Nonparametric Analysis of Immune System Markers in the HIV Epidemic

Was there a change in the AIDS epidemic in 1987?

By

Kjell Doksum and Sharon-Lise Normand

Department of Statistics, University of California, Berkeley
Department of Health Care Policy, Harvard Medical School, Boston

Nonparametric Analysis of Immune System Markers in the HIV Epidemic. Was there a change in the AIDS epidemic in 1987?

Kjell Doksum and Sharon-Lise Normand

Department of Statistics, University of California, Berkeley
Department of Health Care Policy, Harvard Medical School, Boston

Abstract. In this paper we analyse CD4, CD8 and CD4/CD8 ratios of HIV infected individuals using nonlinear nonparametric models and methods which are a combination of regression smoothing methods and multivariate techniques. In particular we analyse data obtained from the San Francisco Men's Health Study (SFMHS) and data from the New York Blood Centre (NYBC). Analysis of subgroups of HIV positive subjects obtained by stratifying on initial CD4 counts show a curious dramatic change for the better in mean decline of CD4 count and CD4/CD8 ratio around October 1, 1987 for all of the San Francisco subgroups (see Figure 1(a)). Since AZT was introduced a few months prior to this date, one could speculate that AZT led to the improvement in mean marker values. However, as seen in Figure 1(b), there is also a turnaround in the CD4 count for the San Francisco control group of HIV negative men. After calibrating for the turnaround in the marker values for the control group, the improvement in the infected groups is eliminated. We conjecture that the change in CD4 values around October 1st 1987 is due to the introduction of more sensitive procedures for obtaining CD4 counts. Thus studies showing improvement in marker value decline after the introduction of treatments need to be calibrated using the CD4 counts from control groups. We also analyse the CD4/CD8 ratio of HIV infected men and find similar, but somewhat larger and less regular, percentage drops in this marker process. The raw data for the HIV positive men in the NYBC study do not show a decline in average CD4 counts; however, when we correct for variable entry times and only include individuals that remain in the study for a certan minimum length of time, we find a decline comparable to the decline in San Francisco study.

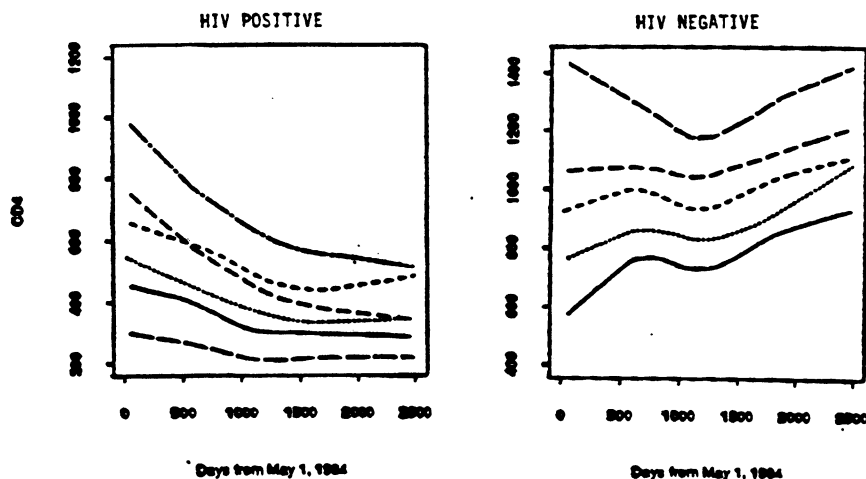*Key words.* HIV positive, HIV negative, stratification, calibration, nonparametric statistical methods.

Figure 1. Estimates of mean CD4 count for groups stratified on initial value of CD4.

1

1 Introduction

In this paper we use nonparametric methods to analyse disease progression in individuals for which we have available observations on surrogate markers that indicate the level of the subject's health. Typical markers will be immune system markers such as CD4 counts, CD8 counts and CD4/CD8 ratios, or some transformation of these. Let t_0 denote the first time the value of a marker is observed. The level of the marker at time $t \geq t_0$ will be represented by the stochastic process $W(t)$. Our focus will be on the mean marker process function $\mu(t) = E(W(t))$.

Our methods are nonlinear and nonparametric in the sense that we do not assume a linear or other parametric form for $\mu(t)$ and we do not assume a Normal or other parametric form for the distribution of the residuals $W(t) - \mu(t)$, $t \geq t_0$. In Section 2 we use locally linear smoothing techniques [1, 2] to produce nonparametric estimates of the mean marker level $\mu(t)$. Since these techniques do not provide standard errors or test procedures for the type of data considered here, in Section 3 we use nonlinear nonparametric multivariate techniques that provide test procedures and standard errors. In this section we use procedures that do not depend on any imputation or censored data techniques and assumptions.

Throughout this article, we will focus on two data sets. The first is the San Francisco Men's Health Study (SFMHS) consisting of waves 1 - 13, corresponding to the time period May 1, 1984 – March, 1991 [3]. This data set includes 381 HIV positive men, 549 HIV negative men and 44 men who seroconverted during the study. Infection times are unknown for the subjects in the sample and consequently, subjects have been infected for variable lengths of time. CD4, CD8 and CD4/CD8 measurements were taken at approximately 6 month intervals. In addition, we have available the calendar time of each visit, the subject's age, sexual orientation as well as, if applicable, date of diagnosis of AIDS and the interval of seroconversion. We removed the 44 seroconverters from our subsequent analyses.

Our second data set consists of a cohort of men and women who donated blood to the New York Blood Center (NYBC) between the period April 1, 1985 and February 1988 [4, 5]. In this article we report on blood donors who tested positive for HIV antibodies. Of the 1181 donors, 888 were contacted for notification of their results. 490 (79% male and 21% female) of the donors agreed to attend the site for follow-up visitsscheduled at 6 month intervals at which time a number of laboratory markers were measured: Leukocytes, CD4, CD8 and CD4/CD8 ratios. As with the SFMHS cohort, infection times are unknown. We report on information collected at a total of 4 visits. The donors also answered a questionnaire which provided demographic and risk information. 48% of those who attended follow-up visits had at least one of the following "risk" factors: had ever used intravenous drugs, had sex with an iv-drug user since 1977 or had sex with a homosexual male since 1977.

Since the infection times are not known, we stratify individuals according to initial CD4 count since, according to current immunological theory, the CD4 count is an indicator of health status for HIV positive individuals. Thus, we group together people with the same health status and examine their decline in CD4 count over time. Moreover, since there is no evidence of tracking [6], that is, there is no evidence that an individual with an initial fast rate of decline of CD4 will persist with a rapid rate of decline of CD4, stratification on initial CD4 count results in groups of individuals that seroconverted at about the same time.

Recent results [7, 8] show a leveling off of the mean CD4 count about 25-35 months after seroconversion as this count reached approximately 500. By stratifying on initial CD4 count for the SFMHS HIV positive group, thereby obtaining groups of men that seroconverted at about the same time, we find that the leveling off of the CD4 count does not have any connection to the level 500 or to the length of time (e.g, 25-35 months) after seroconversion. Rather, the leveling off occured about October 1, 1987, for each strata. We performed a similar analysis for the SFMHS HIV *negative* group and again find a changepoint about October 1, 1987, with the CD4 count changing from decreasing to increasing. To see if the correlation between low CD4 count and missingness could explain the result, we split the data into two further subgroups: those for which there were data throughout a specified period and those for which there were not. The changepoint around October 1, 1987, persisted for both subgroups. The analysis was repeated for the CD4/CD8 ratio and the October 1, 1987 effect persisted. Searching the literature, we found that a change in the incubation period distribution of HIV July, 1987, to stochastically longer period has been postulated [9, 10, 11]. This would indicate a change in the AIDS epidemic to a less severe state on this date, perhaps because of use of AZT. However, the change in the CD4 and CD4/CD8 ratios on this date for the HIV negative group would indicate a change in marker measuring techniques or instruments resulting in more sensitive procedures; or it could be explained by an infectious disease that boosted marker values among both HIV negative and positive individuals; or it could be explained by the HIV negative group using AZT as a "preventative" drug, a very unlikely explanation.

We contacted immunology experts [12, 13] to discuss the plausibility of the hypothesis that procedures for measuring CD4 became more sensitive in mid 1987. This hypothesis is indeed plausible. We then used the HIV negative marker results to calibrate the HIV positive marker results. The changepoint about October 1, 1987 indeed disappears for the CD4 data.

2 Nonparametric Curve Estimation of Marker Means

We describe a smoothing technique which displays average levels of marker values over time. We start by presenting curves that give overall summaries of the data and proceed to discuss ways of handling lack of information on infection times, missing observations and stratification.

The data consist of pairs (t_{ij}, W_{ij}) where t_{ij} is an actual visit time for subject i and $W_{ij} = W(t_{ij})$ is the value of the marker of interest for the ith subject at time t_{ij}. The t_{ij}'s do not have to be distinct across subjects since several subjects may visit on the same day. Let n be the number of individuals in the study and let $N = \sum_i^n n_i$ denote the total number of pairs (t_{ij}, W_{ij}) for a given marker such as CD4. We describe a locally linear smoothing technique which provides a convenient and efficient estimate of the mean level $\mu(t) = \mathrm{E}(W(t))$ of the marker at time t. The basic idea [1, 2, 14, 15] is to produce a locally linear fit to $\mu(t)$ as follows.

Let s be a time point of interest. Assume that in an interval $[s - h, s + h]$, $h > 0$, $\mu(t)$ can be closely approximated by a line $\alpha_0 + \beta_0 t$. Now let $a(s) + b(s)t$ denote the weighted least squares line computed from the data $(t_{ij}, W_{ij}), j = 1, 2, \cdots, n_i,\ i = 1, 2, \cdots, n$ with weights

$$v_{ij} = K\left(\frac{t_{ij} - s}{h}\right), j = 1, 2, \cdots, n_i,\ i = 1, 2, \cdots, n$$

3

where $K\left(\frac{t_{ij}-s}{h}\right)$ is zero for t_{ij} outside the interval $[s-h, s+h]$. A good choice for K would be the quartic kernel [1]

$$K(u) = (\frac{15}{16})(1-u^2)^2 I(\mid u \mid \leq 1)$$

The constant $2h$ is called the window size since the points in the window $[s-h, s+h]$ are the ones used to produce the local linear fit $a(s) + b(s)t$. A good choice for $2h$ would be 15 months = 450 days, since then the locally linear fit is based on at least two waves but typically no more than three waves.

Now the estimate of the mean marker process, $\mu(s)$, is defined to be

$$\hat{\mu}(s) = a(s) + b(s)s$$

The estimate $\hat{\mu}(t)$ can be computed for a sample of m grid $\tilde{t}_1, \cdots, \tilde{t}_m$ along the t-axis using weighted least squares software. Connecting the fitted points, $\tilde{y}_i = (\tilde{t}_i, \hat{\mu}(\tilde{t}_i))$, $i = 1, \cdots, m$, yields a smooth curve estimate of $\mu(t)$.

A robustified version of this locally linear approach, called LOWESS, can be computed using **S** , a statistical software package [16]. LOWESS robustifies the estimate of $\mu(t)$ by downweighting those W_{ij}'s with extreme residuals after a preliminary fit of the data [1, 2].

2.1 Estimated overall marker levels. Results and discussion.

Figure 2.1A presents the LOWESS estimates of mean CD4 count for the 381 HIV positive men and the 549 HIV negative men from the SFMHS. We used the LOWESS smoothing constant 0.50. Other constants gave qualitatively the same results. Note that for the HIV positive group, there is a leveling off of the slope around $t = 1250$ days after the start of the study. After this time the decline of the estimated mean CD4 count is very slow. Note that $t = 1250$ corresponds to the date October 1, 1987, which is three months after the date that it has been hypothesized that the incubation period of AIDS started to lengthen in the Los Angeles cohort study [9] as well as in the SFMHS [11]. However for the HIV negative group, there is also a change in the estimated mean CD4 count level around October 1, 1987. It starts to increase on this date. The introduction of AZT in late 1986 and its increased use in the first half of 1987 would be a possible explanation for the change in mean level for the HIV positive group but not for the HIV negative group. Two possible explanations for a change in both groups on the same date are (1) a change in how CD4 counts are obtained or measured and (2) an infectious disease, such as the common cold, triggered an immune response in the population under study. Note that Figure 2.1C shows the same trend for the CD4/CD8 ratio: Both the HIV positive and negative groups have a change point close to October 1, 1987. On the other hand, the CD8 graph Figure 2.1B shows that the estimated mean CD8 count for both HIV positive and negative groups do not change much over time, with a higher level for the HIV positive group (about 1000) than for the HIV negative group (about 760).

Figure 2.2.A shows smooth estimates of marker means for HIV positive NYBC men and NYBC women. The time axis scale on this graph has been calibrated with the SFMHS data so that the time point 600 corresponds to January 1, 1986, which is approximately 600 days from May 1, 1984. There is a turnaround point for the NYBC males where the estimated mean CD4

4

count changes from decreasing to increasing around $t = 1200$, which corresponds to September 1, 1987. The estimated mean ratio has a weak turnaround point near this date, but starts to decline again about three months later. The estimated mean CD8 count increases from about 790 to 1010 over the time period considered.

There is a changepoint in estimated mean marker values around $t = 840$ for the NYBC women with the estimated mean CD4 and CD4/CD8 ratio higher for NYBC females than for the NYBC males. There is no observable change point around October 1, 1987, however only 9% of the 103 NYBC women had a visit after this date compared to 15% (n = 55), of the NYBC men and 50% (n = 191 for HIV postive and n = 275 for HIV negative) of the SFMHS men.

2.2 Stratification on the initial marker value.

The estimated mean CD4 count and CD4/CD8 ratio processes for the HIV positive SFMHS group both have a change point near October 1, 1987 (figure 2.1). However, we could speculate that this date is irrelevant. Rather, a possible explanation is that the rate of decline in the mean CD4 level and mean ratio level diminishes after a certain threshold value is obtained (e.g., from figure 2.1, threshold of 450 and 0.42 for the observed CD4 and ratio processes, respectively). However, the threshold explanation was not borne out by our analysis.

We consider the new marker process $W(t|w_0)$ whose distribution is the distribution of the conditional distribution of $W(t)$ given $W(t_0) = w_0$. In other words, we stratify on the initial marker value, and $W(t|w_0)$ gives the marker values for those subjects whose initial marker values at the start of the study was w_0. The parameter of interest is now the conditional mean marker function

$$\mu(t|w_0) = E(W(t)|W(t_0) = w_0)$$

There are (at least) two reasons for this stratification:
(i) We can address the question of whether the change point is due to the marker reaching a certain threshold value or because it reached a certain date.
(ii) It addresses the problem that the infection dates for the subjects are unknown. Thus we regard subjects with the same marker values at the beginning of the study as being nearly equally affected by the HIV infection. This means that we can use an estimate $\hat{\mu}(t|w_0)$ of $\mu(t|w_0)$ to do backward projection. We can predict when an individual with initial marker value w_0 was infected by finding the intersection of $\hat{\mu}(t|w_0)$ with the mean marker value for HIV negative subjects.

Let $I(w_0)$ denote an interval around w_0. Our estimate $\hat{\mu}(t|w_0)$ of $\mu(t|w_0)$ will be the estimate $\hat{\mu}(t)$ of Section 2.1 computed for subjects with initial marker values in $I(w_0)$. Again, we use LOWESS to produce a robust version of $\hat{\mu}(t|w_0)$.

2.3 Results and discussion for stratified data.

The SFMHS HIV positive group

Figure 2.3 shows the results for the HIV positive men in SFMH study stratified according to initial CD4 count as indicated. The shapes of the curves are remarkably similar. They all look roughly piecewise linear with changepoint around October 1, 1987. The conjectured threshold value around CD4 = 450 plays no role. When we stratify on initial CD4, the mean curves for the ratio of CD4 to CD8 show the same behaviour as the mean CD4 curves: they are piecewise linear with changepoint near the middle of 1987.

Figure 2.4 is the corresponding graph when we condition the CD4/CD8 ratio rather than the CD4 count. The shapes of the estimated mean ratios corresponding to initial values $w_0 = 0.4$, 0.6, 0.8, 1.0 and 1.3 are also remarkably similar. They are piecewise linear with changepoint near the middle of 1987. Again, when conditioning on the ratio, the mean CD4 shows a changepoint at a similar time.

All the stratified CD4 and CD4/CD8 ratio curves have the same shape as the overall unstratified curves of Figures 2.1 and 2.2. There is not much change in the mean CD8 curves. However, the graphs show that stratification greatly reduces variability.

The SFMHS HIV negative group

FigureS 2.5A and 2.6C show the estimated mean marker curves when we condition on the initial CD4 count and initial CD4/CD8 ratio, respectively. Figure 2.6C, where we condition on initial marker values 1.0, 1.4, 1.6 and 2.1, is the most dramatic. The shapes are remarkably similar with a striking turnaround point from decreasing to increasing around October 1, 1987. The curve starting at 1.0 is shaped as a dampened version of the curve starting at 2.1. Figure 2.5A show a similar but less striking story when we condition an initial CD4 count rather than initial CD4/CD8 ratio.

The NYBC data

Figures 2.7A and 2.8A show the corresponding results for the NYBC HIV positive men. Remarkly, the mean marker curves for the groups with high initial CD4 and CD4/CD8 ratio have shapes similar to the curves for the SFMHS HIV negative groups except the changepoint comes three months earlier. However the increase in mean marker values after the changepoint is less for the NYBC HIV positive men with high initial marker values than for the SFMHS HIV negative men.

2.4   Rescaled marker processes.

We consider two adjustments to the marker processes:
1) Divide the marker process values in the ith stratum by the average of initial marker values in that stratum. This turns the marker process in the ith stratum into a process which gives percentage values relative to the initial stratum mean. We would expect the resulting mean marker curves to be nearly identical and thus combinable. Figure 2.9 shows how the strata mean adjusted marker processes corresponding to CD4 and CD4/CD8 values on a square root scale are much more parallel than the unadjusted processes in Figure 2.3.

Note that in Figure 2.9, we have, after adjusting the slope by dividing by the initial strata mean, added the average strata mean in order to separate the graphs. Otherwise, the

curves fall nearly on top of each other and we are unable to distinguish to which strata they correspond. We also tried dividing by actual initial values rather than by averages over strata. The results were similar but more erratic due to the great variability of initial values. They are not presented here.

We also plotted strata adjusted marker processes for processes stratified according to initial CD4/CD8 ratio values rather than initial CD4 values and found the same results as in Figure 2.9. They are not shown here.

2) Our second adjustment has already been mentioned. It is to take the square root of the strata adjusted marker processes. This is done since the distribution of these markers are skew to the right and the distribution of the logarithms are skewed to the left. The distribution of the square root values are nearly symmetric and the shape resembles that of a normal distribution.

2.5   Recombination of strata.

In Section 2.4 we saw that plots of rescaled marker processes for different strata are nearly identical. In this section we combine the rescaled processes across strata by replacing the observed process for each individual by {marker/initial strata mean marker value}$^{1/2}$ and smooth this series.

The SFMHS data

The results are presented in Figure 2.10 for the SFMHS HIV positive and negative groups. The October 1, 1987, effect is evident. This graph is based on stratification on initial CD4 values. Stratification on initial CD4/CD8 ratio values led to nearly indistinguishable results to those based on stratification by initial CD4. Note that for the HIV positive group the square root of CD4 and CD4/CD8 drop on the average by 24 and 30 percent over the 2500 days of the study, respectively. This corresponds to a 42 and 51 percent drop in the average CD4 and CD4/CD8 values, respectively. The average CD8 counts for the HIV positive group remains remarkably constant and, as expected, higher than that of the HIV negative group.

The NYBC data

The rescaled, recombined mean marker processes for the NYBC men are presented in Figures 2.11 - 2.12 and for the NYBC women, in Figure 2.13. The resulting smoothed curves are surprising. For the NYBC men, the estimated mean adjusted CD4 does not decrease over the time interval of the study (1035 days) in fact, in increases slightly (2.5%). This compares to a 16% drop in mean square root CD4 counts ovr the first 1035 days of the SFMH study. On the untransformed (no square root) scale (graph not shown) the SFMH study CD4 curve drops by 30% over the first 1035 days compared to a 2.5% increase in the NY CD4 count for the same period. Note that the estimated mean CD8 increases and the estimated mean CD4/CD8 ratio decreases in a nonlinear fashion.

The percentage drop in mean square root CD4/CD8 ratio over the 1035 days for the men in the NYBC study was 15%. This compares with a 22% drop in the mean square root CD4/CD8 ratio over the first 1035 days of the SFMH study.

Since, for the NYBC men, stratification on CD4 produced different curves than stratification on CD4/CD8 ratio (see Figures 2.7 and 2.8), we give the results for both types of stratification here. However, the strata mean adjusted curves of Figures 2.11 and 2.12 are not very different. Note that the NYBC women have not been stratified on initial values due to the small sample size. The smaller sample size also means that these curves are less reliable.

2.6 The calibrated rescaled curve.

We next calibrate the SFMHS HIV positive curve of Figure 2.10 using the HIV negative curve of Figure 2.10. More precisely, if the HIV$^-$ negative marker process has changed by a certain percentage from time t_0 to t, we calibrate the HIV$^+$ marker process by subtracting the corresponding predicted percentage change in the HIV$^+$ marker process. Formally, the calibrated process $W_c^+(t)$ is

$$W_c^+(t) = W^+(t) - \{\bar{W}^+(t_0)[\bar{W}^-(t) - \bar{W}^-(t_0)]/\bar{W}^-(t_0)\}.$$

Since the processes have been rescaled to start at one, we set $\bar{W}^+(t_0) = \bar{W}^-(t_0) = 1$ and arrive at the calibration formula

$$\{HIV^+ \text{ marker/initial } HIV^+ \text{ strata mean}\}^{1/2}$$

$$-[\{HIV^- \text{ marker/initial } HIV^- \text{ mean}\}^{1/2} - 1].$$

The results are given in Figure 2.14. The calibrated rescaled CD4 curve is nearly linear. The October 1, 1987, effect disappears after calibrating for changes in the HIV negative group. The calibrated curve drops steadily by 30% over the 2500 days of the study. This drop in square root values correspond to a drop of 51 percent in the average CD4 value over the 2500 study days. The drop in CD4/CD8 ratio values is not linear due to the strange behaviour of mean CD8 counts. It has a curious "ski-jump" shape. The CD4/CD8 ratio on the square root scale drops by 41 percent, which corresponds to a drop of 65 percent on the original scale. The curves based on stratification on initial CD4/CD8 ratios were very similar and are not presented here.

3 A Multivariate Approach

For the SFMHS, the exploratory nonparametric analysis of Section 2 indicated a dramatic change in the rate of decline of CD4 and ratio markers around October 1, 1987. This change persisted after various stratifications of the data. However, even though these graphs are extremely informative, statistical inference procedures based on this type of analysis that are accurate for these small sample sizes have not yet been developed. In this section, we convert the data and parameters to a form suitable for multivariate analysis where statistical inference is possible without making any parametric assumptions.

We are interested in $\mu(t)$ at one or more given time points. However, the observed time points vary across subjects. One way to deal with this is to use a marked point process framework [16]. A detailed examination of marker process models is given in [17]. In this section we use a different approach which consists of using linear interpolant marker processes.

Let t_{ij} denote the number of days from the start of the study to the jth visit for the ith individual, $j = 1, 2, \cdots, n_i$, for $i = 1, 2, \cdots, n$ individuals. For the ith individual in the study we connect the points $(t_{i1}, W(t_{i1})), \cdots, (t_{in_i}, W(t_{in_i}))$ by linear interpolation thereby obtaining the n independent processes $W_1(t), \cdots, W_n(t)$, where

$$W_i(t) = W_i(t_{ij}) + \frac{W_i(t_{ij+1}) - W_i(t_{ij})}{t_{ij+1} - t_{ij}}(t - t_{ij}), \quad t_{ij} < t \le t_{ij+1}, \quad j = 1, \ldots, n_i.$$

We assume that $W_1(t), \cdots, W_n(t)$ are independent and identically distributed. That is, at time zero we regard the n subjects in the study as being selected from some pool of subjects, and the future times that the markers are to be measured as well as the marker values at these times are independently and identically distributed random vectors across subjects.

The introduction of the linear interpolant processes makes it possible to study the marker processes at fixed time points, say at six months intervals. The original time points $\{t_{ij}\}$ come in waves that are on the average six months apart. However, they are quite spread out and can not be assumed to be at six month intervals. The introduction of the interpolant processes allows a simple nonparametric analysis based on the multivariate central limit theorem.

At this point we make a convenient reparametrization which consists of replacing the mean marker level $\mu(t)$ by the mean level of the interpolant process, i.e. by

$$m(t) = E(W_i(t)) \tag{1}$$

Note that $m(t)$ is very close to $\mu(t)$, in fact, if $\mu(t) = \alpha + \beta t$, then $m(t) = \mu(t)$. In general, when comparing risk groups or assessing the effect of covariates, differences in $m(t)$ between risk groups or for different covariate values reflect differences in $\mu(t)$ between risk groups and different covariate values. Thus we can make $m(t)$ the object of our inference. Now our estimate of $m(t)$ is

$$\hat{m}(t) = n^{-1} \sum_{i=1}^{n} W_i(t). \tag{2}$$

Clearly $\hat{m}(t)$ is unbiased with variance $\text{var}(\hat{m}(t)) = n^{-1}\sigma^2(t)$ where $\sigma^2(t) = \text{var}(W_i(t))$. In fact, let $t_1, \cdots, t_k$ denote time points of interest, then $\hat{m}(t_1), \cdots, \hat{m}(t_k)$ have, by the multivariate central limit theorem, approximately a k dimensional normal distribution with mean vector $m(t_1), \cdots, m(t_k)$. The covariance matrix Σ of $\hat{m}(t_1), \ldots, \hat{m}(t_k)$ can be estimated using the sample covariance matrix. Thus, once we have introduced the interpolant processes and the reparametrization to $m(t)$ we can use the usual multivariate procedures and software packages designed to analyse longitudinal and other multivariate data.

3.1 Overall marker levels: Correcting for missingness

Next we considered the question of whether there is a difference in mean marker processes between HIV positive individuals who stay in the study for essentially the entire time period of the study and those who do not. This, in turn, is to answer the question of whether the change

in average CD4 and ratio values around Oct. 1, 87 observed in Section 2 can be explained by a strong association between low marker values and missingness since those with very low marker values are more likely to be very sick or deceased. In other words, the change around Oct. 1, 1987 could be due to individuals with low marker values dropping out of the study around this date. Thus we are asking whether the results obtained when individuals with missing values are included biases the results since, possibly, low marker values and missingness are highly correlated. To address this question we split the observations into two groups: Those that stayed in the study the entire time period from 180 days to 2340 days and those that did not. The first group was defined by selecting those who had at least one visit before 180 days and at least one visit past 2340 days. Thus their interpolant processes $\{W_i(t)\}$ spanned the time interval $[180, 2340]$.

Figure 3.1 compares the group of subjects that had observation periods spanning $[180, 2340]$ with those that did not. It shows that those that did not remain in the study for the given time period had lower CD4 counts and CD4/CD8 ratios. Thus we will in this section consider only the group with observations spanning the entire time period $[180, 2340]$.

Tables 1 and 2 present the estimated mean marker values at eight time points spanning from 180 to 2340 days from May 1, 1984 for the SFMHS HIV positive and HIV negative men respectively. The results are consistent with Figure 2.1. The decline in mean CD4 and CD4/CD8 values for the HIV$^+$ subjects in the SFMHS tapers off around October 1, 1987, while for the HIV$^-$ subjects these marker processes go from decreasing to increasing around this time. Note that for the HIV$^+$ group the drop in mean CD4 is statistically significant at the 5% level only for one interval, from 720 to 1080 days. The drop in mean CD4/CD8 HIV$^+$ ratio values is significant for the two time intervals leading up to Oct. 1, 1987 ($t = 1248$). The drop in the mean HIV$^-$ CD4 count in the interval prior to Oct. 1, 87, is significant as is the increase in the two intervals following this date. A similar result holds for the HIV$^-$ CD4/CD8 values.

It is also instructive to examine changes over time intervals. Thus we let $\Delta_i =$ expected change per year in marker values over the time interval (t_i, t_{i+1}), $i = 1, 2, \ldots, 7$, that is

$$\Delta_i = 360 \frac{E[W(t_{i+1}) - W(t_i)]}{t_{i+1} - t_i}.$$

Table 3 gives the estimated changes. All the downward changes in mean CD4 for the HIV$^+$ SFMHS group before Oct. 1, 87 are significant at the 5% level while the changes over the two time intervals after this date are not. For the HIV$^-$ group the downward change in the last interval before Oct. 1, 87 is significant as is the upward change in the first interval after this date.

The above significance statements are "one at the time" statements. To get simultaneous significance across time intervals, we turn to the 90% Bonferroni confidence intervals. They are given by

$$\Delta_i = \hat{\Delta}_i \pm 2.45 \text{ (standard error } \hat{\Delta}_i), \quad i = 1, \ldots, 7$$

Table 4 gives these intervals and confirms the results of Table 3. That is, it shows the leveling off of the decline in mean CD4 counts for the HIV$^+$ group around Oct. 1, 87, and the significant increase in mean CD4 counts for the HIV$^-$ group after this date.

10

3.2   Rescaled multivariate marker processes

In Section 2.3 it was shown that stratification on initial marker values reduces data variability and that the marker processes corresponding to different strata have remarkably similar shapes. Thus, as in Section 2.3, we rescale the marker process values for an individual in strata i by dividing that persons marker process values by the ith strata mean of initial marker values. We also considered log and square root transformations of the adjusted marker values and found that the square root transformations gave nearly symmetric distributions of the values at given time points. Thus we present the analysis on the square root scale; however, the results on the untransformed scale were very close to what is presented here. Table 5 gives the results for

$$\{\text{marker/initial strata mean}\}^{1/2}$$

Table 5 shows the same pattern as Tables 1, 3 and 4 as well as Figure 2.10. The Oct. 1, 87, effect is striking. The drop in estimated mean rescaled CD4 and CD4/CD8 on the square root scale is 20% and 28%, respectively. This compares to 24% and 30% for the slightly longer time period of Figure 2.10. The careful analysis of this section which avoids problems due to missingness and provides standard errors largely confirms the results of Figure 2.10. Note however, that it confirms the pattern of the decline, but not the level of the marker values. When subjects that did not stay in the study over the entire time period from 180 to 2340 days are included, the level is lower.

The NY data

Table 6 shows a similar analysis of the NY men. However, here the time axis is no longer from May 1, 84. For each subject we measured time from the first visit. This was done because the times at which the subjects were measured varied greatly. Thus when we used calender time there was no reasonably long time interval for which there was data available on a reasonable number of subjects. Using the time from first visit, the subjects with observation period spanning the time interval from 180 to 450 days, and the men that attended all four scheduled visits, we found the results of Table 6. This is very different from the results of Section 2. That is, when we have removed the subjects with missing data and used the time from first visit as our time unit, there is a decline in the rescaled CD4 count on the square root scale. Moreover, the declines in both the CD4 and CD4/CD8 processes for the NY men are comparable to the declines in the SFMHS over "time" intervals of the same length.

For the CD4 process of Table 6, the p-values corresponding to the drop in rescaled mean marker values on the square root scale for the two time intervals from 180 to 360 days and 360 to 450 days were 0.0448 and 0.1244, respectively. For the CD4/CD8 process, the corresponding p values were and 0.0001 and 0.0006.

3.3   Multivariate calibrated processes

We next use the fluctuations in the marker processes for the HIV negative SF men to calibrate the marker processes for the SFMHS HIV positive group using the calibration formula of Section 2.6. Table 7 shows how calibration has removed the Oct. 1, 87 effect. Note that the calibrated

CD4 results are very close to linear except for a dip at 720 days. Also note that the curious ski-jump shape of the calibrated CD4/CD8 curve first noticed in Figure 2.14 persists. The calibrated mean processes drop by 25% (CD4) and 38% (CD4/CD8), respectively, as compared to 30% and 51% for the slightly longer time scale of Figure 2.14.

4   Discussion

In this article we used nonlinear nonparametric techniques that do not rely on any parametric assumptions concerning the mean structure of a process or the form of the distribution of residuals to analyse marker processes over time. In particular, we examined the CD4 count, the CD8 count and the CD4/CD8 ratio values for HIV positive and negative subjects from the San Francisco mens health study and HIV positive men from the New York Blood Center. For the SF group, we found that in the raw data there was a leveling off of the decline in mean CD4 and CD4/CD8 values around the middle of 1987, which corresponds to the conjectured date for a change in the incubation period of HIV (9, 10, 11). We separated the subjects into groups of comparable health status by stratifying on the initial CD4 and CD4/CD8 values. The patterns of marker value decline were remarkably similar across strata with change points near Oct. 1, 1987. When recording the within strata marker processes on a percentage scale (percent of predicted initial value), they were nearly identical. Figure 2.10 and Table 5 give summaries resulting from combining strata. They give the percentage change in mean marker values over time and they clearly show the changepoint around Oct. 1, 1987.

Since there is also a changepoint around Oct. 1, 1987 for the SFMHS HIV negative group we calibrated the SFMHS HIV positive marker processes using the pattern of change in the HIV negative marker processes. The results are given in Figure 2.14 and Table 7. They show a nearly linear decline in the mean calibrated CD4 count except for a downward blip around May, 1986. The mean calibrated CD4/CD8 ratio values decline in a nonlinear "ski-jump" fashion. Table 8 below show how these patterns are reflected in the p-values (based on Wilk's lambda multiple comparison test).

As a final summary of the SFMHS, we give in Figure 4.1 the calibrated CD4 and CD4/CD8 mean marker process points that correspond to Tables 7 and 8. The May 86 dip is present in both groups and the ski-jump pattern is evident in the CD4/CD8 marker points.

For the New York group, the raw data shows a puzzling lack of downward trend in the mean CD4 marker process. However, after the careful analysis of Section 3 where subjects with an excess of missing values are excluded and the time axis is adjusted for variable entry times, a downward trend in mean CD4 marker values comparable to that in the SF study is evident.

Our analysis in Section 2 is data analytic with no standard errors or p-values available. Our analysis in Section 3 is inferential in that it provides standard errors and p-values.

References

[1] Cleveland, WS Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 1979; **74**, 829-836.

[2] Cleveland W.S. and Devlin S.J. Locally weighted regression: An approach to Regression Analysis by Local Fitting. *J. Amer. Statist. Assoc.*, 1988; **83**, 596-610.

[3] Winkelstein W. Jr., Lyma DM, Padian N, Grant R, Samuel M, Wiley M, Anderson RE, Lang W, Riggs J, and Levy JA. Sexual practices and risk of infection by the human immunodeficiency virus. *J. Am. Med. Assoc.*, **257**, 321-325 (1987).

[4] Cleary, PD, Singer, E et al. Sociodemographic and behavioural characteristics of HIV antibody-positive blood donors. *American Journal of Public Health*, 1988; **78**, 953-957.

[5] Cleary, PD, Van Devanter, N et al. Trends in sociodemographic and behavioural characteristics of HIV antibody-positive blood donors. *AIDS Educ. and Prev.*, 1991; **3**(1), 60-71.

[6] Taylor, JMG, Cumberland, WG, Sy JP. A stochastic model for analysis of longitudinal AIDS data. Technical Report, 1992; Department of Biostatistics, UCLA.

[7] Vittinghoff, E, Malani, HM, and Jewell, NP Estimating Patterns of CD4+ lymphocyte decline using data from a prevalent cohort of HIV infected individuals. *Biostatistics Technical Report #30 Group in Biostatistics*, 1992; Univ. of Calif., Berkeley.

[8] Zeger, SL, Diggle, PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Technical Report #714, 1992; Department of Biostatistics, The John Hopkins University.

[9] Taylor, JMG, Kuo, J-M, Detels, R. Is the incubation period of AIDS lengthening? *J. Acquired Immune Deficiency Syndromes*, 1991; **4**, 69-75.

[10] Brookmeier, R. Reconstruction and future trends of the AIDS epidemic in the United States. *Science*, 1991; **253**, 37-42.

[11] Bacchetti, PR, Segal, M, and Jewell, NP Uncertainty about the incubation period of AIDS and its impact on backcalculation. *AIDS Epidemiology: Methodological Issues*, 1992. N. Jewell, K Kietz and V Farewell (eds), Boston: Birkhäuser-Boston.

[12] Cambrosio, A and Keating, P. A matter of FACS: Constituting novel entities in immunology. *Med. Antrop. Quart.*, 1992.

[13] Kagan, J. Personal communication. NIAID, NIH, 1992.

[14] Stone, CJ. Consistent nonparametric regression. *Ann. Statist.* 1977; **5**, 595-620.

[15] Fan, J. Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, 1992; **20**. To appear.

[16] Becker, RA, Chambers, JC, Wilks, AR. The new S language: a programming environment for data analysis and graphics. Wadsworth and Brooks/Cole Advanced Books and Software. Pacific Grove, CA, 1988.

[17] Scheike, TH. Nonparametric kernel regression: a martingale approach. Technical Report, Department of Statistics, U.C. Berkeley, 1992.

[18] Jewell, NP, and Kalbfleisch, JD. Marker models in survival analysis and applications to issues associated with AIDS, 1992. N. Jewell, K. Kietz and V. Farewell (eds), Boston: Birkhäuser-Boston.

Table 1. Mean marker process values $\bar{W}(t)$ with standard errors (se's). HIV positive men. SFMHS. Subjects with observation period spanning [180, 2340]. $n = 99$

Days from May 1/84	180	360	720	1080	1440	1800	2160	2340
$\bar{W}(t) = \overline{CD4}$	721.9	684.0	640.7	545.9	511.9	505.3	514.5	484.5
$se(\overline{CD4})$	26.2	27.0	25.8	22.0	23.6	24.2	29.8	28.6
$\bar{W}(t) = \overline{CD8}$	994.9	1039.9	1094.4	1115.9	1114.1	1074.7	1169.8	1183.0
$se(\overline{CD8})$	42.48	49.6	52.7	44.4	45.5	45.3	55.1	54.6
$\bar{W}(t) = \overline{\text{ratio}}$	0.79	0.73	0.65	0.56	0.53	0.54	0.51	0.47
$se(\overline{\text{ratio}})$	.030	.034	.029	.030	.029	.031	.033	.032

Table 2. Mean marker process values $\bar{W}(t)$ with standard errors (se.'s) HIV negative men. SFMH. Subjects with observation period spanning [180, 2340]. $n = 137$.

Days from May 1/84	180	360	720	1080	1440	1800	2160	2340
$\bar{W}(t) = \overline{CD4}$	1110.3	1123.5	1171.7	1017.6	1040.0	1135.2	1227.3	1218.9
$se(\overline{CD4})$	26.0	28.5	29.9	25.4	28.2	27.8	30.9	29.1
$\bar{W}(t) = \overline{CD8}$	868.3	779.5	780.6	805.4	791.1	757.5	804.6	813.9
$se(\overline{CD8})$	26.2	26.5	28.9	26.4	25.5	23.5	28.0	26.3
$\bar{W}(t) = \overline{Ratio}$	1.41	1.59	1.65	1.38	1.42	1.63	1.68	1.64
$se(\overline{Ratio})$	.043	.048	.048	.042	.039	.046 ⸱	.050	.047

Table 3. Estimated change $\hat{\Delta}_i$ per year in mean marker values for SFMHS HIV negative and positive subjects for the ith time interval. Subjects with observation period spanning [180, 2340].

Time interval	1	2	3	4	5	6	7
HIV$^+$; change per year of mean CD4	-76	-43	-95	-34	-7	10	-60
$n = 99$; standard errors	32.0	17.5	16.5	19.0	16.8	14.9	20.5
HIV$^-$; change per year of mean CD4	26	49	-54	22	95	92	-16
$n = 137$; standard errors	34.6	24.8	23.1	23.0	24.1	22.7	28.5

Table 4. Simultaneous 90% confidence intervals for the change Δ_i per year in mean marker values over the indicated time intervals. SFMHS HIV positive ($n = 99$) and negative ($n = 137$) subjects. Subjects with observation period spanning [180,2340]. Significance at the 10% level is indicated with a "*".

Interval	1	2	3	4	5	6	7
HIV$^+$	-76 ± 78.4	$-43^* \pm 42.9$	$-95^* \pm 40.4$	-34 ± 46.6	-7 ± 41.2	10 ± 36.5	$-60^* \pm 50.2$
HIV$^-$	26 ± 84.8	49 ± 60.8	-54 ± 56.6	22 ± 56.4	$95^* \pm 59.0$	$92^* \pm 55.6$	-16 ± 69.8

14

Table 5. Mean rescaled marker process values with standard errors. Square root scale. HIV positive men from the SFMH study. Subjects with observation period spanning $[180, 2340]$. $n = 99$.

Percentage of Marker Value at Visit 1								
Days from May 1/84	180	360	720	1080	1440	1800	2160	2340
$\sqrt{CD4_0}$	99.5	97.6	94.8	87.3	84.3	83.6	83.3	80.1
$se\sqrt{CD4_0}$	0.96	1.71	1.82	1.74	1.85	1.97	2.42	2.62
$\sqrt{\text{Ratio}}$	97.4	93.2	87.8	80.5	77.9	78.4	75.6	71.8
$se\sqrt{CD4_0}$	1.65	1.99	1.83	1.96	2.00	2.15	2.39	2.47

Table 6 NYBC Men that attended all 4 visits (n = 56)

Days from First Visit	180	360	450
$\sqrt{CD4}$	21.4	20.7	20.3
$se\sqrt{CD4}$	0.64	0.73	0.76
$\sqrt{\text{Ratio}}$	0.78	0.73	0.71
$se\sqrt{\text{Ratio}}$	0.025	0.024	0.024
Percentage of Marker Value at Visit 1			
$\sqrt{\frac{CD4}{CD4_1}}$	98.0	95.1	93.8
se	1.62	2.64	2.88
$\sqrt{\frac{\text{Ratio}}{\text{Ratio}_1}}$	95.6	90.5	88.0
se	1.35	2.22	2.19

Table 7. Mean calibrated rescaled marker process values with standard errors. Square root scale. SFMHS HIV positive men. Subjects with observations spanning $[180,2340]$. n = 99.

Percentage of marker value at visit 1								
Days from May 1/84	180	360	720	1080	1440	1800	2160	2340
$\sqrt{CD4_0}$	99.5	97.2	92.2	91.7	87.7	82.5	78.2	75.3
$se\sqrt{CD4_0}$	0.96	1.71	1.82	1.74	1.85	1.98	2.42	2.62
$\sqrt{\text{Ratio}}$	95.9	85.5	77.8	79.9	75.9	69.0	64.3	62.1
$se\sqrt{\text{Ratio}}$	1.65	1.99	1.83	1.96	2.00	2.15	2.39	2.47

Table 8. Multiple comparison p-values for testing that there is no change in mean marker values across adjacent intervals. The marker processes are the calibrated CD4 and CD4/CD8 marker processes described in Section 3.3.

Intervals	1-2	2-3	3-4	4-5	5-6	6-7	7-8
CD4	.0789	.0005	.6666	.0091	.0002	.0003	.0014
CD4/CD8	.0001	.0001	.0578	.0015	.0001	.0001	.0012

FIG. 1. Estimates of mean CD4 count for groups stratified on initial value of CD4. The curves are for: 1(a), the HIV positive men in the San Francisco Mens Health Study; 1(b), the HIV negative men in the San Francisco Mens Health Study.

FIG. 2.1. SFMHS HIV positve men (overall)

FIG. 2.2. NYBC men and women (overall and scaled to may 1/84)

FIG. 2.3. SFMHS HIV positive men stratified by initial CD4 count

FIG. 2.4 SFMHS HIV positive men stratified by initial ratio value

FIG. 2.5. SFMHS HIV negative men stratified by initial CD4 count

FIG. 2.6. SFMHS HIV negative men stratified by initial ratio value

FIG. 2.7. NYBC men stratified by initial CD4 count

FIG. 2.8. NYBC men stratified by initial ratio value

FIG. 2.9. Marker processes rescaled by dividing by the initial strata mean. Stratification on initial CD4 count. Square root scale. The curves have been separated by adding the strata means as the last step before plotting. The curves are for the SFMHS HIV positive groups.

FIG. 2.10. Plots of rescaled mean marker processes for the SFMHS. The curves represent average values of $\{marker/initial\ marker\ strata\ mean\}^{1/2}$. The solid curve is the HIV positive men and is based on the six strata of Figures 2.3 and 2.9. The dotted curve is for the unstratified HIV negative men.

FIG. 2.11. Plots of rescaled mean marker processes for the NY men. The curves represent average values of $\{marker/initial\ marker\ strata\ mean\}^{1/2}$. Recombined strata after stratification on initial CD4 counts.

FIG. 2.12. Plots of rescaled mean marker processes for the NY men. The curves represent average values of $\{marker/initial\ marker\ strata\ mean\}^{1/2}$. Recombined strata after stratification on initial CD4/CD8 ratio values..

FIG. 2.13. Plots of rescaled mean marker processes for the NY women. The curves represent average values of $\{marker/initial\ marker\ mean\}^{1/2}$.

FIG. 2.14. The calibrated rescaled mean marker curves for the SFMHS HIV positive group. The curves represent the mean level as percent of initial level on the square root scale.

FIG. 3.1. A comparison of "full span" (those with observation period spanning [180, 2340])

and "partial span" (those with observation period not spanning [180, 2340]) individuals in the SFMHS. The white box plots corresponds to the "full span" subjects and the black box plots corresponds to the "partial" span individuals.

FIG. 4.1. The calibrated CD4 and CD4/CD8 mean marker processes for the SFMHS. The points are mean marker levels as percentage of initial marker levels. Square root scale. The arrows indicate ± two standard errors.

San Francisco Mens Health Study Data: SMOOTHED MARKERS
Smoothing Fraction = 0.50
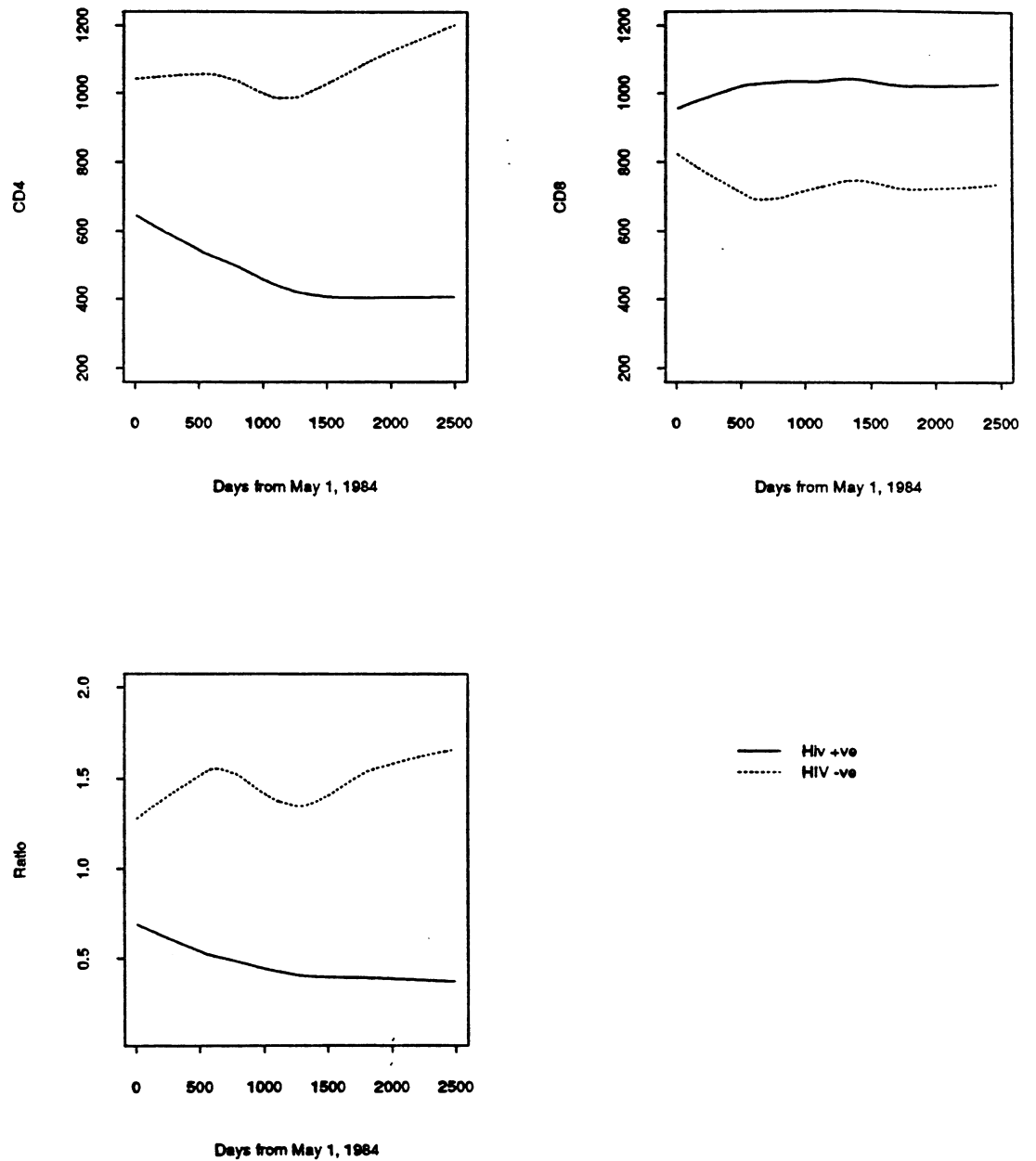
FIG. 2.1. SFMHS HIV positve men (overall)

New York Blood Donor Data: HIV +ve Men and Women: SMOOTHED MARKERS
Smoothing Fraction = 0.50
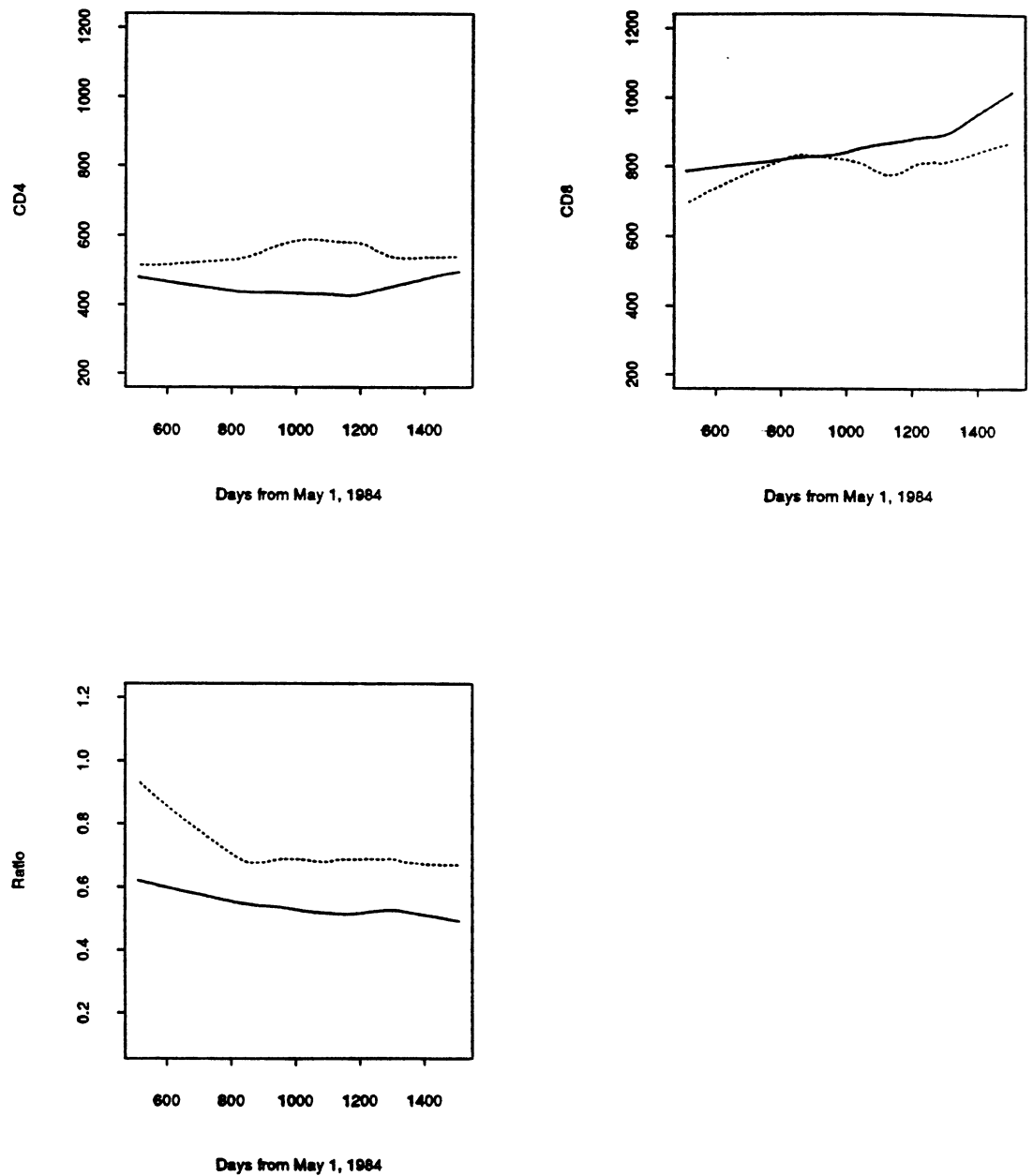
FIG. 2.2. NYBC men and women (overall and scaled to may 1/84)

San Francisco Mens Health Study Data: HIV +ve Men: SMOOTHED MARKERS
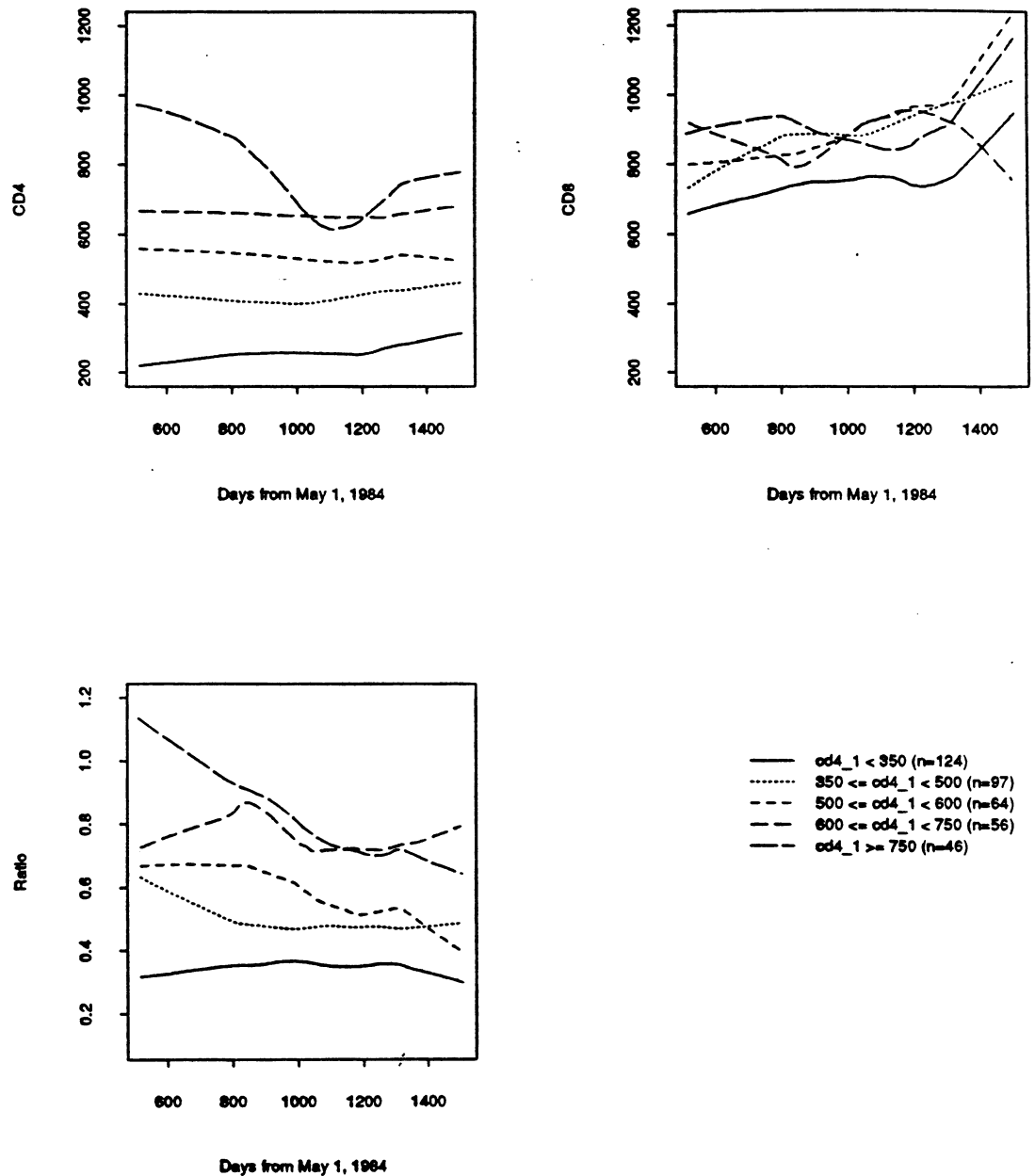Stratified by Cd4 Value at Visit 1 (Smoothing Fraction = 0.5)

FIG. 2.3. SFMHS HIV positive men stratified by initial CD4 count

San Francisco Mens Health Study Data: HIV +ve Men: SMOOTHED MARKERS
Stratified by Ratio Value at Visit 1 (Smoothing Fraction = 0.5)

CD4

Days from May 1, 1984

CD8

Days from May 1, 1984

RATIO

Days from May 1, 1984

——— ratio_1 < 0.5 (n=95)
·········· 0.5 <= ratio_1 < 0.7 (n=103)
– – – 0.7 <= ratio_1 < 0.9 (n=97)
— — 0.9 <= ratio_1 < 1.1 (n=45)
—·— ratio_1 >= 1.1 (n=41)
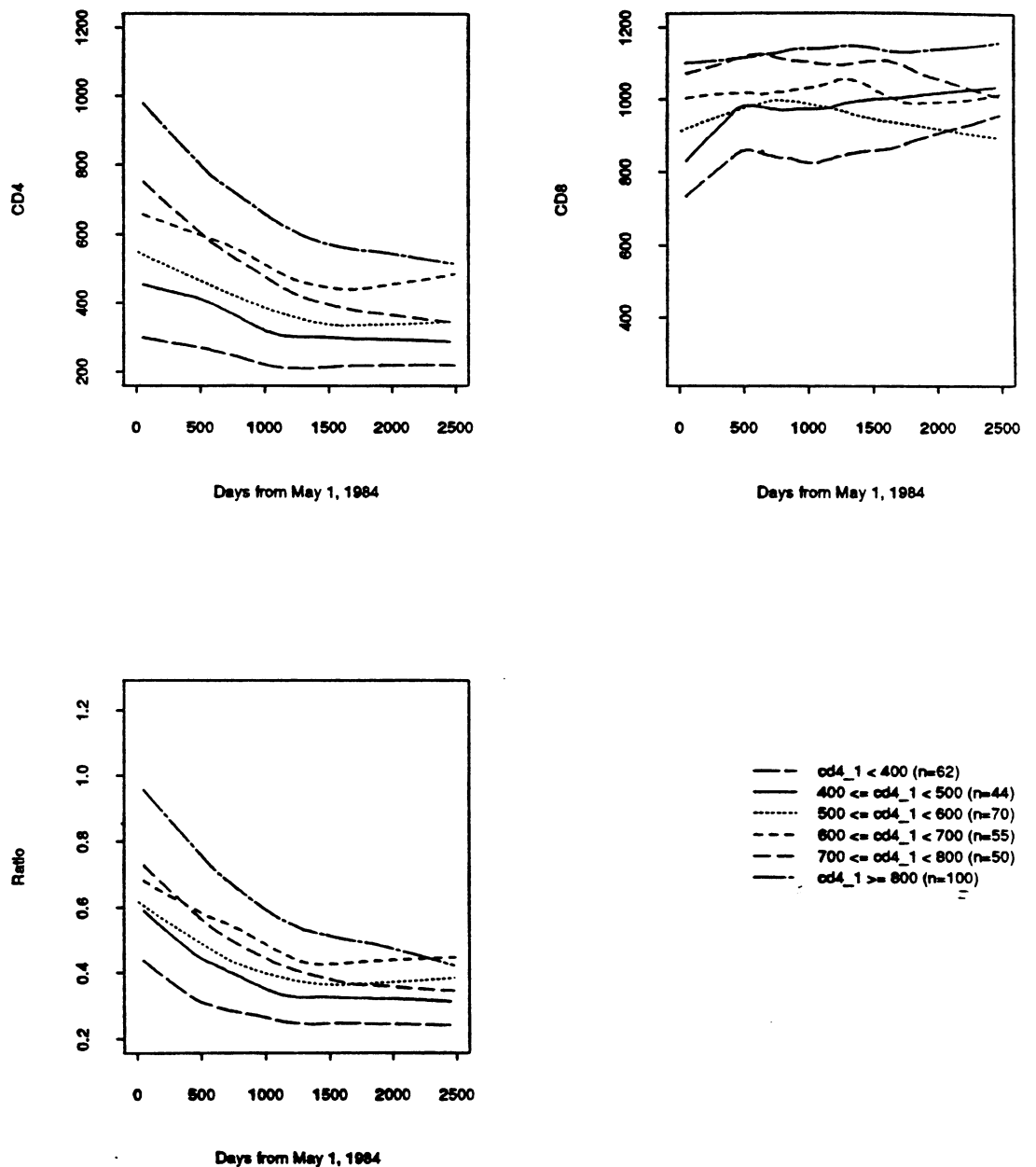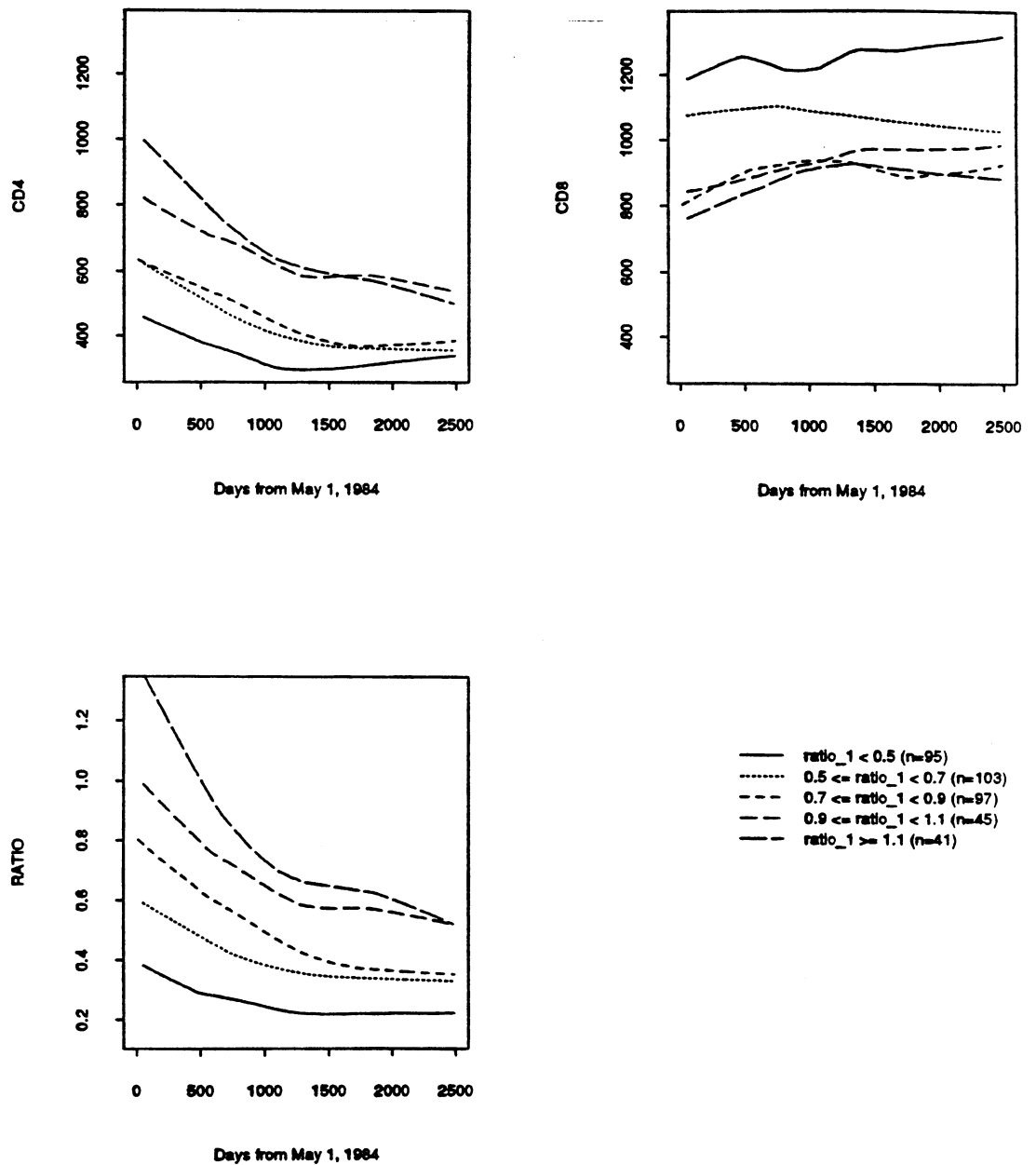
Fri May 22 15:23:15 EDT 1992

FIG. 2.4 SFMHS HIV positive men stratified by initial ratio value

FIG. 2.5

San Francisco Mens Health Study Data: HIV -ve Men: SMOOTHED MARKERS
Stratified by CD4 Value at Visit 1 (Smoothing Fraction = 0.5)

FIG. 2.5. SFMHS HIV negative men stratified by initial CD4 count

San Francisco Mens Health Study Data: HIV -ve Men: SMOOTHED MARKERS
Stratified by Ratio Value at Visit 1 (Smoothing Fraction = 0.5)

ratio_1 < .9 (n=89)
.9 <= ratio_1 < 1.1 (n=92)
1.1 <= ratio_1 < 1.3 (n=109)
1.3 <= ratio_1 < 1.5 (n=101)
1.5 <= ratio_1 < 1.7 (n=60)
ratio_1 >= 1.7 (n=98)
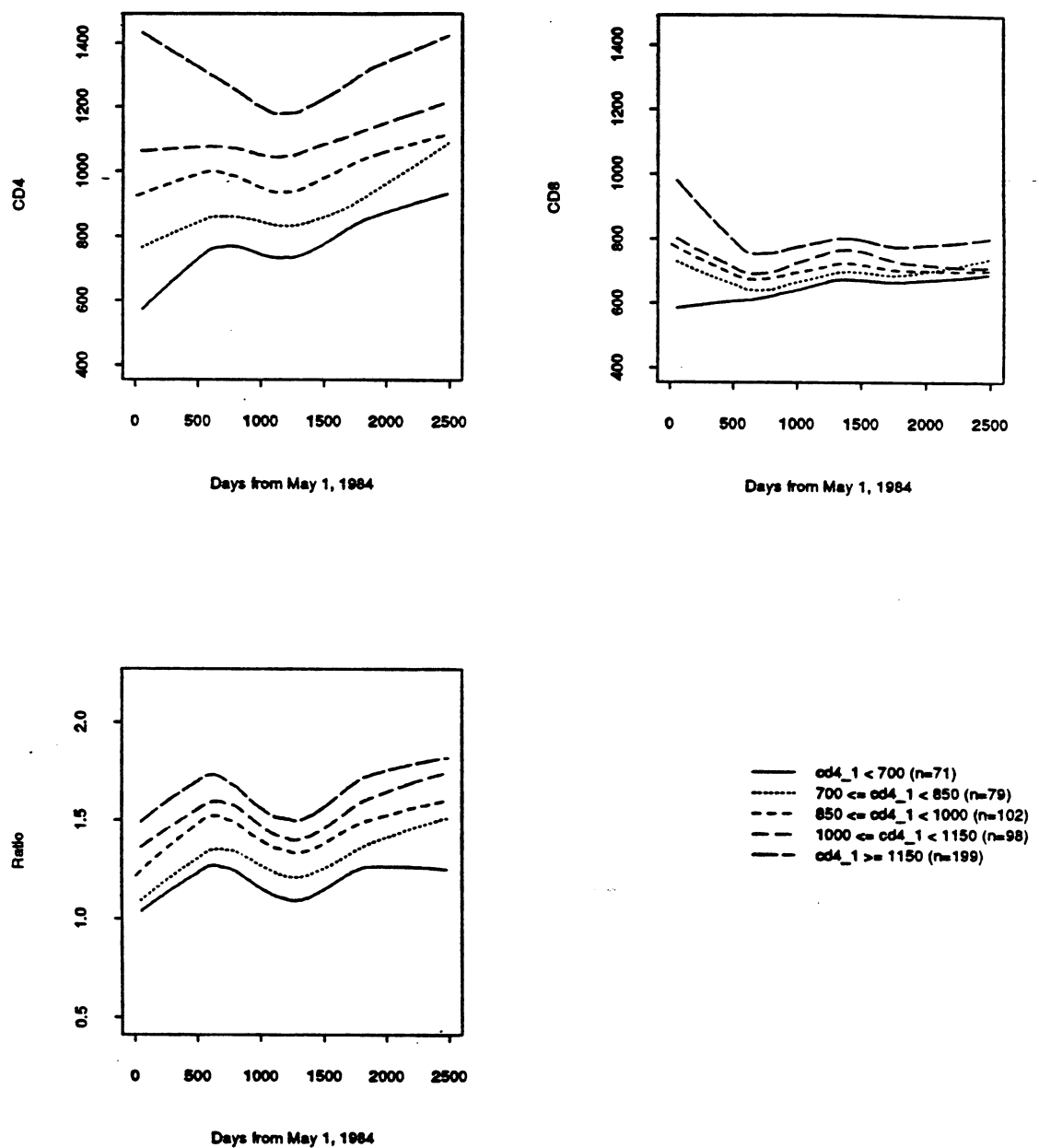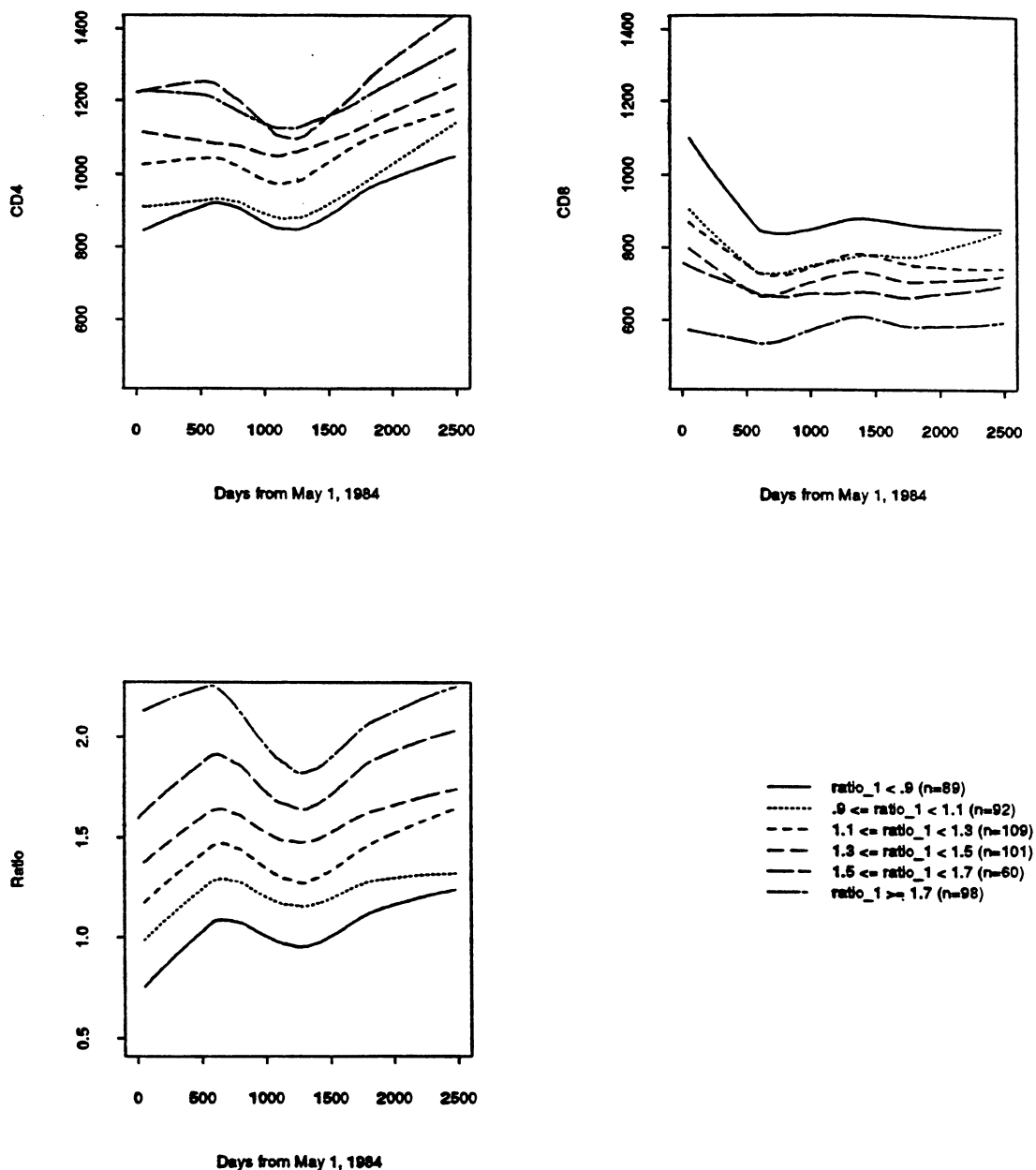
Fri Jun 5 10:37:16 EDT 1992

FIG. 2.6. SFMHS HIV negative men stratified by initial ratio value

New York Blood Donor Data: HIV +ve Men: SMOOTHED MARKERS
Stratified By Cd4 Value at Visit 1 (Smoothing Fraction = 0.50)

FIG. 2.7. NYBC men stratified by initial CD4 count

New York Blood Donor Data: HIV +ve Men: SMOOTHED MARKERS
Stratified by Ratio Value at Visit 1 (Smoothing Fraction = 0.50)

ratio_1 < 0.3 (n=64)
0.3 <= ratio < 0.5 (n=104)
0.5 <= ratio < 0.7 (n=86)
0.7 <= ratio < 0.9 (n=57)
0.9 <= ratio < 1.1 (n=33)
ratio >= 1.1 (n=43)
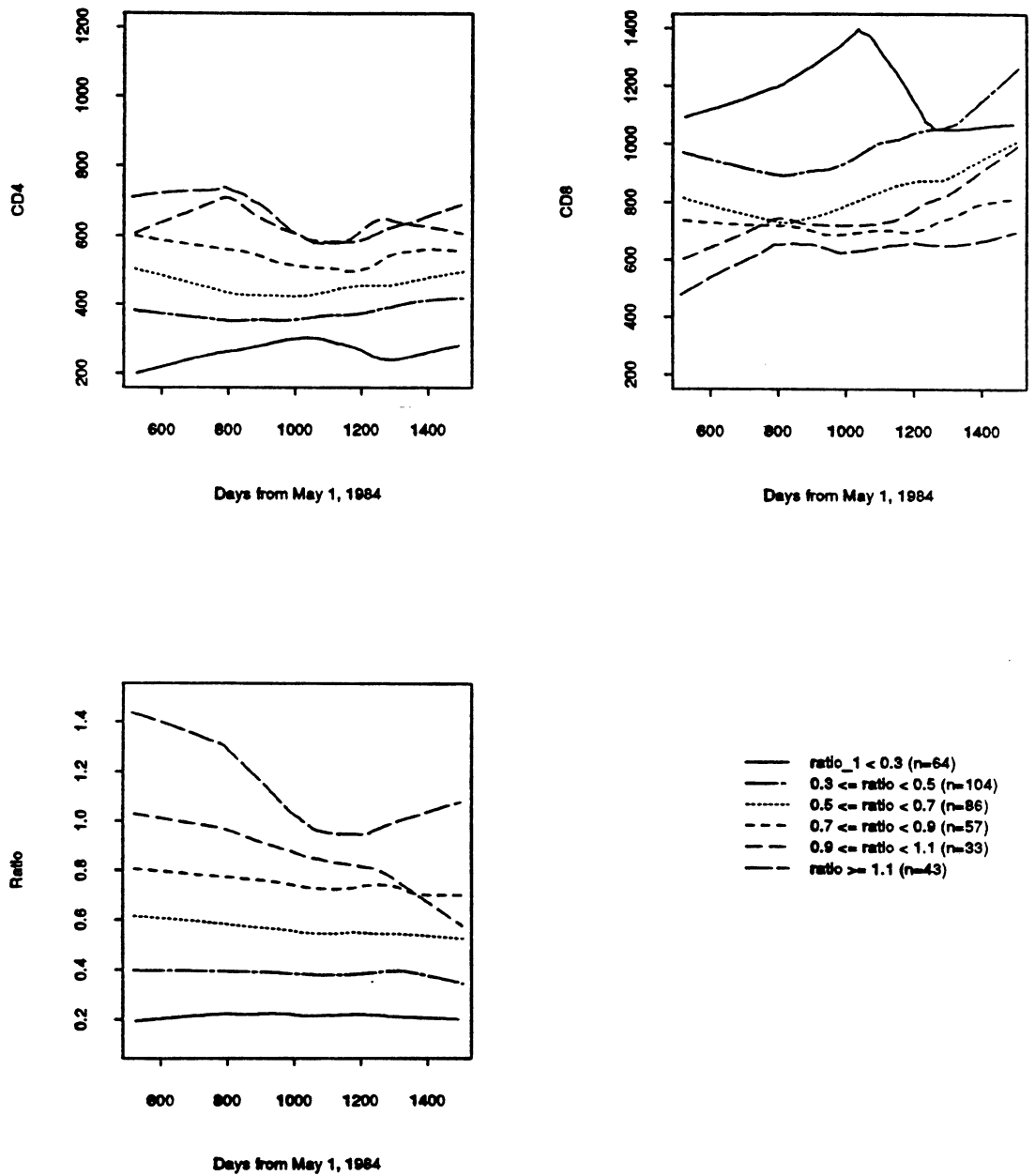
Tue Jun 16 19:00:09 EDT 1992

FIG. 2.8. NYBC men stratified by initial ratio value

San Francisco Mens Health Study Data: HIV +ve Men: SMOOTHED MARKERS
Stratified by CD4 Value at Visit 1 (Smoothing Fraction = 0.5)
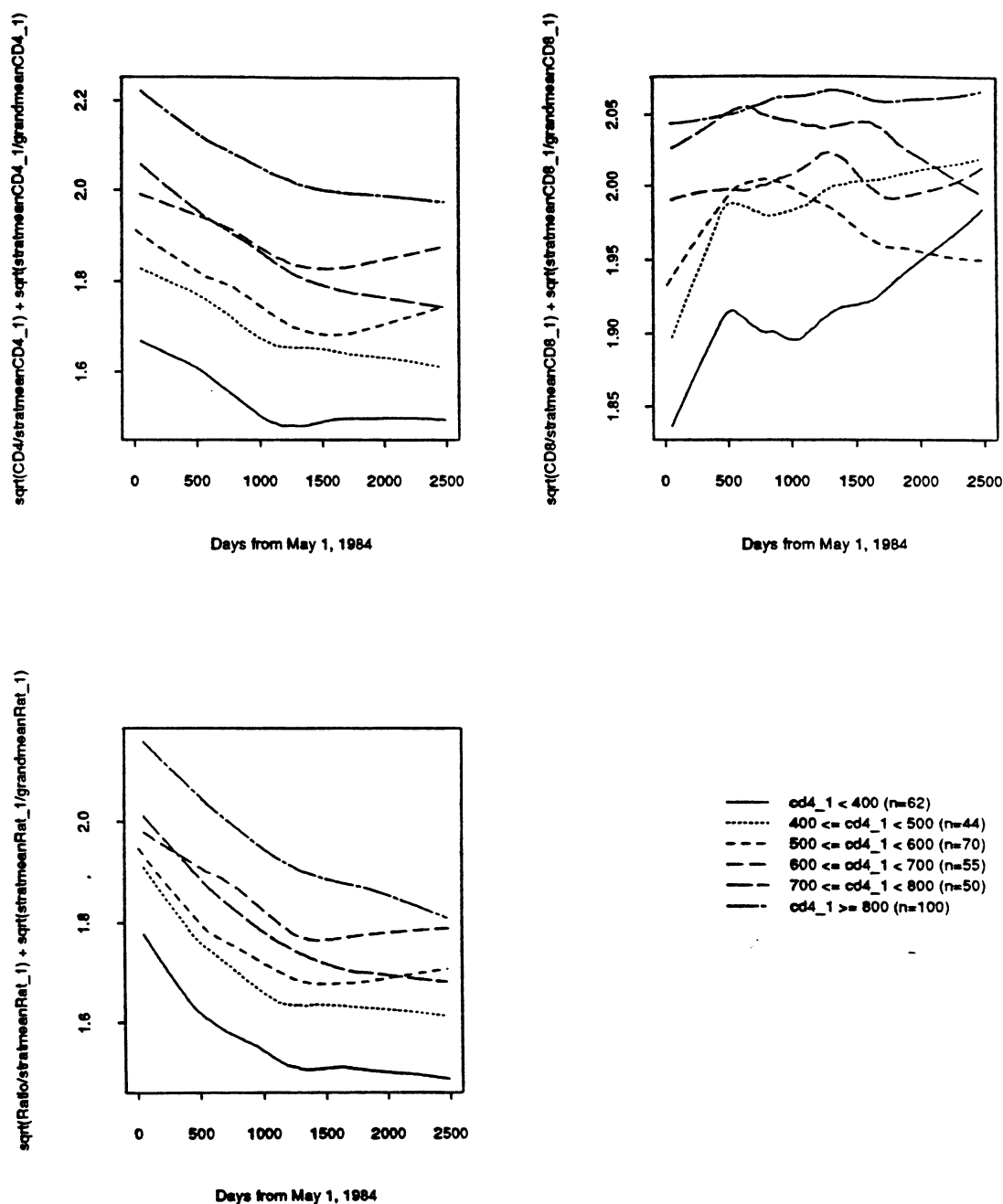
FIG. 2.9. Marker processes rescaled by dividing by the initial strata mean. Stratification on initial CD4 count. Square root scale. The curves have been separated by adding the strata means as the last step before plotting. The curves are for the SFMHS HIV positive groups.

San Francisco Mens Health Study Data: SMOOTHED MARKERS
Stratified by CD4 Value at Visit 1:
6 strata for HIV +ve Men, 1 stratum for HIV -ve Men
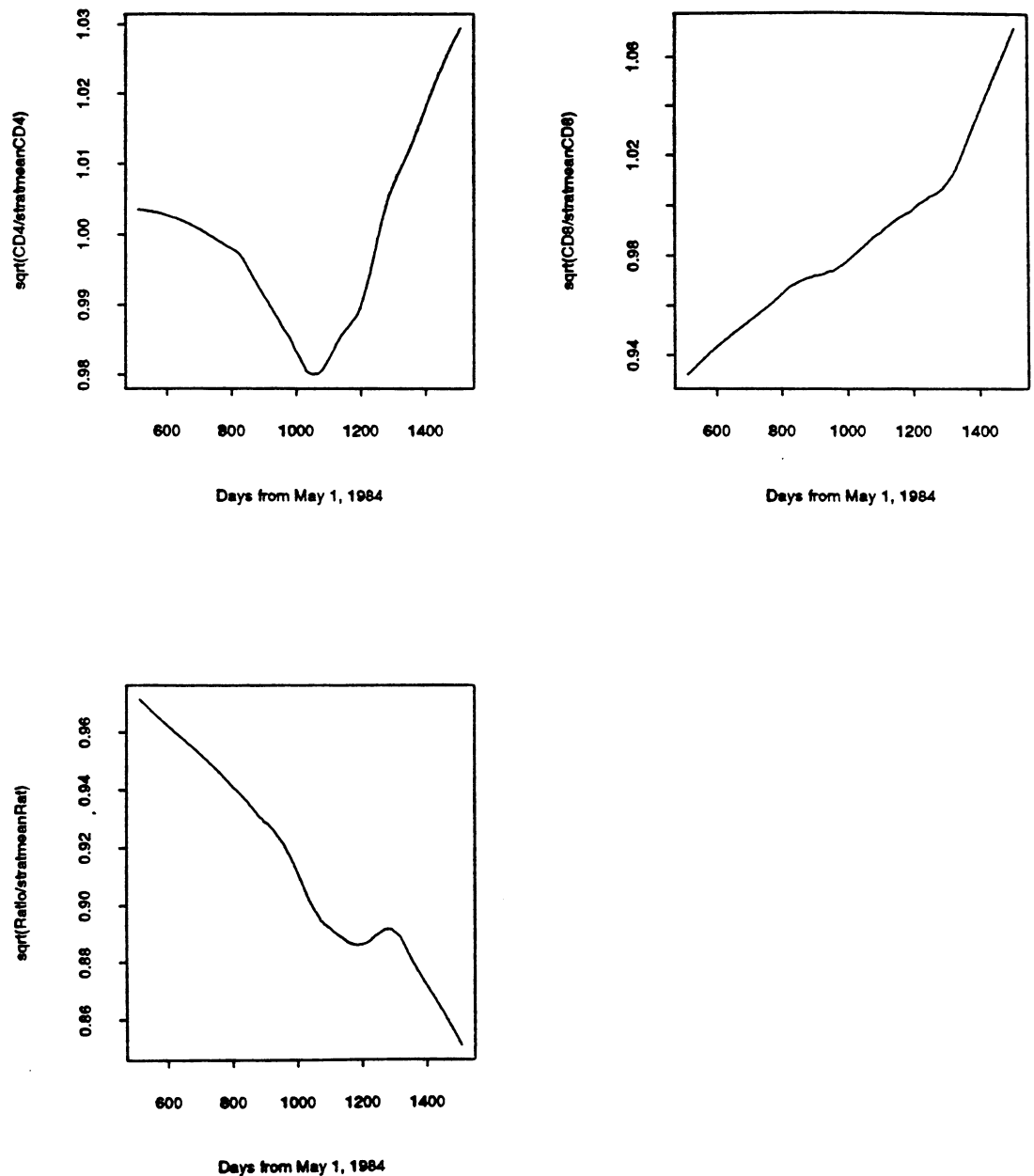(Smoothing Fraction = 0.5)
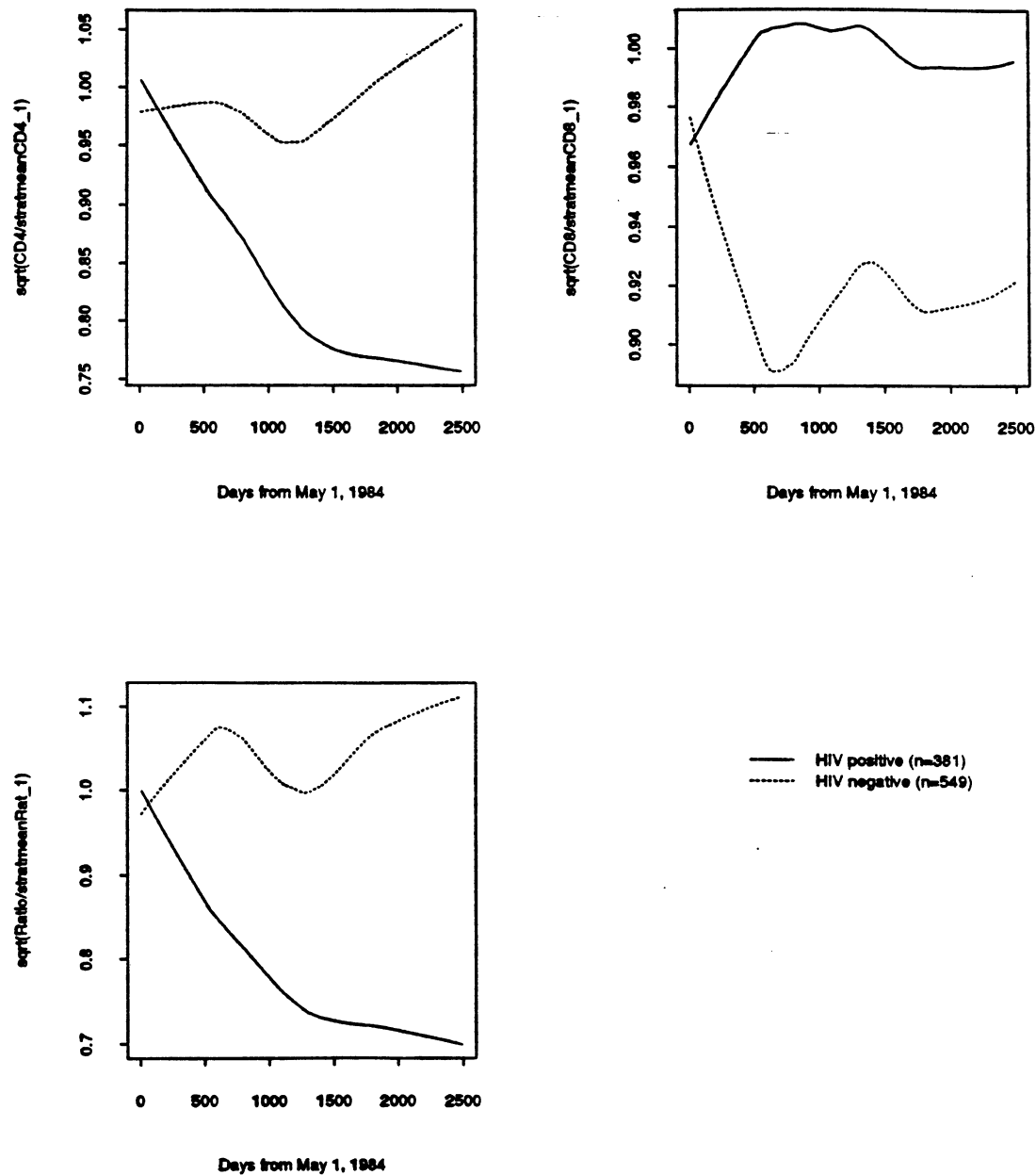
Fri Jul 3 11:28:18 EDT 1992

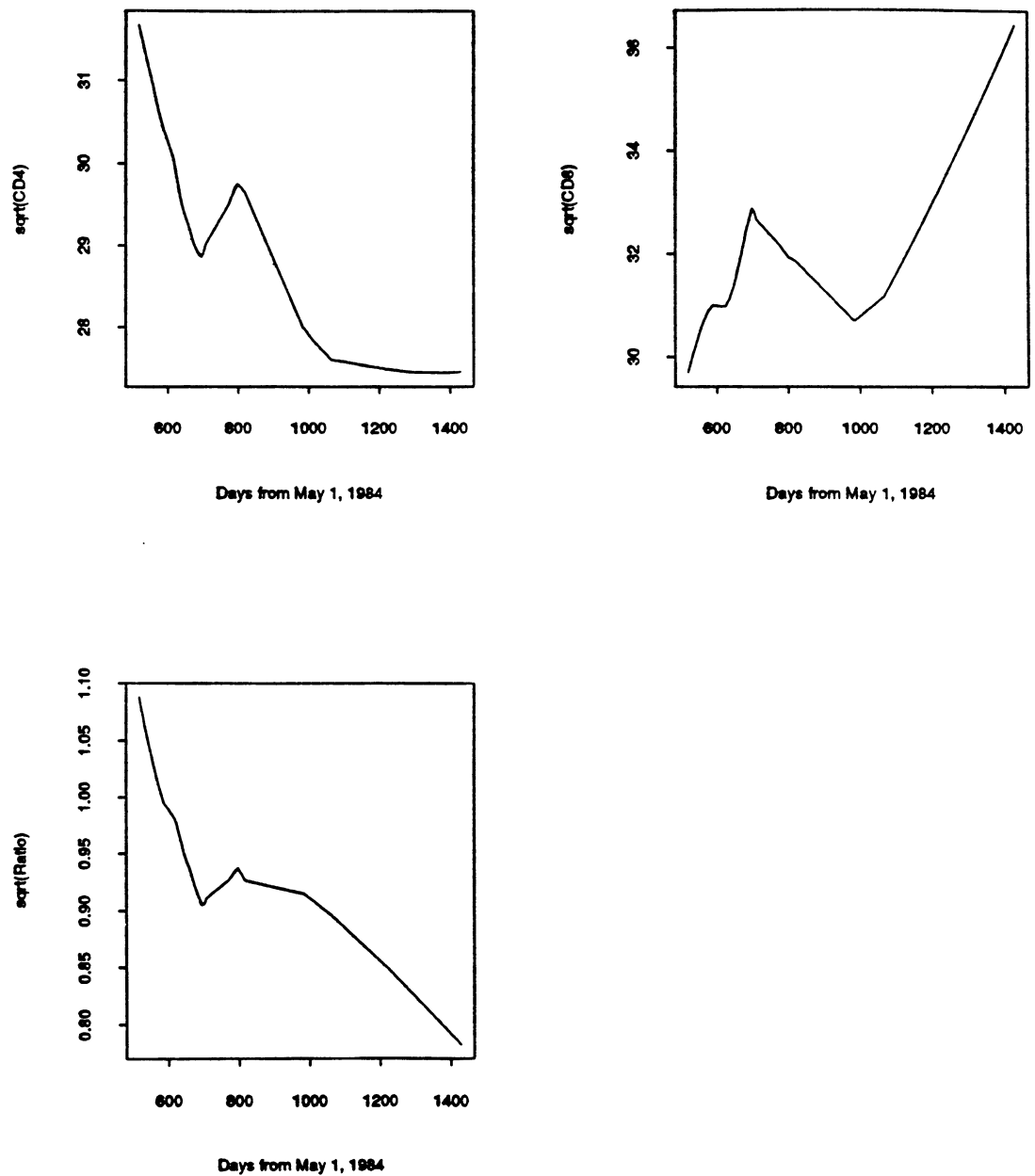**FIG. 2.10.** Plots of rescaled mean marker processes for the SFMHS. The curves represent average values of $\{\text{marker/initial marker strata mean}\}^{1/2}$. The solid curve is the HIV positive men and is based on the six strata of Figures 2.3 and 2.9. The dotted curve is for the unstratified HIV negative men.

FIG. 2.11. Plots of rescaled mean marker processes for the NY men. The curves represent average values of $\{marker/initial\ marker\ strata\ mean\}^{1/2}$. Recombined strata after stratification on initial CD4 counts.

New York Blood Donor Data: HIV +ve Men: SMOOTHED MARKERS
Stratified By Cd4 Value at Visit 1 (Smoothing Fraction = 0.50)

FIG. 2.12. Plots of rescaled mean marker processes for the NY men. The curves represent average values of $\{marker/initial\ marker\ strata\ mean\}^{1/2}$. Recombined strata after stratification on initial CD4/CD8 ratio values..

New York Blood Donor Data: HIV +ve Women: SMOOTHED MARKERS
(Smoothing Fraction = 0.50)

FIG. 2.13. Plots of rescaled mean marker processes for the NY women. The curves represent average values of $\{\text{marker/initial marker mean}\}^{1/2}$.

San Francisco Mens Health Study Data: CALIBRATED CURVES
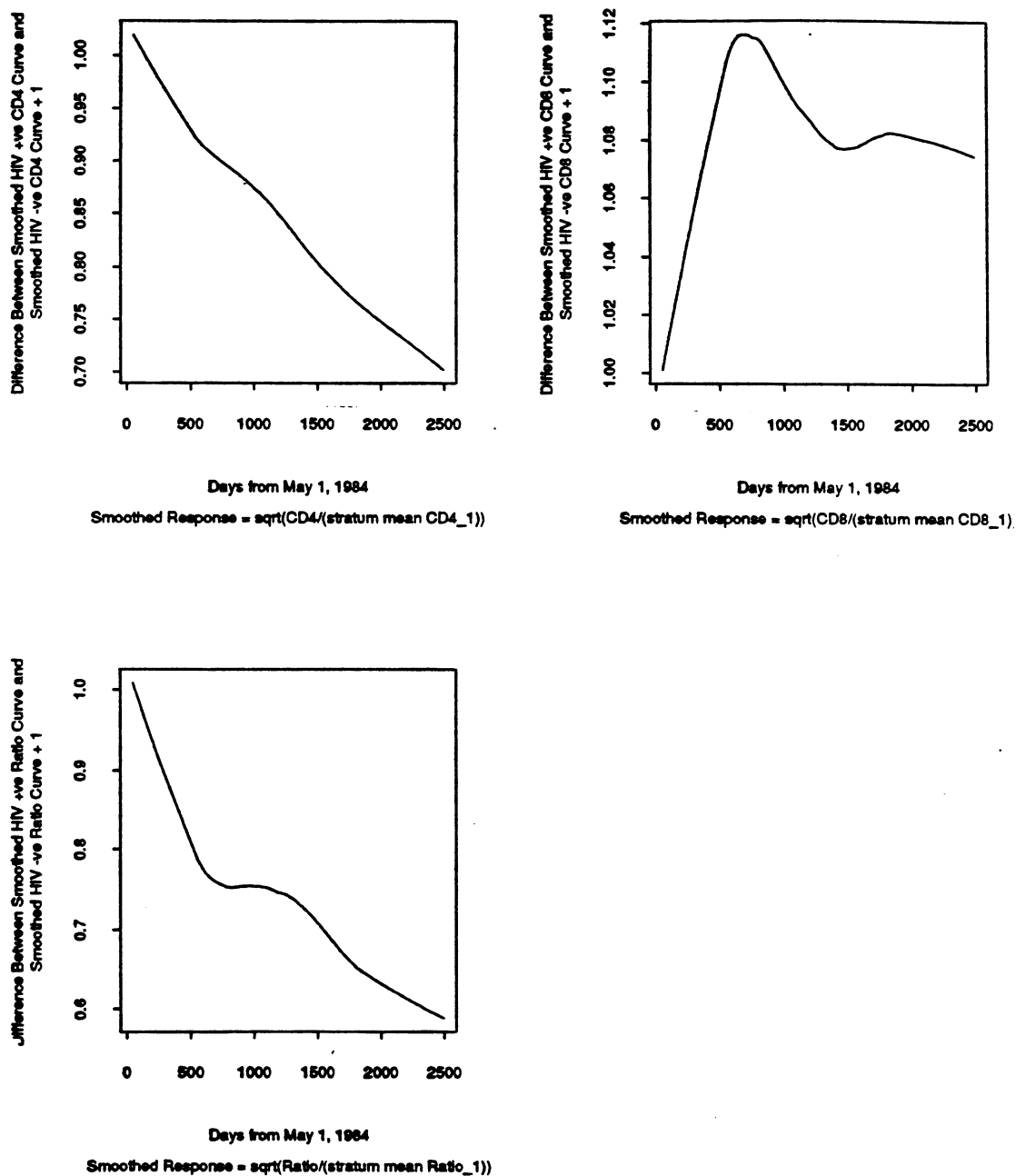6 strata (by cd4_1) for HIV +ve Men, 1 stratum for HIV -ve Men
(Smoothing Fraction = 0.5)

Days from May 1, 1984
Smoothed Response = sqrt(CD4/(stratum mean CD4_1))

Days from May 1, 1984
Smoothed Response = sqrt(CD8/(stratum mean CD8_1))

Days from May 1, 1984
Smoothed Response = sqrt(Ratio/(stratum mean Ratio_1))

FIG. 2.14. The calibrated rescaled mean marker curves for the SFMHS HIV positive group. The curves represent the mean level as percent of initial level on the square root scale.
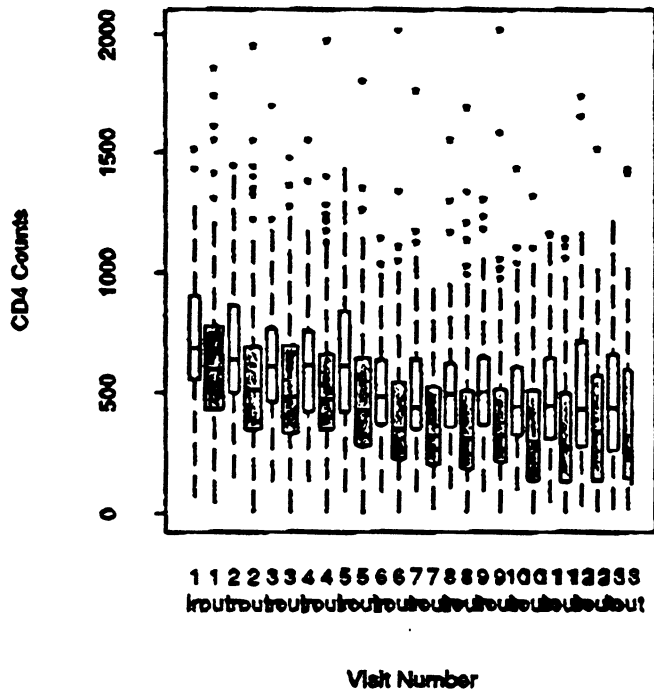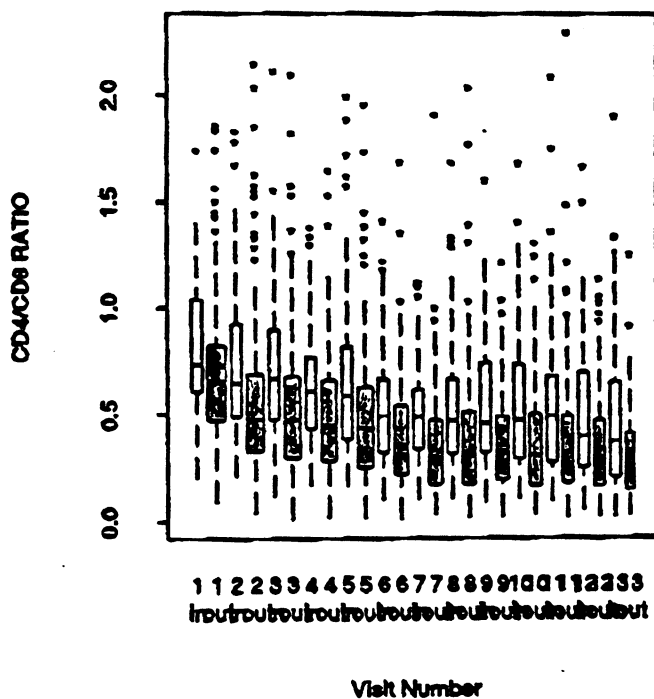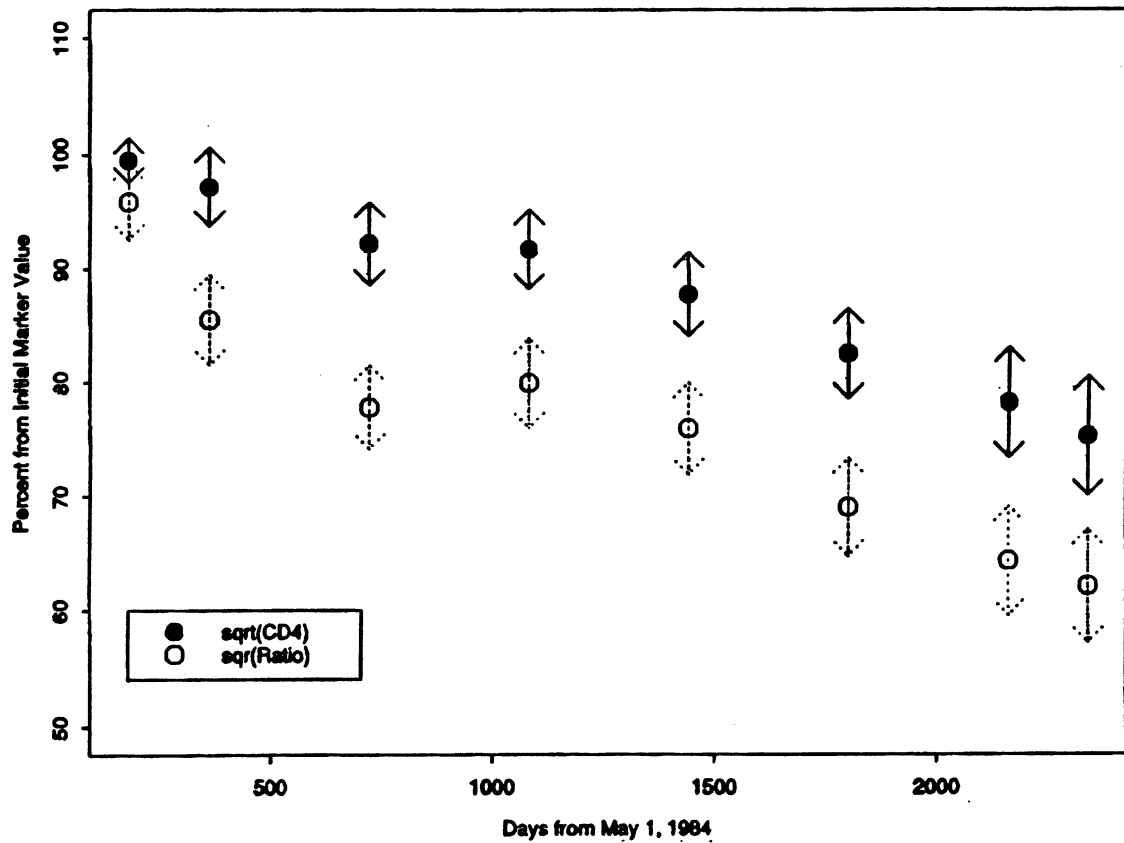
FIG. 3.1. A comparison of "full span" (those with observation period spanning [180, 2340]) and "partial span" (those with observation period not spanning [180, 2340]) individuals in the SFMHS. The white box plots corresponds to the "full span" subjects and the black box plots corresponds to the "partial" span individuals.

FIG. 4.1. The calibrated CD4 and CD4/CD8 mean marker processes for the SFMHS. The points are mean marker levels as percentage of initial marker levels. Square root scale. The arrows indicate ± two standard errors.