eAppendix: Additional Methodological Detail

A.1. ADJUSTMENT OF DATA FOR FREQUENCY OF CSV VISITS

Since our population of interest in the mathematical modeling activities is the broad commercial sex venue (CSV)-going population—for example, all men who visit a CSV during a given year—selection of subjects in 2004 CSV survey may be considered biased towards the inclusion of frequent CSV visitors. Subjects may have reflected a typical CSV population over a short period, such as a day or perhaps across a week, but since the vast majority of infrequent yearly visitors would have had little chance of encountering study recruiters during the 2004 survey (as compared with frequent attendees), their chances of recruitment were relatively low.

Evidence of this is provided by the population-based RDD survey, conducted in 2003, which also served as an external comparison population used in adjusting the CSV sample. This RDD sample included 400 men who have sex with men (MSM), 73 of who reported attending a CSV in the last year. Of the 73 CSV visitors, 71 responded to a survey question regarding the number of times visiting a CSV in the last year. These 71 men also reported the approximate number of times they had attended a CSV in the past year—a number that we categorized into the frequency levels shown in Table A.1.

We categorized respondents of the 2004 CSV survey in similar fashion, by frequency of visit (see ‘Reported’ in Table A.1). Even a glancing comparison of the distributions of subjects by visit frequency reveals a large disparity between the two groups—with far more members of the CSV survey population (62%) than the RDD population (21%) indicating monthly or more frequent CSV visits.

To adjust the CSV survey data to match the visit frequency distribution observed in the RDD survey, we calculated a probability weight for each subject, defined by the subject’s visit frequency group. Also shown in Table A.1, this probability weight for each group was calculated as simply equal to the proportion of all respondents that fall in the group in the RDD survey, divided by the same proportion in the 2004 CSV survey.

This weighting factor was subsequently used in analyses conducted in SPSS to adjust for the sampling bias we describe here. All parameter values and initial conditions used in our models, therefore, reflect this adjustment.

A.2. ADJUSTMENT OF DATA FOR PRESENCE OF UNDIAGNOSED HIV INFECTIONS

We adjusted both the initial conditions (i.e., the initial size of the populations in CSV patron compartments) and the partnership formation rate parameters for the presence of undiagnosed HIV-infected men in the King County MSM population. In the adjustment of initial conditions, our aim was to identify the proportion of King County MSM with undiagnosed HIV infection—this proportion could then be used to reapportion men from uninfected compartments to the corresponding infected compartments. We assumed that the proportion of undiagnosed men was similar for the four activity-level combinations. Our estimate of the number of undiagnosed HIV cases is based on a recent study of MSM in five US cities conducted by the US Centers for Disease Control and Prevention (79). In this study, the investigators found that the probability of an undiagnosed HIV infection was heterogeneous across racial/ethnic groups (Table A.3). We used this information to estimate the percentage of undiagnosed HIV cases among the 2004 CSV survey population, which we presumed reflected the demographic makeup of the overall King County CSV population as well as the King County MSM population that does not attend CSV. To calculate the total number of HIV infected by racial/ethnic group, we used the following formula:

Total HIV infected = (Diagnosed HIV infected) / (1 – Proportion undiagnosed)

The total HIV infected and diagnosed HIV infected values for each racial/ethnic group appear in Table A.3. The proportion undiagnosed for each group is shown in Table A.2. The calculated number of total HIV infected is shown in Table A3 as well.

Overall, 31.3% ((73 actual cases – 50 diagnosed cases) / 73 actual cases) of infections were estimated to be undiagnosed. The final percentage that we applied to our data was 7.5% ((73 actual cases – 50 diagnosed cases) / 303 previously thought to be uninfected)—which is the percentage of self-described uninfected men we now presume were infected. We therefore reduced each susceptible compartment by 7.5%, and shifted these men to the corresponding infected compartment within the activity level group.

In adjusting the partnership formation rates for undiagnosed infections, we assumed that the mean rates for self-reported uninfected men in the 2004 survey were, in truth, a blend of rates for uninfected men and undiagnosed HIV-infected men. The actual behaviors of this latter group of undiagnosed HIV-infected men, we assumed, reflected the behaviors of the corresponding known HIV infected men within the given activity level. Our calculations indicated that approximately 7.5% of survey respondents who did not report a previous HIV-positive test were, in fact, HIV-infected. Therefore, we presumed that the UAI partnership formation rates for each susceptible compartment reflected a group consisting of 92.5% uninfected men and 7.5% infected men. We further assumed that the UAI rates for 7.5% of undiagnosed infected men in each of the four susceptible compartments were equal to the rates for the known HIV-infected men within each corresponding activity level. (For example, the partnership rate for undiagnosed infected men in n_{10} was assumed to be equal to the rate reported by for men in n_{11}.)

With these assumptions in place, simple calculations were used to adjust the partnership rates for each compartment. Reported UAI partnership formation rates are shown in Table A.4. for infected subjects (R_{a1}) and for subjects with no previous HIV-positive test (R_{a0}). The adjusted

UAI partnership formation rates for susceptible men, also shown in table A.4., were calculated using the following equation:

$$\text{Adjusted rate} = (R_{a0} - 0.075R_{a1}) / 0.925.$$

The adjusted partnership rates shown in Table A.4 were the rates ultimately used in our mathematical models (shown in Table 2 of the published manuscript).

A.3. CALCULATION OF PER-UAI ACT HIV TRANSMISSION PROBABILITY

Vittinghoff et al. (1999) provide an estimate of per-contact HIV transmission probability for unprotected receptive anal intercourse (URAI) between a known HIV-infected person and an uninfected partner (0.8%), but no estimate for transmission given unprotected *insertive* anal intercourse (UIAI) between such partners (80). The authors did, however, provide estimates of per-contact HIV transmissibility for both URAI (0.27%) and UIAI (0.06%) between uninfected men and all partners who are either HIV-infected or of unknown HIV status. Following Goodreau (108), we assumed that the ratio between estimates for UIAI and URAI using this latter criterion (0.06/0.27=22%) is similar to the ratio between UIAI and URAI probabilities between uninfected men and known HIV-infected partners. Applying this ratio to the Vittinghoff, et al. estimate of 0.82% for per-URAI contact HIV transmission probability between a known HIV-infected person and an uninfected partner, we then estimated that the per-UIAI contact of transmission between such partners is 0.18% (0.82% x 22%). We therefore assigned per-act transmission probabilities in our mathematical model of 0.82% for URAI and 0.18% for UIAI.

A.4. CALCULATION OF DISCORDANT UAI PARTNERSHIPS

This process consisted of two broad steps: 1) identification of the daily number of UAI partnerships for men in each compartment (i.e. activity-level/HIV-status combination), by partner activity level, by location (CSV vs. outside CSV); and 2) calculation of the number of discordant

UAI partnerships for each activity-level/HIV-status combination, accounting for serosorting preference among men.

(1) We begin by determining the number of UAI partnerships initiated in location *l* for all men of activity level *a*, HIV status *i*, at time *t* (d_{lah}, i.e., total "degree") for each compartment:

$$d_{lah}(t) = n_{ah}(t) \bullet c_{lah}$$

where $n_{ah}(t)$ is the number of men in activity class *a* with serostatus *h* at time *t*, and c_{lah} is the rate of new partner acquisition by men in activity class *a* with serostatus *h* in location *l*. We then assumed proportional mixing by activity level (within location), and random selection of role (insertive/receptive) with regard to activity status and HIV status. We define $p_{la_ih_ia_rh_r}$ as the number of new partnerships per unit time in location *l*, between an insertive partner of activity class a_i and HIV status h_i and a receptive partner of activity class a_r and HIV status h_r. Using a dot to signify a quantity summed across all values of a given subscript, we have:

$$p_{la_ih_ia_r\cdot} = \frac{d_{la_ih_i} d_{la_r\cdot}}{2d_{l\cdot\cdot}}$$

$$p_{la_i\cdot a_rh_r} = \frac{d_{la_rh_r} d_{la_i\cdot}}{2d_{l\cdot\cdot}}$$

The two in the denominator reflects the fact that the total number of contacts for men in a given compartment is partitioned into half insertive and half receptive contacts.

<u>(2) Calculation of number of discordant UAI partnerships:</u> The values from the previous step, in conjunction with the use of an odds ratios reflecting serosorting in CSV2004, allowed for the calculation of the number of discordant partnerships of each type. This calculation may be conceptualized by arraying our many contact rates into a series of 2x2 contingency tables (see Table A.5) showing the distribution of partnerships in location *l* involving members of activity group a_i as insertive UAI partners and members of activity group a_r as receptive partners, by HIV

status. The cells labeled $p_{la_i 0 a_r 1}$ and $p_{la_i 1 a_r 0}$ show the numbers of discordant contacts per unit time in which the susceptible is insertive and receptive, respectively. Because our model assumes no seropositioning, these values are by definition always equal to one another.

We know the margins of these totals and the odds ratios for the cell entries; given these, there is a unique solution for the cell entries (see reference 30). Defining $\alpha_{la_i a_r}$ as the total number of serodiscordant contacts between men in activity classes a_i and a_r in location l per unit time, we have:

$$\alpha_{la_i a_r} = p_{la_i 0 a_r 1} + p_{la_i 1 a_r 0} = \frac{OR \times p_{la_i \cdot a_r 0} - OR \times p_{la_i 0 a_r \cdot} - p_{la_i 1 a_r \cdot} - p_{la_i \cdot a_r 0} \pm \sqrt{\left(OR \times p_{la_i 0 a_r \cdot} + p_{la_i 1 a_r \cdot} + p_{la_i \cdot a_r 0} - OR \times p_{la_i \cdot a_r 0}\right)^2 + 4(OR - 1)\left(p_{la_i \cdot a_r 0} \times p_{la_i 1 a_r \cdot}\right)}}{(OR - 1)}$$

A.5. SENSITIVITY ANALYSIS

We designed a simple process to allow for an exploration of the CSV partnership formation rate parameters in both the main model and counterfactual models in a parallel fashion. Since the CSV "replacement" partnership formation rates in the counterfactual models were derived from the number of overall CSV sex partners, and the main model CSV rates reflected the number of reported UAI partners in CSV (effectively a subset of the number of overall CSV sex partners), we based the LHS parameter value selection on the mean number of overall CSV sex partners per group. A multiplier value was calculated for each main model (Table A6) and counterfactual (Table A7) version of the of the CSV partnership formation rates, which was equal to the UAI partnership rate divided by the overall CSV partnership formation rate for the activity level. (In other words, each multiplier was equal to the proportion of all sex partners that were UAI partners in each scenario.) This multiplier was then used to recalibrate the rates selected in the LHS from overall sex partner rates to UAI partnership rates, through calculating the product of the multiplier and the overall CSV sex partner rate from the LHS.

While outside of CSV UAI partnership rates for CSV patrons were identical in the main and counterfactual models, we did perform a similar process for the outside of CSV partnership rates in the sensitivity analysis. An overall sexual partnership rate was calculated for each HIV status-activity level group, and a multiplier was calculated using this number and the UAI partnership rate identified previously. The overall sexual partnership rate for each group (also shown in Table A6) was used as the mean value in the LHS, and the values selected in the LHS process were recalibrated back to UAI partnership rates, using the multiplier as described above.

The distributions used in the LHS were intended to capture the distribution of parameter estimates that would be derived from multiple large samples. Thus, following the central limit theorem, each parameter was modeled using a normal distribution, The standard error of the mean for each value, also calculated from the 2004 CSV survey data, was entered as a measure of variability in the LHS for all parameters except the non-CSV partnership rate. We did not have immediate access to the RDD2003 data, which provided the non-CSV partnership rate estimate. However, since the standard deviation of the eight CSV patron partnership rates showed a striking pattern of similarity with the mean partnership rate values for the groups, we assumed that the standard deviation for this parameter was equal to its mean. This standard deviation was converted to a standard error for use in the LHS.

A.6. PARTIAL CORRELATION ANALYSIS

Findings for the partial correlation analysis (Table A8) provide information regarding the linear associations between the parameter variables listed and the 10-year attributable number of HIV cases . Although we observed relatively consistent outcomes across the many runs within each replacement scenario, the partial correlation analysis informs us that the variation that does exist across runs is most strongly associated with three of the model inputs examined: HIV-negative, low CSV activity rate; HIV-positive, low CSV activity rate; and HIV-positive, high CSV activity rate. The former captures the CSV activity of the largest subpopulation, while the later two collectively define the CSV activity of the positive men, who might be expected to drive the infection process disproportionately. If one wished to identify the number of HIV cases attributable to bathhouses more precisely than we have here, collecting better data on these three parameters would be of greatest value.

Table A.1. Frequency of CSV visits, RDD and 2004 CSV survey populations

	RDD survey		CSV survey				
Frequency			Reported		Probability	Weighted	
	n	%	n	%	weight	n	%
Infrequent	30	42.3%	78	21.7%	1.95	152	42.3%
< monthly	26	36.6%	57	15.8%	2.31	132	36.6%
Monthly	9	12.7%	58	16.1%	0.79	46	12.7%
Every other week	4	5.6%	71	19.7%	0.29	20	5.6%
Almost wkly	1	1.4%	30	8.3%	0.17	5	1.4%
Weekly	1	1.4%	66	18.3%	0.08	5	1.4%
Total	71	100.0%	360	100.0%		360	100.0%

Note: Infrequent=1-2 visits in a year

Table A.2. Proportion with undiagnosed HIV infection by racial/ethnic group

Group	Proportion undiagnosed
White	0.18
Black	0.67
Latino	0.48
Multiracial	0.50
Other	0.50
Total	0.48

Source: CDC (2005)

Table A.3. Diagnosed and estimated undiagnosed HIV cases by racial/ethnic group

	n	%	Diagnosed HIV infection (n)	No diagnosed HIV infection(n)	Total HIV infected (n)
White	262	70.4%	37	211	45
Asian/PI	33	8.9%	3	29	6
Black	21	5.6%	2	19	6
Latino	32	8.6%	5	24	10
AK Nat/Am Ind	6	1.6%	2	3	4
Other/Mixed	18	4.8%	1	17	2
Total	372	100%	50	303	73

Source: 2004 CSV survey
Note: Shaded cells indicate estimated values

Table A.4. UAI partnership formation rates, self-reported HIV infected men and adjustment of UAI partnership formation rates for undiagnosed HIV infections

	Susceptible		Infected		
Rate	Activity level group (n_{a1})	Reported rate (R_{a1})	Activity level group (n_{a0})	Reported rate (R_{a0})	Adjusted rate
CSV	n_{11}	0.00217	n_{10}	0.00085	0.00075
	n_{21}	0.00217	n_{20}	0.00085	0.00075
	n_{31}	0.10357	n_{30}	0.03946	0.03426
	n_{41}	0.10357	n_{40}	0.03946	0.03426
Outside	n_{11}	0.00115	n_{10}	0.00101	0.00100
	n_{21}	0.07240	n_{20}	0.04150	0.03900
	n_{31}	0.00115	n_{30}	0.00101	0.00100
	n_{41}	0.07240	n_{40}	0.04150	0.03900

Source: 2004 CSV survey. Adjusted rate=R_{a0}-0.075R_{a1}/0.925.

TABLE A.5. Concordant and discordant partnerships between partners in activity classes a_i and a_r, in location l

Insertive partner in activity class a_i:	Receptive partner in activity class a_r: HIV-	HIV+	Total
HIV-	$p_{la_i-a_r-}$	$p_{la_i-a_r+}$	$p_{la_i-a_r\square}$
HIV+	$p_{la_i+a_r-}$	$p_{la_i+a_r+}$	$p_{la_i+a_r\square}$
Total:	$p_{la_i\square a_r-}$	$p_{la_i\square a_r+}$	

TABLE A.6. Main model: Means, standard errors of means, and multipliers for sensitivity analysis

Parameter	CSV			Outside of CSV		
	Mean[a]	Standard error of mean	Multiplier[b]	Mean[a]	Standard error of mean	Multiplier[b]
HIV-negative: Low CSV activity	0.08	0.01	0.01	0.04	0.00	0.03
High CSV activity	0.29	0.05	0.12	0.15	0.03	0.25
HIV-positive: Low CSV activity	0.12	0.02	0.02	0.04	0.01	0.03
High CSV activity	0.31	0.06	0.34	0.14	0.03	0.52
All non-CSV MSM				0.003	0.0002	
Serosorting log odds ratio[c]	1.04	0.14				

[a]Mean partnership formation rates – all types of sex, main model. All assumed to be normally distributed.
[b]Multipliers equal to probability that sexual partnerships will involve UAI
[c]Serosorting odds ratio applies to both CSV and outside of CSV

TABLE A.7. Counterfactual model: Multipliers for sensitivity analysis

Variable	Multiplier
HIV-negative: Low outside activity	0.10
High outside activity	0.37
HIV-positive: Low outside activity	0.12
High outside activity	0.70

Note: Means and standard errors of means presented in Table 2.8 used for counterfactual models. Multipliers used to convert selected values to UAI partnership formation rates.

TABLE A.8. Partial correlation coefficients: correlation with 10-year attributable number of HIV infections, by percent replacement of CSV partnerships

Variable	Counterfactual - Percent replacement				
	100%	75%	50%	25%	None
HIV-negative, low CSV activity rate	0.88	0.67	0.58	0.69	
HIV-negative, low outside activity rate	(0.04)	(0.12)	0.01	0.07	(0.02)
HIV-negative, high CSV activity rate	(0.33)	(0.30)	(0.48)	(0.96)	
HIV-negative, high outside activity rate	0.12	0.05	0.09	0.06	(0.06)
HIV-positive, low CSV activity rate	0.71	0.51	0.36	0.66	
HIV-positive, low outside activity rate	0.21	0.12	0.05	0.00	0.06
HIV-positive, high CSV activity rate	0.86	0.59	0.31	(0.09)	
HIV-positive, high outside activity rate	(0.07)	(0.29)	(0.03)	0.08	(0.07)
Non-CSV rate	0.02	(0.06)	0.14	(0.01)	(0.16)
Serosorting OR	(0.14)	(0.13)	(0.07)	0.28	(0.05)

Note: Parentheses indicate negative numbers. All variable coefficients were adjusted for all other variables listed in table, except for in the no replacement scenario, in which CSV activity rate variables were not applicable.