

1 **Title:**

2 A new method for estimating the number of undiagnosed HIV infected based on HIV
3 testing history, with an application to men who have sex with men in Seattle/King
4 County, WA
5

6 **Authors:**

7 Ian E. Fellows PhD 1
8 Martina Morris PhD 2
9 Julia Dombrowski MD, MPH 2
10 Susan Buskin PhD 3
11 Amy Bennett MPH 3
12 Matthew R. Golden MD, MPH 2,3
13

14 **Affiliations:**

- 15 1. Fellows Statistics, San Diego, CA, USA
16 2. University of Washington, Seattle, WA, USA
17 3. Public Health Seattle/King County, Seattle, WA, USA
18

1 **Abstract**

2 We develop a new approach for estimating the undiagnosed fraction of HIV cases,
3 the first step in the HIV Care Cascade. The goal is to address a critical blindspot in HIV
4 prevention and treatment planning, with an approach that simplifies data requirements and
5 can be implemented with open-source software. The primary data required is HIV testing
6 history information on newly diagnosed cases. Two methods are presented and compared.
7 The first is a general methodology based on simplified back-calculation that can be used to
8 assess changes in the undiagnosed fraction over time. The second makes an assumption of
9 constant incidence, allowing the estimate to be expressed as a simple closed formula
10 calculation. We demonstrate the methods with an application to HIV diagnoses among men
11 who have sex with men (MSM) from Seattle/King County. The estimates suggest that 6% of
12 HIV-infected MSM in King County are undiagnosed, about one-third of the comparable
13 national estimate. A sensitivity analysis on the key distributional assumption gives an
14 upper bound of 11%. The undiagnosed fraction varies by race/ethnicity, with estimates of
15 4.9% among white, 8.6% of African American, and 9.3% of Hispanic HIV-infected MSM
16 being undiagnosed.

1 Introduction

2 HIV prevention is increasingly focused on identifying infected persons and ensuring
3 that they initiate and continue treatment[1]. The success of these efforts can be tracked
4 through the HIV care cascade or care continuum[2], and many public health authorities now
5 seek to measure each step in that continuum as a routine epidemiologic monitoring activity
6 [3].

7 The first step in the care continuum – the proportion of HIV infected persons who
8 have not yet been diagnosed – is perhaps the most challenging to estimate. Directly
9 estimating this number requires HIV testing a representative sample of persons at risk for
10 infection and identifying the fraction of infected persons who are unaware of their status.
11 In the U.S., Centers for Disease Control and Prevention (CDC) investigators have adopted
12 this direct approach as part of the National HIV Behavioral Surveillance system (NHBS) and
13 their experience highlights some of the difficulties involved. The NHBS of men who have
14 sex with men (MSM) relies on venue-based sampling, and the representativeness of the
15 surveyed population is unknown. The system measures awareness of HIV status based on
16 self-report, which is vulnerable to under-reporting by respondents who may be reluctant to
17 reveal their HIV positive status to interviewers[4,5]. In addition, this approach is costly,
18 labor intensive and not replicable by many local health jurisdictions.

19 Alternatively, estimates of the undiagnosed fraction can be indirectly obtained by
20 using some form of back-calculation to estimate underlying HIV incidence. Variants of this
21 method are increasingly being used for this purpose [6]. CDC has published back-
22 calculation estimates for the undiagnosed fraction of MSM in the US of 19-26% (the range
23 reflects differences by race/ethnicity) [7]. These estimates are based on an “extended back
24 calculation” methodology and relies on surveillance data on the number of AIDS and HIV

1 diagnoses in each year. The methodology requires a set of reasonably complex assumptions
2 regarding testing frequency and the distributions of time from infection to AIDS diagnosis
3 (see the web appendix to [8]), though the complete mathematical details of the model have
4 not been published. As a result, local health departments currently do not have an accepted
5 method for estimating the undiagnosed fraction of HIV infections in their area.

6 We present in this paper two new methods for estimating the undiagnosed fraction
7 that are designed for use by local health departments. The first is a relatively simple back-
8 calculation based approach which requires no assumptions about trends in HIV incidence,
9 and the second is an even simpler formula that can be used when HIV incidence counts are
10 stable. Both methods use data on HIV testing history for newly diagnosed cases, which
11 already is, or can be, collected as part of routine public health activities. The calculations do
12 not rely on AIDS diagnoses, so do not require assumptions regarding the distribution of
13 time from infection to AIDS, which is a major potential point of model fragility for previous
14 back-calculation approaches. The testing frequency distribution is estimated from patient
15 reported testing history data, and thus avoids the need to assume a constant testing rate.
16 Calculations are done using a package written in the programming language R, and both the
17 R software and this package are open source and publicly available. We demonstrate the
18 methods by estimating the undiagnosed fraction of HIV cases among MSM in King County,
19 WA, overall and broken down by race.

20 **Materials and Methods**

21 The methods developed here rely on the date of a last negative test for all newly
22 diagnosed HIV positive cases to estimate the distribution of time from infection to HIV
23 diagnosis (TID), and then use the TID to estimate the undiagnosed fraction. Since the

1 methods were developed specifically to utilize the type of data that are often available in
2 health departments, we start by describing the sources of HIV testing data available in our
3 application setting in King County, WA. Approximately 85% of all cases of HIV in King
4 County occur in MSM, so for this application we restrict our analysis to MSM.

5 **Sources of HIV testing data: King County, WA**

6 HIV testing data were all collected as part of routine public health activities
7 undertaken by Public Health – Seattle & King County (Public Health). Public Health has
8 three sources of HIV testing history data: The enhanced HIV/AIDS reporting system
9 (eHARS), the CDC treatment and testing history questionnaire (HIS), and data collected
10 through HIV partner services. The eHARS data only include dates of HIV negative tests that
11 have verified documentation ascertained through review of medical records. HIS and
12 partner services testing history data are collected by Public Health Disease Intervention
13 Specialists (DIS) using standardized instruments, and the date of last test may be based on
14 self-report by persons with newly diagnosed HIV infection and/or medical record reviews.
15 Partner services data were only available for the period 2010-2012 and are based on client
16 self-report. eHARS data has the advantage of including only verified information, but in
17 some instances persons with HIV report having recent HIV negative tests that cannot be
18 verified, and eHARS then records older but verifiable HIV test results. As part of our
19 evaluation of these different sources of data, we used Pearson's correlations to assess the
20 agreement between the three data sources for time since last HIV negative test.

1 **Methods and Assumptions**

2 **Estimating the time of last negative HIV test**

3 This is directly estimated from the testing history data. We used the eHARS date
4 when available, because it was a chart validated (if somewhat conservative) measure. In
5 the absence of eHARS data, we used the most recent date observed in the HIS or partner
6 services. If the day and/or month of the last HIV negative test was missing, we assumed
7 that testing occurred in the middle of the month or on July 1, respectively.

8 **Estimating the possible infection interval**

9 For each individual, the period during which they were possibly infected was
10 defined as the time between subjects' last negative HIV test and their HIV diagnosis, with a
11 maximum assigned value of 18 years. This maximum value is consistent with prior
12 analyses suggesting that 95% of persons will develop AIDS within 18 years of infection [9].
13 When testing history data are missing, we defined the period of possible infection as their
14 age at time of HIV diagnosis minus 16, the median age of sexual debut in the U.S.; here again
15 we assume that no one is HIV infected for more than 18 years prior to diagnosis.

16 **Estimating the distribution of time from infection to diagnosis (TID)**

17 The methods start by estimating the distribution of time from infection to diagnosis.
18 This requires an assumption about the distribution of infection within the possible infection
19 interval, so we conducted a sensitivity analysis using two different assumptions:

- 20 • Base Case Estimate: Here we assume that infection occurred at a random point
21 distributed uniformly during the period of possible infection. If some men test for
22 HIV in response to specific risks, their TID will shift toward the end of the possible

1 infection interval (as was observed in [10]), so this estimate is likely to be somewhat
2 conservative.

- 3 • Upper bound Estimate: Here we assume that HIV acquisition occurred at the
4 beginning of the possible infection interval, i.e., the day after the last negative test.
5 This represents the most conservative possible estimate of the time from infection
6 to diagnosis consistent with the data.

7 Both estimated distributions assume that the distribution of time from infection to
8 diagnosis does not change appreciably from year to year, though testing frequency for
9 individuals may vary. We tested this assumption using a Welsh ANOVA.

10 **Back calculation allowing for time-varying incidence**

11 We use a back-calculation based method to estimate quarterly HIV incidence using
12 observed testing history, with progression from infection to diagnosis defined by either the
13 base case or upper bound TID assumption. Incidence counts are modeled as a (possibly
14 time-varying) Poisson process with rate λ_i per quarter, using the standard convolution
15 equation [11]:

$$16 \quad E(D_t) = \sum_{i < t} \lambda_i f_{i,t-i}$$

17 where D_t is the number diagnosed with HIV during quarter t , and $f_{i,t-i}$ is the
18 probability that an individual infected during quarter i is diagnosed $t - i$ quarters later, at
19 quarter t . The unknown incidence rates λ_i are traditionally estimated using an
20 expectation-maximization (EM) algorithm [8]. Once the quarterly incidence rates have
21 been estimated, the number undiagnosed at quarter t (U_t) may be estimated as

$$22 \quad E(U_t) = \sum_{i < t} \lambda_i \left(\frac{1}{2} f_{i,t-i} + \sum_{k < t-i} f_{i,k} \right)$$

Each term in the outer sum represents the expected number infected during quarter i who are diagnosed after quarter t plus half the expected number of cases diagnosed during quarter t .

We made two refinements to the usual EM algorithm for fitting this model. First, since our data do not extend back to the beginning of the HIV epidemic but only back to 2006, we adjusted for our limited surveillance window. Second, we added a quadratic smoothing penalty to the likelihood to stabilize the parameter estimates and enforce the *a priori* assumption that any changes in incidence would be relatively gradual. The details of the fitting algorithm are described in the supporting document (File S1).

In contrast to the back calculation methods reviewed in [6], this method does not rely on AIDS diagnoses or other biomarkers, and it does not require assumptions about the rates of testing, since the relevant testing interval is directly estimated from data. This considerably simplifies the back-calculation itself, and reduces the impact of assumptions related to disease severity and progression on the estimates.

Direct calculation assuming constant incidence

If both HIV incidence (λ) and the distribution of TID are constant over time, a much simpler formula for estimating the number of undiagnosed can be used:

$$E(U) = \lambda \int_0^{\infty} P(TID > t) dt \quad (1)$$

Under these conditions the expected number of new diagnoses (D) will be equal to the expected incidence, so equation (1) can be estimated by substituting the average number diagnosed in each quarter for λ , and either the base case or upper bound estimate of TID for P .

1 **Subpopulation estimates**

2 Some sub-populations, such as groups defined by race or ethnicity, may have
3 different testing behaviors and risks for infection. We tested for sub-population differences
4 in time from last negative HIV test to diagnosis using Welsh ANOVAs, and estimated the
5 number of undiagnosed persons by applying the methods outlined above to the subset of
6 the data containing only those in the sub-population.

7 **Estimates of the undiagnosed fraction of cases**

8 The above equations provide an estimate of the number of undiagnosed cases of HIV
9 infection. The proportion of cases with undiagnosed HIV infection is the estimated number
10 of persons with undiagnosed HIV divided by the sum of the number persons with diagnosed
11 and undiagnosed HIV infection. We used Public Health HIV surveillance data to estimate the
12 number of persons with diagnosed HIV infection living in the area. Medical providers in
13 Washington State are legally required to report all new cases of HIV infection to Public
14 Health – Seattle & King County, and laboratories are required to report all positive HIV test
15 results, all CD4 lymphocyte counts and all HIV RNA test results. Public health surveillance
16 staff investigate all newly identified cases, and surveillance data integrate both in- and out-
17 migration of persons to the area [12].

18 **Results**

19 From 2006 through the end of 2012, 1522 MSM were diagnosed with HIV in King
20 County, WA. Multiple sources of testing history data were often available, as shown in Table
21 1. HIS, HIV partner services and eHARS data on date of last HIV negative test were available
22 on 1080 (71%), 476 (31%) and 382 (25%) men, respectively. Combining data from the
23 three sources, among 1522 men, information on date of last HIV negative test was available

1 for 1233 (81%), including 101 (6.6%) men who were diagnosed with HIV at their first test,
2 and no data were available on 289 (19%). The HIS showed excellent agreement with the
3 eHARS and partner services measures (correlations of 0.76 and 0.85 respectively). Among
4 the 113 subjects with both eHARS and partner services data the agreement was less strong
5 (correlation: 0.37), largely due to a cluster of outliers showing much more recent eHARS
6 test date than the date reported in partner services interview. Given the hierarchical
7 assignment, the last negative date was taken from eHARS for 382 men (31% of those with
8 an observed last test date), from HIS for 152 men (12%), and from partner services for 80
9 men (6%). 518 men (42%) had no eHARS data and had identical HIS and partner services
10 dates

11 The estimated mean and median period of possible infection were 3.12 years and
12 1.25 years, respectively, with no statistically significant difference across years ($F= 1.64$ (6,
13 523), $p > 0.1$). Figure 1 shows the estimated TID distributions for the base case
14 (Median=0.5 years) and upper bound (Median=1.3 years).

15 The back-calculation estimates of HIV incidence yielded almost identical quarterly
16 counts for both the base case and upper bound TID models: 49.7-57.5 and 49.6-56.8
17 respectively. Figure 2 (upper panel) shows the estimated incidence count curves for each
18 model, along with the observed quarterly diagnosis counts. Both models find a relatively
19 stable incidence over the 2006 to 2012 period, suggesting that the simpler constant
20 incidence model can be used to estimate the undiagnosed fraction for these data.

21 Figure 2 (lower panel) shows the estimated number of undiagnosed cases of HIV
22 among MSM based on each model. The number remains stable throughout the period of
23 observation, ranging from 333-368 in the base case, and about double that, 662-713 for the
24 upper bound. This base case estimates correspond to about 6% of HIV-infected MSM being

undiagnosed; the upper bound to about 11%. Table 1 compares the back calculation based estimates to those generated by applying Equation 1 assuming constant incidence.

The testing history data show significant variation by race/ethnicity in the estimated time from infection to diagnosis. The mean possible infection interval estimates were 2.8, 4.3 and 3.4 years, for Whites, African Americans and Hispanics respectively ($F=4.5$ (2,199), $p < 0.05$). The medians are 1.2, 1.8 and 1.5 respectively. As shown in Figure 3, this contributes to relatively large differences in estimates of the undiagnosed fraction between White MSM, and MSM of color.

Using the base case model and assuming constant incidence, we estimate that 4.9% of White HIV-infected MSM, 8.6% of African American HIV-infected MSM, and 9.3% of Hispanic HIV-infected MSM in King County are undiagnosed (the full table of estimates for each model broken down by race/ethnicity is provided in Table S1.). It may seem counterintuitive that the undiagnosed fraction would be slightly higher among Hispanics than African Americans when their mean TID is lower. The reason is that shape of the TID for Hispanic MSM is distinctively different: their early diagnoses are later than both Whites and African Americans, while their later diagnoses are more frequent than whites, but less extreme than African Americans. For example, 11% of African Americans had periods greater than 15 years, versus 6% among Hispanics, but 25% of African Americans had tested negative within 6 months of diagnosis, versus 20% for Hispanics.

Discussion

We present a new approach to estimating the proportion of persons with undiagnosed HIV infection based on HIV testing history data. Applying the methods to data from King County, WA, our best estimate suggests that about 6% of HIV-infected MSM are

undiagnosed. If we use the most conservative assumptions that are consistent with the observed data, the upper bound overall estimate is 11%. Under both models, the estimated undiagnosed fraction is almost twice as high among African Americans and Hispanics than among Whites.

Our estimates for King County, WA are much lower than the most recent national estimates: using an extended back calculation method, CDC investigators estimated that in 2008 19% of HIV-infected MSM were undiagnosed [7]. The extent to which our findings reflect true differences between King County and the U.S. as a whole, versus differences due to methodology is not clear. Ideally, we would compare our results to estimates generated using CDC's extended back calculation method [8,12] [7]. However, at present, the details of that method are not available in sufficient detail to undertake such a comparison. However, two other findings suggest that at least some part of the difference may be real. First, based on laboratory reported surveillance data, Public Health currently estimates that 68% of HIV diagnosed persons in the area are virologically suppressed compared to national estimate of 31%[3,12]. Second, among MSM participants in the 2008 cycle of NHBS, 15% of HIV-infected MSM in Seattle vs. 44% of MSM nationally were reported to be unaware of their HIV status[13]. While the validity of NHBS findings on this subject have been questioned[4], the very large difference observed within NHBS suggests that there is a real difference of some magnitude between King County, WA and the U.S.

Our findings have several important limitations. First, our estimate is based on HIV testing history data with varying levels of reliability. Just over 30% of the cases draw the last negative test date from the eHARS system, and eHARS records the last test date that can be validated against medical records. This places an upper bound on the last negative HIV test date, but it may not be the most recent test, so this may bias the TID toward

conservative estimates. For the remainder of the cases, TID is based on self-reported date of last test. The validity of these data are not known; we found good agreement between some measures, but not all; this source of potential error could result in either an over estimate or an underestimate of the undiagnosed fraction. Second, our base case model assumes that individuals are infected uniformly between last negative test and first positive. If some people test in response to a recent risk exposure [10], our base case assumption is also conservative, and the undiagnosed fraction would be lower than what we report here. It would be possible to extend this method to incorporate such data, if the surveillance system collected such information. Finally, over 80% of the cases in our data had some information on the date of last negative test; the estimates from this method are likely to be more uncertain, and less useful, in populations for which testing prior to diagnosis is uncommon, or data on HIV testing history are unavailable. However, in other analyses, we have found the uncertainty to be modest even when half of the cases lack a test prior to diagnosis.

These methods could be useful to public health agencies interested in monitoring their local HIV care continuum. We developed them in response to a request from our Center for AIDS Research Northwest Regional Public Health Consortium, a group convened to advance the integration of public health practice and university research in our region. The approach is designed to be relatively simple to implement and understand, and to integrate local data to create locally relevant estimates. Compared to the range of back calculation methods currently used for estimating the undiagnosed fraction [6], our approach depends on fewer, and more straightforward assumptions, and requires simpler data and statistical methods. It does not use data on AIDS diagnoses, so it does not incur the well-known uncertainties associated with assumptions about the distribution of time from infection to AIDS. The testing history data it does use can be routinely collected as part of

HIV surveillance and partner services, which allows for these local data, which may vary substantially from national data, to be integrated into the model. The model uncertainty arising from the key assumption – the distribution of time from infection to diagnosis in the test window interval – can be explored with sensitivity analysis to identify a true upper bound that is consistent with the observed data. We have not presented confidence intervals (which would be conditional on the statistical model), choosing to focus instead on the model uncertainty. Bootstrap confidence intervals can be developed, but we expect that in most contexts, there will be a sufficient number of diagnosed cases that the uncertainty associated with the model will be substantially larger than the uncertainty that may come from sampling.

The statistical methods presented here are implemented using a package written in R, an open source software environment (will be available on GitHub when this paper is published). We are developing a simple graphical user interface for the package based on Shiny [14] that will also be open source, and will allow end users to access the package functionality via their web browser.

Acknowledgements

We would like to thank the members of the Northwest Regional Public Health Consortium, Jason Carr from the Washington State Department of Health, and Jeanette Birnbaum for many helpful comments and suggestions.

References

1. White House Office of National AIDS Policy (2010) National HIV/AIDS Strategy for the United States.

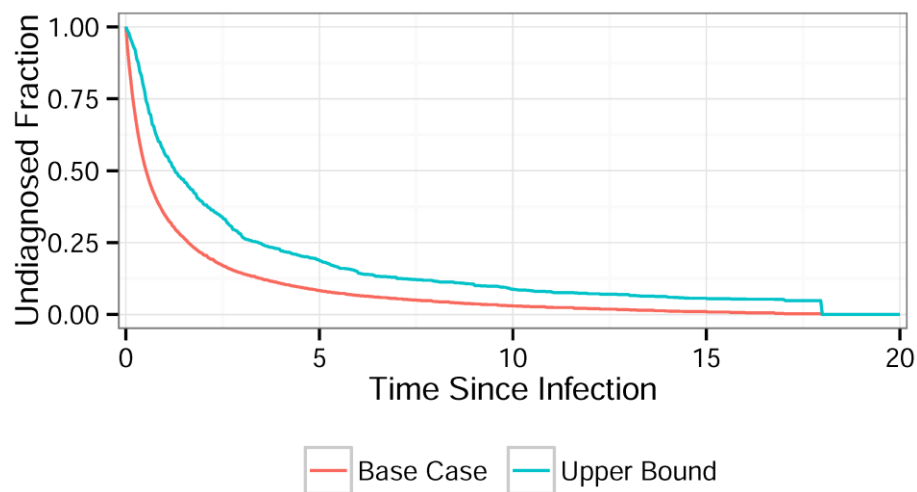
- 1 2. Mugavero MJ, Amico KR, Horn T, Thompson MA (2013) The State of Engagement
2 in HIV Care in the United States: From Cascade to Continuum to Control. Clin
3 Infect Dis.
- 4 3. HIV/AIDS Epidemiology Unit PHSKC, and the Infectious Disease and Reproductive
5 Health Assessment Unit WSDoH (2013) HIV/AIDS Epidemiology Report, Second
6 Half 2012: Volume 81. In: Health PH-SKCaWSDo, editor. Seattle, WA.
- 7 4. Sanchez TH, Kelley CF, O'Hara M, Whiteside YO, Lambert R, et al. Lack of
8 Awareness of HIV Infection: Problems and Solutions with Self-reported HIV
9 Serostatus or Men Who Have Sex with Men; 2012 2012; Washington DC.
- 10 5. Marzinke MA, Clarke W, Wang L, Cummings V, Liu T-Y, et al. (2013) Nondisclosure
11 of HIV Status in a Clinical Trial Setting: Antiretroviral Drug Screening Can Help
12 Distinguish Between Newly Diagnosed and Previously Diagnosed HIV Infection.
13 Clinical Infectious Diseases.
- 14 6. Lodwick R, Alioum A, Archibald C, Birrell P, Commenges D, et al. (2011) HIV in
15 hiding: methods and data requirements for the estimation of the number of
16 people living with undiagnosed HIV Working Group on Estimation of HIV
17 Prevalence in Europe. Aids 25: 1017-1023.
- 18 7. Chen M, Rhodes PH, Hall IH, Kilmarx PH, Branson BM, et al. (2012) Prevalence of
19 undiagnosed HIV infection among persons aged ≥ 13 years--National HIV
20 Surveillance System, United States, 2005-2008. MMWR Morb Mortal Wkly Rep 61
21 Suppl: 57-64.
- 22 8. Hall HI, Song RG, Rhodes P, Prejean J, An Q, et al. (2008) Estimation of HIV
23 incidence in the United States. Jama-Journal of the American Medical Association
24 300: 520-529.
- 25 9. Lui KJ, Lawrence DN, Morgan WM, Peterman TA, Haverkos HW, et al. (1986) A
26 model-based approach for estimating the mean incubation period of transfusion-
27 associated acquired-immunodeficiency-syndrome. Proceedings of the National
28 Academy of Sciences of the United States of America 83: 3051-3055.
- 29 10. Skar H, Albert J, Leitner T (2013) Towards estimation of HIV-1 date of infection:
30 a time-continuous IgG-model shows that seroconversion does not occur at the
31 midpoint between negative and positive tests. PLoS One 8: e60906.
- 32 11. Brookmeyer R, Gail MH (1988) A method for obtaining short-term projections
33 and lower bounds on the size of the aids epidemic. Journal of the American
34 Statistical Association 83: 301-308.
- 35 12. Hall HI, Frazier EL, Rhodes P, Holtgrave DR, Furlow-Parmley C, et al. (2013)
36 Differences in human immunodeficiency virus care and treatment among
37 subpopulations in the United States. JAMA Intern Med 173: 1337-1344.

1 13. Centers for Disease C, Prevention (2010) Prevalence and awareness of HIV
2 infection among men who have sex with men --- 21 cities, United States, 2008.
3 MMWR Morb Mortal Wkly Rep 59: 1201-1207.

4 14. R Studio (2014) Shiny: A web application framework for R.
5 <http://shiny.rstudio.com/>
6

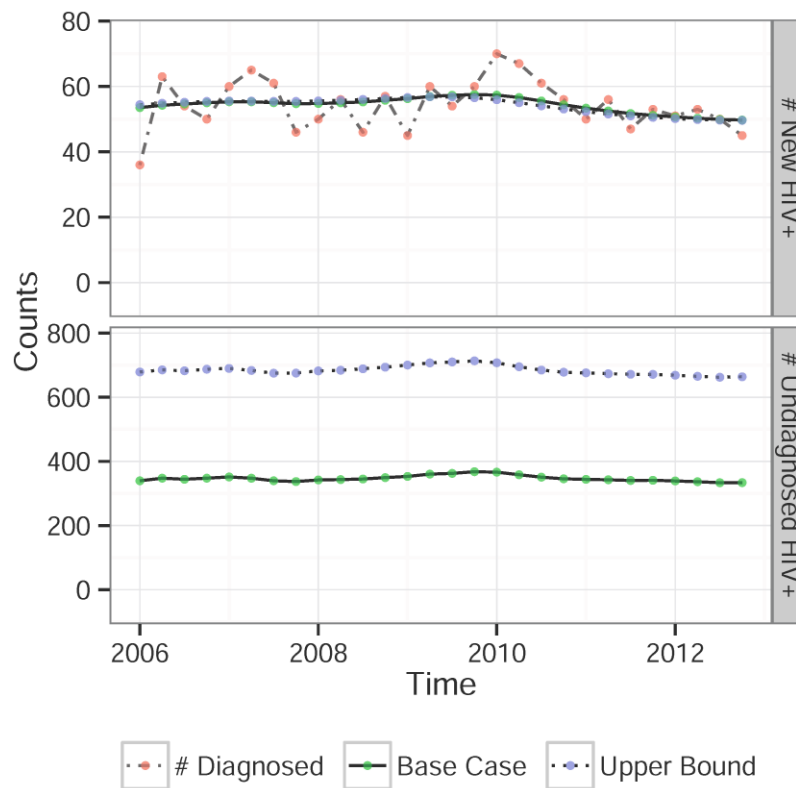
7 Figure Legends

8 **Figure 1: Distribution of time between infection and diagnosis (TID).** This shows the
9 two estimates of the population TID: the base case, which assumes a uniform distribution
10 of the probability of infection in the interest interval, and the upper bound, which assumes
11 infection occurs at the start of the intertest interval, i.e., the day after the last negative test.



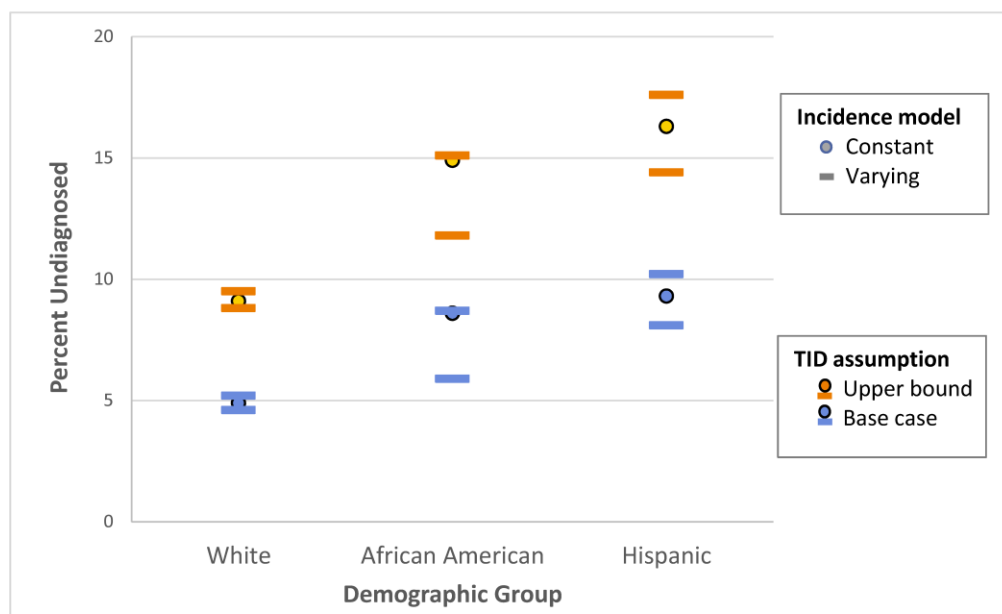
12

- 1 **Figure 2: Observed and estimated quarterly HIV incidence among MSM in King**
- 2 **County by quarter.** The upper panel displays the HIV incidence estimates along with the
- 3 observed number of diagnosed cases. The estimation uses a quadratic smoothing
- 4 parameter of 0.1. The lower panel displays the model-based estimates of the total number
- 5 of undiagnosed HIV+ cases by quarter.



6

1 **Figure 3. Racial/Ethnic disparities in the undiagnosed fraction with HIV.** The plot
2 shows the group-specific estimates under different assumptions: constant or time-varying
3 incidence, and base case or upper bound estimate of the TID from Figure 1. The time
4 varying incidence is summarized by the lowest and highest observed values from 2006-
5 2012.



6

7

1 **Supporting Information Legends**

2 **Table S1. HIV Incidence and undiagnosed fraction estimates broken down by**
3 **race/ethnicity.** Estimates of the number of undiagnosed HIV cases among MSM in King
4 County stratified by ethnicity. (* Sum of cases thought to reside in King County based on HIV
5 surveillance data (N= 4188, 458, and 572 respectively) and the estimated number of
6 undiagnosed cases).

7

Under Review

Tables

Table 1. Overlap in HIV testing data sources from King County WA, and source used for analysis.

Case has:	Case also has:†			Used for analysis
	HIS	PS	eHARS	
Testing history questionnaire (HIS)	1080	456	346	670
Partner Services data (PS)††		476	113	80
Enhanced HIV/AIDS reporting system (eHARS)			382	382
Total with previous negative test				1132
Never Tested				101
Total with testing history information				1233
No testing history Information				289
Total diagnosed				1522

Public Health Seattle/King County has three sources of testing history information. For cases that have more than one source of data, we use the most reliable source in this analysis (see text for more detail).

† Diagonal elements represent the total number of newly diagnosed cases who had this source of testing information available.

†† Partner services data collection began in 2010, and were available for 69% of the 659 MSM newly diagnosed with HIV after this time. 518 cases had identical PS and HIS dates leading to ambiguity in the “Used for Analysis” column. We count these as HIS for the purposes of this table.

1 **Table 2: Estimates of the number and fraction of undiagnosed HIV cases among MSM**
2 **in King County.**

Distribution Assumption	Incidence Model	Number of MSM with HIV infection†	Number of MSM with undiagnosed HIV infection	Fraction Undiagnosed
Base case	Varying	5850-5884	333.5-367.8	5.7%-6.3%
	Constant	5863	346.7	5.9%
Upper bound	Varying	6178-6229	662.2-713.3	10.7%-11.4%
	Constant	6203.2	687.2	11.1%

3
4 The different estimates reflect the sensitivity to the TID assumption (Base Case vs. Upper
5 Bound) and the trend assumption (allowed to vary vs. constant).

6 † Population size estimated as the sum of HIV-infected MSM thought to reside in King
7 County, WA based on HIV surveillance data (n=5516) plus the estimated number of
8 undiagnosed cases.