Improving the use of long infection windows for undiagnosed estimates

Martina Morris and Jeanette Birnbaum

May 19, 2016

1 Overview

Those with long infection windows have the greatest potential for narrowing the window using CD4 and/or VL data, given that so many of them did not have a BED+ result. This report investigates the potential for CD4 count and viral load to improve our use of the long infection window data.

2 Infection window lengths

Table 1 shows the median window lengths for those with an LNT and those with no LNT, for two time periods. The No LNT group is split up into those who get the 18-yr window and those who get the age minus 16 window.

Windows for those with observed LNT are all quite short regardless of the subgroup. They are greater for non-MSM than MSM, but the difference is smaller for inside vs outside KC. The median window lengths for those with the age minus 16 assumption are substantially longer, and of course the 18-yr groups has median window lengths of 18 years.

The last row shows the distribution of cases across the LNT and the two No LNT groups for each time period. The percent of cases who have an observed LNT is high, remembering that this is a percent among those with either an LNT or reported no LNT. We do not know a lot about how reporting is evolving over time and whether the characteristics of the missing LNT population are changing. In both WA and Philadelphia, trends show rising reports of missing LNT and declining reports of no LNT.

Table 1: Median window lengths, in years, among the 3016 cases with non-missing LNT. For pre-2012 and 2012-2014 separately, columns define three groups: those with a recorded LNT, those who reported no LNT and received the 18-yr assumption, and those who reported no LNT and received the age-16 assumption. The final row shows the distribution of the cases across these three groups as percents within the two time periods.

	Before 2012			2012-2014		
	LNT	No LNT (18y)	No LNT (age-16)	LNT	No LNT (18y)	No LNT (age-16)
MSM	1.1	18.0	9.0	0.9	18.0	9.0
non-MSM	2.7	18.0	9.0	2.9	18.0	10.0
Inside KC	1.1	18.0	9.0	1.0	18.0	10.0
Outside KC	1.7	18.0	10.0	1.4	18.0	8.0
Percent of Population	80.1	12.7	7.3	77.9	14.1	8.0

Table 2 shows how the window lengths, both the observed ones and the assumed ones for those with no LNT, distribute across several interval lengths: 0-1 year, 1-2 years, 2-5 years, 5-17 years, and 17-18 years. The last row shows that overall, about 50% of those with non-missing testing history have a window greater than 2 years.

Table 2: Distribution of sample with non-missing testing history. Values are row percents

	Window Lengths (Years)				
	(0,1]	(1,2]	(2,5]	(5,17]	(17,18]
MSM	42.8	15.5	18.8	14.4	8.5
non-MSM	13.5	10.3	19.0	25.7	31.5
Total	35.6	14.2	18.9	17.2	14.2

3 CD4 and Viral Load

3.1 Time of measurement

CD4 counts and viral loads are measured within 6 months of diagnosis for more than 75% of cases (Figure 1). Almost all cases have non-missing CD4 and VL measurements (Figure 2 panel 1). By 2010 and beyond, about 75% or more of cases have their CD4 and VL measurements taken within 30 days of diagnosis (Figure 2 panel 2). This pattern holds among those with long (>5y) infection windows (Figure 2 panel 3). Unlike the BED measurements, CD4 and VL are readily available for most of the sample, including those with long infection windows.

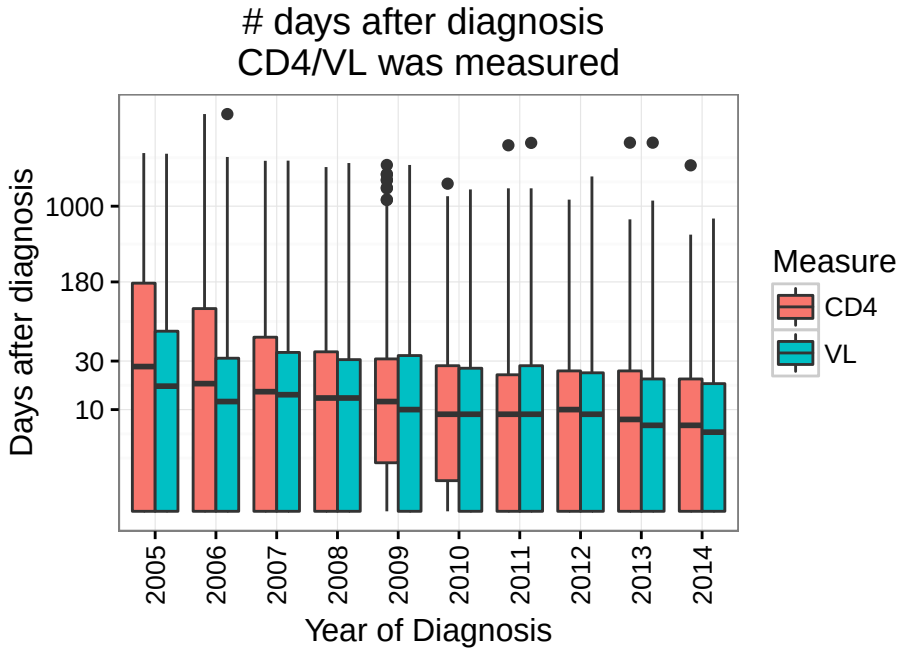

Figure 1: days after diagnosis that CD4 or VL was measured

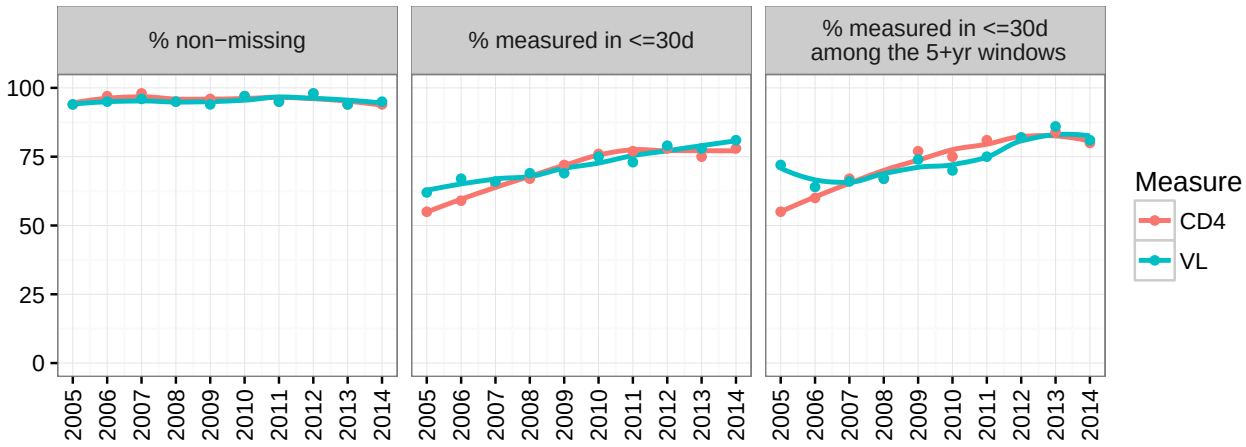

Figure 2: By year, percent of cases who have (left) a non-missing result, (middle) a measurement ¡=30 days of diagnosis, and (right) a measurement ¡=30 days of diagnosis among those with infection windows of 5y or more

3.2 VL and CD4 distributions

In the left panels, Figure 3 shows the distribution of (top) viral loads and (bottom) CD4 counts for the entire population by LNT status. The right panels show only those with known recent infections, defined as having an LNT within 6 mos of diagnosis and the respective VL or CD4 measurement taken on the day of diagnosis.

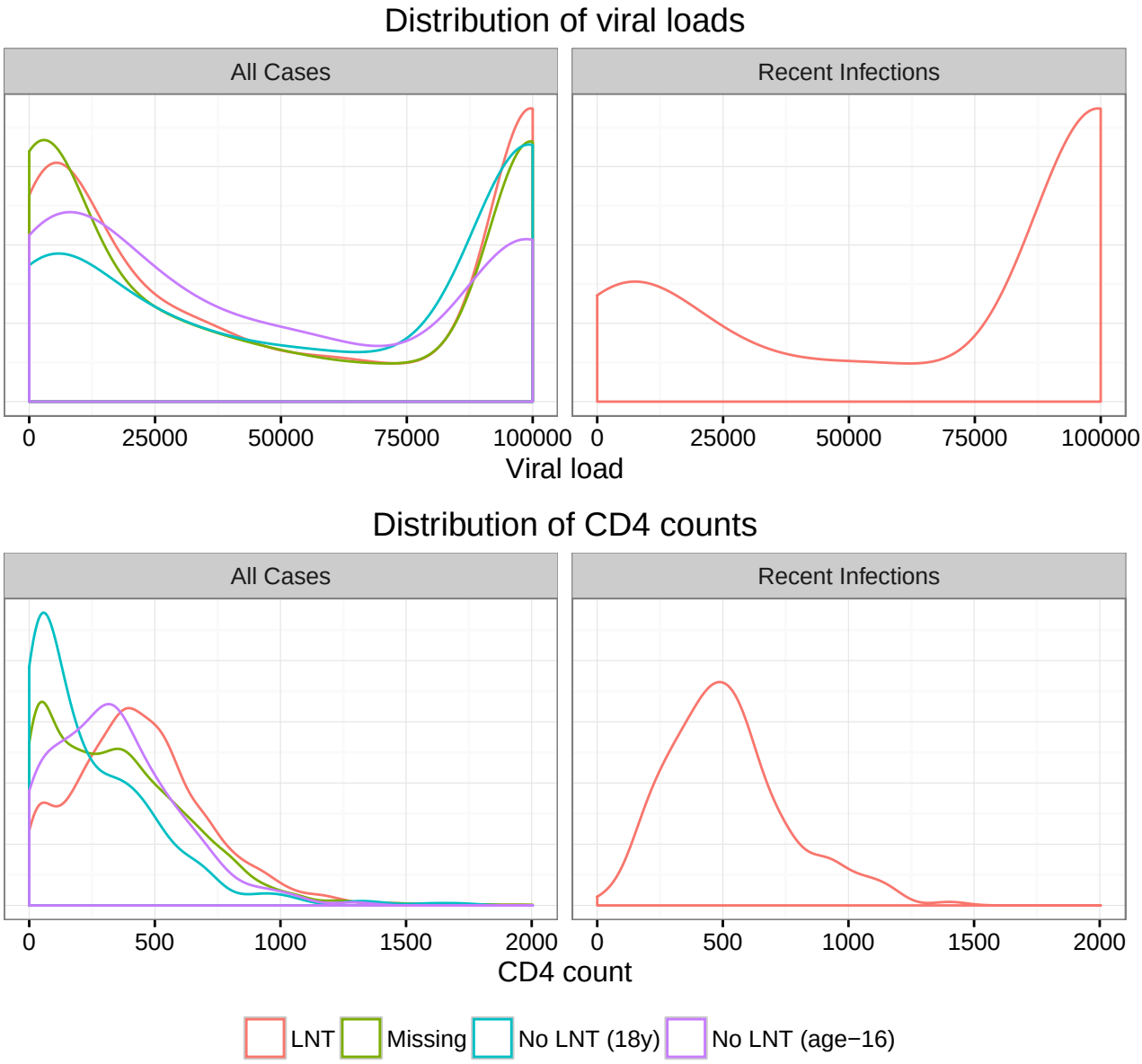

Figure 3: Distribution of (top) first VL and (bottom) first CD4 values among (left) all cases and (right) recent infections (N=211 for VL, N=105 for CD4)

3.3 Using CD4 and VL to identify recent infections

Figure 3 shows that viral load peaks at very low and very high loads, to varying degrees for the four different groups. CD4 count is strongly skewed towards very low counts in the No LNT-18 yr window group in particular. The distributions in the recently infected cases are most similar to the LNT subgroup, but in the case of viral load even more skewed towards very high viral loads.

These differences clearly indicate that there is relevant information in the viral load and CD4 data. These data are also strong in that they are measured on the majority of the sample, including those with long infection windows. Integrating them into the method would still be a complex process, however. To use some combination of very low or very high viral load plus a high CD4 count as an indicator of recent infection, we would need to consider:

- **Treatment:** Was the CD4/VL measured close enough to diagnosis that treatment has probably not started/affected the measurements? How many days is "close"?
- **Concurrent diagnosis:** If AIDS was diagnosed soon after, was that late presentation or fast progression?
- **CD4 threshold:** What threshold for CD4 count suggests recent infection for an untreated case?
- **Viral load threshold:** What thresholds for viral load suggest recent infection for an untreated case? Viral load starts out low, peaks, and then falls to a "set-point" value during the long asymptomatic period that ensues for untreated cases. This set-point level varies substantially across cases. Figures 4 and 5 provide some data regarding the variation in set-point and peak viral loads. We could investigate using thresholds that are so high or low that they are unlikely to be setpoints. We may also need to consider the typical timeframe for reaching set point viral load when assessing the time of measurement (Figure 5), which seems to be on the order of 40-60 days.
- **Method for classifying recent infections using VL and CD4:** Given the diversity in immune response and measurement time, static thresholds may be too simplistic. We could consider looking into the literature for, or validating ourselves, a prediction model for classifying recent infections based on VL and CD4 and other relevant factors such as concurrent diagnosis. We would need to select an acceptable probability-of-recent-infection cutoff for classifying cases using the model.
- **Reconsidering the missing testing histories:** Thinking along the lines of a prediction model, we could also explore imputing broad LNT categories for those with missing data, rather than assuming MAR, or using propensity score weights. This might inform us as to whether our MAR assumption is pushing our results towards greater or fewer undiagnosed.
- **Potential impact:** It may also be worth considering a simulation study to determine the potential that shortening long infection windows has for impacting the undiagnosed results. For example, if about 30% of those with non-missing testing history have a window >5 years, and we are able to shorten 50% of those windows to <1 year, what impact would that have on the estimates? What if we can only shorten 10% of those long windows?

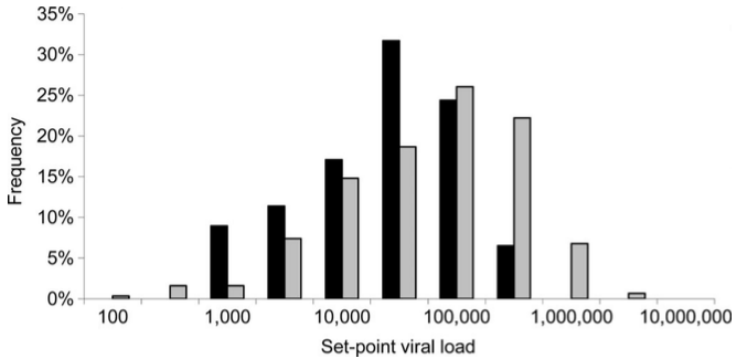

Fig. 1. The distribution of set-point viral loads. The distribution of viral loads (copies per milliliter of peripheral blood) is plotted for untreated individuals in the Amsterdam Seroconverters Cohort (black bars) and the Zambian Transmission Study (7) (gray bars). The bars represent bins 0.5 $\log_{10}$ wide and are labeled by their midpoint viral load.

Figure 4: Set-point viral load distributions for two untreated cohorts, the Amsterdam Seroconverters Cohort (black bars) and the Zambian Transmission Study (gray bars). Set-point viral load was defined as the geometric mean viral load between 6 mos after diagnosis and the first AIDS-defining event or censoring. From Fraser 2007, PNAS, Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis

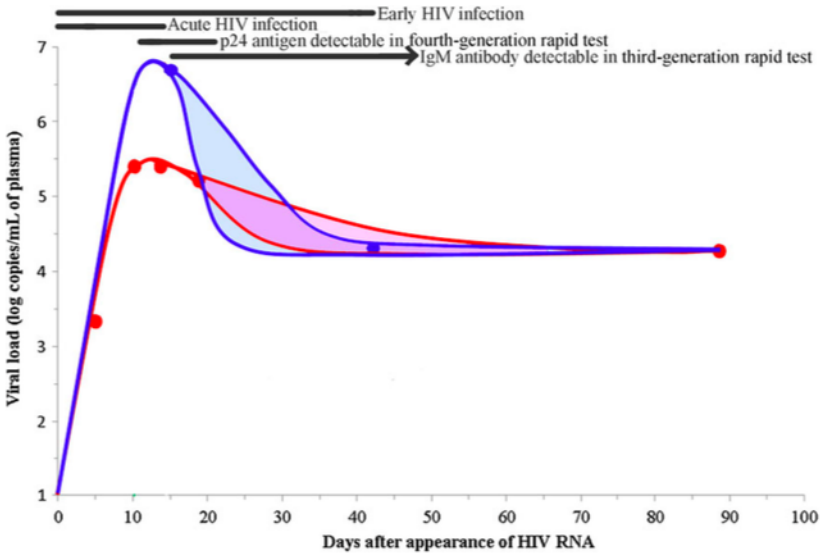

Figure 1. Appearance of diagnostic and viral markers during acute and early human immunodeficiency virus (HIV) infection. Median HIV loads are up to 100 days after the appearance of HIV RNA in individuals from Kenya, Uganda, Tanzania, and Thailand (top blue curve) [2] and the United States (bottom red curve) [1]. The pairs of lines indicate the range of trends that are consistent with the data as published. It is not possible to define the timing of the fall to a set point with precision by using the US data (follow-up points were approximately 3, 8, 11, 16, 54, and 112 days after infection). The eclipse period, ie, the 11 days from infection to detectable viral load, is not included in this figure [12]. Abbreviation: IgM, immunoglobulin M. This figure is available in black and white in print and in color online.

Figure 5: Viral dynamics measured in two cohorts in which blood was collected biweekly and, with viral measurements once HIV infection was detected. From Suthar 2015, J Infect Dis, Programmatic Implications of Acute and Early HIV Infection

4 Using CD4 and VL to make smarter assumptions

Let's look at the cases for which their infection window is 2+ years (either due to reporting never having a LNT or due to reporting a long-ago LNT) and for whom CD4 was measured within 30 days of diagnosis. For those cases, what is the breakdown of CD4 into <200, 200-500 and 500+?

The sample size is 5,148. Of those, 3,016 have non-missing testing history. Of those, 1,522 have an infection window greather than 2 years. Of those, 1,134 have a CD4 measurement within 30d of diagnosis. Thats about 22% of the entire sample and 37% of the non-missing testing histories.

Table 3: Distribution of CD4 counts among those with non-missing testing history, infection window of 2+y, and a CD4 count within 30d of diagnosis. Values are row percents

	Window Lengths (Years)		
	(0,200]	(200,500]	(500,1.72e+03]
MSM	38.2	38.6	23.3
non-MSM	47.5	34.6	17.9
Total	41.3	36.9	21.2

If we're roughly planning to shorten about half of the windows with CD4 count higher than 500, we're looking at editing half of 21.2% of the 37% of cases contributing to the TID, or about 4% of the cases contributing to the TID.

We should think not just about magnitude of impact, however, but also about increase of accuracy. Considering CD4 counts could help tell a story about the true undiagnosed time of those cases with long infection windows. A population with higher CD4 counts among those with no/long infection windows probably has more risk-based testing than one that has low CD4 counts–indicating more late diagnoses.

David's research has found that progressing from HIV to CD4<200, the AIDS definition, happens surprisingly frequently even among those reporting short infection windows. However, I think the key element to using CD4 data wisely is accounting for a distribution of set-point viral loads. As long as we only assume that half (or another data-driven estimate) of those with CD4>500 were likely to have been infected within the last 2 years, we allowing for the other half to still receive our conservative assumption. And if people are actually progressing to low CD4 counts *faster* than to CD4<500 in 2 years, then we're still being conservative.