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

Martina Morris and Jeanette Birnbaum

February 19, 2016

1 Overview

This report contains the initial results comparing the standard testing history model results for WA State undiagnosed estimates to an extended model incorporating stage of infection data. This version of the report focuses on the impact of BED results rather than dual diagnosis, in order to understand why the BED data have only a minor impact.

This report differs from version 4 in that we now use the BED window of 162 days to define the infection window for BED+ cases with missing testing history (did not answer Yes or No to "Have you ever had a negative test?"). Previously we left those infection windows as missing, effectively treating them as missing at random with respect to the stage of infection groups by which we stratified the estimation. This report also includes violin plots of the distributions of reported and imputed (for those reporting no testing history) infection windows. Additionally, the definition of dual diagnosis was corrected from using timeAids-timeHIVDx<1 year to using ≤1 year, which increased the % with dual diagnoses from 33 to 34%.

2 Understanding the Data

2.1 Construction of analytic sample

Data from the advanced HIV/AIDS reporting system (eHARS) and the CDC treatment and testing history questionnaire (HIS) provided records for 26,134 HIV cases in WA state.[1]

Figure 1 diagrams the construction of the analytic sample. We first restricted to cases diagnosed in WA state in the years 2005-2014. We further excluded cases diagnosed at age 16 or younger if their date of last negative test was missing, because the assumptions we use when date of last negative test is missing are not applicable to this age group.

The final sample includes 5,176 cases. In the 2014 report there were 4744 cases in the final sample across diagnosis years 2005-2013. Of the additional 447 diagnoses reported in 2014 eligible for this anaylsis, 432 met all our inclusion criteria.

2.2 Frequencies of and trends in BED and dual diagnosis

Table 1 shows the sample breakdown by BED status and dual diagnosis (DD) status over 2005-2014, where a dual diagnosis is defined as AIDS within 1 year of the HIV diagnosis. From the "Total" column we see that the BED+ population is 17% of the sample. In 2014, the BED+ drop to 7% of the sample (Table 2).

Within the BED+, 11% are DD+ (Table 1). We are assuming these are true recent infections of individuals who have a fast disease progression, although this is a matter for further research. The current assumption maximizes the impact of the BED information.

Of those with a missing BED result, 43% have a dual diagnosis, which is substantially higher than the percentages among those with a BED+ or BED- result (Table 1). This suggests those with a BED result do not well-represent those with missing BED.

[1] Provided by Jason Carr, Washington State Department of Health, June 2015

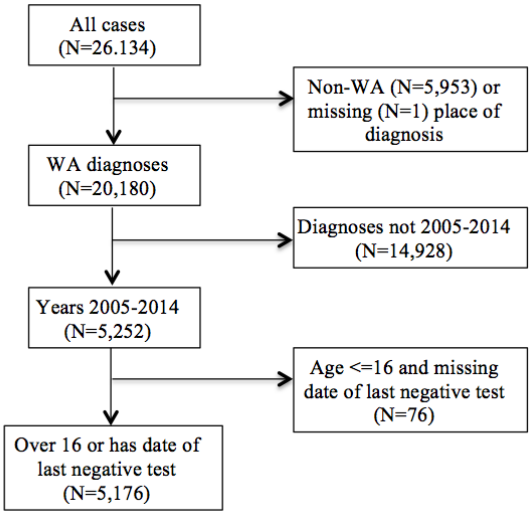

Figure 1: Construction of analytic sample

Table 1: Cross-tabulation of BED and dual diagnoses, 2005-2014. Row percents show the percent of dual diagnoses (DD) within each BED group or the total sample (last row)

	DD+		DD-		Total	
	N	Row %	N	Row %	N	Col %
BED +	110	12	770	88	880	17
BED -	402	28	1009	72	1411	27
BED Miss	1249	43	1636	57	2885	56
Total	1761	34	3415	66	5176	100

Table 2: Cross-tabulation of BED and dual diagnoses, 2014 only

	DD+		DD-		Total	
	N	Row %	N	Row %	N	Col %
BED +	2	6	29	94	31	7
BED -	2	6	32	94	34	8
BED Miss	119	32	257	68	376	85
Total	123	28	318	72	441	100

2.3 Time trends

The graphs in Figure 2 show how the percentages of cases with missing BED (left) and cases with dual diagnosis (right) vary over time.

2.4 Monthly distribution of dual diagnoses

2.5 Relationship between BED and testing histories

Tables 3 and 4 show the breakdown of testing histories within BED groups, overall and for MSM vs non-MSM transmission groups, respectively. Individuals who have a BED+ result are much more likely to also have reported a LNT (71%) than those with a BED- result (50%) or those with missing BED (37%).

We see this correlation in the MSM/non-MSM subgroups as well (Table 4). It is not ideal, since having a BED+ result would be most useful for those who reported never having an LNT. Among non-MSM, however, there are relatively more BED+ with no LNT than among MSM, so the BED+ information will be more influential for them. The BED+ non-MSM only comprise 4% of the population, however, so the impact of this group on the total estimates will be small.

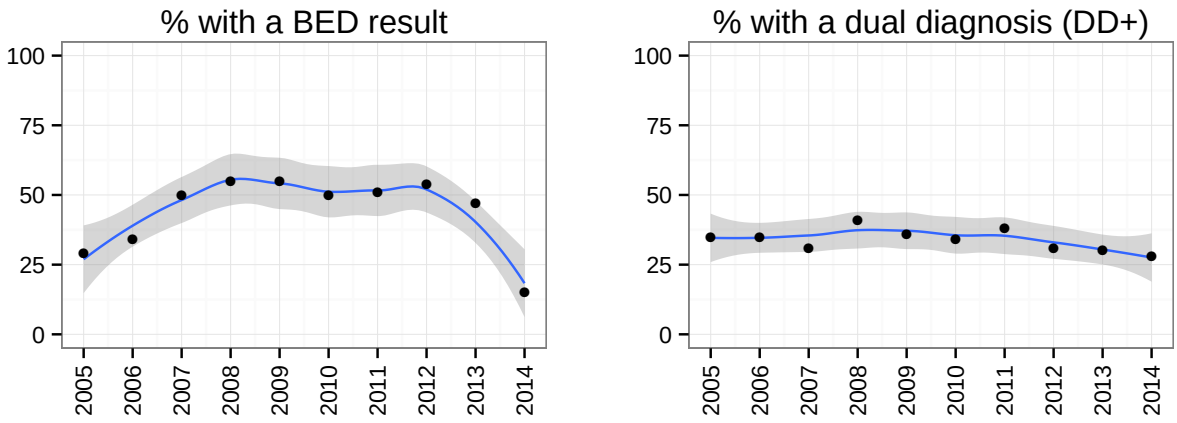

Figure 2: Percent of cases who have non-missing BED (left) or a dual diagnoses (right), by year

Table 3: Testing history responses by BED status. Column % sums to 100 across all rows. Availability of testing history data within each subgroup level is shown as row percents

				Ever Had a Negative Test		
Characteristic	Subgroup	N	Column %	% Yes	% No	% Missing
BED Result	BED +	880	17	71	7	21
	BED -	1411	27	50	16	34
	BED Miss	2885	56	37	12	52
Total		5176	100	46	12	42

3 Stage of infection impact on infection window

3.1 Numbers of cases affected by a BED+ result

The top bar plot in Figure 4 shows the breakdown of the cases with non-missing testing history into those with a LNT and those who reported no LNT. For each of these groups, the lower violin plots show the distribution of the infection windows using a log scale. The violin plots have boxplots overlayed to show, when possible, the medians and interquartile ranges. The infection windows for those reporting no LNT come from the assumption $min(age - 16, 18)$.

This plot shows that for the >75% of cases with non-missing testing history who report a LNT, the median infection window is just over 1 year and about 25% of cases are diagnosed within 6 mo nths of their LNT. For those reporting no LNT, at least half of them get the maximum infection window of 18 years.

Figure 5 displays the same information stratified by MSM vs non-MSM. MSM (who comprise 66% of the sample) have a median infection window of 1 year when the LNT is reported. For non-MSM, the median infection window is over twice that. Non-MSM also have more than double the proportion of reports of no LNT.

There are three types of BED+ cases whose BED+ result influences our model: those who report having no LNT, those who report a LNT that is greater than the BED window (162 days), and those who have a missing answer to the testing history question. In all cases, we shorten the infection window to the BED window.

Table 5 summarizes the distribution of cases by type of change. Those "Changed by being BED+" who have a LNT are not only BED+ but also have a reported infection window that is greater than the BED window of 162 days.

3.2 Degree of impact of the BED+ result

We saw above that 12.9% of all cases will be affected by the BED+ result. Of these 667 cases, 413 have reported LNTs that will be shortened to the BED window, 65 have assumed LNTs of $min(age - 16, 18)$ that

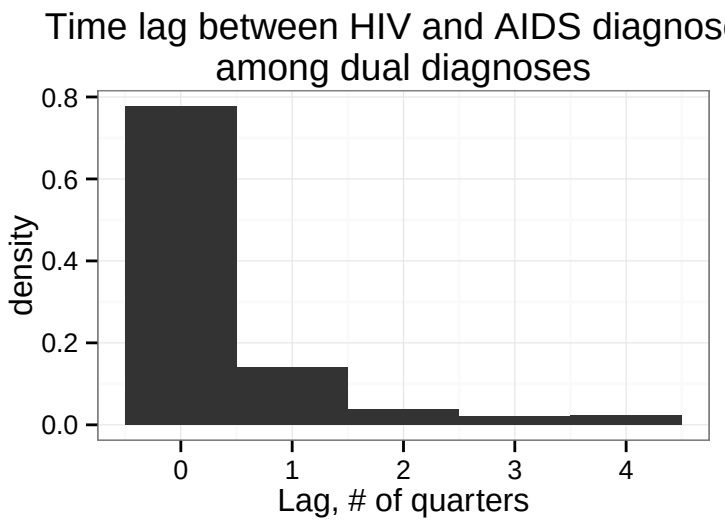

Figure 3: Density of time lag between HIV and AIDS diagnoses among the dually diagnosed, measured in quarters (e.g, 4 quarters = 1 year)

Table 4: Testing history responses by BED status, separately for MSM and non-MSM. Column % sums to 100 across all rows. Availability of testing history data within each subgroup level is shown as row percents

				Ever Had a Negative Test		
Transmission Mode	BED Result	N	Column %	% Yes	% No	% Missing
MSM	BED +	685	13	80	5	16
	BED -	872	17	63	12	24
	BED Miss	1846	36	46	9	45
non-MSM	BED +	195	4	42	17	41
	BED -	539	10	29	21	50
	BED Miss	1039	20	20	16	64
Total		5176	100	46	12	42

will be shortened to the BED window, and 189 have missing testing history and will be given an infection window of the BED window.

The infection windows of the BED+ before modification reflect the reported and assumed LNTs. When we perform the modifications specified above, we see that the long tail of infection windows is reigned in (Table 6). However, the median drops only slightly when the BED+ information is used, indicating only minor changes in the lower 50% of the data. The first quartile increases due to the inclusion of cases with missing testing history who are given an infection window of the BED window.

The 1st quartile infection window increases slightly due to the use of the BED window for the 189 missing testing histories that were given a BED window for their infection window.

Table 7 shows how this translates into the various cases for the TID estimates.

4 Four stage-of-infection groups

We originally wanted to stratify our estimation by all six subgroups defined by the three BED categories and two dual diagnosis categories. However, due to small samples of BED+/- DD+ (Table 3), we aggregated into four subgroups: BED+, BED-, BED missing DD+ (BEDmDD+), and BED missing DD- (BEDmDD-).

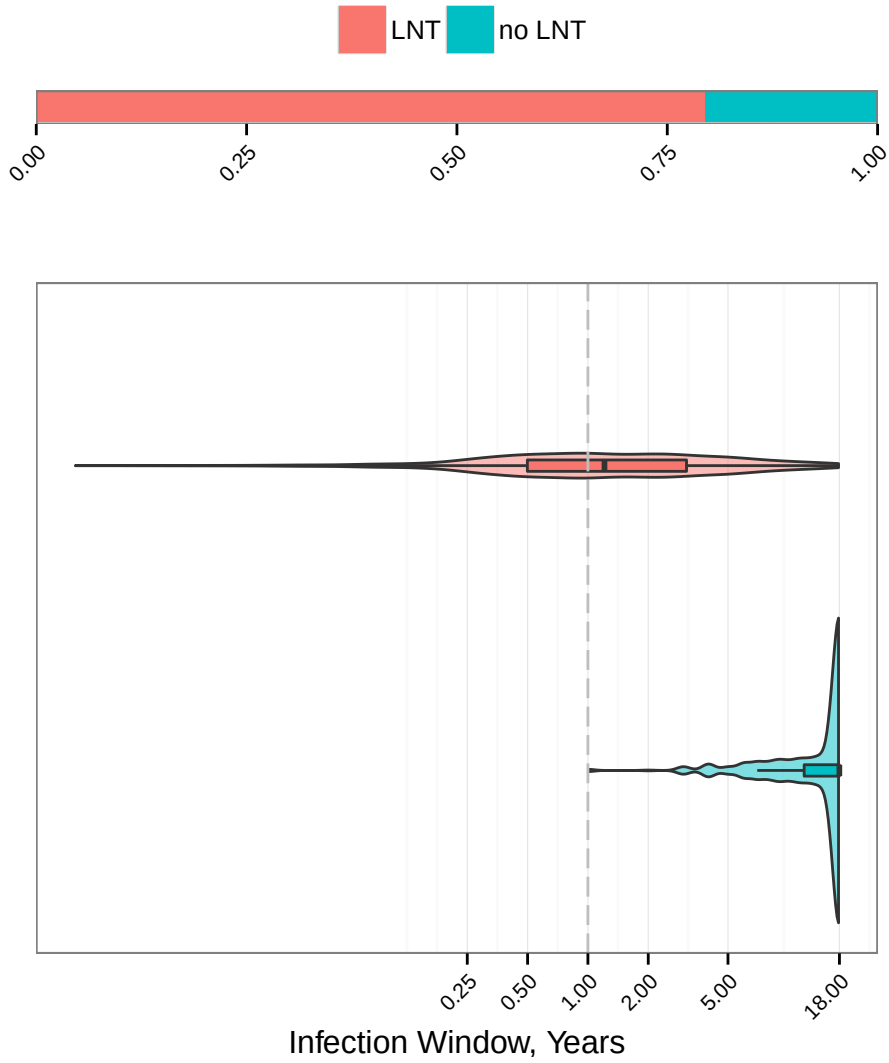

Figure 4: Distribution of infection windows among the 3013 cases with non-missing testing history

5 Undiagnosed Results

5.1 Time from infection to diagnosis (TID)

Figure 6 shows the original and modified base case TIDs for each of the four stage-subgroups.

5.2 Unstratified, Without-Stage Results

The estimated incidence and undiagnosed counts for each scenario are shown as quarterly counts in Figure 8 and summarized over all quarters in Table 8. These results are not stratified by any group, although we do have a version of the total results that reflects stratification by MSM and non-MSM.

5.3 Stratified, Without- and With-Stage Results

When we run the model allowing stage (so far, just BED) to impact the TID, we also stratify by stage subgroups in order for the missing testing histories to be missing conditional on stage subgroup.

Quarterly incidence and undiagnosed counts are plotted in Figure 9. The summary results over 2005-2015

Table 5: Impact of BED+ results on the entire sample, by testing history (TH) status. Cases have no change if they are not BED+ or they are BED+ but with a reported LNT that is shorter than the BED window of 162 days. Cases are impacted by BED+ results if they are BED+ with an LNT farther back than 162 days, they report no LNT, or they have missing testing history.

	No Change		Changed by BED+		
	N	Row %	N	Row %	Total
No LNT	554	90	65	10	619
LNT	1981	83	413	17	2394
Missing TH	1974	91	189	9	2163
Total	4509	87	667	13	5176

Table 6: Among the BED+, summary of infection windows pre- and post- modification by BED+, in years

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Original BED+ infection windows	0.003	0.366	0.748	2.514	1.970	17.980	189.000
After modifying windows for BED+	0.003	0.444	0.444	0.397	0.444	0.444	0.000

and for 2014 alone are given in Table 10. The mean without- and with-stage undiagnosed estimates for those two time periods are compared in Table 11. Table 12 shows the 2014 undiagnosed fraction results as well.

Regarding the impact on uncertainty, from Table 11 we can additionally calculate that the undiagnosed range in 2014 was 1,274-2,493 and adding stage decreased that to 1,192-2,335, which amounts to a difference of 76 cases. From Table 12 that amounts to a decrease from 7.2% to 6.9% for the range of the mean undiagnosed fraction.

Table 7: Among the BED+, the fraction remaining undiagnosed at quarter-year time steps, using the different cases. Time indicates the left/lower bound of the time interval

Time	Original Base Case	Modified Base Case	Original Upper Bound	Modified Upper Bound
0.000	0.579	0.365	0.861	0.891
0.250	0.386	0.000	0.641	0.000
0.500	0.290	0.000	0.499	0.000
1.000	0.201	0.000	0.337	0.000

Table 8: Observed diagnoses and estimated quarterly incidence and undiagnosed counts over 2005-2014 in WA state

Diagnoses/Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
# Diagnosed	Diagnoses	91	120	129	129	140	163
Base Case	Incidence	108	115	126	124	134	138
Base Case	Undiagnosed Cases	1236	1303	1401	1371	1435	1461
Upper Bound	Incidence	105	109	121	120	130	135
Upper Bound	Undiagnosed Cases	2473	2575	2739	2704	2818	2870

Table 9: Estimated true prevalence and the undiagnosed fraction in WA state, limited to just 2014

Year	Diagnoses/Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2014.0	PLWHA	PLWHA				12691.0		
2014.0	Base Case	Undiagnosed Cases	1236.0	1243.0	1253.0	1251.0	1261.0	1262.0
2014.0	Base Case	True Prevalence	13927.0	13934.0	13944.0	13942.0	13952.0	13953.0
2014.0	Base Case	Undiagnosed Fraction (%)	8.9	8.9	9.0	9.0	9.0	9.0
2014.0	Upper Bound	Undiagnosed Cases	2473.0	2480.0	2494.0	2492.0	2507.0	2509.0
2014.0	Upper Bound	True Prevalence	15164.0	15171.0	15185.0	15183.0	15198.0	15200.0
2014.0	Upper Bound	Undiagnosed Fraction (%)	16.3	16.3	16.4	16.4	16.5	16.5

Table 10: Observed diagnoses and estimated quarterly incidence and undiagnosed counts over 2005-2014 and just 2014 in WA state, using stage-subgroup strata

Year	Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2005-2014	Base Case	Undiagnosed Cases	1267	1371	1435	1403	1443	1466
2005-2014	Base Case using Stage	Undiagnosed Cases	1185	1287	1405	1350	1415	1434
2005-2014	Upper Bound	Undiagnosed Cases	2478	2633	2844	2751	2868	2899
2005-2014	Upper Bound using Stage	Undiagnosed Cases	2320	2441	2673	2595	2742	2758
2014	Base Case	Undiagnosed Cases	1267	1272	1277	1274	1278	1279
2014	Base Case using Stage	Undiagnosed Cases	1185	1190	1194	1192	1196	1198
2014	Upper Bound	Undiagnosed Cases	2478	2489	2500	2493	2501	2502
2014	Upper Bound using Stage	Undiagnosed Cases	2320	2330	2340	2335	2343	2346

Table 11: Impact of using BED result to modify the TID on mean undiagnosed estimates

Year	Case	With Stage	Without Stage	Difference	Percent Change
2005-2014	Base Case	1350.0	1403.0	-53.0	-4.0
2005-2014	Upper Bound	2595.0	2751.0	-156.0	-6.0
2014	Base Case	1192.0	1274.0	-82.0	-6.0
2014	Upper Bound	2335.0	2493.0	-158.0	-6.0

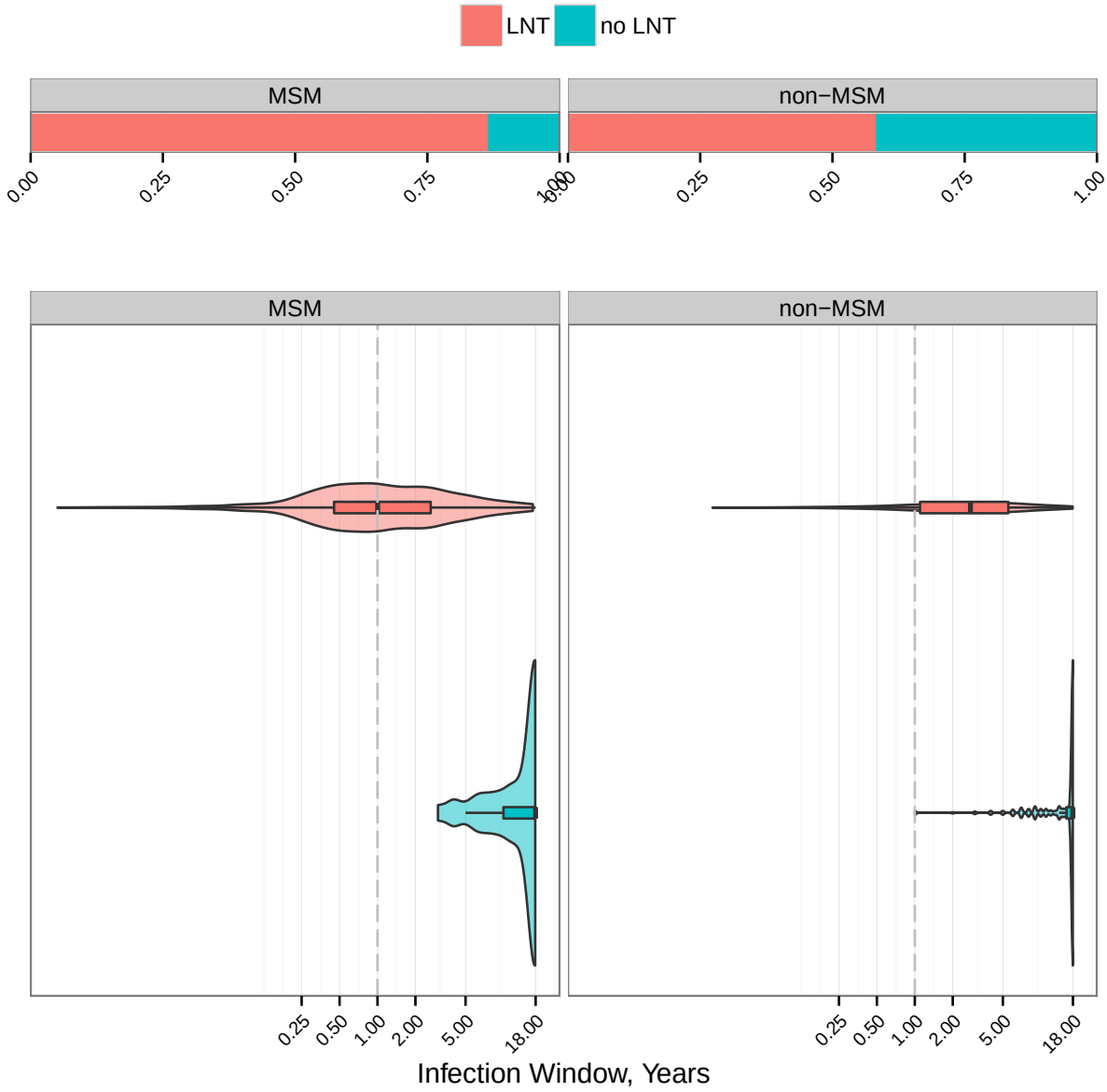

Figure 5: Distribution of infection windows among the 3013 cases with non-missing testing history, stratified by MSM vs non-MSM

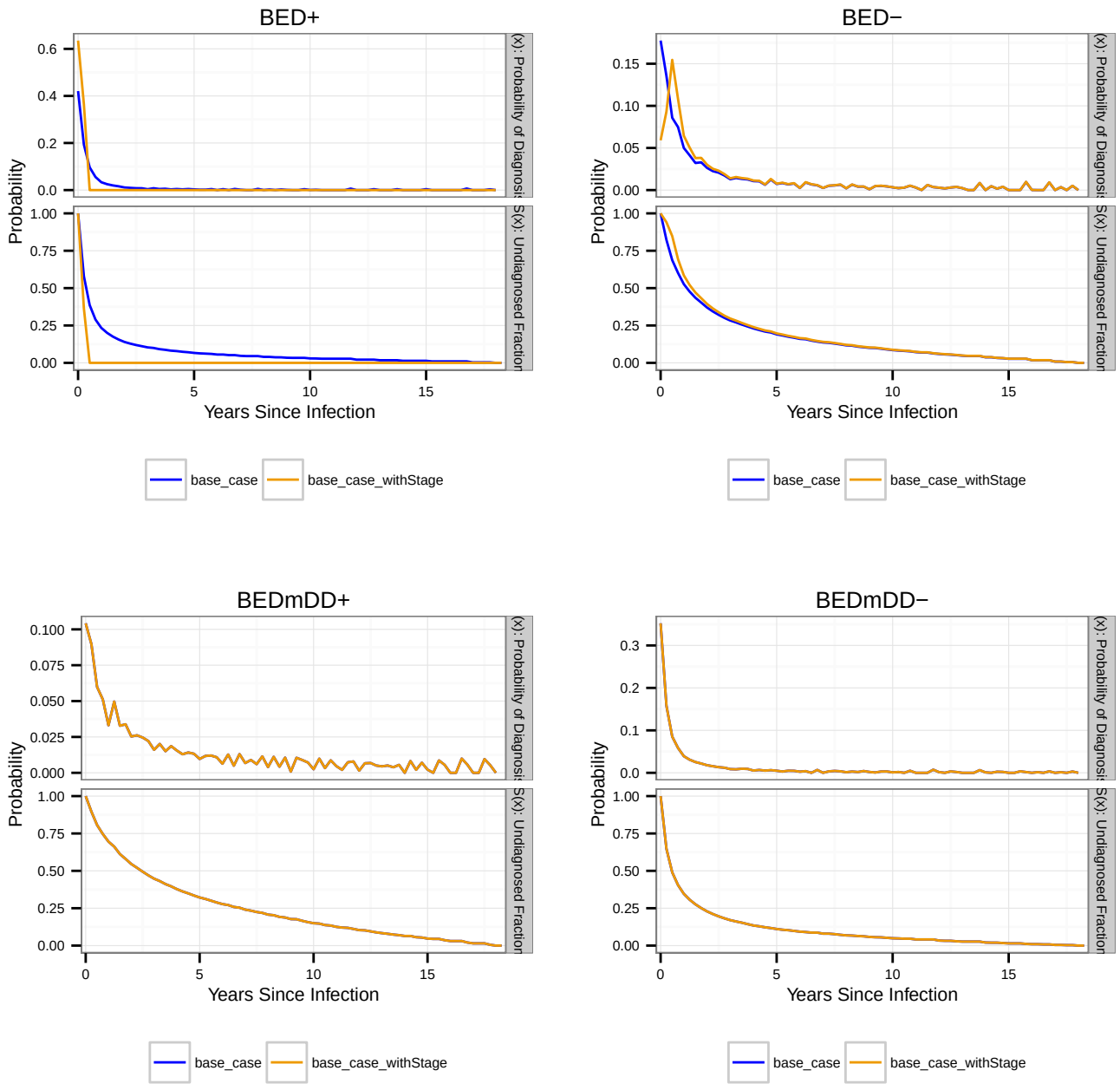

Figure 6: Time from infection to diagnosis (TID) for base case without and with stage, 4 groups

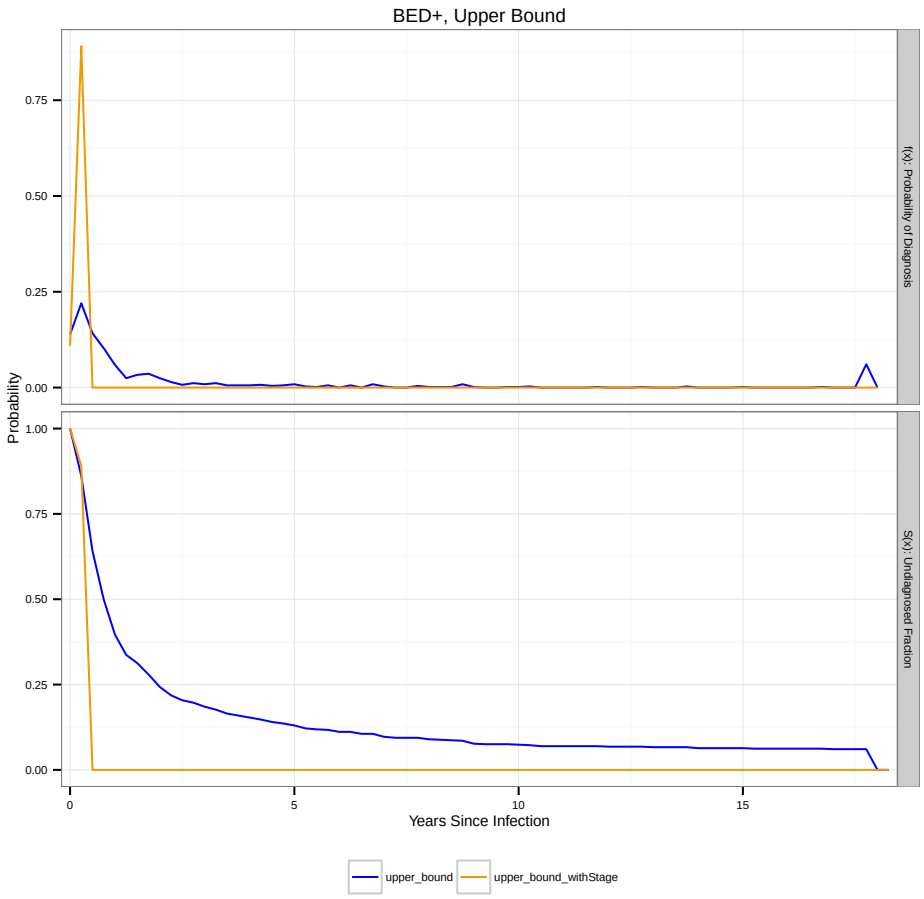

Figure 7: Time from infection to diagnosis (TID) for upper bound without and with stage, BED+

Table 12: Estimated true prevalence and the undiagnosed fraction for 2014 in WA state, using stage-subgroup strata

Year	Diagnoses/Case	Estimate	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2014.0	PLWHA	PLWHA				12691.0		
2014.0	Base Case	Undiagnosed Cases	1267.0	1272.0	1277.0	1274.0	1278.0	1279.0
2014.0	Base Case using Stage	Undiagnosed Cases	1185.0	1190.0	1194.0	1192.0	1196.0	1198.0
2014.0	Upper Bound	Undiagnosed Cases	2478.0	2489.0	2500.0	2493.0	2501.0	2502.0
2014.0	Upper Bound using Stage	Undiagnosed Cases	2320.0	2330.0	2340.0	2335.0	2343.0	2346.0
2014.0	Base Case	True Prevalence	13958.0	13963.0	13968.0	13965.0	13969.0	13970.0
2014.0	Base Case using Stage	True Prevalence	13876.0	13881.0	13885.0	13883.0	13887.0	13889.0
2014.0	Upper Bound	True Prevalence	15169.0	15180.0	15191.0	15184.0	15192.0	15193.0
2014.0	Upper Bound using Stage	True Prevalence	15011.0	15021.0	15031.0	15026.0	15034.0	15037.0
2014.0	Base Case	Undiagnosed Fraction (%)	9.1	9.1	9.1	9.1	9.1	9.2
2014.0	Base Case using Stage	Undiagnosed Fraction (%)	8.5	8.6	8.6	8.6	8.6	8.6
2014.0	Upper Bound	Undiagnosed Fraction (%)	16.3	16.4	16.5	16.4	16.5	16.5
2014.0	Upper Bound using Stage	Undiagnosed Fraction (%)	15.5	15.5	15.6	15.5	15.6	15.6

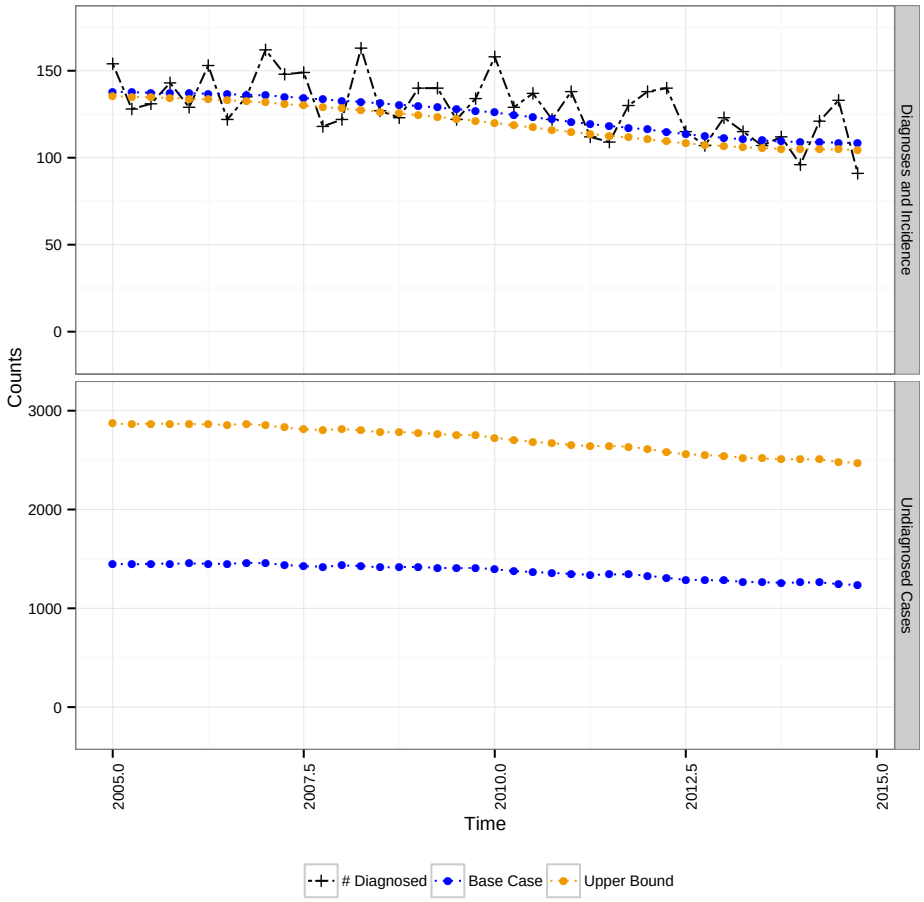

Figure 8: Observed diagnoses and estimated quarterly and undiagnosed counts over 2005-2014 in WA state

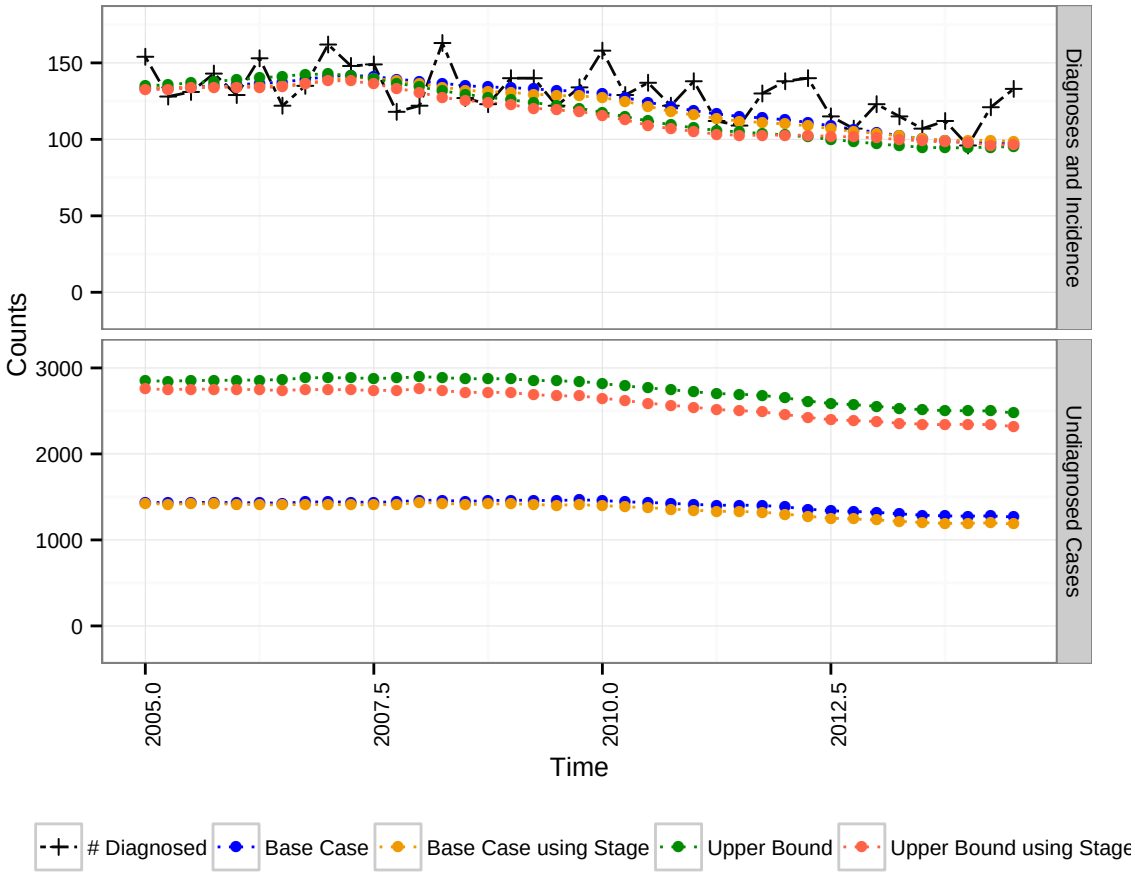

Figure 9: Observed diagnoses and estimated quarterly and undiagnosed counts over 2005-2014 in WA state, using stage-subgroup strata

A Appendix of Additional Results

Table 13: 2005-2014 breakdown of MSM vs non-MSM diagnoses

	mode2
MSM	3403
non-MSM	1773

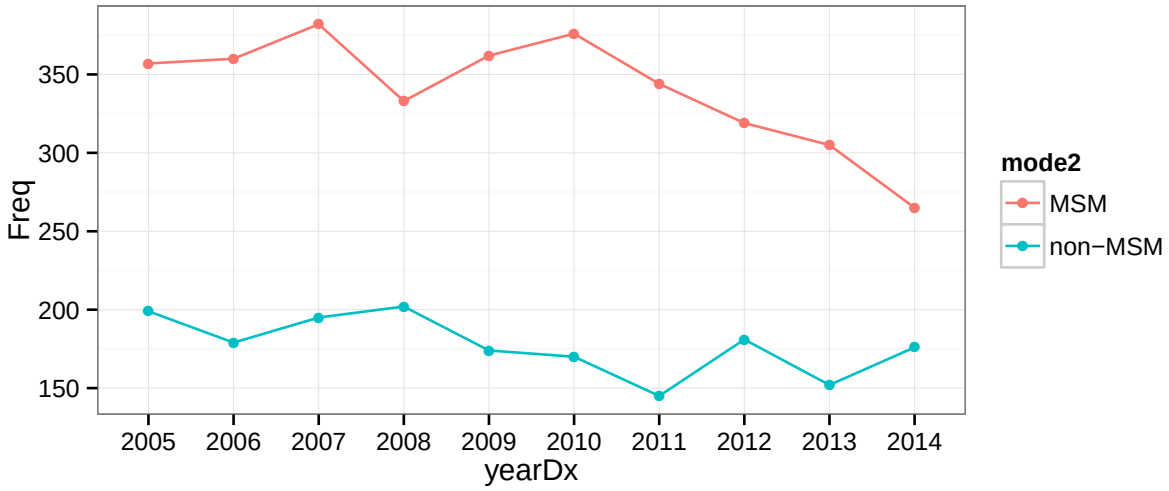

Figure 10: MSM vs non-MSM diagnoses over time

Table 14: LNT by BED

	BED +	BED -	BED Miss
No LNT	65	222	332
LNT	626	707	1061
Missing	189	482	1492

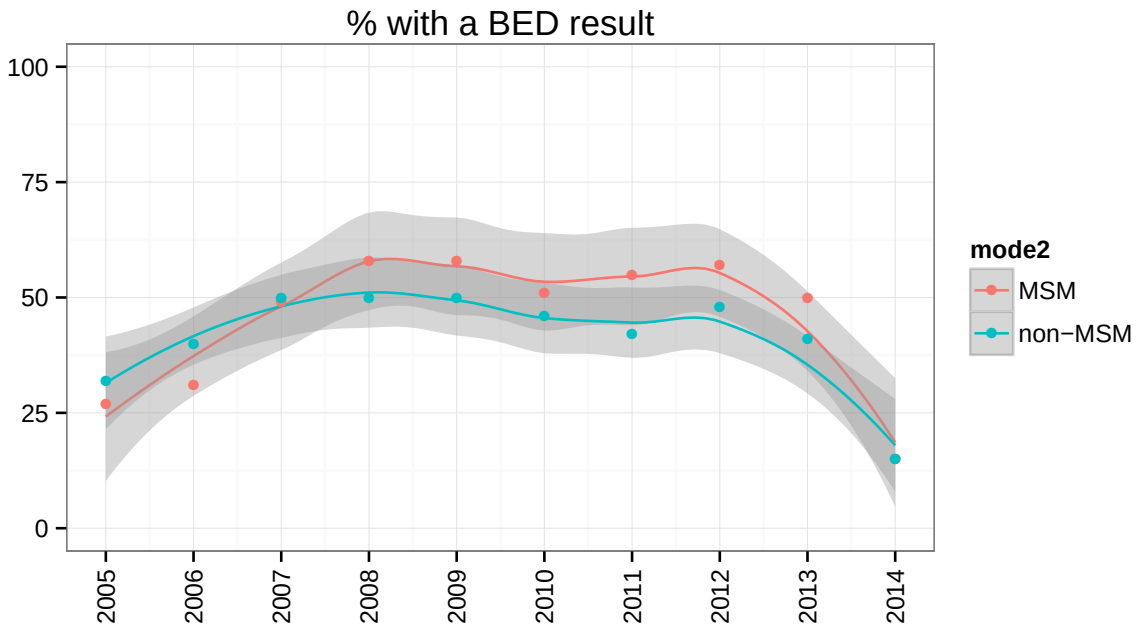

Figure 11: Percent of cases who have a non-missing BED result, by year

Table 15: LNT by BED+ Impact

	No Change	Changed by BED+	Total
No LNT	554	65	619
LNT	1981	413	2394
Missing	1974	189	2163
Total	4509	667	5176

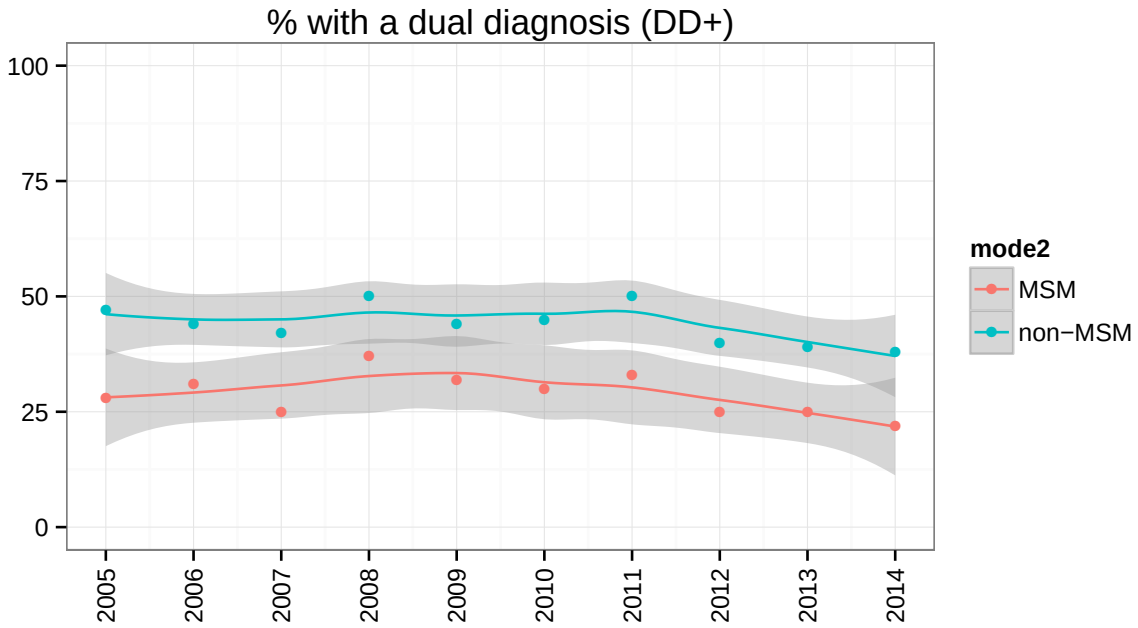

Figure 12: Percent of cases who have a dual diagnoses, by year

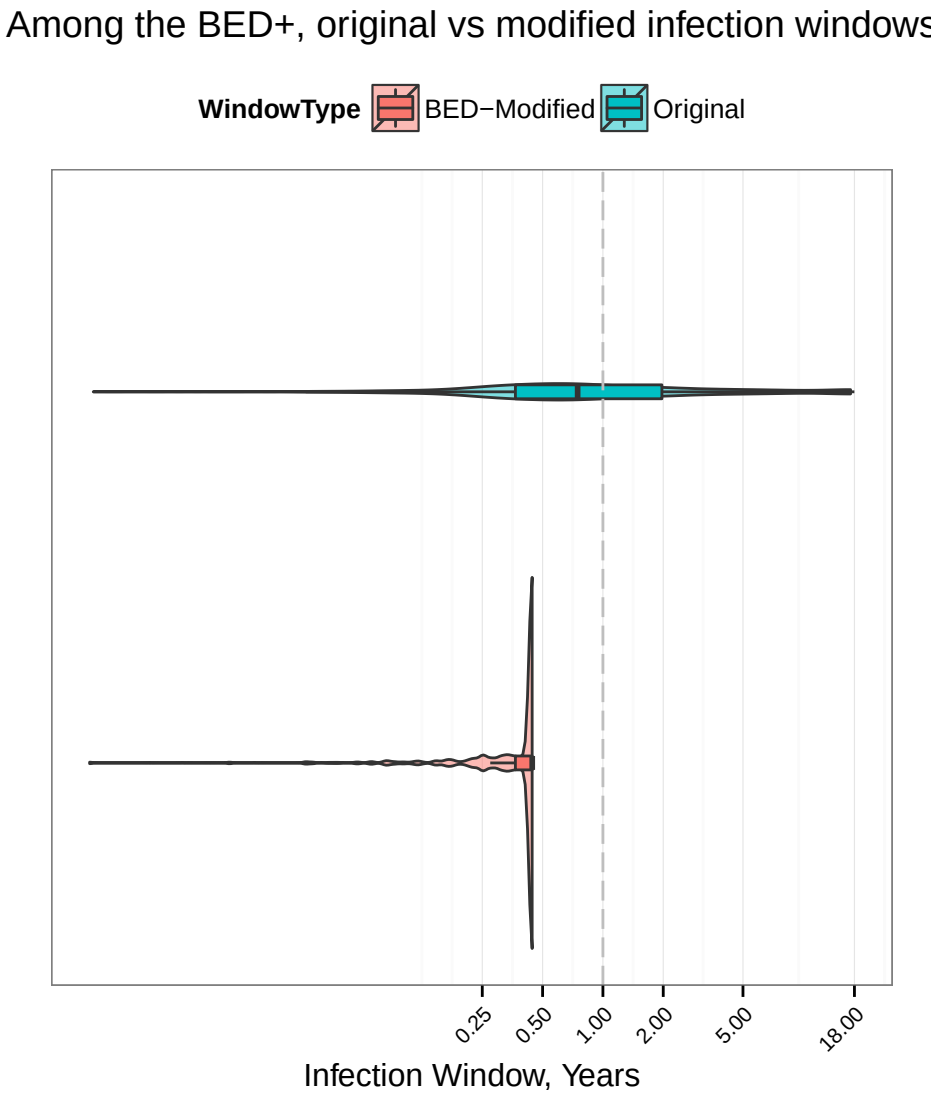

Figure 13: Distribution of infection windows among the BED+ (691 Original, 880 Modified)

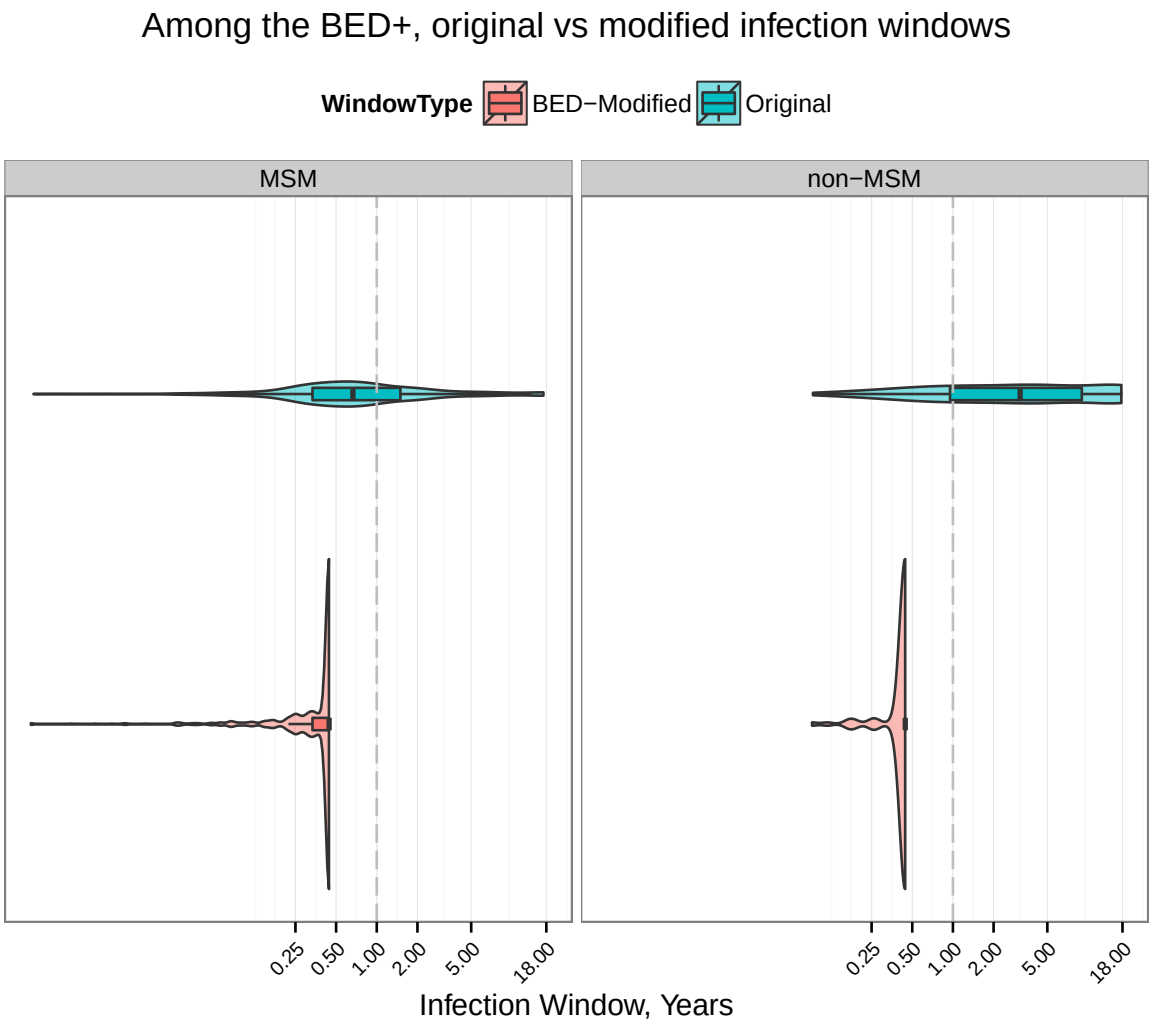

Figure 14: Distribution of infection windows among the BED+ (691 Original, 880 Modified), by MSM vs non-MSM