also to evaluate the proportion of subjects who would start treatment with latent CD4 counts which might be considered much too low. For this purpose, we use a threshold of 350 cells/mm³ as this level is used in some treatment management guidelines, recognizing though that such guidelines are generally written based on gold standard measurements rather than underlying latent values. Interestingly, using this threshold, among patients who would start treatment, the percentage starting with latent CD4 count < 350 cells/mm³ is reasonably constant at just below 6% across the values of δ studied. Obviously this has the caveat that the proportion of patients who would not have started treatment by 5 years is higher for the gold standard technology and decreases for the SimpleCD4 technology with higher δ .

However, considering all patients, the percentage who either would have started treatment within 5 years with latent CD4 count < 350 cells/mm³ or who had not started but had a latent CD4 count < 350 cells/mm³ at 5 years was also similar at just over 5% across the values of δ studied. Thus, while treatment costs would be increased because patients start treatment earlier on average when the SimpleCD4 technology is used, in this specific application, the percentage of patients who start treatment with latent CD4 counts which might be considered unacceptably low would not be appreciably affected based on what can be ascertained during the 5 years simulated.

3.3.2. The Utility of Confirmatory Measurements—We explored the utility of SimpleCD4-based treatment initiation protocols which make use of a confirmatory measurement (as in the examples in Figure 2), as their use would lead to longer waiting times until treatment initiation, which might compensate for the reduced times induced by increased variance for a SimpleCD4 technology.

As previously mentioned, we considered monitoring protocols which require a confirmatory value taken one month after a CD4 measurement is identified as being below the threshold. The choice of a one-month interval is somewhat arbitrary but in practice needs to be long enough to allow time for testing of the initial sample and reaching a patient for retesting, and long enough that the successive realizations of ε can be considered independent. It is also necessary to assume that a model fitted using semi-annual measurements remains appropriate for monthly measurements, specifically that $W_i(t)$ appropriately captures the short-term variability that occurs from month to month, whereas ε captures the biological variability that occurs over shorter time intervals, e.g. a few days.

Table 2 shows results for the dual confirmatory and mean confirmatory monitoring algorithms. From the table, we see that adding a requirement for a confirmatory measurement delays treatment initiation. For example, the median treatment starting time for a SimpleCD4 technology with $\delta = 2$ is 12 months when no confirmatory value is required but 25 and 31 months for the mean and dual confirmatory protocols, respectively; this contrasts with 18 months for the gold standard. Compared with using a single measurement of the gold standard, this delay leads to an increased number of pre-treatment CD4 tests (due in part to the need for confirmatory measurements) and fewer person-years on treatment. For example, whereas monitoring using the gold standard requires, on average, 5.1 pre-treatment CD4 measurements per patient and 3.1 person-years on treatment during the 5-year period simulated, a dual confirmatory protocol when $\delta = 2$ requires 7.8 CD4 measurements and 2.4 person-years on treatment on average. The mean confirmatory protocol is intermediate between these two extremes. A general conclusion in this case study is that the use of a confirmatory measurement seems to overcompensate for increased variability in SimpleCD4 measurements compared with monitoring based upon semiannual gold standard measurements with no confirmatory value. Indeed, δ needs to be about 5 before use of the mean confirmatory algorithm gives similar distributions of treatment starting times and person-years on treatment. However, this comes with an increase in monitoring costs as measured by the mean number of pre-treatment CD4 tests per person during the period simulated (6.4 vs. 5.1) and an increase in the proportion of patients beginning treatment with latent CD4 count < 350 cells/mm³ (12.3% vs. 5.8%, with 87% of patients starting treatment in this period for both technology/monitoring protocol combinations). Even with such an inflated error variance ($\delta = 5$), it is clear from Table 2 that the dual confirmatory protocol delays treatment initiation compared with using a single value of the gold standard.

3.3.3. Gap Times—To this point, we have evaluated the marginal distributions of treatment start times when using a SimpleCD4 technology compared with the gold standard. These results are particularly informative for understanding the potential implications for

population-based public health programs. It may also be important to evaluate other aspects of the joint distribution of treatment start times for the two technologies as this provides information about how individual subjects within the population might be affected if a SimpleCD4 technology is used rather than the gold standard. One aspect of interest might be the distribution of the differences, or gaps, in treatment start times given by $T_{2i} - T_{1i}$ among subjects in the population. Unlike a comparison of technologies based on the marginal distributions of T_1 and T_2 , the correlation of the errors ε_{1it} and ε_{2it} does affect the distribution of gap times. We present results for the same simulation study in which ε_{1it} and ε_{2it} were assumed to be independent, which, as noted earlier, would tend to increase the gap times compared with when these errors are highly correlated.

Table 3 summarizes the distribution of gap times from the same simulation study. One issue is that the gap times cannot be calculated from the simulation study for subjects for whom the treatment start time for one or both of the SimpleCD4 and gold standard technologies is beyond 5 years; that said, if the treatment start time is known for one of the two thresholds, then the gap time is censored and so there is some information about its possible value. Thus the percentages shown for specific ranges of gap times are lower bounds on the percentages that would be obtained had it been possible to simulate the complete trajectory of CD4 counts for all subjects. Under the no confirmatory protocol with $\delta = 1$, we can assess the gap times between the gold standard and a SimpleCD4 technology that is as accurate as that gold standard but has independent errors (i.e. ε_{1it} and ε_{2it} are independent). As would be expected in this case, the distribution of gap times is symmetric, with (a lower bound of) 50% of subjects having a gap time of no more than 6 months (i.e., one interval between scheduled measurements); this includes 23% of subjects who had the same treatment start times for the two technologies. A further 12% of subjects would start 12 to 18 months earlier when using the SimpleCD4 technology and 12% would start 12 to 18 months later. Similarly, a further 6% would start 24 months or more earlier and 6% would start 24 months or more later.

Thus, even when the two technologies have the same variability, the magnitude of σ_ε^2 induces reasonably varied treatment starting times. In this context, although there is a shift toward more negative gap times with increasing δ , the changes in the percentage of subjects in each category of gap time is modest for values of δ up to 2. When $\delta = 5$, the major effect is to shift the percentage of subjects starting at least 12 months early to 40% from 18% when $\delta = 1$ and 27% when $\delta = 2$. If the confirmatory protocols are used, we notice a modest decrease in percentage of subjects having a treatment start time within 7 months of that for the gold standard and a clear increase in the percentage of subjects with delays of 13 to 19 months and of 25 months or more (with the caveat that the percentage of subjects with unknown gap times is also increased).

3.3.4. The Effect of Bias on Treatment Starting Times—In addition to the possibility of increased variability, a novel SimpleCD4 technology might measure CD4 counts with bias. To assess how this might affect times of treatment initiation, we explore the case of bias that is systematically a constant 10% above or 10% below the latent CD4 count. The results are displayed in Table 4 for when the variability of the SimpleCD4 technology is the same as the gold standard (i.e., $\delta = 1$) and when it is increased ($\delta = 2$). Considering first the case of $\delta = 1$ and the no confirmatory measurement protocol, if the SimpleCD4 technology has a 10% negative bias so that the SimpleCD4 count underestimates the latent CD4 count on average, then the median treatment start time is 12 months instead of 18 months for an unbiased SimpleCD4 technology. Conversely, if the SimpleCD4 technology has a +10% bias, the median treatment start time is increased by 6 months to 24 months. Similar 6-month changes in the median were seen for the mean and dual confirmatory measurement protocols when $\delta = 1$ and also for the no confirmatory and confirmatory protocols when $\delta = 2$ (except the median was not changed for the no

confirmatory protocol with -10% bias). For the measures of potential monitoring/treatment program costs, negative bias leads to a decrease in the mean number of pre-treatment CD4 evaluations and an increase in the mean person-years of treatment during the 5 years simulated. Conversely, positive bias has the opposite effect on each of these two parameters. A consequence of this is that the effects of positive bias and increased variability on these parameters tend to be in opposite directions; for example, the values of the various parameters shown are similar for the case of +10% bias and $\delta = 2$ to those for the case of no bias and no increase in variability (i.e., $\delta = 1$). However, in this study, the effects of bias of the order of $\pm 10\%$ on the percentage of subjects having latent CD4 counts below 350 cells/mm³ at the time of starting treatment is greater than the effects of increases in variability of the order $\delta = 2$. Indeed, when there is +10% bias, the percentages of subjects first reaching the criteria for starting treatment with a latent CD4 count below 350 cells/mm³ are about 20% and 29% for the mean and dual confirmatory measurement protocols, respectively, (for both $\delta = 1$ and 2) whereas it is just under 6% for the no confirmatory measurement protocol when there is no bias (again for both $\delta = 1$ and 2).

4. Discussion

We have described an approach for evaluating the implications of increased variability and/or bias in a new, less expensive (or simpler) technology compared with a gold standard technology on the time taken to observe a measurement (or measurements, if confirmatory values are required) below a defined threshold at which treatment is initiated. In the context of HIV research and ultimately public health policy in resource-limited settings, the problem is of critical current importance. Our case study provides an approach for assessing the potential impact on treatment starting times in a population if longitudinal data about the natural history of CD4 measurements obtained with a gold standard technology are available and cross-sectional data concerning the variability and bias of measurements obtained using a new technology versus the gold standard become available. It is important to appreciate that the latter difference in variability needs to reflect not only pure measurement error (e.g., within a laboratory) but other conditions that may change with the use of new technology, e.g. due to differences in the timing of blood sampling, specimen processing, etc.

Increased variability leads, on average, to earlier threshold crossing times. In our illustrative case study, using a threshold of 500 cells/mm³ for starting treatment, the median treatment starting time was reduced from 18 months to 12 months if variability of the SimpleCD4 technology was double that of the gold standard. From a public health perspective, the major impact of these changes is to increase the expected costs of a treatment program. This was relatively small in our case study: the mean person-years on treatment during the period simulated increased by about 10%. Of note, we found that the use of confirmatory measurements in a SimpleCD4 monitoring protocol might lead to deferral of treatment relative to a protocol based on single gold standard measurements such that an unacceptable proportion of patients might start treatment when the risk of AIDS events is much higher (e.g. when latent CD4 counts are < 350 cells/mm³). We also found that modest amounts of bias (e.g. if SimpleCD4 measurements overestimate the latent CD4 count by 10%) can lead to a marked increase in the proportion of patients starting treatment when risk of AIDS events is much higher. Such considerations should guide the design of cross-sectional studies that evaluate new technologies.

Our case study focused on simulating a population of patients followed from close to the time of infection. In practice, patients will present to an HIV monitoring program at varying times from infection and hence with various levels of immunodeficiency. In such a setting, the impact of inflated variance and bias from using a less expensive technology might be diminished compared with the results of our case study. This is because those patients who

present with underlying CD4 counts well below the threshold would likely be correctly classified as requiring treatment despite the use of a more variable or biased technology. In contrast, if the population predominantly includes patients with underlying CD4 counts above the threshold and this population is enriched with patients with more slowly declining CD4 counts than in our seroincident population, then increased variability or bias in a new technology might increase the difference in timing of treatment initiation versus the gold standard technology. Thus, simulation study of possible seroprevalent populations would be useful. With cross-sectional data about CD4 counts when patients first present in a given population, the simulation study could be adapted to be based on this distribution of presentation times. Longitudinal data would also be valuable to address the possibility that patients in the population with CD4 counts above the threshold might be enriched with slower progressors.

The simulation study might also be modified to evaluate alternative monitoring algorithms, for example one that uses a SimpleCD4 technology as an initial monitoring tool but requires a confirmatory gold standard measurement when the SimpleCD4 technology gives a value below the threshold, or one that leads to a switch in monitoring from the SimpleCD4 technology to the gold standard technology when a SimpleCD4 measurement is first obtained below some higher threshold (such as 650 cells/mm³ in our case study). In addition, it might be important to assess the impact of missing measurements in monitoring patients. As noted earlier, this occurred frequently in the MACS study and would likely be an issue in monitoring programs in practice. Assuming that the decision to initiate treatment depends only on an observed count below the threshold of interest, missingness would delay treatment initiation relative to what we found in our simulation study. How it would affect the difference in distribution of treatment start times for different technologies would, however, also depend on whether missingness patterns might differ between technologies. This could occur, for example, if the location of testing varies between technologies. This issue is further complicated by the fact that the desirability of starting treatment during asymptomatic infection might be reduced if patients who miss scheduled monitoring visits are more likely to not fully adhere to treatment, thereby increasing risk of viral resistance to drugs.

There are some interesting statistical problems that require thought in extending some of the ideas that we have presented. We evaluated the proportion of subjects who started treatment with low latent CD4 counts (specifically < 350 cells/mm³) where risk of developing an AIDS event is more markedly increased. A useful extension would be to directly assess the consequences of using a SimpleCD4 technology on the incidence of AIDS events. From a practical perspective, this would require a model for the short-term risk of an AIDS event over the interval between measurements, as well as a model for the reduction in risk of such events after treatment is initiated. We did not pursue this using the MACS data as the number of AIDS events observed at the higher CD4 counts of interest was small. There are other studies that could provide relevant data; for example, one study [22] used a Poisson regression model to predict six-month risk of AIDS with the observed CD4 count measured using the gold standard technology. However, in the simulation study, it would be important to use a model for predicting AIDS events which has the underlying latent CD4 count at the beginning of the interval as the predictor covariate rather than the gold standard measurement. This is necessary to avoid a biased assessment of risk and, hence, a biased comparison of the gold standard and novel technologies due to the well-known problem of bias when covariates in a regression model are measured with error.

A second problem concerns what natural history data might be available. We used data from a study that was initiated prior to the availability of antiretroviral therapy. For a natural history study conducted today, follow-up would generally be censored when a gold standard

measurement is obtained that is below the threshold suggested in current treatment management guidelines because of an ethical obligation to initiate treatment. This obviously would limit the ability to model the natural history of gold standard measurements below that threshold. Although this does not impact the ability to estimate the distribution of the treatment initiation times, T_1 , for the gold standard technology, it does impact the ability to do so for the corresponding time, T_2 , for the SimpleCD4 technology. Such a data generating process, called semicompeting risks data [23], would also arise in the ideal situation in which a prospective longitudinal study is designed to obtain measurements on both the gold standard and SimpleCD4 technologies in a real-life comparison of the impact on treatment initiation times.

In conclusion, we have described a framework for assessing agreement in the timing of treatment initiation as determined by the longitudinal monitoring of measurements obtained using a novel technology versus a gold standard technology. In particular, we focused on understanding the implications on timing of increased variability and bias of the novel technology versus the gold standard technology. This provides a good basis for evaluating the potential performance of a new technology early in its development using data from cross-sectional studies. The framework should be valuable not only in the context of monitoring when to start antiretroviral treatment for HIV infection but also more broadly in other diseases.

Acknowledgments

Contract/grant sponsor: This work was supported in part by grant numbers AI 24643 and AI 68634 from the United States National Institutes of Health.

References

1. WHO. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. WHO; Geneva, Switzerland: August 72006 [December 3, 2009]. Available at <http://www.who.int/entity/hiv/pub/guidelines/artadultguidelines.pdf>
2. WHO. Rapid advice: antiretroviral therapy for HIV infection in adults and adolescents. WHO; Geneva, Switzerland: November 302009 [December 3, 2009]. Available at http://www.who.int/hiv/pub/arv/rapid_advice_art.pdf
3. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services; Washington DC, USA: December 12009 [December 3, 2009]. Available at <http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf>
4. Rodriguez W, Christodoulides N, Floriano P, Graham S, Mohanty S, Dixon M, Hsiang M, Peter T, Zavahir S, Thior I, et al. A microchip CD4 counting method for HIV monitoring in resource-poor settings. *PLoS Medicine*. 2005; 2(7):e182. [PubMed: 16013921]
5. Forum for Collaborative HIV Research. Transfer of HIV monitoring technologies into resource-poor settings: moving the field forward. February 262005 [April 10, 2009]. Available at <http://www.hivforum.org/storage/hivforum/documents/Febeport.pdf>
6. Mandy F, Janossy G, Bergeron M, Pilon R, Faucher S. Affordable CD4 T-Cell enumeration for resource-limited regions: a status report for 2008. *Cytometry Part B (Clinical Cytometry)*. 2008; 74(1):S27–S39.
7. Spacek L, Shihab H, Lutwama F, Summerton J, Mayanja H, Ronald A, Margolick J, Nilles T, Quinn T. Evaluation of a low-cost method, the Guava EasyCD4 Assay, to enumerate CD4-positive lymphocyte counts in HIV-infected patients in the United States and Uganda. *Journal of Acquired Immune Deficiency Syndromes*. 2006; 41(5):607–610. [PubMed: 16652034]
8. Srithanaviboonchai K, Rungruengthanakit K, Nouanthong P, Pata S, Sirisanthana T, Kasinrerak W. Novel low-cost assay for the monitoring of CD4 counts in HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes*. 2008; 47(2):135–139. [PubMed: 18091046]

9. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1(8476):307–310. [PubMed: 2868172]
10. Madec Y, Japhet C. First passage time problem for a drifted Ornstein-Uhlenbeck process. *Mathematical Biosciences*. 2004; 189(2):131–140. [PubMed: 15094316]
11. Multicenter AIDS Cohort Study (MACS). Public Dataset: Release PO4. Springfield, VA: National Technical Information Service; 1995 [September 14, 2009]. Available at <http://www.statepi.jhsph.edu/macspdt.html>
12. Kaslow R, Ostrow D, Detels R, Phair J, Polk B, Rinaldo C Jr. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*. 1987; 126(2):310–8. [PubMed: 3300281]
13. Boscardin W, Taylor J, Law N. Longitudinal models for AIDS marker data. *Statistical Methods in Medical Research*. 1998; 7(1):13–27. [PubMed: 9533259]
14. Taylor J, Cumberland W, Sy J. A stochastic model for analysis of longitudinal AIDS data. *Journal of the American Statistical Association*. 1994; 89(427):727–736.
15. Zhang D, Lin X, Raz J, Sowers M. Semiparametric Stochastic Mixed Models for Longitudinal Data. *Journal of the American Statistical Association*. 1998; 93(442):710–719.
16. Fitzmaurice, G.; Laird, N.; Ware, J. *Applied Longitudinal Analysis*. Wiley-Interscience; 2004.
17. Berman S. A stochastic model for the distribution of HIV latency time based on T4 counts. *Biometrika*. 1990; 77(4):733–741.
18. Braithwaite R, Roberts M, Chang C, Goetz M, Gibert C, Rodriguez-Barradas M, Shechter S, Schaefer A, Nucifora K, Koppenhaver R, et al. Influence of alternative thresholds for initiating HIV treatment on quality-adjusted life expectancy: a decision model. *Annals of Internal Medicine*. 2008; 148(3):178–185. [PubMed: 18252681]
19. Kitahata M, Gange S, Abraham A, Merriman B, Saag M, Justice A, Hogg R, Deeks S, Eron J, Brooks J, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. *New England Journal of Medicine*. 2009; 360:1–12. [PubMed: 19118299]
20. Diggle P. Testing for random dropouts in repeated measurement data. *Biometrics*. 1989; 45(4): 1255–1258.
21. Ridout M. Reader reaction: Testing for random dropouts in repeated measurement data. *Biometrics*. 1991; 47(4):1617–1621. [PubMed: 1786335]
22. Phillips A, Pezzotti P. Short-term risk of AIDS according to current CD4 cell count and viral load in antiretroviral drug-naïve individuals and those treated in the monotherapy era. *AIDS*. 2004; 18(1):51–58. [PubMed: 15090829]
23. Fine J, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88(4):907–919.

Figure 1. Cumulative probability of threshold crossing in simulated data compared with the corresponding Kaplan-Meier estimate (with 95% confidence bands) based on data from the MACS.

Figure 2. Simulated gold standard, SimpleCD4 ($\delta = 2$), and latent CD4 trajectories for three randomly-chosen patients showing the corresponding treatment starting times.

Figure 3.

Bar chart of the cumulative probability of threshold crossing time (and hence starting treatment) estimated from simulated data for gold standard and SimpleCD4 technologies when $\delta = 2$ or 5.

Table 1

Estimates for the parameters in the model $Y_{1i}(t_{ij}) = \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + W_i(t_{ij}) + \varepsilon_{1it_{ij}}$, where Y_{1i} is the square-root of CD4 count measured by the gold standard and t is the time (years) from first seropositive HIV test. In the MACS data used to fit this model, times ranged up to 5 years.

Intercept	Linear	Brownian motion	Error	Random effects
β_0 (SE)	β_1 (SE)	σ_w^2 (SE)	$\sigma_{\varepsilon 1}^2$ (SE)	$R = \begin{bmatrix} \text{Var}(b_{i0}) & \text{Cov}(b_{i0}, b_{i1}) \\ \text{Cov}(b_{i0}, b_{i1}) & \text{Var}(b_{i1}) \end{bmatrix}$
27.0303 (0.2918)	-2.0821 (0.1320)	1.0976 (0.5034)	10.0937 (0.5377)	$\begin{bmatrix} 17.3047 & -0.3546 \\ -0.3546 & 2.2293 \end{bmatrix}$

SE: Standard error

Table 2

Operating characteristics of different HIV monitoring protocols to determine when to start treatment based on SimpleCD4 technologies having various values of δ compared to the gold standard. Results are based on simulations of CD4 trajectories for 50,000 patients measured semiannually for 61 months (5.08 years) after first seropositive HIV test.

Technology/Monitoring protocol	δ	Quantiles of treatment starting times (months)				% starting treatment after 61 months	Mean number over 61 months			% of qualified subjects starting treatment with latent CD4 < 350 ^c
		10%	50%	90%	IQR ^a		Pre-treatment CD4 measurements	(SE) ^b	Person-years on treatment	(SE) ^b
Gold standard	1 ^d	0	18	>60	30	13	5.1	(.017)	3.1	(.008)
	1.1	0	18	>60	30	13	5.0	(.016)	3.2	(.008)
	1.25	0	18	>60	30	12	4.9	(.016)	3.2	(.008)
	1.5	0	18	>60	30	11	4.7	(.016)	3.3	(.007)
	2	0	12	54	24	9	4.4	(.015)	3.4	(.007)
SimpleCD4/No confirmatory	5	0	12	36	24	4	3.4	(.012)	3.9	(.006)
	1	1	25	>61	36	22	7.1	(.016)	2.6	(.008)
	1.1	1	25	>61	36	21	7.0	(.016)	2.6	(.008)
	1.25	1	25	>61	36	21	7.0	(.016)	2.6	(.008)
	1.5	1	25	>61	36	20	7.0	(.016)	2.6	(.008)
SimpleCD4/Mean confirmatory	2	1	25	>61	36	18	6.9	(.016)	2.7	(.008)
	5	1	19	>61	30	13	6.4	(.016)	3.1	(.007)
	1	7	31	>61	48	25	7.8	(.016)	2.3	(.008)
	1.1	7	31	>61	42	25	7.8	(.016)	2.3	(.008)
	1.25	7	31	>61	42	25	7.9	(.016)	2.3	(.008)
SimpleCD4/Dual confirmatory	1.5	7	31	>61	42	24	7.9	(.017)	2.4	(.008)
	2	1	31	>61	42	23	7.8	(.017)	2.4	(.008)
	5	1	25	>61	36	18	7.6	(.018)	2.6	(.008)
	1	7	31	>61	48	25	7.8	(.016)	2.3	(.008)
	1.1	7	31	>61	42	25	7.8	(.016)	2.3	(.008)
	1.25	7	31	>61	42	25	7.9	(.016)	2.3	(.008)
	1.5	7	31	>61	42	24	7.9	(.017)	2.4	(.008)
	2	1	31	>61	42	23	7.8	(.017)	2.4	(.008)
	5	1	25	>61	36	18	7.6	(.018)	2.6	(.008)

^aIQR = Interquartile range.

^bSE = Standard error.

^cCalculated only for those who met the criteria for starting treatment (with an initial qualifying measurement for confirmatory protocols) by 61 months.

^dThe marginal distribution for SimpleCD4 with $\delta = 1$ is the same as for gold standard.

Distribution of “gap times” (differences in treatment starting times within a subject) for HIV monitoring protocols based on SimpleCD4 technologies having various values of δ compared to the gold standard. Results are based on simulations of CD4 trajectories for 50,000 patients measured semiannually for 61 months (5.08 years) after first HIV seropositive test.

Table 3

SimpleCD4 protocol		% starting treatment $\times$ months later than gold standard						
δ	$x \leq -23$	$-18 \leq x \leq -11$	$-6 \leq x \leq 7$	$12 \leq x \leq 19$	$x \geq 24$	unknown ^d		
No confirmatory	1	6	12	50	12	6	13	
	1.1	7	13	50	12	6	12	
	1.25	8	13	49	12	6	12	
	1.5	10	14	48	11	5	11	
	2	12	15	47	10	5	10	
Mean confirmatory	5	22	18	44	8	3	5	
	1	2	7	46	16	12	17	
	1.1	2	7	46	16	12	17	
	1.25	2	8	45	16	12	17	
	1.5	3	8	45	16	11	17	
Dual confirmatory	2	4	9	45	15	11	16	
	5	10	12	43	15	8	12	
	1	1	5	40	19	16	19	
	1.1	1	5	41	19	16	18	
	1.25	1	5	40	19	17	18	
Dual confirmatory	1.5	2	5	40	19	17	18	
	2	3	6	39	19	16	18	
	5	6	9	38	18	15	15	

^aThe gap time is unknown for those subjects who cross the threshold by either SimpleCD4 or the gold standard after 61 months and whose gap time is not known to be 23 months or more in absolute value.

Table 4

Operating characteristics of different HIV monitoring protocols to determine when to start treatment based on SimpleCD4 technologies having various values of δ when bias may be present, compared to the gold standard. Results are based on simulations of CD4 trajectories for 50,000 patients measured semiannually for 61 months (5.08 years) after first seropositive HIV test.

SimpleCD4 protocol	δ	bias % ^a	Quantiles of treatment starting times (months)				% starting treatment after 61 months	Mean number over 61 months			% of qualified subjects starting treatment with latent CD4 < 350 ^d
			10%	50%	90%	IQR ^b		Pre-treatment CD4 measurements	SE) ^c	Person-years on treatment	
No confirmatory	1	-10	0	12	60	30	9	4.2	(.015)	3.5	3.6
		0	0	18	>60	30	13	5.1	(.017)	3.1	5.7
		+10	0	24	>60	36	17	5.8	(.017)	2.7	9.5
	2	-10	0	12	42	24	6	3.7	(.014)	3.8	4.1
		0	0	12	54	24	8	4.4	(.015)	3.4	5.9
		+10	0	18	>60	30	12	5.1	(.016)	3.1	8.8
Mean confirmatory	1	-10	1	19	>61	36	16	6.3	(.016)	3.0	5.7
		0	1	25	>61	36	22	7.1	(.016)	2.6	10.8
		+10	7	31	>61	>42	27	7.7	(.015)	2.2	19.2
	2	-10	1	19	>61	30	13	6.1	(.016)	3.1	7.2
		0	1	25	>61	36	18	6.8	(.016)	2.7	11.9
		+10	1	31	>61	42	24	7.5	(.016)	2.4	19.5
Dual confirmatory	1	-10	1	25	>61	42	19	7.1	(.017)	2.7	8.7
		0	7	31	>61	48	25	7.8	(.016)	2.3	17.0
		+10	7	37	>61	>42	30	8.4	(.015)	2.0	28.8
	2	-10	1	25	>61	36	17	7.1	(.017)	2.8	11.5
		0	1	31	>61	42	22	7.8	(.017)	2.4	19.5
		+10	7	37	>61	>42	28	8.4	(.016)	2.1	29.4

^aBias is a percentage of latent CD4 count.

^bIQR = Interquartile range.

^cSE = Standard error.

d Calculated only for those who met the criteria for starting treatment (with an initial qualifying measurement for confirmatory protocols) by 61 months.