Impact of using CD4 to inform undiagnosed estimates

Martina Morris and Jeanette Birnbaum

September 13, 2016

1 Overview

This version of the report is for internal use only.

2 Methods: Proofs of Equivalence

2.1 Base Case versus Base Case Continuous

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. In addition, this approach treats the Base Case TID as a continuous function that has a unique value at every time point, rather than a step function in which the steps are determined by the observed infection windows.

So first let's check that Base Case Alt is the same as Base Case (Table 1 and Figure **??**).

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

2.2 Fake CD4 Case versus Base Case Continuous

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

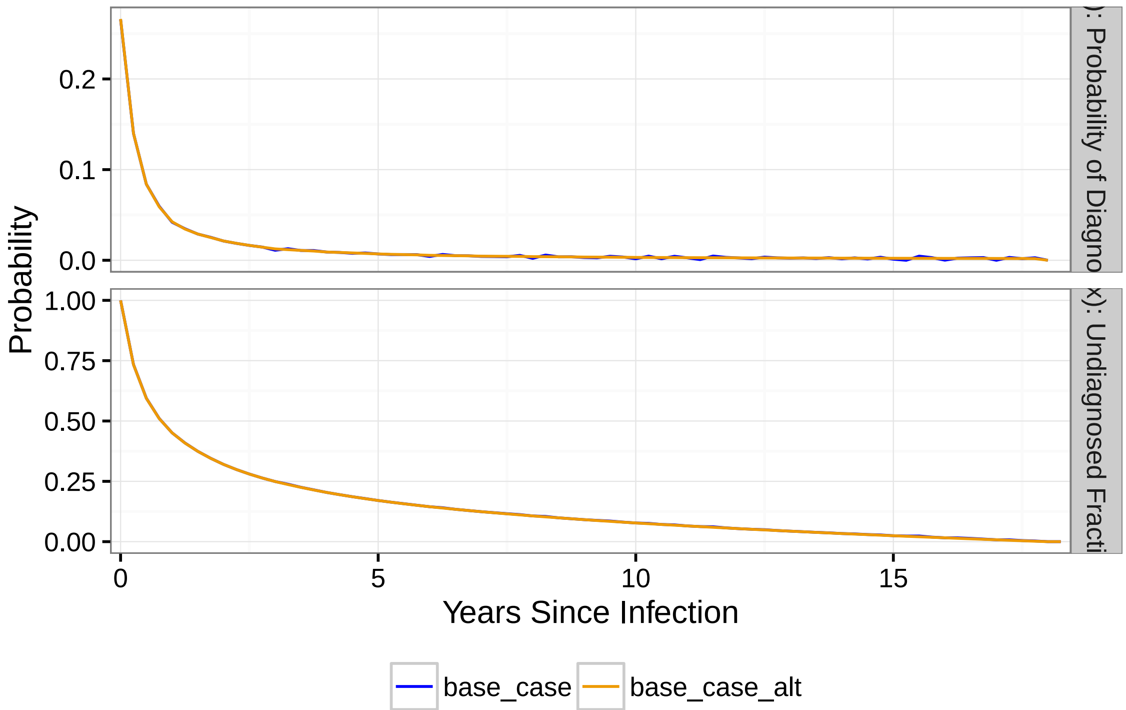


Figure 1: Base Case versus Base Case Continuous TIDs

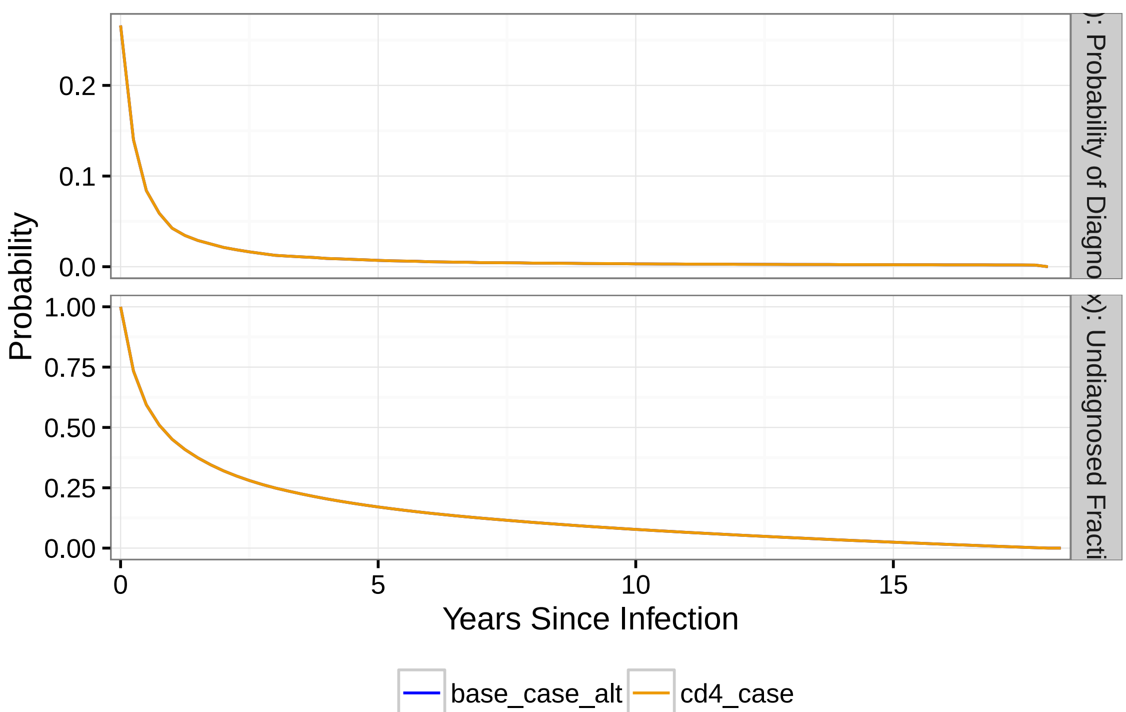


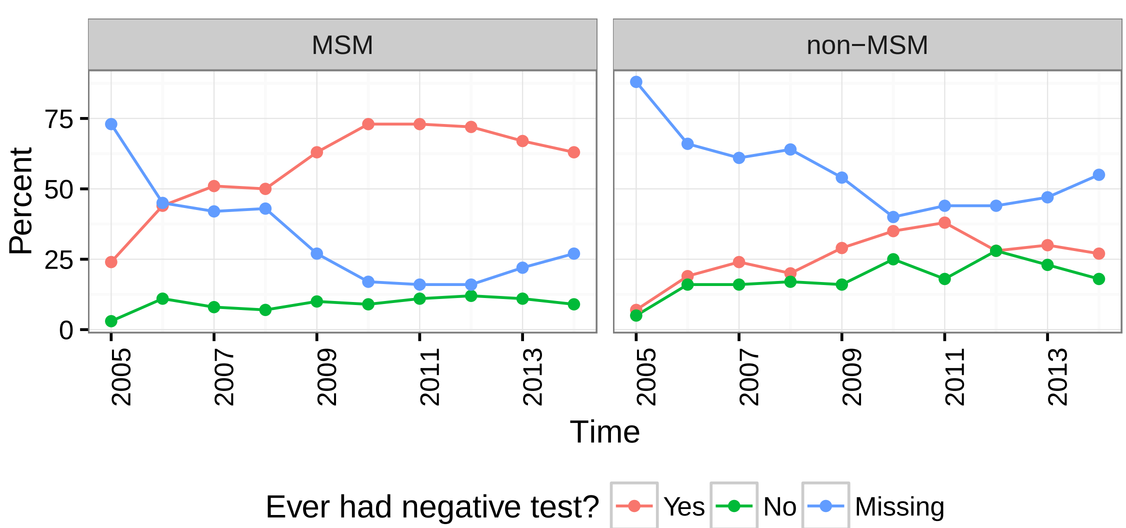
Figure 2: Fake CD4 Case versus Base Case Continuous TIDs

3 Methods: setting up real CD4-based medians

```r
# 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) 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))

# ********************* 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, impacted = 0)

for (i in 1:nrow(cd4meds)) {
    dataf <- transform(dataf, temp = cd4within30 & firstcd4cnt >= cd4meds[i, "cd4lower"] &
        firstcd4cnt < cd4meds[i, "cd4upper"] & infPeriod >= 2 * cd4meds[i, "medWindow"])
    dataf <- transform(dataf, impacted = ifelse(temp == 1, 1, impacted))
    dataf <- within(dataf, {
        medWindows[hasTestHist & cd4within30 & firstcd4cnt >= cd4meds[i, "cd4lower"] &
            firstcd4cnt < cd4meds[i, "cd4upper"] & infPeriod >= 2 * cd4meds[i, "medWindow"]] <- cd4meds[i,
            "medWindow"]
    })
}

# Was expecting 296 cases impacted; need to find the 6
with(dataf, sum(medWindows != infPeriod/2, na.rm = TRUE))

## [1] 290

# Ok
with(dataf, table(mode2, impacted))

##          impacted
## mode2        0    1
##   MSM     3232  171
##   non-MSM 1620  125
```

```
with(dataf, table(mode2, impacted)/rowSums(table(mode2, impacted)))

##          impacted
## mode2              0          1
##   MSM     0.94975022 0.05024978
##   non-MSM 0.92836676 0.07163324

# Now look among the 3016 with testing history
with(subset(dataf, !is.na(everHadNegTest)), table(mode2, impacted))

##          impacted
## mode2        0    1
##   MSM     2098  171
##   non-MSM  622  125

with(subset(dataf, !is.na(everHadNegTest)), table(mode2, impacted)/rowSums(table(mode2,
    impacted)))

##          impacted
## mode2             0         1
##   MSM     0.9246364 0.0753636
##   non-MSM 0.8326640 0.1673360

# Show old and new median windows AMONG the 3016 contributing to testing
# histories
ddply(subset(dataf, !is.na(everHadNegTest)), .(mode2, cd4cat), summarise, N_impacted = sum(impacted),
    avgOldMedian = round(mean(infPeriod/2, na.rm = TRUE), 1), avgNewMedian = round(mean(medWindows,
        na.rm = TRUE), 1), Difference = avgOldMedian - avgNewMedian)

##      mode2      cd4cat N_impacted avgOldMedian avgNewMedian Difference
## 1      MSM     [0,200)          0          3.9          3.9        0.0
## 2      MSM   [200,350)         24          1.9          1.9        0.0
## 3      MSM   [350,500)         35          1.4          1.2        0.2
## 4      MSM [500,2e+03]        112          1.3          0.8        0.5
## 5      MSM        <NA>          0          1.0          1.0        0.0
## 6  non-MSM     [0,200)          0          6.0          6.0        0.0
## 7  non-MSM   [200,350)         34          4.4          4.2        0.2
## 8  non-MSM   [350,500)         31          3.9          2.8        1.1
## 9  non-MSM [500,2e+03]         60          3.0          1.8        1.2
## 10 non-MSM        <NA>          0          2.8          2.8        0.0

############################################################ Estimate TIDs
```

4 Results: Reminder of testing histories in MSM versus non-MSM



5 Results: Total percent of probability re-assigned

Looking within mode subgroups, the total percent of probability re-assigned confirms that the relative impact on TID probability is greater for non-MSM.

```
# Compute BC probability assigned within the median window: just 1/infPeriod
# times the medWindow. Then compare that to 0.5, which is how much the CD4 Case
# assigs within the median window

dataf <- within(dataf, {
    medProbBC <- (medWindows) * (1/infPeriod)
    probReassigned <- 0.5 - medProbBC
})

summary(dataf$probReassigned)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.0000  0.0000  0.0000  0.0218  0.0000  0.4166    2132

ddply(subset(dataf, !is.na(everHadNegTest)), .(mode2), summarise, totalReassigned = sum(probReassigned,
    na.rm = TRUE), propReassigned = sum(probReassigned)/length(probReassigned))

##     mode2 totalReassigned propReassigned
## 1     MSM        39.32961     0.01733345
## 2 non-MSM        26.55218     0.03554508

# Look separately among impacted cases
ddply(subset(dataf, !is.na(everHadNegTest) & impacted == 1), .(mode2), summarise,
    totalReassigned = sum(probReassigned), propReassigned = sum(probReassigned)/length(probReassigned))

##     mode2 totalReassigned propReassigned
## 1     MSM        39.32961      0.2299977
## 2 non-MSM        26.55218      0.2124174
```

6 Results: TIDs in MSM versus non-MSM

6.1 TID values at select times

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

6.2 Full TID curves

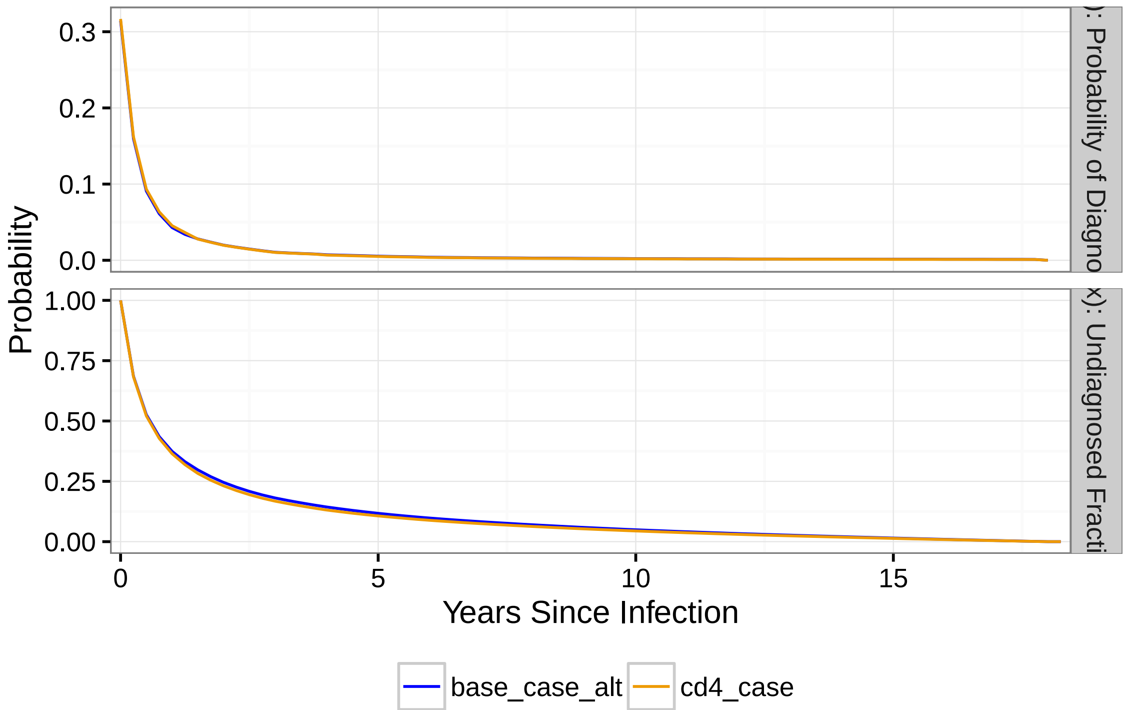


Figure 3: MSM: Real CD4 Case versus Base Case Continuous TIDs

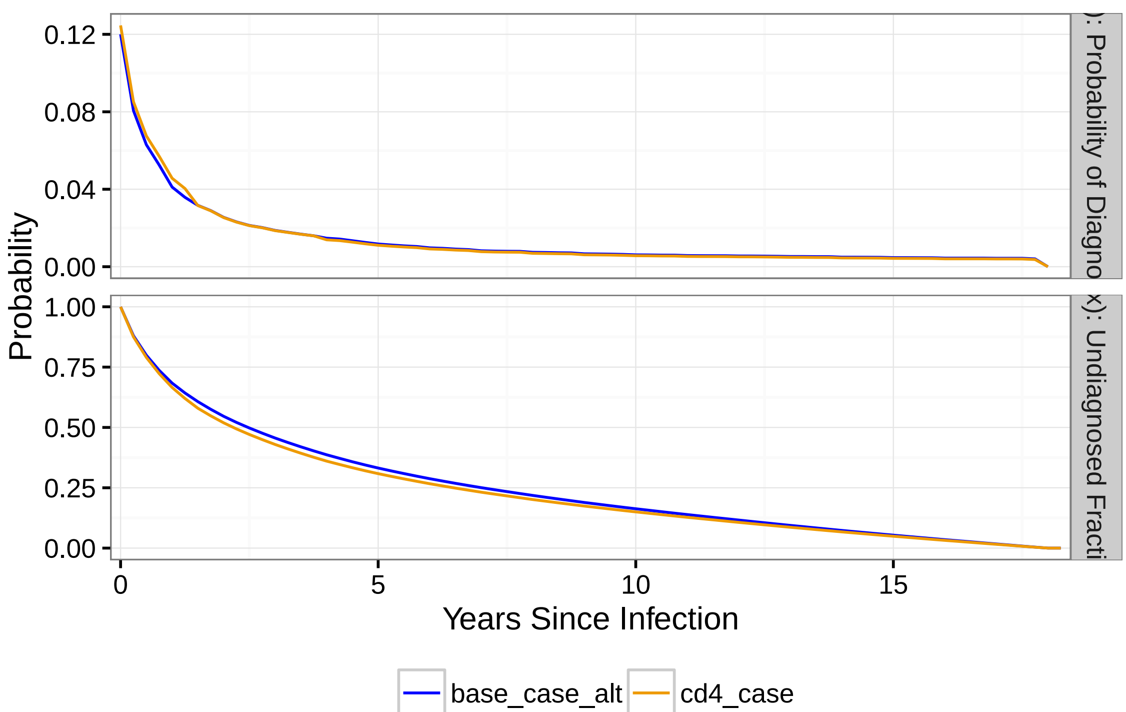


Figure 4: non-MSM: Real CD4 Case versus Base Case Continuous TIDs

6.3 Median time undiagnosed

This is a little annoying to get, I don't think we need it to tell the story

6.4 Mean time undiagnosed

Here's where it gets interesting. Our eyes see the absolute difference, but the relative difference tells a different story.

```
##      mode bc_auc cd4_auc ratio diff
## 1     MSM    1.8     1.7  0.94 0.11
## 2 non-MSM    4.4     4.1  0.94 0.25
```

7 Results: Incidence and undiagnosed cases

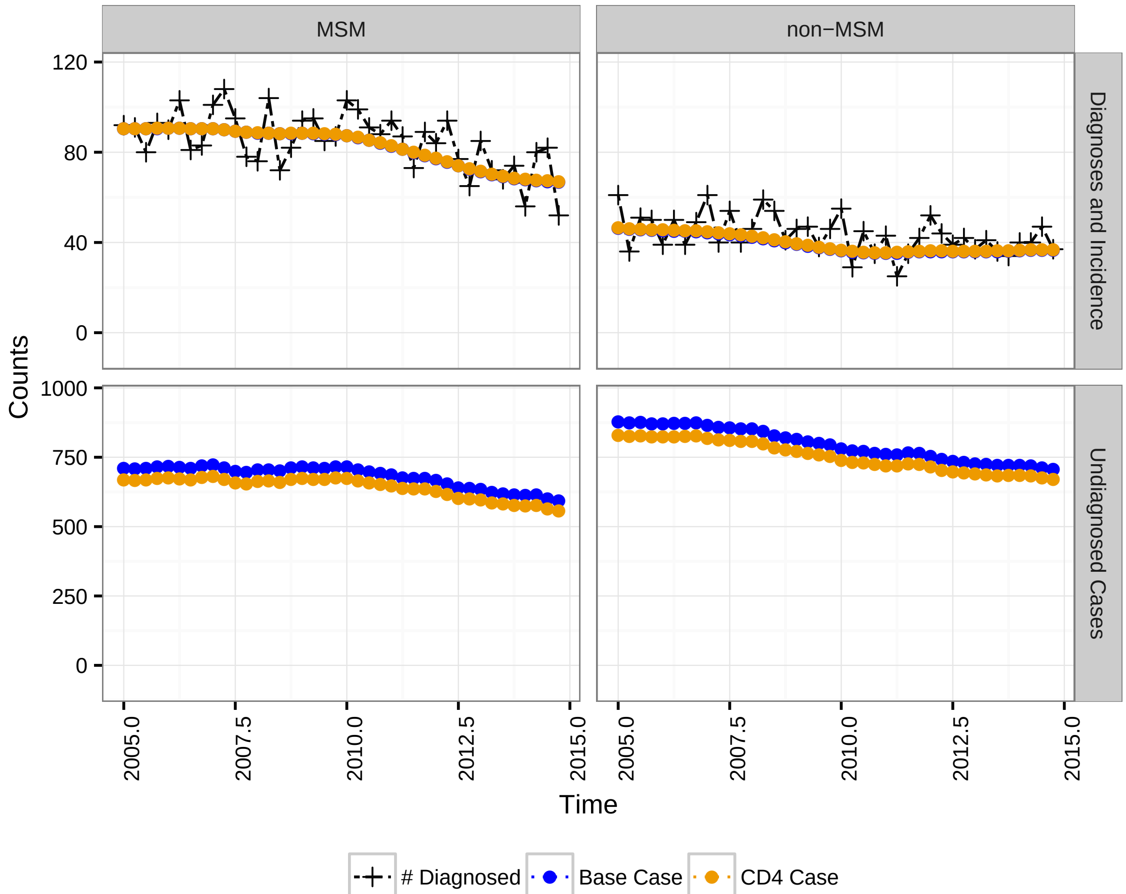


Figure 5: Diagnoses, incidence and undiagnosed counts for MSM and non-MSM

8 Results: Undiagnosed cases and undiagnosed fractions

Table 4: Impact of CD4 Case on mean undiagnosed estimates and fractions

Group	Year	Estimate	Base Case	CD4 Case	Difference	Percent Change
Total	2010.0	Undiagnosed Cases	1474.0	1393.0	81.0	5.0
Total	2010.0	Undiagnosed Fraction (%)	11.2	10.6	0.6	5.0
Total	2011.0	Undiagnosed Cases	1440.0	1361.0	79.0	5.0
Total	2011.0	Undiagnosed Fraction (%)	10.9	10.4	0.5	5.0
Total	2012.0	Undiagnosed Cases	1390.0	1314.0	76.0	5.0
Total	2012.0	Undiagnosed Fraction (%)	10.5	9.9	0.6	6.0
Total	2013.0	Undiagnosed Cases	1345.0	1271.0	74.0	6.0
Total	2013.0	Undiagnosed Fraction (%)	9.9	9.4	0.5	5.0
Total	2014.0	Undiagnosed Cases	1319.0	1247.0	72.0	5.0
Total	2014.0	Undiagnosed Fraction (%)	9.4	8.9	0.5	5.0
MSM	2010.0	Undiagnosed Cases	701.9	662.2	39.7	6.0
MSM	2010.0	Undiagnosed Fraction (%)	7.4	7.0	0.4	5.0
MSM	2011.0	Undiagnosed Cases	677.7	638.8	38.9	6.0
MSM	2011.0	Undiagnosed Fraction (%)	7.2	6.8	0.4	6.0
MSM	2012.0	Undiagnosed Cases	649.5	611.5	38.0	6.0
MSM	2012.0	Undiagnosed Fraction (%)	6.8	6.4	0.4	6.0
MSM	2013.0	Undiagnosed Cases	622.2	585.0	37.2	6.0
MSM	2013.0	Undiagnosed Fraction (%)	6.4	6.1	0.3	5.0
MSM	2014.0	Undiagnosed Cases	604.7	568.4	36.3	6.0
MSM	2014.0	Undiagnosed Fraction (%)	6.2	5.8	0.4	6.0
non-MSM	2010.0	Undiagnosed Cases	772.0	731.1	40.9	5.0
non-MSM	2010.0	Undiagnosed Fraction (%)	20.6	19.7	0.9	4.0
non-MSM	2011.0	Undiagnosed Cases	761.9	722.1	39.8	5.0
non-MSM	2011.0	Undiagnosed Fraction (%)	20.3	19.5	0.8	4.0
non-MSM	2012.0	Undiagnosed Cases	740.6	702.2	38.4	5.0
non-MSM	2012.0	Undiagnosed Fraction (%)	19.7	18.8	0.9	5.0
non-MSM	2013.0	Undiagnosed Cases	723.1	685.8	37.3	5.0
non-MSM	2013.0	Undiagnosed Fraction (%)	18.2	17.4	0.8	4.0
non-MSM	2014.0	Undiagnosed Cases	714.3	678.2	36.1	5.0
non-MSM	2014.0	Undiagnosed Fraction (%)	17.1	16.4	0.7	4.0