Impact of using CD4 to inform undiagnosed estimates

Martina Morris and Jeanette Birnbaum

September 12, 2016

1 Overview

We use a simple approach to incorporate CD4 data into our testing history method for estimating undiagnosed HIV, creating a "CD4 Case" alternative to our Base Case. We compare the results and discuss implications.

2 Methods: Base Case versus CD4 Case

In the Base Case, we uniformly distribute probability of infection across the infection window.

In the CD4 Case, we uniformly distribute 50% of the probability within the median window defined by what the literature says about typical median times to CD4 bins.

However, in the CD4 Case, we prioritize testing history data over the CD4 data by defaulting to the Base Case when the testing history data indicate shorter times to infection. For example, if CD4 indicates that 50% probability should occur in the 1st 2 years of the window but the window is only 3 years long, we will keep the Base Case assumption that 50% probability of infection occurs within 1.5 years.

This is a very simplified way of reconciling conflicting assumptions. It seems reasonable given that the median times to CD4 count are pretty rough estimates–the data come from other cohorts, and there's population heterogeneity in set-point viral load and thus CD4 trends.

We could make some exceptions to the prioritization of testing history and the Base Case over CD4 data. Perhaps we should distribute probability further away from the time of diagnosis for those cases with low CD4s and long windows? I don't love this idea either. I wonder if we should be doing something for the cases with no testing history - using our CD4-based assumptions for them? The CD4 Case is pretty conservative...

3 Results: TIDs in WA

3.1 Base Case versus Base Case Alt

I coded up an alternate version of the Base Case that is identical theoretically but uses a different computational approach, one that can be easily altered to accommodate the CD4 Case.

So first let's check that Base Case Alt is the same as Base Case:

Table 1: Base Case TIDs using different computational approaches

Time	Original Base Case	Alternate Base Case
0.000	0.734	0.734
0.250	0.594	0.594
0.500	0.510	0.510
1.000	0.409	0.408
5.000	0.164	0.164
18.000	0.000	0.000

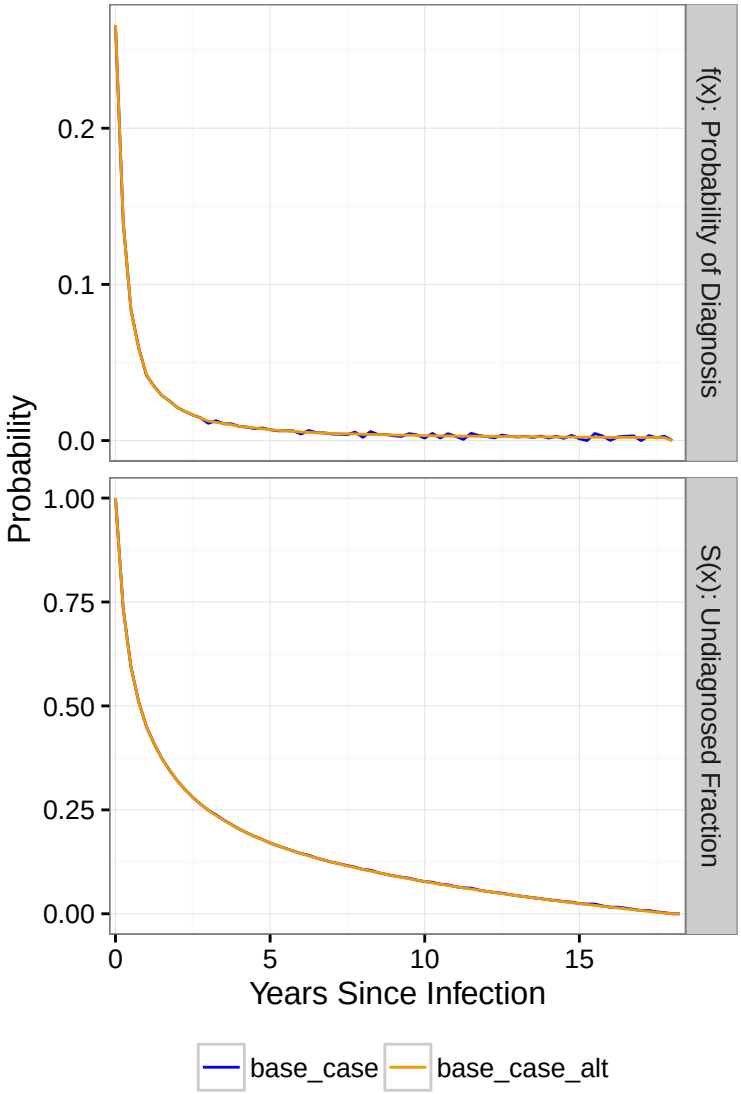

Figure 1: Base Case versus Base Case Alt TIDs

3.2 CD4 Case versus Base Case Alt

3.2.1 Confirm that code is correct

Here I use fake median windows of half the observed infection windows when computing the CD4 Case. If the code is correct, this fake CD4 Case should replicate the Base Case.

Table 2: Base Case versus Fake CD4 Case TIDs

Time	Alternative Base Case	Fake CD4 Case
0.000	0.734	0.734
0.250	0.594	0.594
0.500	0.510	0.510
1.000	0.408	0.408
5.000	0.164	0.164
18.000	0.000	0.000

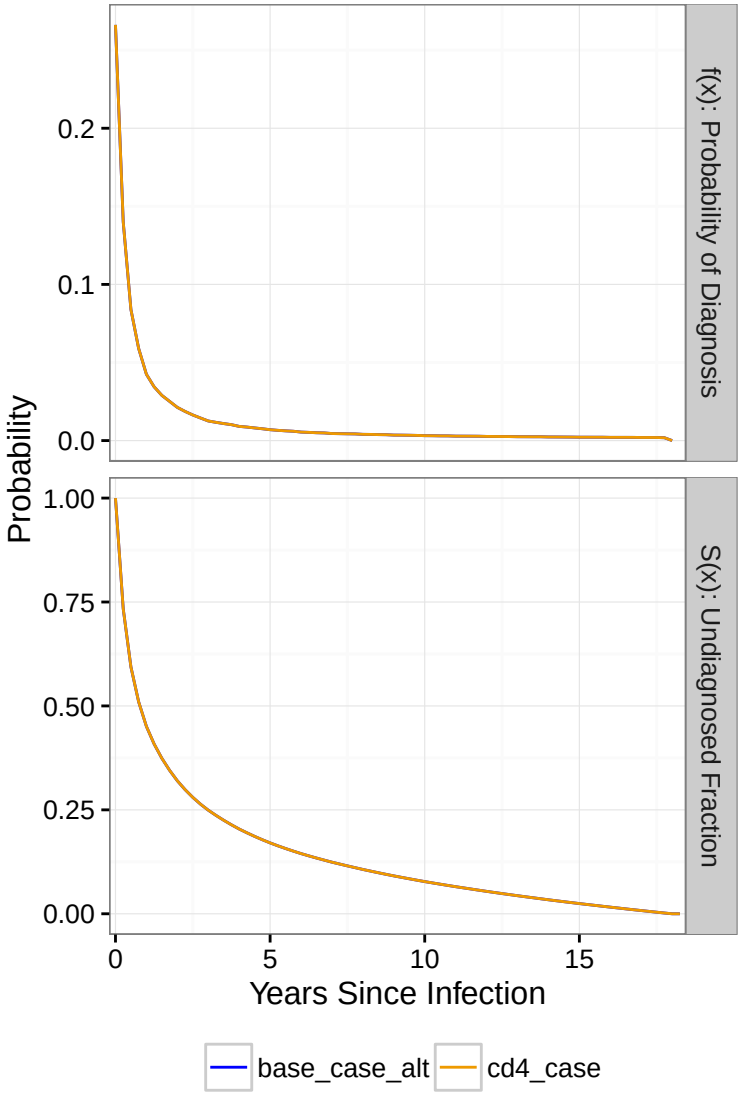

Figure 2: Base Case Alt versus Fake CD4 Case TIDs

3.2.2 Real CD4 Case

The code below assigns CD4-based medians to those cases for whom it is indicated (as discussed in Section

```
# ********************* Define our literature-based median times to infection by
# CD4 bin
(cd4meds <- data.frame(cd4lower = c(500, 350, 200), cd4upper = c(2000, 500, 350),
    medWindow = c(1.5, 4, 8)))

##   cd4lower cd4upper medWindow
## 1      500     2000       1.5
## 2      350      500       4.0
## 3      200      350       8.0

# ********************* Define who should get a CD4-based median
cd4breaks <- c(0, 200, 350, 500, 2000)
windowbreaks <- c(0, 3, 8, 16, 18)

dataf <- within(dataf, {
    # Non-missing testing history
    hasTestHist <- !is.na(everHadNegTest)
    # CD4 measured within 30d
    cd4within30 <- hasTestHist & !is.na(cd4_days) & cd4_days <= 30 & !is.na(firstcd4cnt)
    # Categories
    cd4cat <- cut(firstcd4cnt, breaks = cd4breaks, include.lowest = TRUE, right = FALSE)
})
with(dataf, table(hasTestHist))

## hasTestHist
## FALSE  TRUE
##  2132  3016

with(dataf, table(cd4within30))

## cd4within30
## FALSE  TRUE
##  2970  2178

# ********************* Assign medians

# Start with 1/2 of infPeriod, which is just the Base Case.  Update to CD4-based
# median if indicated by infPeriod (infection window)
dataf <- transform(dataf, medWindows = infPeriod/2)

for (i in 1:nrow(cd4meds)) {
    dataf <- transform(dataf, temp = cd4within30 & firstcd4cnt >= cd4meds[i, "cd4lower"] &
        firstcd4cnt < cd4meds[i, "cd4upper"] & infPeriod >= 2 * cd4meds[i, "medWindow"])
    cat("Row", i, "\n")
    print(summary(dataf$temp))

    dataf <- within(dataf, {
        medWindows[hasTestHist & cd4within30 & firstcd4cnt >= cd4meds[i, "cd4lower"] &
            firstcd4cnt < cd4meds[i, "cd4upper"] & infPeriod >= 2 * cd4meds[i, "medWindow"]] <- cd4meds[i,
            "medWindow"]
    })
}

## Row 1
##    Mode   FALSE    TRUE    NA's
## logical    4976     172       0
## Row 2
##    Mode   FALSE    TRUE    NA's
## logical    5082      66       0
## Row 3
##    Mode   FALSE    TRUE    NA's
## logical    5090      58       0

# 172+66+58 = 296 cases impacted; the other 6 may be right on the border and so
# they evaluate as == infPeriod/2
with(dataf, sum(medWindows != infPeriod/2, na.rm = TRUE))

## [1] 290
```

```
# Show old and new median windows
ddply(dataf, .(cd4cat), summarise, N_impacted = sum(medWindows != infPeriod/2, na.rm = TRUE),
    avgOldMedian = round(mean(infPeriod/2, na.rm = TRUE), 1), avgNewMedian = round(mean(medWindows,
        na.rm = TRUE), 1), Difference = avgOldMedian - avgNewMedian)

##        cd4cat N_impacted avgOldMedian avgNewMedian Difference
## 1     [0,200)          0          4.7          4.7        0.0
## 2   [200,350)         55          2.6          2.5        0.1
## 3   [350,500)         65          1.9          1.5        0.4
## 4 [500,2e+03]        170          1.6          1.0        0.6
## 5        <NA>          0          1.6          1.6        0.0
```

Now, here are the results:

Table 3: Base Case versus CD4 Case TIDs

Population	Time	Alternative Base Case	CD4 Case
All	0.000	0.734	0.731
	0.250	0.594	0.588
	0.500	0.510	0.501
	1.000	0.408	0.394
	5.000	0.164	0.150
	18.000	0.000	0.000
MSM	0.000	0.686	0.683
	0.250	0.526	0.521
	0.500	0.435	0.428
	1.000	0.331	0.319
	5.000	0.112	0.102
	18.000	0.000	0.000
non-MSM	0.000	0.880	0.875
	0.250	0.799	0.790
	0.500	0.736	0.723
	1.000	0.643	0.620
	5.000	0.320	0.297
	18.000	0.000	0.000

Table 4: Observed diagnoses and estimated quarterly incidence and undiagnosed counts over 2005-2014 and just 2014 in WA state, using CD4 Case

Year	Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2005-2014	Base Case	Undiagnosed Cases	1300	1391	1503	1476	1560	1592
2005-2014	CD4 Case	Undiagnosed Cases	1228	1315	1421	1395	1474	1504
2014	Base Case	Undiagnosed Cases	1300	1309	1322	1319	1332	1333
2014	CD4 Case	Undiagnosed Cases	1228	1237	1249	1247	1259	1260

Table 5: Impact of CD4 Case on mean undiagnosed estimates

Year	Base Case	CD4 Case	Difference	Percent Change
2005-2014	1476.0	1395.0	81.0	5.0
2014	1319.0	1247.0	72.0	5.0

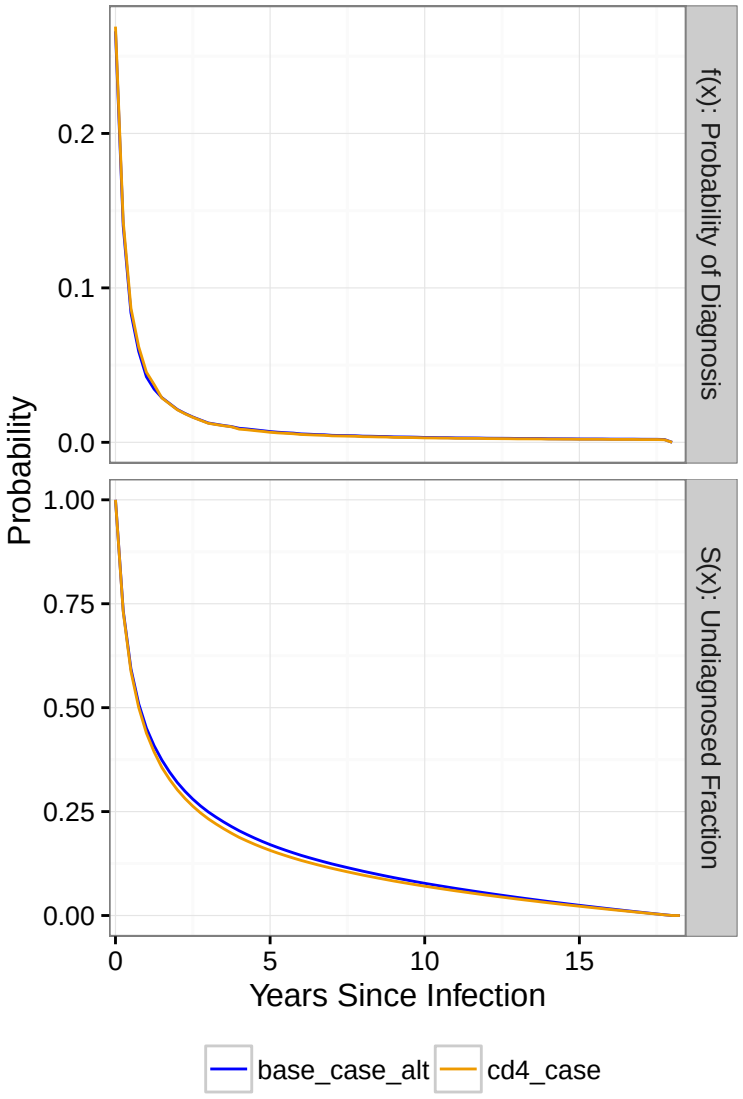

Figure 3: Base Case Alt versus CD4 Case TIDs

Table 6: Estimated true prevalence and the undiagnosed fraction for 2014 in WA state, using CD4 Case

Year	Diagnoses/Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2014.0	PLWHA	PLWHA				12691.0		
2014.0	Base Case	Undiagnosed Cases	1300.0	1309.0	1322.0	1319.0	1332.0	1333.0
2014.0	CD4 Case	Undiagnosed Cases	1228.0	1237.0	1249.0	1247.0	1259.0	1260.0
2014.0	Base Case	True Prevalence	13991.0	14000.0	14013.0	14010.0	14023.0	14024.0
2014.0	CD4 Case	True Prevalence	13919.0	13928.0	13940.0	13938.0	13950.0	13951.0
2014.0	Base Case	Undiagnosed Fraction (%)	9.3	9.3	9.4	9.4	9.5	9.5
2014.0	CD4 Case	Undiagnosed Fraction (%)	8.8	8.9	9.0	8.9	9.0	9.0