# Online Supplemental Information

## Study Sample Selection Procedure

Our study sample came from a parent study, Young Men’s Affiliation Project (YMAP), a longitudinal social network project to examine HIV transmission among MSM between the ages of 16 and 29. YMAP investigates social networks and networks formed through affiliations with social venues and health-promoting venues and determines how these networks affect HIV/STD risk and prevention in Houston, TX (*N* = 378) and Chicago, IL (*N* = 377). YMAP participants were recruited using respondent-driven sampling (RDS) ([Heckathorn, 1997](#_ENREF_2)) between 2014 and 2016. Eligibility criteria include male sex assigned at birth and current male identification, engaging in oral or anal sex with another male in the prior year, residing in and planning to remain in Houston or Chicago for the following year, and English-speaking ability.

Among 755 YMAP participants (*N* = 378 for Houston and *N* = 377 for Chicago), we first restricted our study sample to 468 YMAP participants with self-reported Black race (*n* = 227 for Houston and *n* = 241 for Chicago). Among these 468 YBMSM, our study sample was restricted to 396 YBMSM participants (*n* = 211 for Houston and *n* = 185 for Chicago) with biologic testing results (a rapid plasma reagin (RPR) test and/or a fluorescent treponemal antibody (FTA) absorption test) by excluding 72 (15.4%) of the 468 YBMSM participants. Figure 1 in the main manuscript illustrates an aggregated referral/social/sex network for each city by syphilis-HIV infection status, using these 396 YBMSM. Our sample for statistical analysis was further restricted to 365 YBMSM participants (*n* = 192 for Houston and *n* = 173 for Chicago) by list-wise’s deleting 31 (7.8%) that had missing values for at least one of the study variables defined in the Measures section in the manuscript and excluding missing or invalid information of personal network size (for the purpose of RDS-adjusted estimation). Thus, among the 468 YBMSM participants, a total of 103 (22.0%) were excluded from our study sample, and 365 counterparts (78%) were retained for our analysis.

## Name Generator for Sex Networks

In the YMAP survey, participants were asked how many total different persons, including men, women, and transgender people, in the last six months with whom they had engaged in oral, anal, or vaginal sex. Once this number was collected, participants were asked to nominate up to five people with whom they had anal, oral, or vaginal sex within the past six months in chronological order (starting from the person with whom they had sex most recently). Then, participants were asked some follow-up questions for each nominated sex partner that concerned anal sex, sex position, and condom use. The measures were derived primarily from a population-based study of Black MSM (Schneider et al., 2017).

## Generating Social Networks that Combine Peer-referral, Social, and Sex Ties via Matching Procedure

We combined three network data sources for each city. These network data sources involved a peer-referral network that was generated by the RDS sampling process, participants’ social network (participants were asked to nominate up to five people with whom they shared personal information), and the sexual networks described above. We employed a matching procedure based on a fuzzy matching algorithm that employs participants’ and their partners’ sociodemographic information, such as first name, last name, preferred name, age, gender, and race, to determine whether the pairs listed are the same person ([Shah et al., 2014](file:///C:\work\PrEP_paper\co-infection_Syphilis_HIV\manuscript\STI_revise_resubmission_Sep2017\revised_draft_both_RDS-II_robust_11082017\sent_to_John_11092017\description%20of%20matching%20procedure_online%20supplement_brittedit_v2.docx#_ENREF_3)). Because the names are in a textual format, they were translated to phonetic codes (in terms of numeric values) by using R package “stringdist” ([Van der Loo, 2014](file:///C:\work\PrEP_paper\co-infection_Syphilis_HIV\manuscript\STI_revise_resubmission_Sep2017\revised_draft_both_RDS-II_robust_11082017\sent_to_John_11092017\description%20of%20matching%20procedure_online%20supplement_brittedit_v2.docx#_ENREF_5)), and the values of the phonetic codes for participant and partner were compared. Thus, similar names should have yielded similar or equal codes.

One similarity metric was created to take into account all of the differences from these phonetic codes and each piece of sociodemographic information mentioned above. The similarity metric score becomes lower as the difference becomes more pronounced. The threshold score for the similarity metric was set up to “weight” the information provided on each alter via the order that it appears in the score. For instance, the score is ordered by information on the following: phoneticized last name, phoneticized first name, phoneticized nickname, age, gender, and race (defined as Black/African American versus not Black/African American). We increased the threshold score by assigning more weight to phoneticized last name, first name, and preferred name, as we had many nominations with only initials as their names. Therefore, larger scores are those with more confidence in similarity in the last name or first name versus confidence in race. The final matching was conducted manually; therefore, the threshold score was used as a guide to filter out pairs that were not probable matches. There was no optimal threshold value.

The composite list of paired nodes that met the threshold was reviewed and scored by a coder. The coder manually scored each pair on a 4-point scale, from 3 as indicating that they were “extremely confident that it is the same person” to 0 as indicating that they were “extremely confident that it is not the same person.” Then, the program was run to verify that matched pairs were transitive and to add missing pairs to achieve transitive sets of pairs (e.g., if A matched B and B matched C, if a match between A and C was missing, it was generated). Comparisons with a score of 3 were considered to be a “match” (the same person). A new set of unique IDs was created for all nodes, with matched nodes’ receiving the same the ID (Part 2 of the matching program that we provided). An edge (tie) list was created for all egos (respondents) and alters (social and sexual network partners) based on the new, unique IDs. The complete network was generated and checked for coherence (i.e., respondents’ being matched). This allowed us to identify a small number of incorrect matches that were then removed, and the renumbering with unique IDs was redone.

Our algorithm was applied to the entire YMAP Wave 1 dataset (Houston *n* = 378; Chicago *n* = 377). We computed the number of new sex or/and social pairs matched, using a matching algorithm, and divided it by the total number of possible matched sex and/or social pairs by name only, without using a matching algorithm, which yielded 20.8% for Houston data, and 25.4% for Chicago data.

## Computation of Network Exposure Variables

In this study, we defined a one-mode adjacency matrix, , with = 1 if an actor *i* (ego)is directly connected to an actor *j* (alter) in any relationships (peer referral/personal/sex ties), and = 0 otherwise. Diagonal values are set to zero. By multiplying by a vector of with an element of alter ’s status of disease infection (coded 1 for infected, and 0 otherwise), i.e.,, a resulting vector yields the count of disease-infected alters who are directly connected to an actor *i* (or egocentric network)*.* This computation is equivalent tothe numerator part of network exposure model ([Valente, 2010](#_ENREF_4)). Our study measured three types of network exposure variables, i.e., network exposure to co-infected members, HIV-infected members, and syphilis-infected members. These network exposure variables were recoded as a dichotomous variable by coding 1 for having at least one disease-infected alter (), and 0 otherwise.

## Formulation and Specification of Multinomial Logistic Regression Models

This study used multinomial logistic regression analysis, as our outcome variable is nominally scaled with more than two levels. Our outcome variable of syphilis-HIV infection status, *Y*, consists of four sub-group categories that represents the following sub-groups, , that are defined in the following and shown in the table:

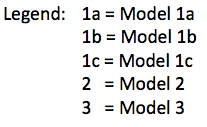
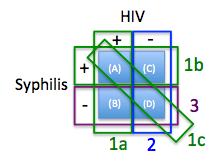


Table 1

With these four outcome categories, we specified two sets of multinomial logistic regression models by choosing a specific sub-group category as a reference category, with each model’s consisting of three logit functions that are estimated simultaneously. Note that different choices of a reference category generate different estimated coefficients but that predicted probabilities will not be affected ([Hardin, Hilbe, & Hilbe, 2012](#_ENREF_1)).

Our first set of multinomial logit models (represented in the green boxes in Figure 1) covers three logit functions for Model 1a, Model 1b, and Model 1c. This was specified by setting the reference outcome category as *Y* = *A*, and the three logit functions are defined in the following:

where represents each logit function, *x* represents a vector of *p* covariates, and represents the corresponding parameters for each logit function. Then, we re-parametrized these logit functions for Outcome Sub-group A as a numerator and redefined them as follows:

The first logit function, , represents Model 1a and compares *Y* = *A* versus *Y* = *B*; the second logit function, , represents Model 1b and compares *Y* = *A* versus *Y* = *C*; and the third logit function, , represents Model 1c and compares *Y* = *A* versus *Y* = *D*. In this first set of multinomial logistic regressions, parameters were estimated (i.e., 3 logit functions and *p* parametersplus 1 constant term).

Our second set of multinomial logistic regressions covers three logit functions for Model 1c (also defined above), Model 2 (represented by the blue box in Figure 1), and Model 3 (represented in the purple box in Figure 1) by setting the reference outcome category as *Y* = *D*. These three logit functions are defined as follows:

where , *x*, and are defined above. The first logit function, , represents Model 1c and compares *Y* = *A* versus *Y* = *D*, which is equivalent to the re-parametrized logit function in the first set of multinomial logit models. The second logit function, , represents Model 2 and compares *Y* = *C* versus *Y* = *D*, and the third logit function, , represents Model 3 and compares *Y* = *B* versus *Y* = *D*.

### Reference

### Hardin JW, Hilbe JM, Hilbe J. *Generalized Linear Models and Extensions* 2012. College Station, Texas: Stata Press.

### Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems* 1997;44(2):174-199.

### Shah, NS, Iveniuk J, Muth SQ, et al. Structural bridging network position is associated with HIV status in a younger Black men who have sex with men epidemic. *AIDS Behav 2014;18*(2):335-345.

Schneider JA, Cornwell B, Jonas A, et al. Network dynamics of HIV risk and prevention in a population-based cohort of young Black men who have sex with men. *Network Science (Cambridge University Press)* 2017;5(3):381-409.

### Valente TW. *Social networks and health: Models, methods, and applications*; 2010. Oxford University Press.

### Van der Loo MP. The stringdist package for approximate string matching. *The R*, 2. 2014

**Web References**

w1. Fujimoto K, Wang P, Kuhns L, Ross MW, et al. Multiplex competition, collaboration, and funding networks among social and health

organizations: towards organization-based HIV interventions for young men who have sex with men. *Med Care* 2017;*55*(2):102–110.

w2. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006;1695(5):1–9.

Table 1(S): Results of Estimated Relative Risk Ratios (SE; lower, upper 95% CIs) of Syphilis and HIV Co- and Mono-infection among Young Black Men Who Have Sex with Men (YBMSM) from the YMAP Cohort (*N* = 378), 2014–2016.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Outcome category | Syphilis-HIV co-infection | | | Syphilis mono-infection | HIV mono-infection |
| Model specification | Model 1a | Model 1b | Model 1c | Model 2 | Model 3 |
| Sub-group outcome category relative to a reference sub-group in Figure 1 | Syphilis+ HIV+ (A)  relative to  Syphilis– HIV+ (B) | Syphilis+ HIV+ (A)  relative to  Syphilis+ HIV– (C) | Syphilis+ HIV+ (A)  relative to  Syphilis– HIV– (D) | Syphilis+ HIV– (C)  relative to  Syphilis– HIV– (D) | Syphilis– HIV+ (B)  relative to  Syphilis– HIV– (D) |
| Relative Risk (RR) | RR of co-infection vs. HIV mono-infection | RR of co-infection vs. syphilis mono- infection | RR of co-infection vs. neither infection | RR of syphilis mono-infection vs. neither infection | RR of HIV mono- infection vs. neither infection |
| Age | 1.06 (0.06; 0.94, 1.19) | 0.97 (0.07; 0.84, 1.12) | 1.04 (0.05; 0.94, 1.15) | 1.07 (0.07; 0.94, 1.21) | 0.98 (0.05; 0.89, 1.08) |
| Education (high school or less) | 0.71 (0.25; 0.36, 1.40) | 0.66 (0.29; 0.27, 1.58) | 0.80 (0.24; 0.44, 1.46) | 1.22 (0.48; 0.56, 2.65) | 1.13 (0.30; 0.66, 1.94) |
| Lifetime housing instability | 0.87 (0.35; 0 .39, 1.91) | 1.27 (0.66; 0.46, 3.53) | 0.87 (0.32; 0.42, 1.79) | 0.68 (0.30; 0.29, 1.61) | 1.00 (0.30; 0.56, 1.80) |
| Sex partners (prior 6 months | 1.29 (0.23; 0.91,1.84) | 1.17 (0.25; 0.76,1.79) | 1.00 (0.14; 0.76, 1.30) | 0.85 (0.17; 0.58, 1.25) | 0.77 (0.12; 0.56, 1.05) |
| Sex partners with drug use | 0.84 (0.13; 0.62,1.14) | 1.27 (0.35; 0.74, 2.17) | 0.93 (0.14; 0.69, 1.24) | 0.73 (0.19; 0.43, 1.22) | 1.10 (0.14; 0.86, 1.40) |
| Number of condomless bottom partners | 0.70 (0.18; 0.42, 1.18) | 0.79 (0.24; 0.43, 1.45) | 0.82 (0.19; 0.52, 1.31) | 1.04 (0.27; 0.63, 1.74) | 1.17 (0.21; 0.82, 1.65) |
| Number of condomless top partners | 1.27 (0.20; 0.93,1.73) | 1.31 (0.31; 0.82, 2.09) | 1.51\* (0.24; 1.10, 2.07) | 1.15 (0.27; 0.73, 1.82) | 1.19 (0.19; 0.86, 1.64) |
| Syphilis-infected network member(s) | 1.38 (0.49; 0.69, 2.78) | 1.61 (0.68; 0.70, 3.69) | 1.88† (0.61; 1.00, 3.54) *p* = 0.050 | 1.17 (0.44; 0.56, 2.44) | 1.36 (0.40; 0.77, 2.42) |
| HIV-infected network member (s) | 2.31\* (0.82; 1.15, 4.64) | 1.61 (0.67; 0.71, 3.65) | 2.67\*\* (0.85; 1.44, 4.97) | 1.66 (0.63; 0.79, 3.48) | 1.16 (0.36;0.63, 2.12) |
| Syphilis-HIV co-infected member(s) | 1.78 (0.68; 0.84, 3.77) | 2.04 (0.96; 0.81, 5.14) | 2.60\*\* (0.91; 1.30, 5.17) | 1.28 (0.57; 0.53, 3.07) | 1.46 (0.49; 0.75, 2.83) |
| Number of health venues attended | 1.02 (0.08; 0.88, 1.19) | 1.12 (0.10; 0.94, 1.32) | 1.08 (0.07; 0.95, 1.23) | 0.97 (0.08; 0.83, 1.13) | 1.05 (0.07; 0.92, 1.21) |
| Number of social venues attended | 1.01 (0.06; 0.91, 1.13) | 0.97 (0.05; 0.88, 1.07) | 0.96 (0.04; 0.89, 1.04) | 0.99 (0.04; 0.91, 1.08) | 0.95 (0.05; 0.86, 1.05) |

Note: †*p* < 0.10; \**p* < 0.05; \*\**p* < 0 .01 for two-tailed test. Parentheses indicate standard errors. The variable of city was controlled. Active syphilis infection is defined as having RPR 1:4 or greater. The results for sociodemographic, behavioral, and venue attendance variables are based on the model with the inclusion of the network variable of having Syphilis-HIV co-infected member(s). The results are based on analysis without RDS-II weights (using Huber-White robust sandwich variance estimator).