Adding Stage of Infection to HIV Back-Calculation in WA State, 2005-2014

Martina Morris and Jeanette Birnbaum

April 22, 2016

1 Overview

This version of the report investigates the potential for CD4 count and viral load to inform our assumptions regarding infection windows. Total N=5,148.

2 Infection window lengths

Those with long infection windows (time between LNT and diagnosis) 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. 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, at 78%, by 2012-2014. It is important to note that this is a percent among those with either an LNT or reported no LNT, however. 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 - 18yrs	No LNT - Age minus 16	LNT	No LNT - 18yrs	No LNT - Age minus 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 testing history, distribution across several interval lengths: 0-1 year, 1-2 years, 2-5 years, 5-17 years, and 17-18 years. Of those with non-missing testing history, about 50% have a window length that is greater than 2 years.

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). The distributions of these measurements

Table 2: Distribution of sample with non-missing testing history. Values are percents of the total population. MSM and non-MSM rows add to the Total percents in the final row.

	Window Lengths (Years)				
	(0,1]	(1,2]	(2,5]	(5,17]	(17,18]
MSM	32.2	11.6	14.2	10.8	6.4
non-MSM	3.3	2.6	4.7	6.4	7.8
Total	35.6	14.2	18.9	17.2	14.2

are shown in Figures 5 and 6, stratified by LNT status: observed, No LNT with the 18-yr window, No LNT with the age-16 window, and missing LNT.

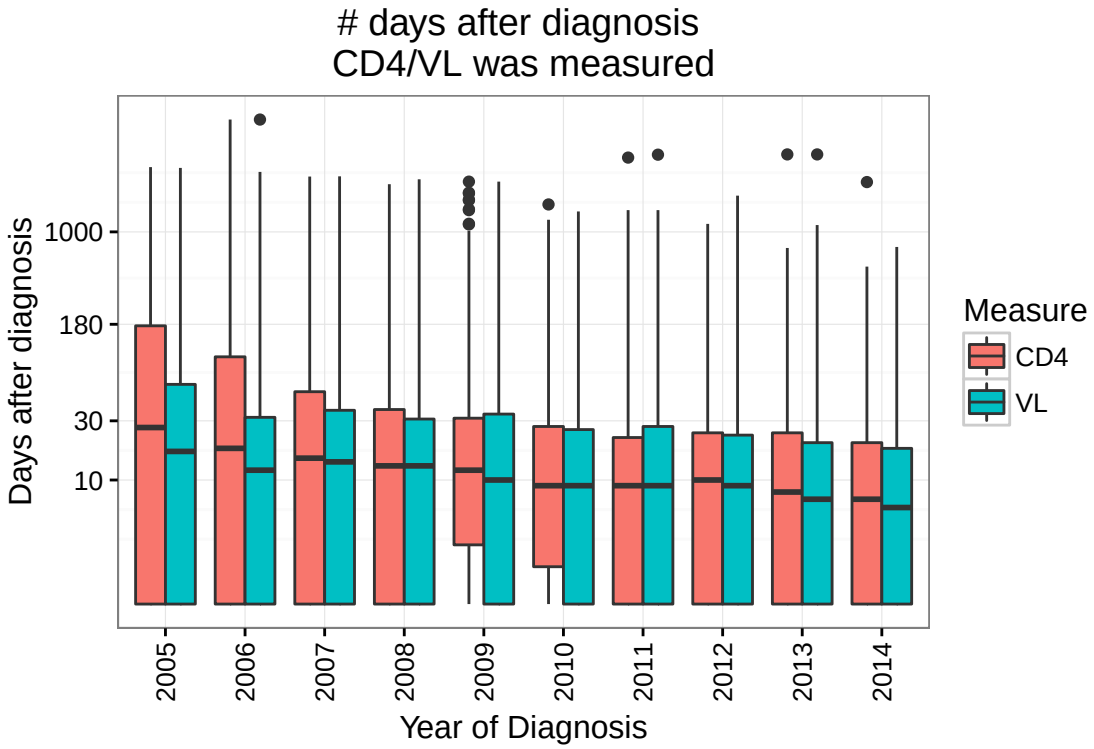

Figure 1: days after diagnosis that CD4 or VL was measured

3.2 Values for those with known recent infections

Figures 3 and 4 display the distribution of viral loads and CD4 counts for individuals who had an LNT within 6 mos of diagnosis and had their measurements taken the same day they were diagnosed.

3.3 Values by LNT subgroups

Figures 5 and 6 show the distribution of VL and CD4 counts (regardless of time of measurement) for the LNT subgroups: those with an observed LNT, those with missing LNT, those with no LNT who get the 18yr assumption, and those with no LNT who get the age-16 assumption.

3.4 Using CD4 and VL to identify recent infections

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.

To use some combination of very low or very high viral load plus a high CD4 count could be as an indicator of recent infection, we would need to consider:

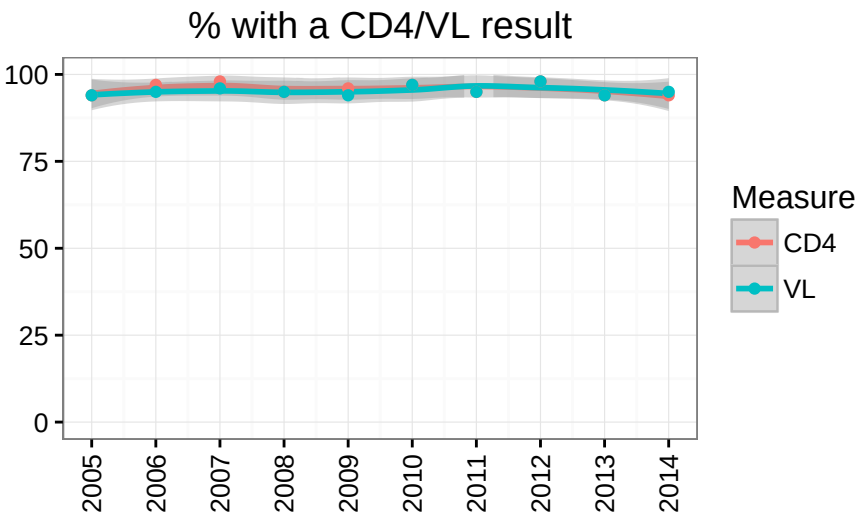

Figure 2: Percent of cases who have a non-missing result, by year

- 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?
- What cutoffs for CD4 count could suggest recent infection for an untreated case?
- What cutoffs for viral load could suggest recent infection for an untreated case? Figures 7 and 8 provide some data regarding the distribution of set point viral load. Perhaps VL's that are so low or high that they are unlikely to be set points could indicate recent infection.
- It seems there is much potential for error, given the diversity in immune response and measurement time. Would we need to validate a prediction model for recent infection and determine an acceptable probability-of-recent-infection cutoff for classifying cases as recent infections based on CD4 count, viral load and concurrent diagnosis? Is this in the literature already? I've seen these measures used in conjunction with an incidence assay like BED, but not alone. Although I haven't done a thorough search, yet.
- Thinking along the lines of a prediction model, one thing we could explore is imputing broad LNT categories for those with missing data rather than assuming MAR. This might inform us as to whether our MAR assumption is pushing our results towards greater or fewer undiagnosed.

It may also be worth considering a simulation study to determine the potential magnitude of impact on the undiagnosed results. For example, if about 50% of those with non-missing testing history have a window >2 years, and we were 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?

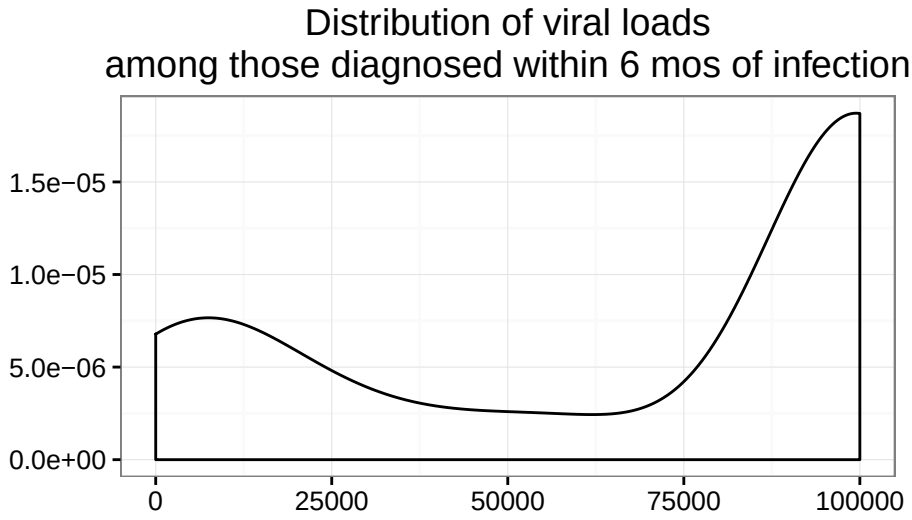

Figure 3: Distribution of first VL values among those with an LNT within 6 mos of diagnosis and measurements on their diagnosis day

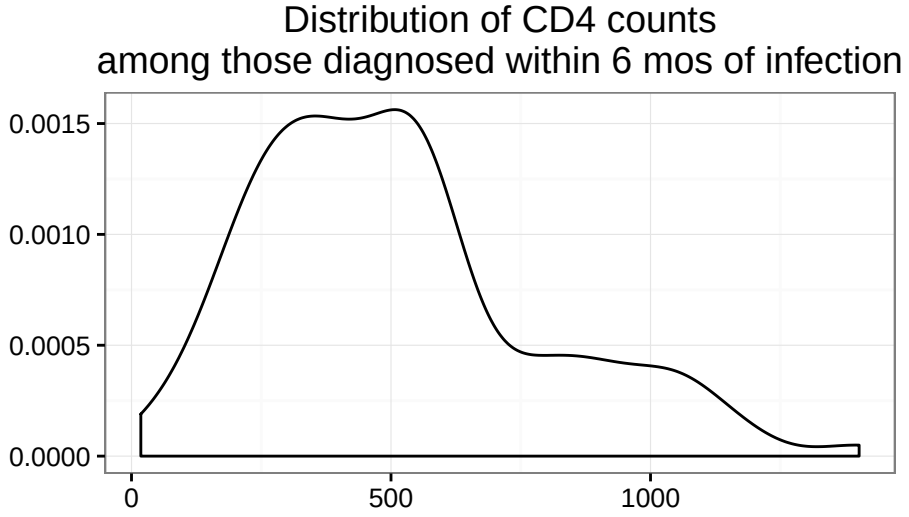

Figure 4: Distribution of first CD4 values among those with an LNT within 6 mos of diagnosis and measurements on their diagnosis day

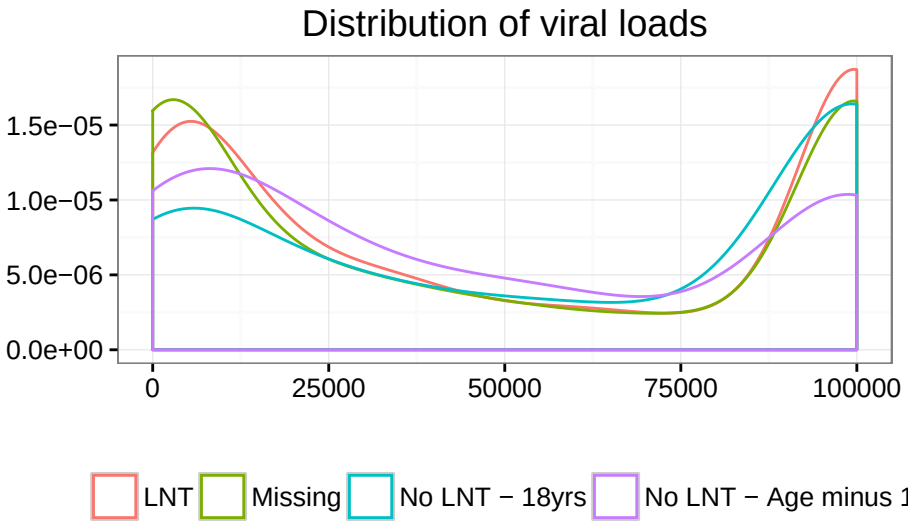

Figure 5: Distribution of first VL values

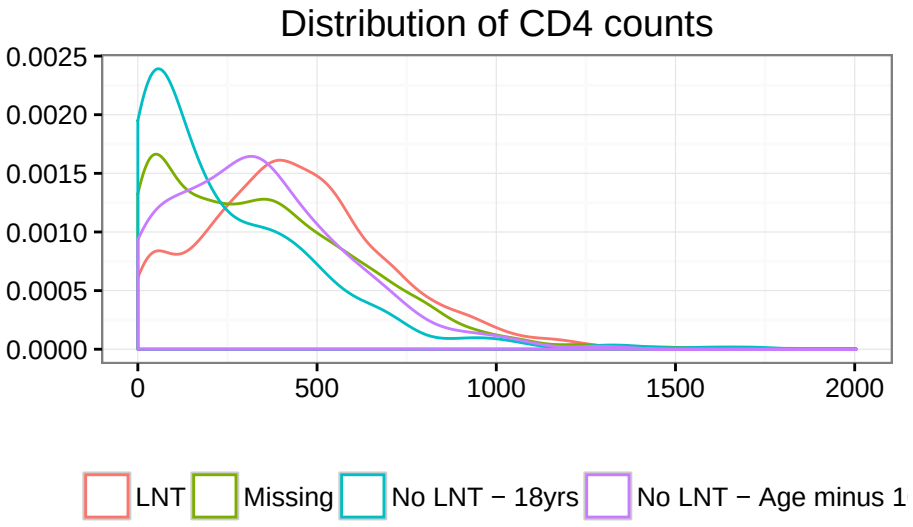

Figure 6: Distribution of first CD4 counts

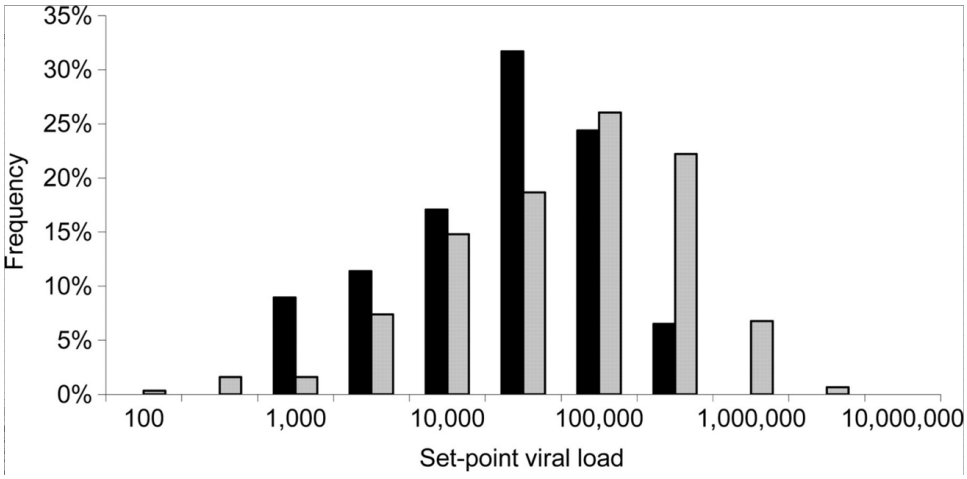

Figure 7: "The distribution of viral loads (copies per milliliter of peripheral blood) is plotted for untreated individuals in the Amsterdam Seroconverters Cohort (black bars) and the Zambian Transmission Study (7) (gray bars). The bars represent bins 0.5 log10 wide and are labeled by their midpoint viral load." From Fraser 2007, PNAS, Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis

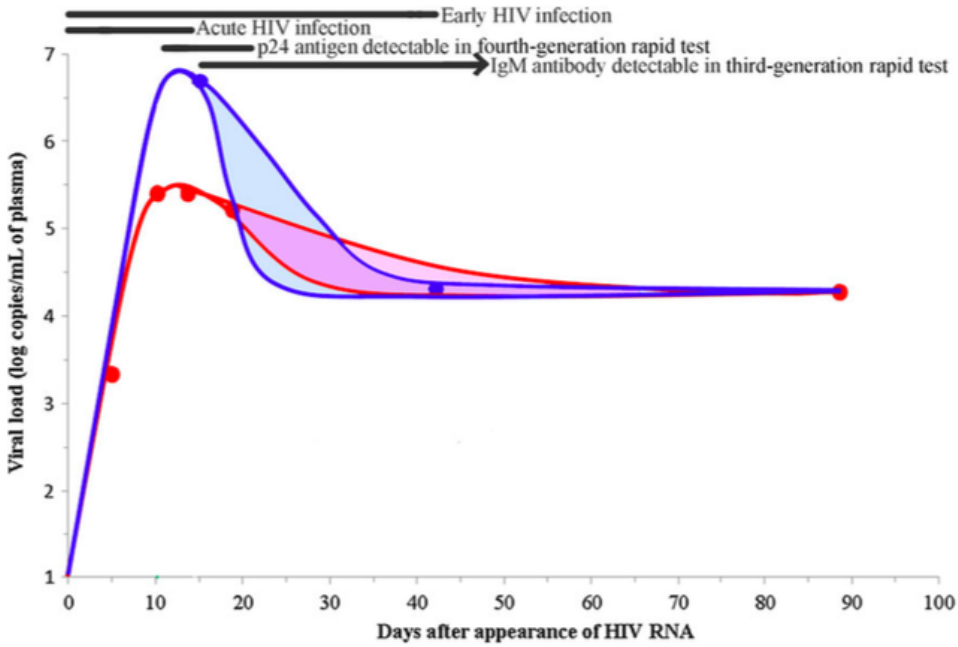

Figure 1. Appearance of diagnostic and viral markers during acute and early human immunodeficiency virus (HIV) infection. Median HIV loads are up to 100 days after the appearance of HIV RNA in individuals from Kenya, Uganda, Tanzania, and Thailand (top blue curve) [2] and the United States (bottom red 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 8: From Suthar 2015