Supplemental appendix: Technical details of the mathematical model

MODEL STRUCTURE AND ASSUMPTIONS

The model represents a cohort of 100,000 individuals from age 16 to 64. The population is stratified into four groups with varying HIV risk and testing practices: low- and high-risk men who have sex with men (MSM), men who have sex with women only (MSW) and women. Figure S1 presents a simplified model schematic. Individuals in each of the four population groups enter the model at age 16 and progress in yearly time-steps through age 64, exiting at age 65 or with HIV diagnosis. All persons start in the susceptible state, and age- and risk group-specific HIV incidence rates are applied to determine the number who become infected at each age. HIV-positive individuals are stratified by year of infection, and HIV diagnosis occurs in one of five ways: through risk-based screening of MSM, prenatal screening of pregnant women, upon progression to symptomatic HIV/AIDS, as a result of partner notification, or with routine screening. Those who remain undiagnosed at the end of the year progress to the next year of infection up through year 16, at which point all are assumed to experience symptoms leading to diagnostic testing. The model was built using Microsoft Excel (Version 16.9).

As a static linear model, it does not represent transmission dynamics or account for the impact of screening and diagnosis on downstream infection. We assume an implicit ordering of events, with HIV infections occurring at the beginning of each year, diagnoses through risk-based, prenatal, and routine screening at the mid-point of the year, and diagnoses prompted by symptom onset or partner notification at the end of the year. The rate of symptom development was estimated from data on time from HIV seroconversion to CD4+ cell count <200 cells/mm^3,[1] which suggest an 8.8% probability of reaching this threshold after 1 year and an approximately linear increase up through a cumulative 32% probability by year 5. Consistent with Golden et al.,[2] we model a

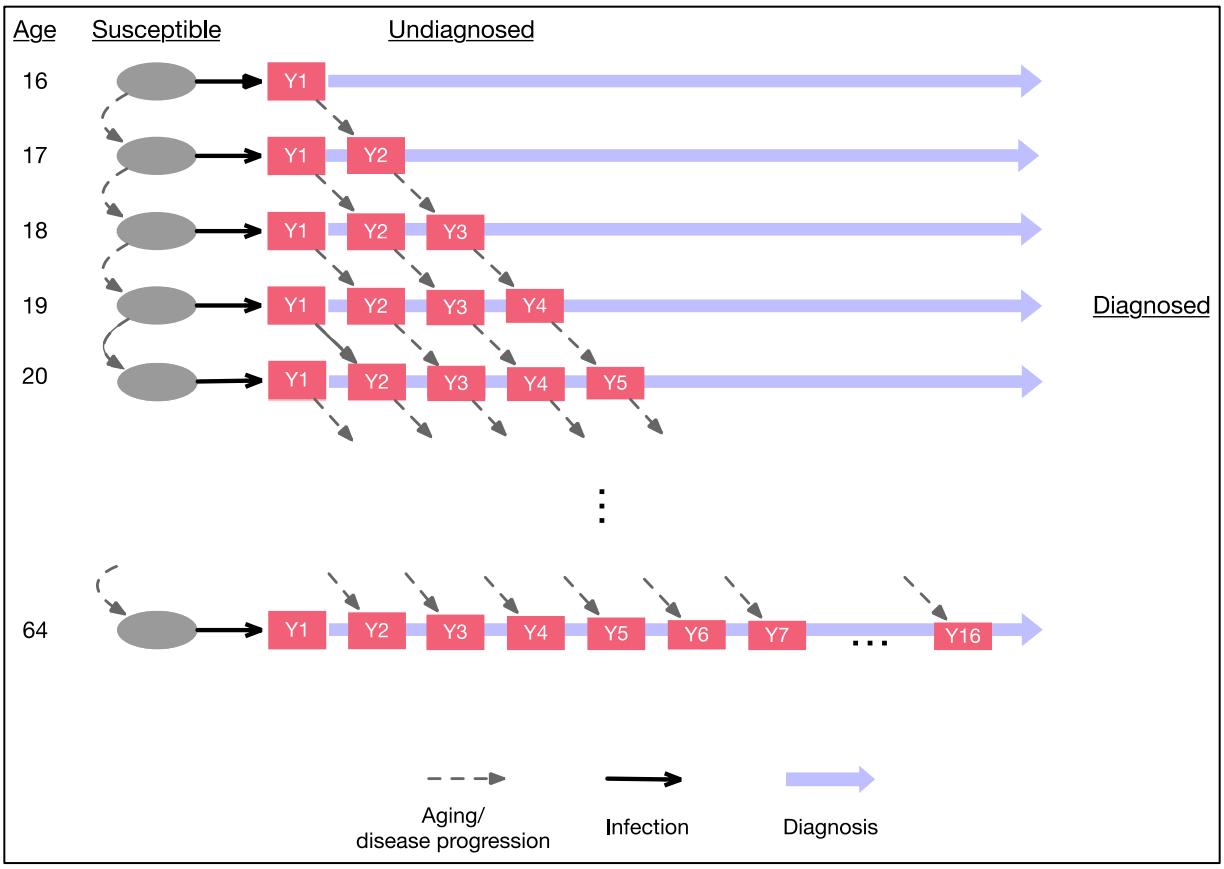

Figure S1: Model schematic showing the processes of infection, aging, disease progression, and diagnosis for one of the four modeled population groups.

6.08% risk of progression in each subsequent year, with all individuals progressing to CD4+ <200 cells/mm^3 after 16 years. We assume that HIV/AIDS associated symptoms present upon reaching this threshold and prompt diagnosis. While some people develop HIV/AIDS-associated symptoms before CD4+ cell depletion, others may experience symptoms later. Our model suggests a median time to symptom onset of 8 years, aligning with estimates from cohort studies.[3,4]

Risk-based, prenatal, and routine screening occur at the midpoint of the year, such that individuals diagnosed through these strategies contribute half a year of undiagnosed infection in the year they are diagnosed. Persons not screened for HIV remain undiagnosed until they develop symptoms or are notified of potential exposure to an infected partner. For partner notification, we

assume no bridging of the MSM and heterosexual populations: the number of diagnoses resulting from partner notification among MSW and women depend on the number of diagnoses from other modes of testing among females and MSW, respectively, and partner notification diagnoses among MSM are calculated based on the number of other diagnoses among MSM in both high- and low-risk groups. In allocating these diagnoses by age, we make a simplifying assumption of perfect age matching in partnerships. All HIV tests are assumed to be 100 sensitive and specific, and individuals exit the model once they become diagnosed.

MONTE CARLO SIMULATION: UNCERTAINTY DISTRIBUTIONS

To account for uncertainty in model inputs, we conducted Monte Carlo simulation using @RISK software (version 7.6, Palisade Company, Ithaca, NY). For parameter estimates for which a 95% confidence interval was available in the literature, we sampled from normal distributions with standard deviations equal to the range of the confidence interval divided by (2×1.96). Where estimates from different sources were inconsistent, we defined normal distributions with standard deviations equal to the range of estimates divided by (2×1.96). As data were not available on the relative probability of risk-based screening for high- and low-risk men, we defined uniform uncertainty intervals with ranges informed by expert opinion.

For the concentrated and diverse epidemic models (informed by data from King County, WA and Philadelphia County, PA, respectively), we lacked confidence interval data on the size of the MSM population and the number of incident cases. We applied a range of $\pm 20\%$ to both sets of parameters, as this range encompassed an alternate estimate of the size of the MSM population in King County and was larger than the margin of error in incidence estimates at the national level, accounting for greater uncertainty with smaller counts. We used this range to define normal

distributions around point estimates of incident cases as described above. However, because we did not have reason to believe that either estimate for the size of the MSM population in King County was closer to the truth,[5,6] we sampled values for this parameter from a uniform distribution.

To account for dependence in parameters defining the proportion of MSM who are high-risk, the incidence rate ratio for high- to low-risk MSM, and the relative risk-based screening frequencies, we defined a correlation matrix to ensure that sampled values for these parameters preserved expected relationships between these variables. We assumed that as the size of the high-risk population decreased, the differences in incidence rates and HIV testing relative to low-risk MSM would become more extreme. We also assumed that, as the relative likelihood of screening annually for high-risk MSM increased, the relative likelihood of not screening would decrease. For each of these associations, we defined moderate correlations, shown in Table S1.

Table S1: Correlation coefficients for Monte Carlo Simulation

	Proportion high risk	Incidence rate ratio (high:low risk)	Relative risk of annual screening	Relative risk of no screening
Proportion high risk	1			
Incidence rate ratio (high:low risk)	-0.6	1		
Relative risk of annual screening	-0.5	0.5	1	
Relative risk of no screening	0.5	-0.5	-0.6	1

MODEL PROCESSES AND FORUMLAE

Table S2 defines model parameters, and the sections that follow detail the formulae and assumptions used to construct and execute the model.

Table S2: Parameter definitions

Parameter/ variable	Definition
a	Age
y	Year of infection (1-16)
r	Risk group (high-risk MSM, msm_h; low-risk MSM, msm_l; non-MSM males, msw; females, fem)
p_r	Proportion of the population in risk group r
f_a	Fertility rate for women of age a
x	Targeted screening group. Among MSM, groups are never, annual, and non-annual risk-based screeners ($ann, nann$, and nev). Among females, groups define pregnant and non-pregnant women ($preg$ and $npreg$). MSW are not stratified by targeted screening group.
n	Non-annual risk-based screening interval (years)
$px_x^{r,a}$	Proportion of low- and high-risk MSM of age a in each risk-based testing group x (never, annual, and non-annual)
$prtest$	Proportion of pregnant women screened for HIV in the pre- or perinatal periods
$dxpn$	Number of new diagnoses from partner services resulting from one index case
δ	Age of routine screening
$rtcov$	Routine screening acceptance rate
α	Proportion of HIV-infected persons who develop symptoms in years 2-16 of infection
β	Proportion of HIV-infected persons who develop symptoms in year 1 of infection
$S_{x,a}^r$	Number of susceptible individuals in risk group r, targeted screening group x, and age a
$I_{x,a}^r$	Incident infections in risk group r and targeted screening group x occurring at age a
IR_a^r	Incidence rate among persons of risk group r and age a
$U_{x,a,y}^r$	Number of undiagnosed infections in risk group r, screening group x, and age a at the start of year of infection y
$RB_{x,a,y}^r$	Number of diagnoses due to risk-based screening among MSM in high- or low-risk group r, screening group x, and age a in year of infection y
$PR_{a,y}$	Number of new diagnoses due to prenatal screening among females at age a in year of infection y

$RT^r_{x,a,y}$	Number of new diagnoses due to routine screening among persons in risk group r, screening group x, and age a in year of infection y
$SX^r_{x,a,y}$	Number of new diagnoses due to symptomatic testing among persons in risk group r, screening group x, and age a in year of infection y
$PN^r_{x,a,y}$	Number of new diagnoses due to partner notification testing among persons in risk group r, screening group x, and age a in year of infection y
$U^r_{x,a,y}$	Number of persons in risk group r, screening group x, and age a who are HIV-infected and undiagnosed at the start of year of infection y
SY	Total number of undiagnosed cases that develop symptoms
YU	Cumulative population-level years of undiagnosed infection

1. Initial population size for each risk group r and screening group x

$for\ r \in \{msm_l, msm_h\},\ S^r_{x,16} = N \times p_r \times px^{r,a}_x$
$S^{fem}_{preg,16} = N \times p_r \times f_{16}$

$S^{fem}_{npreg,16} = N \times p_r \times (1 - f_{16})$

$S^{msw}_{x,16} = N \times p_r$

Assumptions: All persons enter the model at age 16 in the susceptible state.

2. Number susceptible in risk group $r$and screening group x at age a>16

$for\ r \in \{msm_l, msm_h\}, \qquad S^r_{x,a>16} = \sum_x \left(S^r_{x,a>(16-1)} - I^r_{x,a>(16-1)}\right) \times px^{r,a}_x$

$S^{fem}_{preg,a>16} = \sum_x \left(S^{fem}_{x,a>(16-1)} - I^{fem}_{x,a>(16-1)}\right) \times f_a$

$S^{fem}_{npreg,a>16} = \sum_x \left(S^{fem}_{x,a>(16-1)} - I^{fem}_{x,a>(16-1)}\right) \times (1 - f_a)$

$S^{msw}_{x,a>16} = S^{msw}_{x,a>(16-1)} - I^{msw}_{x,a>(16-1)}$

Assumptions: No mortality and no new entries to the population.

3. Incident infections in risk group $r$and screening group x at age a

$I^r_{x,a} = S^r_{x,a} \times IR^r_a$

4. Undiagnosed infections in risk group r, screening group x, and age a at the start of year of infection y

$$for\ r \in \{msm_l, msm_h\},$$
$$U^r_{x,a,y} = U^r_{x,a-1,y-1} - RB^r_{x,a-1,y-1} - RT^r_{x,a-1,y-1} - SX^r_{x,a-1,y-1} - PN^r_{x,a-1,y-1}$$

$$U^{fem}_{preg,a,y} = f_a \times \left(U^{fem}_{npreg,a-1,y-1} - RT^{fem}_{npreg,a-1,y-1} - SX^{fem}_{npreg,a-1,y-1} - PN^{fem}_{npreg,a-1,y-1}\right)$$

$$U^{fem}_{npreg,a,y} = (1 - f_a) \times \Big(\left(U^{fem}_{npreg,a-1,y-1} - RT^{fem}_{npreg,a-1,y-1} - SX^{fem}_{npreg,a-1,y-1} - PN^{fem}_{npreg,a-1,y-1}\right) + \left(U^{fem}_{preg,a-1,y-1} - RT^{fem}_{preg,a-1,y-1} - SX^{fem}_{preg,a-1,y-1} - PN^{fem}_{preg,a-1,y-1}\right)\Big)$$

$$U^{msw}_{x,a,y} = U^{msw}_{x,a-1,y-1} - RT^{msw}_{x,a-1,y-1} - SX^{msw}_{x,a-1,y-1} - PN^{msw}_{x,a-1,y-1}$$

5. Number of MSM in high- or low-risk group r, screening group x, age a, and year of infection y diagnosed through risk-based screening

$$RB^r_{ann,a,1} = I^r_{ann,a}$$
$$RB^r_{nann,a,1} = \left(\frac{1}{n}\right) \times I^r_{nann,a}$$
$$for\ y \in \{2, \ldots, 16\},$$
$$RB^r_{nann,a,y} = \left(\frac{1}{n - (y-1)}\right) \times U^r_{nann,a,y}$$

Assumptions: Risk-based screening is modeled for MSM only. Of MSM who test every n years, $1/n$ of those not yet diagnosed are assumed to test in years 1:n.

6. Number in risk group r, screening group x, age a and year of infection y diagnosed through routine screening

$$RT^r_{x,\delta,y} = \left(U^r_{x,\delta,y} - RB^r_{x,\delta,y} - PR_{\delta,\gamma}\right) \times rtcov$$

Assumptions: Routine screening is implemented only for those in the specified age who were not already screened through prenatal or risk-based testing in that year.

7. Number of women at age a and year of infection y diagnosed through prenatal screening

$$PR_{a,y} = U^{fem}_{preg,a,y} \times prtest$$

Assumptions: Pregnancy occurs at the start of the year. Women cannot be pregnant two years in a row, such that all women pregnant at age a were not pregnant at age $a-1$.

8. Number in risk group r, screening group x, age a and year of infection y diagnosed through symptom-based testing

$$SX^{r}_{x,a,1} = \left(U^{r}_{x,a,1} - RB^{r}_{x,a,1} - RT^{r}_{x,a,1} - PR_{a,1}\right) \times \beta$$

$$for\ y \in \{2,\dots,16\}, \qquad SX^{r}_{x,a,y} = \left(U^{r}_{x,a,y} - RB^{r}_{x,a,y} - RT^{r}_{x,a,y} - PR_{a,y}\right) \times \left(\frac{\alpha}{(1-\alpha(y-2)-\beta)}\right)$$

Assumptions: The proportion of individuals who progress to symptomatic HIV/AIDS is assumed to be constant for years 2-16 of infection at 0.608. All persons who develop symptoms are diagnosed, and all persons living with undiagnosed HIV/AIDS are assumed to develop symptoms resulting in diagnosis by year 16.

9. Number in risk group r, screening group x, age a and year of infection y diagnosed through partner notification

$$for\ r \in \{msm_{high}, msm_{low}\},$$
$$PN^{r}_{x,a,y} = dxpn \times U^{r}_{x,a,y} \times \sum_{r' \in \{msm_h, msm_l\}} \sum_{x'} \sum_{y'} \left(\frac{\left(RB^{r'}_{x,a,y} + RT^{r'}_{x,a,y} + SX^{r'}_{x,a,y}\right)}{\left(U^{r'}_{x',a,y'} - RB^{r'}_{x',a,y'} - RT^{r'}_{x',a,y'} - SX^{r'}_{x',a,y'}\right)}\right)$$

$$PN^{msw}_{x,a,y} = dxpn \times \left(\frac{U^{msw}_{x,a,y}}{\sum_{x'}\sum_{y'} U^{msw}_{x',a,y'} - RT^{msw}_{x',a,y'} - SX^{msw}_{x',a,y'}}\right) \times \sum_{x'} \sum_{y'} \left(PR^{fem}_{a,y'} + RT^{fem}_{x'a,y'} + SX^{fem}_{x',a,y'}\right)$$

$$PN^{fem}_{x,a,y} = dxpn \times \left(\frac{U^{fem}_{x,a,y}}{\sum_{x'}\sum_{y'} U^{fem}_{x',a,y'} - RT^{fem}_{x',a,y'} - SX^{fem}_{x',a,y'}}\right) \times \sum_{x'} \sum_{y'} \left(RT^{msw}_{x'a,y'} + SX^{msw}_{x',a,y'}\right)$$

Assumptions: MSM are assumed to only partner with other MSM, non-MSM males partner exclusively with females, and females partner exclusively with non-MSM males. Notified partners are of the same age as index cases, reflecting an assumption of perfect age matching in sexual partnerships. For persons in risk group r and age a, the number

of new diagnoses resulting from partner notification is a function of the number of diagnoses from risk-based, prenatal, routine, or symptomatic testing among members of the group with whom they partner who are in the same age a, multiplied by the diagnostic yield from partner notification (dxpn). These partner notification diagnoses are distributed across year of infection and screening group according to the proportion of cases that remain undiagnosed at the end of the year in risk group r and age a that are in each year of infection and screening group.

10. Total number of cases that progress to symptomatic HIV/AIDS prior to diagnosis

$$SY = \sum_{r}\sum_{x}\sum_{a}\sum_{y} SX^{r}_{x.a,y}$$

Assumptions: All symptomatic HIV/AIDS cases are diagnosed at the time of symptom onset. Persons diagnosed before symptoms arise exit the model and do not contribute to the count of symptomatic cases.

11. Cumulative population-level years of undiagnosed infection

$$YU = \sum_{r}\sum_{a}\left(U^{r}_{x,a,y} - RB^{r}_{x,a,y} - PR_{a,y} - RT^{r}_{x,a,y}\right) + \left(\frac{1}{2}\right) \times \left(RB^{r}_{x,a,y} + PR_{a,y} + RT^{r}_{x,a,y}\right)$$

Assumptions: Persons diagnosed through risk-based, prenatal, and routine screening contribute half a year of undiagnosed infection in the year of diagnosis. Persons diagnosed through symptomatic testing or partner notification contribute 1 year of undiagnosed infection, as do those who remain undiagnosed at the end of the year.

REFERENCES

1 Lodi S, Phillips A, Touloumi G, *et al.* Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 Cells/mm(3): assessment of need following changes in treatment guidelines. *Clin Infect Dis* 2011;**53**:817–25. https://doi.org/10.1093/cid/cir494.

2 Golden MR, Hughes JP, Dombrowski JC. Optimizing the Timing of HIV Screening as Part of Routine Medical Care. *AIDS Patient Care STDS* 2017;**31**:27–32. https://doi.org/10.1089/apc.2016.0185.

3 Lee CA, Phillips AN, Elford J, *et al.* Progression of HIV disease in a haemophilic cohort followed for 11 years and the effect of treatment. *BMJ (Clinical Research Ed)* 1991;**303**:1093–6.

4 Goedert JJ, Kessler CM, Aledort LM, *et al.* A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N Engl J Med* 1989;**321**:1141–8. https://doi.org/10.1056/NEJM198910263211701.

5 Grey JA, Bernstein KT, Sullivan PS, *et al.* Estimating the Population Sizes of Men Who Have Sex With Men in US States and Counties Using Data From the American Community Survey. *JMIR Public Health and Surveillance* 2016;**2**:e14. https://doi.org/10.2196/publichealth.5365.

6 HIV/AIDS Epidemiology Unit, Public Health – Seattle & King County, Infectious Disease Assessment Unit, Washington State Department of Health. *HIV/AIDS Epidemiology Report 2018*. 2018.