

A new mixed agent-based network and compartmental simulation framework for joint modeling of related infectious diseases- Application to sexually transmitted infections

Chaitra Gopalappa, PhD, *Hari Balasubramanian, PhD, *Peter J. Haas, PhD

¹University of Massachusetts Amherst, 160 Governors Drive, Amherst, MA 01003

*HB and PJH equally contributed to the work and are listed in alphabetical order.

Corresponding author: Chaitra Gopalappa, chaitrag@umass.edu, Associate Professor, University of Massachusetts Amherst, 160 Governors Drive, Amherst, MA 01003

Keywords: joint-modeling diseases; multi-disease modeling; simulation modeling; HIV and STIs; diseases and social determinants; SDOH

ABSTRACT

Background

A model that jointly simulates infectious diseases with common modes of transmission can serve as a decision-analytic tool to identify optimal intervention combinations for overall disease prevention. In the United States, sexually transmitted infections (STIs) are a huge economic burden, with a large fraction of the burden attributed to HIV. Data also show interactions between HIV and other sexually transmitted infections (STIs), such as higher risk of acquisition and progression of co-infections among persons with HIV compared to persons without. However, given the wide range in prevalence and incidence burdens of STIs, current compartmental or agent-based network simulation methods alone are insufficient or computationally burdensome for joint disease modeling. Further, causal factors for higher risk of coinfection could be both behavioral (i.e., compounding effects of individual behaviors, network structures, and care behaviors) and biological (i.e., presence of one disease can biologically increase the risk of another). However, the data on the fraction attributed to each are limited.

Methods

We present a new mixed agent-based compartmental (MAC) framework for jointly modeling STIs. It uses a combination of a new agent-based evolving network modeling (ABENM) technique for lower-prevalence diseases and compartmental modeling for higher-prevalence diseases. As a demonstration, we applied MAC to simulate lower-prevalence HIV in the United States and a higher-prevalence hypothetical Disease 2, using a range of transmission and progression rates to generate burdens replicative of the wide range of STIs. We simulated sexual transmissions among heterosexual males, heterosexual females, and men who have sex with men (men only and men and women). Setting the biological risk of co-infection to zero, we conducted numerical analyses to evaluate the influence of behavioral factors alone on disease dynamics.

Results

The contribution of behavioral factors to risk of coinfection was sensitive to disease burden, care access, and population heterogeneity and mixing. The contribution of behavioral factors was generally lower than observed risk of coinfections for the range of hypothetical prevalence studied here, suggesting potential role of biological factors, that should be investigated further specific to an STI.

Conclusions

The purpose of this study is to present a new simulation technique for jointly modeling infectious diseases that have common modes of transmission but varying epidemiological features. The numerical analysis serves as proof-of-concept for the application to STIs. Interactions between diseases are influenced by behavioral factors, are sensitive to care access and population features, and are likely exacerbated by biological factors. Social and economic conditions are among key drivers of behaviors that increase STI transmission, and thus, structural interventions are a key part of behavioral interventions. Joint modeling of diseases helps comprehensively simulate behavioral and biological factors of disease interactions to evaluate the true impact of common structural interventions on overall disease prevention. The new simulation framework is especially suited to simulate behavior as a function of social determinants, and further, to identify optimal combinations of common structural and disease-specific interventions.

1. INTRODUCTION

A model that jointly simulates infectious diseases that have common modes of transmission will be a useful decision-analytic tool to identify optimal intervention combinations for overall disease prevention. In the United States, sexually transmitted infections (STIs) continue to impose high disease and economic burdens. The lifetime medical costs for treatment of STIs and sequelae, for incident infections in 2018, were estimated at \$15.9 billion, with a large fraction attributed to human immunodeficiency virus (HIV) infection (\$13.7 billion)(1). Further, there are considerable interactions between HIV and other STIs, including human papillomavirus (HPV), hepatitis C (HCV), hepatitis B (HBV), gonorrhea (NG), chlamydia (CT), and syphilis, in terms of HIV acquisition, transmission, or progression as seen in the following examples. The odds of HPV infection, infection with high-risk oncogenic HPV types that lead to cervical cancer, HBV infection, and liver-related mortalities from HCV or HBV are higher in women living with HIV compared to women without an HIV infection (2–5). Among persons with HIV (PWH), the risk of cervical cancer (3) and mortality from liver cancer (6) were directly correlated with HIV stage. Persons with syphilis, CT, and NG have higher risk of HIV transmission and acquisition (7,8). Studies also show higher risk of CT and NG among HIV infected persons who also had a recent HCV infection or syphilis infection(9). Further, though there is sufficient evidence associating risk and severity of coinfections to biological factors (i.e., presence of one disease can biologically increase the risk of another), these estimates are confounded by behavioral factors (combined effects of individual sexual behavior, partnership network structure, and care behaviors). Observational studies alone are insufficient to quantify risk attributed to each factor(10,11).

A model that can jointly simulate STIs can quantify risk of disease attributable to behavioral and biological factors, determine intervention needs, and jointly evaluate interventions for overall disease prevention. Interventions to address biological factors include disease management interventions, such as pharmaceutical and care support programs. On the other hand, structural interventions, such as health care coverage, subsidized housing, childcare and food programs, access to mental healthcare, and early

childhood academic enrichment programs (12–16) are key part of behavioral interventions. This is due to the fact that social and economic conditions are among key drivers of behaviors associated with STI transmission, e.g., higher number of partners, higher rates of condomless sex, higher rates of substance abuse, and lower care access among people experiencing homelessness compared to those stably housed (17–20). Consequently, social determinants are key correlates of disease burden. Among persons living with diagnosed HIV infection, 44% had a disability (including physical, mental, and emotional disabilities), 41% were unemployed, 43% had household incomes at or below the federal poverty threshold, and 10% were experiencing homelessness (21,22). A joint disease model, that simulates disease interactions through the behavioral factors that are common modes of transmission, will be ideal to simulate behaviors as functions of social determinants, and eventually, serve as a decision-analytic tool for identifying the most cost-effective combinations of disease-specific and common structural interventions for overall STI prevention.

Multi-disease models in the literature are limited, and most focus on chronic diseases prediction(23,24) or management (25–27). Some focus on infectious diseases, but use a single simulation technique (10,11,28,29) which is not computationally sufficient for modeling diseases of widely varying incidence and prevalence. Other models only do combined health infrastructure costing of diseases without modeling disease interactions (30). Among commonly used simulation techniques, compartmental modeling is sufficient for fast-spreading or high-prevalence infections, whereas agent-based network modeling (ABNM) is preferred for slower-spreading, lower-prevalence infections, where granular representations of network structures and individual-level characteristics are key for accurate estimations (31). Even ABNM, however, becomes computational challenging to use for diseases with very low prevalence, such as HIV in the United States. Thus, in our previous work, we developed an agent-based evolving network modeling (ABENM) technique, which uses a hybrid ABNM and compartmental structure for single disease modeling of lower-prevalence diseases (32). This simulation technique was applied to develop PATH 4.0 (Progression and Transmission of HIV), a comprehensive simulation model

of HIV in the United States that was validated against metrics from the National HIV Surveillance System (NHSS) for the period 2006 to 2017(33). While HIV, HBV, HCV, and syphilis are slower-spreading, lower-prevalence diseases, HPV, NG, and CT are faster-spreading higher-prevalence diseases(34). Therefore, ABNM or ABENM alone will not be sufficient for co-modeling of these diseases , due to computational challenges (discussed in Methods) arising from co-modeling widely varying disease burdens in a national context. We developed a mixed agent-based compartmental (MAC) framework that uses ABENM for lower-prevalence diseases and compartmental model for higher-prevalence diseases. This paper presents the MAC mathematical framework for multi-disease modeling. We demonstrate the framework by applying it to the case of two-disease modeling, representing Disease 1 via ABENM and Disease 2 via compartmental modeling. We adopt the PATH 4.0 model to represent HIV as lower-prevalence Disease 1, and we construct a hypothetical higher-prevalence Disease 2, evaluating it with a range of values for epidemiological parameters to generate incidence and prevalence representative of the wide range of STIs. Further, we conduct numerical analyses to quantify the risk of coinfection attributable to behavioral factors alone, assessing its sensitivity to disease burden, care access, and population heterogeneity. This paper serves as proof-of-concept for the MAC framework, to demonstrate its feasibility and to highlight, through the numerical analyses, the potential significance of joint modeling of diseases.

2. METHODS

Figure 1 gives an overview of the MAC framework, which employs the ABENM-based PATH 4.0 model of HIV(33). The main concept of ABENM as used in PATH 4.0 is to simulate persons infected with HIV and their immediate contacts as individual agents and all other persons using a compartmental model. Immediate contacts are defined as all sexual partners a person will have over their lifetime; at the current time-step they could either be infected or susceptible. In-turn, as these susceptible contacts in the network become infected with HIV, their immediate contacts are added as agents to the network (transitioning

from the compartmental portion of the model to the network portion of the model), thus *evolving the contact network*. An Evolving Contact Network Algorithm (ECNA) maintains the network dynamics.

The MAC simulation framework expands on the above concepts of ABENM, where all persons in a population are either in a network or in a compartmental model. To accommodate the multiple diseases, it simulates only persons infected with at least one lower-prevalence disease and their immediate contacts as agents in a network, and it simulates all other persons, including those infected with only higher-prevalence diseases, via a compartmental model (see Figure 1). As the exposed immediate contacts of agents in the network become infected with a lower-prevalence disease, their contact network is generated by moving persons from the compartmental model to the network using the ECNA. As HIV or HCV are slower-spreading lower-prevalence diseases in the United States, with an annual incidence at 13 and 1, respectively, per 100,000 persons in 2018 (Figure 2a)(34), they are suitable candidates for ABENM. As HPV, CT, or GN are faster-spreading higher-prevalence diseases in the United States, with annual incidence at 1820, 640, and 212, respectively, per 100,000 persons in 2018 (Figure 2a)(34), they are suitable candidates for compartmental modeling. The MAC framework would be computationally advantageous over ABNM for jointly modeling these diseases that have varying epidemiological features.

For example, in ABNM, simulating 100,000 nodes in the network representative of the sexually active population in the United States will generate sufficient samples for the higher-prevalence disease, but for lower-prevalence HIV it would generate 490 persons with HIV with 13 new cases each year, which is not a sufficient sample size for our analyses. In contrast, in ABENM, as the network will consist of only persons with the lower-prevalence disease and their partners, the population size in the network can be controlled to meet computational needs without compromising the sample size of either disease. As an example, we can generate the network with a sufficient but computationally efficient sample of HIV infected persons, say 12,000 (as was the case in our numerical analyses), as a scaled representation of the 0.4% HIV prevalence in 2017 in the United States(35), which is equivalent to simulating a population size of ~3 million (compartmental plus network).

2.1 Overview of MAC

We present an overview of the MAC simulation framework, using, without loss of generality, a two-disease example, lower-prevalence Disease 1 and higher-prevalence Disease 2. An overview of the MAC simulation framework is presented in Figure 1 using, for illustration, two diseases, Disease 1 is tracked in a network model and Disease 2 is tracked through a compartmental model. We describe below an overview of the model and present its mathematical formulations in the Appendix.

2.1.1 Computational structure of MAC

We present below a brief description of the MAC computational structure and present its mathematical representation in Appendix S1.1 and S1.2. We track Disease-1-infected persons and immediate contacts using a dynamic graph $G_t(\mathcal{N}, \mathcal{E})$, with the number of nodes $Q_t = |\mathcal{N}(G_t)|$ and the number of edges $|\mathcal{E}(G_t)|$ in the graph dynamically changing over time t as persons become newly infected with Disease 1 and their immediate contacts are added to the network. Each person in the network has attributes such as age, HIV transmission category (heterosexual males, heterosexual females, men who have sex with men), degree (number of lifetime partnerships), geographic jurisdiction, and health status ($\{\text{stage of Disease 1, stage of Disease 2}\}$, including $\{0, 0\}$ to indicate uninfected with either disease).

We track all other persons using an array S_t of size $A \times R \times D \times G \times H$ (our numerical analyses uses $7 \times 3 \times 9 \times 1 \times 2$), where, A is the number of age-groups, R is the number of risk-groups, D is the number of degree-bins (degree is the number of contacts per person, degrees are grouped into bins analogous to age grouped into age-groups), G is the number of geographic jurisdictions (in our numerical analyses we assumed $G = 1$, corresponding to a national jurisdiction), and H is the number of health states (note: here the size would be equal to the number of Disease 2 health states because, by design, persons infected with Disease 1 will be in the network and thus all persons in the compartmental model will have a value of 0 for Disease 1). Each element of the array ($S_t[\bar{a}, r, \bar{d}, g, h]$) is the number of people in that specific category. We use a dash for age-group and degree-bin notations to indicate that they are grouped intervals in the

compartmental model, unlike in the network where each node has a discrete value. A summary list of notations is presented in Appendix Table S1.

Thus, $Q_t + \sum_{i \in [\bar{a}, \bar{r}, \bar{d}, \bar{g}, \bar{h}]} S_{t,i}$, denotes the total number of people in the population at time t . Because all Disease 1 infected persons and exposed partners are in the network, Disease 1 transmissions and state progression are modeled at the individual level (see network transmission and network disease progression modules below and in Appendix Section S2.2). Disease 2 transmissions and progression are modeled using differential equations (as typically done in compartmental modeling) but with the consideration that people in both the network $G_t(\mathcal{N}, \mathcal{E})$ and compartmental array S_t can be infected with Disease 2 (see compartmental module below and in Appendix Section S2.1).

The mathematical challenge to address for the above MAC framework to work is to maintain the dynamics between the network $G_t(\mathcal{N}, \mathcal{E})$ and compartmental array S_t , including transitioning people from S_t to $G_t(\mathcal{N}, \mathcal{E})$ when a node becomes newly infected. Specifically, the simulation algorithm must determine ‘who’ are to be added as the immediate contacts of the node newly infected with Disease 1. Here we determine ‘who’ according to their degree, transmission group, age, and geographical location, as these characteristics of infected persons and their contacts are known to be correlated (36–41). Upon determining and adding all lifetime partners of nodes newly infected with Disease 1, we need to determine when the partnerships would initiate and dissolve, including the age of both partners at that time, such that the overall dynamics of age-mixing and transmission-group-mixing of the resulting network match that of the population being simulated. As network generation techniques in commonly used agent-based models are designed to generate the network between the full population (susceptible and infected persons), they cannot be adopted here. We developed an evolving contact network algorithm (ECNA(32,33)) for modeling the transitions from S_t to $G_t(\mathcal{N}, \mathcal{E})$ (see ECNA module below and in Appendix Section S2.3). For diseases that are chronic (such as HIV) once persons enter the network $G_t(\mathcal{N}, \mathcal{E})$ they do not transition back to S_t . For

diseases with recovery, recovered persons with no infected contacts can be added back to the compartments corresponding to their categorical group (see (32) for stability of network dynamics under different epidemic profiles).

2.1.2 The four main modules of MAC

The overall epidemiological, demographical, and network dynamics are maintained through simulation of four main modules that are run at every time-step (monthly) of the simulation: compartmental module, network transmission module, ECNA network generation module, and network disease progression module.

The formulation of the compartmental module for a hypothetical Disease 2 is presented in Appendix S2.1, we provide a brief overview here. The compartmental module updates the demographic features (births, aging, and deaths) and transmission and progression features of higher-prevalence diseases (Disease 2 here) among persons tracked through the array S_t and in the network $G_t(\mathcal{N}, \mathcal{E})$, as follows. As in typical compartmental modeling, it uses difference equations to transition persons between the compartments. It also determines transitions of the network nodes from one state to another. It does so by converting transition rates to transition probabilities by assuming that sojourn times follow an exponential distribution, as typically done in Markov chains. Then, for every state transition, the module uses the corresponding probability as a parameter of a binomial distribution to determine the number of persons to transition, and it randomly samples that many people from the network who are in the reference state and moves them to the next state. With this approach, the same level of granularity is applied to persons in the network and compartmental models, e.g., in the above representation, the granularity in the compartmental model is $[\bar{a}, r, \bar{d}, g, h]$ and thus the rates would be applied specific to age-group, transmission-group, degree-bin, geographic location, and health state. A key feature of this setup is approximating network features into the compartmental modeling structure. This is done by splitting the compartments into degree-bins ($\bar{d}$, a

dimension in array S_t) and using a degree-mixing matrix to simulate partnership mixing between people in different degree-bins. As noted earlier, degree between partners are correlated (36), and thus this feature helps better capture the network dynamics even in the compartmental model. Whereas the distributions for partnership mixing by age, transmission group, and degree are applied at the individual-level in the network, they are applied at the aggregated-level in the compartmental model.

The network transmission module determines if nodes exposed to a lower-prevalence disease (Disease 1 here) become infected using an individual-level Bernoulli transmission equation. Transmissions are determined at the individual-level using the network structure and individual-level sexual behaviors and transmission risk factors. Thus, the granularity of transmissions can be controlled by modifying the individual-level attributes that are tracked among agents. As an example, Appendix S2.2 presents a Bernoulli transmission equation using the level of granularity applied in the PATH 4.0 model. Note that, by definition, as persons in the compartmental model are not partners of any person infected with the lower-prevalence diseases (Disease 1 here), their chance of infection is zero. Further note that persons can move from the compartmental model to the network (see Figure 1) if their partners become infected with Disease 1, modeled using the ECNA module (below), which would then expose them to Disease 1.

The ECNA module controls the overall network dynamics of partnerships between nodes. Specifically, for every node newly infected with Disease 1 in the network, it determines the number of new partnerships to generate and the features of each of those new partners, including their degree (number of lifetime partners), their transmission-group, and their current age-group. The module then randomly selects susceptible persons who meet these criteria and moves them from the compartmental model to the network. The ECNA module also determines partnership details, such as the age of both partners and simulation times at partnership initiation and termination. Three main algorithms were used in the development of this module, which were presented elsewhere (32,33), and are briefly summarized below and discussed in Appendix S2.3. The first algorithm determines the degree of the new partner using a neural network prediction model,

270 a machine learning method, developed based on the assumption that sexual partnerships are scale-free
 271 networks with power-law distribution. For scale-free networks, the degree of node-neighbors are correlated,
 272 i.e., in the context here, the degree of the new partner is conditional on the degree of the newly infected
 273 node. While the literature presents analytical methods for estimation of the conditional distribution, they
 274 are developed for static networks, and thus, the degree of any node is conditional on the degree of ‘all’ its
 275 neighbors. In the evolving network here, the degree of the new partner is conditional on only the degree of
 276 the newly infected node, i.e., dependent on the network path taken to reach this person, and thus, the degree
 277 correlation is also influenced by the stochastic process of disease transmissions. Our previous work showed
 278 that, given a specific network, the degree correlations are not influenced by variations in the probability of
 279 transmission but influenced by the prevalence of disease, and thus, trained the neural network by generating
 280 the data through multiple simulations of hypothetical diseases, characterized by different values of
 281 probability of transmission (32). Upon determining the degree of the new partner, the second algorithm is
 282 applied to determine the age at which each of those partnerships are active, including the age of the other
 283 person in each partnership (one of those partnerships is with the newly infected node). Direct data for this
 284 would be a longitudinal survey over the duration of life of an individual, where the individual reports the
 285 number of partnerships they initiated at every age points of their life, and the age of their partner. Such
 286 surveys, however, are unavailable. Typical survey data only collect the number of partners up to the current
 287 age of the surveyed individual, and age of partners at a cross-sectional time-point. By assuming the number
 288 of partners up to the current age are steady state distributions of a time-invariant Markov chain with a
 289 multivariate state space of age-group - degree-bin combinations (time-invariant from assumption of no
 290 generational changes in partnership behavior), the transition probabilities were solved using simulation-
 291 based optimization to determine the number of new partnerships initiated in each age-group, for each
 292 degree-bin. The third algorithm determines the age-groups of both partners at the time of partnership
 293 activation by formulating the problem as a variant of an unbalanced assignment problem, a category of
 294 optimization problems with a classic example being assigning n jobs (here partners) to m machines (here

age-groups), applying age-mixing matrix and number of partners in each age-group as constraints. Further details of the three algorithms and data assumptions are presented in Appendix S2.3 and S3, respectively. Finally, the network disease progression module updates the individual-level demographic and disease dynamics for every person infected with Disease 1 (and other low-prevalence diseases as the case may be) in the network. The level of granularity for disease progression could be dependent on the analyses and diseases of interest.

Appendix S4 provides further details on model initialization and S5 gives a more detailed description of the steps of the simulation.

2.2 Numerical analyses

Using the MAC simulation framework, we conducted numerical analyses, using *relative prevalence (RP)* metrics, to quantify the risk of coinfection attributable to behavioral factors (individual sexual behaviors, partnerships networks, and care behaviors) alone. Further, we evaluated the sensitivity of relative prevalence to variations in care access, disease burdens (incidence and prevalence), and population heterogeneity by estimating *RP* metrics specific to transmission group and under varying hypothetical assumptions of care and epidemiological assumptions. For these analyses, we adopted the validated HIV model from PATH 4.0 as Disease 1 and constructed a hypothetical Disease 2 using a compartmental model. To keep the focus on behavioral factors, we assume no biological risk of coinfection, i.e., Disease 1 does not biologically increase the risk of Disease 2, and vice-versa. We first give a brief overview of Disease 1 (HIV) and Disease 2 modeling, and then discuss the numerical analysis in detail.

2.2.1 Overview of HIV (Disease 1) model

For the development of MAC, we directly adopted the HIV model from PATH 4.0, which has been validated to match well against data from the CDC's National HIV Surveillance System (NHSS) for both population epidemic features and HIV-network features. Details of PATH 4.0 and its validation are presented elsewhere (33); we give a brief description below. PATH 4.0 simulates sexual transmission of HIV in the United States in three transmission categories: heterosexual females (HETF), heterosexual males

(HETM), and men who have sex with men (MSM). (Note that MSM includes men who have sex with men only and men who have sex with men and women). To validate the model, we first generated an initial population that is representative of PWH in the United States in 2006 using data from several studies. These include demographical, sexual behavioral, clinical, and HIV care and treatment behavioral studies that originated from multiple large national surveillance and survey systems in the United States, along with other small studies. The surveillance and survey systems include the National HIV Surveillance System (NHSS), the Medical Monitoring Project (MMP), the HIV Outpatient Study (HOPS), the National HIV Behavioral Surveillance (NHBS), the National Survey for Family Growth (NSFG), and the National Survey for Sexual Health and Behavior (NSSHB)(42–47). The model was validated for the period 2006 to 2017 by calibrating to 2006 data, simulating the epidemic from 2006 to 2017 in monthly-time steps, and comparing simulated estimates for multiple epidemic features and HIV-network features against data from NHSS. Details of this validation study are presented elsewhere (33).

2.2.2 Overview of Disease 2 model

Individual sexual behaviors and partnership networks do not change specific to disease, and thus, sexual behavioral data for Disease 2 are the same as in HIV, though in aggregated form for the compartmental model structure. Specifically, individual sexual behaviors such as the number of lifetime partners, distribution of these partnerships over the lifetime of the person (as they transition across age-groups), number of sex acts, and condom use, most parameters specific to transmission group (HETF, HETM, and MSM) and age-group, and sexual network structures such as degree mixing matrix, age-group mixing matrix, and transmission-group mixing matrix to model mixing between partnerships, were adopted from PATH 4.0 (42–47). Note that, behaviors that change specific to disease can be added over general population behaviors, e.g., change in condom-use behavior upon awareness of HIV status was added to HIV-agents. While behaviors were simulated at the individual-level for HIV, they were simulated at the aggregated-level for Disease 2 in the compartmental model, using corresponding principles of each type of simulation method. That is, in the network, each person was assigned characteristics such as age,

transmission -group, and degree, matrices defining mixing between these groups were used for generating their network, and behavior modeled as a function of these characteristics to determine the probability that a susceptible person becomes infected. Whereas, in the compartmental model, compartments define age-group, transmission-group, and degree-bin combinations, and behaviors specific to these compartments and matrices defining mixing between these compartments were included into infection rate estimations (see Appendix S2.1 for methodological details). Infection rates were then used for determining the number of people to transition from susceptible to infected. Demographic features, such as the population sizes of transmission groups (HETF, HETM, MSM), population distributions by age-group, birth rate and natural mortality rates (48–50), which are not disease-specific, are a feature of the population modeled, and thus the overall model (network +compartmental) would match population features. As noted under Disease 1 overview, these data were comprehensively informed through numerous data sources, estimation methods, and validation processes and are presented elsewhere (32,33).

The only change for Disease 2 would be its epidemiology. While HIV is a Susceptible-Infected epidemiology structure, i.e., persons live with the infection for the remaining duration of life, we assumed Disease 2 to be a Susceptible-Infected-Susceptible epidemiology structure, i.e., persons can recover from Disease 2 and become susceptible for reinfection, which is representative of higher-prevalence STIs. Transition from Susceptible to Infected stage for Disease 2 would use the same sexual behaviors as in HIV, as noted above, except for the per contact transmission rate which is an epidemiologic parameter specific to a disease. The duration of the Infected stage (transition from Infected to Death) for HIV is an outcome of simulating HIV-disease stage progressions using care data specific to the United States (such as rates of diagnosis, linkage to care, and treatment), in addition to natural disease progression rates. For Disease 2, we used an overall rate of recovery that transitions persons from the Infected to the Susceptible stage, but we did not model sequelae. Thus, the inverse of the recovery rate represents the average duration of Infected stage, reflective of the natural disease progression and care access.

Therefore, there are only two parameters specific to Disease 2, transmission rate and recovery rate, which can be varied to generate diseases of differing incidence and prevalence. As described in the next section, we evaluated 16 scenarios using different combinations of these rates, to generate incidence and prevalence replicative of the range corresponding to STIs observed in the United States.

2.2.3 Metrics and scenarios

For the numerical analysis, we calculated the following metrics:

- The *relative prevalence of D2 given D1* ($RP_{D2|D1}$), estimated as the prevalence of Disease 2 among persons with HIV compared to persons without HIV, i.e.,

$$RP_{D2|D1} = \frac{\# \text{ of persons with Disease 2 and HIV} / \# \text{ of persons with HIV}}{\# \text{ of persons with Disease 2 only (no HIV)} / \# \text{ of persons with no HIV}}, \text{ and therefore, if}$$

$RP_{D2|D1} > 1$, then persons with HIV have a higher burden of Disease 2 than persons without HIV.

- The *relative prevalence of D1 given D2* ($RP_{D1|D2}$), estimated as the prevalence of HIV among persons with Disease 2 compared to persons without Disease 2, i.e.,

$$RP_{D1|D2} = \frac{\# \text{ of persons with Disease 2 and HIV} / \# \text{ of persons with Disease 2}}{\# \text{ of persons with HIV only (no Disease 2)} / \# \text{ of persons with no Disease 2}}, \text{ and therefore, if}$$

$RP_{D1|D2} > 1$, then persons with Disease 2 have a higher burden of HIV than persons without Disease 2.

- The average degree (average number of partners) among all persons with HIV (d_{HIV+}), average degree among all persons with Disease 2 (d_{D2+}), and average degree in the overall population ($d_{Overall}$), which are metrics representing network features.

When estimating $RP_{D2|D1}$ and $RP_{D1|D2}$, we assume no biological risk of coinfection, i.e., Disease 1 does not biologically increase the risk of Disease 2 acquisition, transmission, or progression (and vice-versa).

Thus, if $RP_{D2|D1} > 1$ or $RP_{D1|D2} > 1$, it would be attributed to behavioral factors alone, which would be

the combined effects of individual-behaviors, partnership network structures, and care behaviors. Note that the assumption of no biological risk is made for purposes of evaluating risk attributed to behavioral factors alone, during application of model in future work, biological risk can be easily modeled by using a multiplier for transmission or disease progression rates.

We report average results for the last year of the simulation (year 2017). Note, over the period 2006 to 2017 (simulation timeline) due to variations in care access (proportion of PWH on treatment with viral suppression increased from ~20% in 2006 to ~56% in 2017(43)) incidence from sexual transmissions decreased from ~44,000 in 2006 to ~33,100 in 2017 (51,52). However, hypothetical D2 is in equilibrium due to the nature of its SIS epidemiology structure.

Prevalence of HIV and care access in the United States has been continuously increasing over the past few decades as noted above (35,53,54). Further, there is considerable heterogeneity in disease burden, care access, and sexual behavior across transmission risk-groups that mix with each other, thus generating cross-over effects. Therefore, in addition to estimating the risk of co-infections attributable to behavioral factors, we evaluate the sensitivity of relative prevalence to care metrics, disease burden, and population heterogeneity under mixing. The effects of each subcomponent on the relative prevalence metrics are not independent, nor is their effect static over time. Thus, fully evaluating the sensitivity of each will not be feasible. Here, we attempted to gain general insight by utilizing naturally observed variations of these components in the United States. Specifically, we estimated relative prevalence metrics specific to transmission group (sexual behaviors and network structures are inherently different across HETF, HETM, and MSM), under varying epidemiological assumptions for HIV (utilizing the inherent increases in HIV disease burden (35,53,54) and care access over the past two decades(55–58)), under hypothetical assumptions for Disease 2 epidemiology (generating disease burdens replicative of the range observed in STIs in the United States), and under varying assumptions of epidemiology across transmission-groups (utilizing the inherent heterogeneity across risk-groups and mixing of MSM with other MSM and women).

In total, we simulated four scenarios (Scenarios 1 to 4) related to HIV disease burden, and under each, sixteen scenarios related to hypothetical Disease 2 burden, simulating both diseases among three transmission-groups, HETF, HETM, and MSM who have varying individual sexual behaviors and network structures. However, across scenarios, transmission-group specific individual sexual behaviors and network structures, and mixing between transmission-groups remain unchanged, and were comprehensively informed through national surveillance and survey systems in the United States, as noted earlier. We simulated each scenario for a 12 year period. Scenario 1 is a status-quo representation of HIV in the United States for the period 2006 to 2017, and Scenarios 2 to 4 are hypothetical, as follows.

HIV scenarios

- *Scenario 1 (status-quo HIV)* was a status-quo representation of HIV in the United States between 2006 and 2017, calibrated to HIV prevalence and HIV care metrics (such as proportions aware, linked to care, and on treatment) specific to transmission group (HETF, HETM, MSM). The model was initiated in 2006, with an HIV prevalence of about 0.1%, 0.05%, and 6% for HETF, HETM, and MSM, respectively, and the proportion of PWH on treatment with viral suppression of about 20%, 18%, and 19%, for HETF, HETM, and MSM, respectively. By 2017, HIV prevalence in the U.S. increased to 0.12%, 0.06%, and 8% for HETF, HETM, and MSM, respectively, and the proportion of PWH on treatment with viral suppression increased to 56%, 50%, and 57%, for HETF, HETM, and MSM, respectively.
- *Scenario 2 (low HIV prevalence low care)* initiates the model with a lower HIV prevalence (taking data from 1990s) of 0.07%, 0.04%, and 3.2% for HETF, HETM, and MSM, respectively, and for the full duration of the simulation (12 years) maintains care metrics to keep the proportion on treatment with viral suppression at a constant value of 20%, 18%, and 19%, for HETF, HETM, and MSM, respectively (corresponding to 2006 care data).
- *Scenario 3 (low HIV prevalence high care)* is similar to Scenario 2 in initiating the model with the low HIV prevalence, but assumes higher care by using care data from 2017, thus resulting in

lower HIV incidence compared to Scenario 2. Specifically, the model initiates with a HIV prevalence of 0.07%, 0.04%, and 3.2% for HETF, HETM, and MSM, respectively, and for the full duration of the simulation, maintains care metrics to keep the proportion on treatment with viral suppression at a constant value of 56%, 50%, and 57%, for HETF, HETM, and MSM, respectively.

- *Scenario 4 (equal and low HIV prevalence for all transmission groups)* is similar to Scenario 2 in the use of the 2006 care data, but the model is initiated with equal and low HIV prevalence(1990s HETF prevalence) for all three transmission-groups. Specifically, the model is initiated with a prevalence of 0.07% for HETF, HETM, and MSM, and for the full duration of the simulation, maintains care metrics to keep the proportion on treatment with viral suppression at a constant value of 20%, 18%, and 19%, for HETF, HETM, and MSM, respectively. As the actual prevalence of HIV among MSM has been significantly higher than heterosexuals (HETF and HETM), and as MSM mix with men and women, comparing Scenario 4 with Scenario 2 will provide insight into the sensitivity of the relative prevalence metrics to heterogeneity in populations that mix.

Disease 1 scenarios are summarized in Table 1. In addition to a dry run, that initializes network dynamics and individual-level event history and populate the initial HIV prevalence and care metrics (Appendix S4), we ran each scenario for a period of 12 years and report results for the last year of the simulation.

Disease 2 scenarios

Under each of Scenarios 1 to 4, we simulated 16 scenarios for Disease 2 using combinations of transmission rates (0.04, 0.06, 0.1, 0.2 per contact), and recovery rates (0.042, 0.083, 0.16, and 0.0083 per month, corresponding to an average infection duration of 1, 2, 5 and 10 years, respectively). These values generate a wide range of estimated incidence and prevalence for Disease 2 and were chosen to mimic the range of STIs observed in the United States population (Figure 2).

3. RESULTS

As expected by design of scenarios, the range of transmission rates and recovery rates used for Disease 2 created epidemics of varying Disease 2 burden, but for a specific combination of transmission rate and recovery rate, Disease 2 burden was similar across all Scenarios 1 to 4 (Figure 3a). Also as expected by design, Scenarios 1 to 4 created varying Disease 1 (HIV) burdens (Figure 3b), the values for the last year of the simulation are as follows. Scenario 1 (*status-quo HIV*) had high HIV prevalence for MSM (~8%) compared to heterosexual female (HETF) (~0.12%) and heterosexual male (HETM) (~0.06%), representative of HIV in the United States (Figure 3b). Compared to Scenario 1, Scenario 2 (*low HIV prevalence, low care*) created a lower HIV prevalence (0.09% for HETF, 0.04% for HETM, and 5% for MSM) but similar incidence because of the low care assumption (Figure 3b). Scenario 3 (*low HIV prevalence, high care*) created HIV prevalence similar to Scenario 2 but lower incidence because of the higher care assumption. Compared to Scenario 1, Scenario 4 (*low and equal HIV prevalence for all transmission groups*) reduced HIV prevalence by an order of magnitude for MSM and, because of mixing between MSM with HETF, reduced HIV prevalence in HETF; HETM had the same prevalence as HETF (0.07% for HETF, 0.06% for HETM, and 0.1% for MSM) (Figure 3b). As expected from inherent differences in behaviors, and reflecting data inputs to the simulation, the overall average degree ($d_{overall}$) was higher among MSM than HETF and HETM.

$RP_{D2|D1}$ was sensitive to both HIV and Disease 2 burden (incidence and prevalence) and HIV care (Figure 4a, 4b), and the patterns could be consistently explained through comparison of resulting average degrees of persons with at least HIV, at least Disease 2, and overall (d_{HIV+} , d_{D2+} , $d_{overall}$, respectively) (Figure 4c, 4d). We discuss these below. Keeping HIV prevalence fixed, i.e., observing within each Scenario and transmission-group, as Disease 2 burden increased, $RP_{D2|D1}$ increased. In Scenario 1, $RP_{D2|D1} > 1.2$ for HETF in most cases, $RP_{D2|D1} < 0.8$ for HETM in most cases, and $RP_{D2|D1} < 1.2$ for MSM in most cases (Figure 4). To recollect, HIV prevalence was moderate for HETF, lowest for HETM, and high for MSM. Thus, there was no consistent pattern when comparing $RP_{D2|D1}$ and HIV disease burden (prevalence)

491 alone. However, the pattern in $RP_{D2|D1}$ could be explained by the network structure as measured by
 492 average degree. In Scenario 1, for HETF, $d_{HIV+} > d_{overall}$, and $RP_{D2|D1} > 1.2$ if $d_{D2+} > d_{HIV+}$, and
 493 $RP_{D2|D1} \rightarrow 1$ as $d_{D2+} \rightarrow d_{overall}$ (Figure 4c). For HETM, $d_{HIV+} < d_{overall}$ and $RP_{D2|D1} < 1$ in most
 494 cases. For MSM, $d_{HIV+} \gtrsim d_{overall}$ and $RP_{D2|D1} < 1.2$. This suggests that, when the average degree of
 495 both diseases is greater than the average degree in the overall population (as was the case for HETF in
 496 certain scenarios), $RP_{D2|D1}$ would be greater than 1, and if average degree of HIV is close to or less than
 497 the overall average degree (as was the case for HETM and MSM), then $RP_{D2|D1}$ would be closer to 1 or
 498 below 1.

499 The interpretation of the above results is that, when both diseases have a sufficiently low prevalence, they
 500 are both concentrated in higher-risk networks ($d_{D2+} > d_{overall}$; $d_{HIV+} > d_{overall}$) and thus, persons with
 501 HIV would have higher risk of Disease 2 co-infection. These conclusions are intuitive, and characteristic
 502 of scale-free networks, where the disease first spreads to high-risk networks before spreading to the rest
 503 of the network.

504 This correlation between $RP_{D2|D1}$ and patterns of average degree were consistent across the four scenarios,
 505 as discussed below, supporting the above conclusion. HETF and HETM in Scenario 2 (*low HIV*
 506 *prevalence low care*) had small decreases in HIV prevalence and saw minimal changes in $RP_{D2|D1}$ and
 507 average degrees (Figure 4a). For MSM in Scenario 2, the lower HIV prevalence led to a concentration of
 508 disease burden in higher risk networks ($d_{HIV} > d_{overall}$) (Figure 4c), and thus, the values of
 509 $RP_{D2|D1}$ tended to be higher (more reds) than in Scenario 1 (Figure 4a). Compared to Scenario 2, though
 510 Scenario 3 (*low HIV prevalence high care*) created only minutely lower HIV prevalence, it had
 511 significantly lower HIV incidence because the higher HIV-care led to fewer transmissions (Figure 3b),
 512 leading in-turn to lower $RP_{D2|D1}$ for HETF and MSM. These changes in $RP_{D2|D1}$ could again be explained
 513 through corresponding changes in average degree. As average degree in HIV decreased and got closer to
 514 or equal to overall average degree ($d_{HIV} \sim d_{overall}$), $RP_{D2|D1}$ decreased to 1 or below. This also suggests

that the reduction in transmissions from the higher HIV-care in Scenario 3 helped dissipate the impact of high HIV-prevalence burden in the high-risk networks.

In Scenario 4, HETM saw no change compared to Scenarios 1 to 3, it had a very low HIV prevalence and had $d_{HIV} < d_{overall}$. On the other hand, the $RP_{D2|D1}$ in HETF and MSM in Scenario 4 were opposite to that in Scenario 1, although both HETF and MSM saw a reduction in HIV prevalence in Scenario 4. For MSM, the average degree among HIV tended to be higher than overall ($d_{HIV} > d_{overall}$), suggesting that, because of the significantly lower HIV epidemic than in Scenario 1, it was now mostly concentrated in high-risk networks, and thus, $RP_{D2|D1} > 1$ when Disease 2 was also low and concentrated in high-risk network ($d_{D2} > d_{overall}$) (Figure 4c). For HETF, though Scenario 4 initiated with the same HIV prevalence as Scenario 1, because of the mixing with MSM, it had a large reduction in HIV incidence and prevalence (Figure 3b). The corresponding average degree among HIV was now closer to overall ($d_{HIV} \sim d_{overall}$), and thus the corresponding $RP_{D2|D1} < 1$ (Figure 4c). These consistent patterns between relative prevalence and average degree, under varying values of HIV and Disease 2 burdens and care access, demonstrate the role of network features in burden of coinfection.

The behavior of $RP_{D1|D2}$ was similar to that of $RP_{D2|D1}$, i.e., if $RP_{D2|D1} > 1$ then $RP_{D1|D2} > 1$ (Figure 5). These results again support that when both disease burdens are lower, the infection is more concentrated in higher risk networks.

4. DISCUSSION

The purpose of this manuscript is to present a novel mixed agent-based compartmental (MAC) simulation framework for joint modeling of diseases with common modes of transmission but varying epidemiological features. The numerical analyses assessed the contribution of behavioral factors to joint-disease outcomes, and its sensitivity to disease burden, care metrics, and population heterogeneity and mixing. This work serves as a proof-of-concept for the feasibility of the proposed method and the

538 sensitivity analyses highlights the potential significance of joint-disease modeling in a heterogeneous
 539 interacting population.

540 The key aspect of the MAC framework is its computational tractability while maintaining sufficient
 541 sample size even for lower prevalence diseases. The computational complexity of network modeling and
 542 compartmental modeling are in the $\mathcal{O}(N^2)$ and $\mathcal{O}(S)$, respectively, where N is the number of people
 543 simulated and S is the number of states. Thus, while compartmental modeling is computationally
 544 efficient, it lacks the granular individual-level features of network modeling, and while networks are
 545 favored for this feature, they are computationally complex and lose tractability as population size
 546 increases. The MAC simulation technique provides an efficient balance, as can be seen by sample
 547 computation times in Table 2. The computational complexity of MAC is in the $\mathcal{O}(Q_t + S + \theta I_t \bar{d}_n^2)$,
 548 where Q_t is the number of nodes in the network at time t , with upper bound equal to the number of
 549 infected nodes in the network (I_t) times the average number of lifetime partnerships per person ($\bar{d}_n$), and
 550 θ is an overall rate of infection, and thus θI_t is the number of newly infected nodes at time t . Note that an
 551 epidemiologically correct expression for new infections is $\frac{\theta I_t S_t}{N}$, where S_t is the number of susceptible
 552 persons, however, as only low prevalence diseases are modeled in the network, $S_t \cong N$. Also note that,
 553 because they are slow spreading diseases, θI_t is very small relative to I_t . Thus, unlike agent-based
 554 modeling where the computational complexity increases with N , here the computational complexity
 555 increases with I_t . This is a useful feature because the selection criteria for population size can be based on
 556 the desired sample size of positive cases without having to worry about its computational feasibility. For
 557 example, we can select the number of infected nodes in the network to be sufficiently large, say 12,000
 558 (I_t). Then, simulating a disease with prevalence of 0.4% will generate a simulated population size of
 559 about 3 million persons (N), whereas, simulating a disease with prevalence of 0.2% will generate a
 560 simulated population size of about 6 million persons (N) but it will not increase the computational
 561 complexity. To achieve the same computational complexity when modeling two low prevalence diseases
 562 in the network, we can specify the total number of agents having one or both of the diseases. The

resulting sample size would then depend on the prevalence of each and the overlap between the two diseases. For example, suppose the prevalence is 0.4% and 0.2% for two low prevalence diseases D1 and D2, respectively, and there is no overlap, i.e., each agent has one and only one of the two diseases. Generating 12,000 infected nodes will generate ~8,000 persons with D1 and ~4,000 persons with D2, and a total population of ~2 million. If these samples are not sufficient, then more nodes can be generated, with a concomitant non-linear increase in computation time that is closer to linear than quadratic. Note that in our numerical analyses we did not remove infected agents who were dead or remove partnerships that happened in the past. Doing so will not impact epidemic projections but can reduce computational time. The simulation technique here could also be useful to further improve efficiency of large-scale simulators that focus on software efficiencies and are capable of simulating millions as nodes in the network with the use of high performance computing (59,60).

In the numerical analyses, the values of relative prevalence were greater than 1 in several scenarios, highlighting the influence of behavioral factors on joint disease outcomes. These values are lower than those reported from observational studies, e.g., a 4 to 8 fold increase in burden of cervical cancer caused by HPV infection was observed among women with HIV compared with women without HIV (61), and a 3 times higher HIV incidence risk was observed among MSM with rectal gonorrhea or chlamydia infection compared with MSM without these STIs (62). The differences are expected because the data from the literature are estimates from cohort studies or case control studies and thus include risk of coinfection attributable to both biological factors and behavioral factors (61), whereas in the model we forced the biological risk to be zero so that relative prevalence is attributable to behavioral factors alone.

Consequently, through joint modeling of diseases, the differences between model estimates and observational studies can be used for determining risk attributed to biological factors alone and behavioral factors alone. These estimations can help inform both the type of interventions and optimal allocation of resources. While disease risk originating from behavioral factors would require behavioral and structural interventions because social determinants are key drivers of high-risk behaviors, disease risk originating

from biological predisposition of pre-existing infections would additionally require disease management and care support programs. Using a model along with observational data to control for behavioral factors would help more realistically estimate the biological risk of acquisition and transmission of infections, and thus inform care management interventions. Modeling work in this area is limited (10,63) and focused on modeling subgroups in isolation, such as MSM only. Our numerical analyses highlights the sensitivity of results in a subgroup to variations in disease burden in persons outside the subgroup due to population mixing, as in the comparison between Scenarios 4 and 2 for HETF.

While estimates of biological risk of co-infection can directly inform care management programs, and the estimates of risk attributable to behavioral factors are necessary for determining those biological risk estimates, the value of risk attributable to behavioral factors is not a sufficient measure for informing the need for behavioral and structural interventions. Indeed, values of relative prevalence closer to 1 or below could be generated because of higher HIV prevalence (MSM compared to HETF in Scenario 1), low HIV prevalence (HETM in all scenarios), mitigation of HIV risk through increased care access (Scenario 3 compared to Scenario 2 for HETF and MSM), or risk mitigation in the higher risk population when two groups mix (HETF in Scenario 4 compared to Scenario 2). Although the relative prevalence values were closer to 1 or below in all these scenarios, structural interventions would be key interventions mainly in scenarios with high-prevalence and low care. Further, relative prevalence of greater than 1 can originate from factors within the population (comparing across Scenarios 1 to 4 for MSM) or from mixing with populations with high-risk of infection (HETF between Scenarios 2 and 4), each requiring different intervention strategies. Thus, although the values of relative prevalence would be necessary for inferring biological risk of infection, they alone are insufficient for informing interventions.

The above results from the numerical analyses justify the need for joint disease modeling for more accurate representation of the behavioral and biological dynamics of disease interactions. Further, social determinants, such as poverty, unemployment, homelessness, and stigma and discrimination, are known correlates of sexual and care behaviors that increase disease risk, such as higher number of partners,

higher condomless sex, and lower treatment uptake among persons experiencing homelessness compared to persons with stable housing (17–20). Therefore, structural interventions, such as healthcare coverage, subsidized housing and food programs, and access to mental healthcare, are key part of behavioral interventions to reduce risk of STI acquisition (12,13,15,16,64). While the costs of these structural interventions can be extrapolated from small cohort studies, the impact of structural interventions on disease burden is infeasible to estimate through controlled trials, because of the intricate disease interactions attributed to behavioral factors and its sensitivity to population and epidemic features (as observed through numerical analyses conducted here). A model would be a fundamental tool for estimation of the impact and thus cost-effectiveness, which are key measures used for allocation of public health resources. The new MAC simulation framework is computationally ideal for the above application, as it can be expanded to simulate sexual and care behavioral factors as a function of social conditions, and thus, subsequently serve as a decision-analytic tool for evaluation of structural interventions.

Joint modeling of STIs in the literature is limited, and most focus on sub population groups (10,63). While sub-population modeling could help compare impact of alternate interventions on the sub-population, a national model that additionally simulates mixing between sub-populations and thus cross-over effects of interventions, would help identify optimal combination of interventions for overall reduction of diseases and disparities. Such a decision-analytic model would be suitable for informing public health guidelines, such as through the HealthyPeople2030 plan (65). Such a joint-disease joint-population model would also be suitable during emergence of new infection outbreaks, such as the recent 2022 Mpox outbreak. It could help shed light on populations at greatest risk and those that would benefit from medical countermeasures, to help inform interventions for containment of transmissions.

Our work is subject to limitations. The above analyses were conducted using hypothetical epidemiological and care assumptions, and thus, our work is limited to evaluating sensitivity of joint disease dynamics to key behavioral factors and epidemiological factors. Thus, the results should not be used to infer actual burden of coinfection. We assumed biological risk of coinfection to be zero to

evaluate the sensitivity of behavioral factors alone. However, computational changes needed for modeling the biological risk are minimal, through use of a factor multiplied to the rates of transmission or progression to represent the increased risk, and calibrating the factor specific to a disease by matching simulated cases of coinfection with surveillance data. The scope of this work was limited to presenting a new simulation framework for joint modeling of diseases, and we did not model all behavioral changes driven from epidemic awareness. While some of these behavioral changes such as partnership selection for serosorting behaviors could be added to the current model structure through some modifications, changes in network structure, such as generational changes in the number of partners that would change the overall network statistics from the ones used in model calibration, would require recalibration. This would be a general challenge for any disease model and is outside the scope of this study. However, changes in network structure for evaluating impact of interventions could be carried out with the current model, by evaluating potential changes to network structure specific to interventions.

5. CONCLUSIONS

The study contributes a new simulation technique that is uniquely suitable for jointly modeling infectious diseases in a heterogeneous population that have common modes of transmission but varying epidemiological features, that a single simulation technique would be insufficient or computationally challenging. The numerical analysis serves as proof-of-concept for its application to STIs. The numerical analysis also demonstrates the influence of behavioral factors on joint disease outcomes, and its sensitivity to disease burden, care access, and population heterogeneity and mixing, which justify the need for joint modeling of related infectious diseases. Social and economic conditions are among key drivers of behaviors that increase STI risk. The new simulation framework is especially suitable for simulating behavioral factors as a function of social determinants, and it can be expanded in future work to subsequently evaluate optimal combinations of common structural interventions and disease-specific interventions for overall reduction of STI burden. The new simulation technique would also be suitable for the joint modeling of other infectious diseases that have common modes of transmissions. This would

663 be especially suitable for early detection and intervention of new or emergent disease outbreaks, when
664 prevalence is still low, but spread within networks of people with other ongoing diseases. Examples in
665 recent years include Mpox for sexually transmitted infections, or COVID-19, SARS, MERS for
666 respiratory infections.

667 **LIST OF ABBREVIATIONS**

668 STIs: sexually transmitted infections
669 HIV: human immunodeficiency virus
670 HPV: human papillomavirus
671 HCV: hepatitis C
672 HBV: hepatitis B
673 NG: Gonorrhea
674 CT: Chlamydia
675 PWH: people with HIV
676 ART: antiretroviral therapy treatment
677 PrEP: pre-exposure prophylaxis
678 NHSS: U.S. National HIV Surveillance Systems
679 SDH: social determinants of health
680 MAC: mixed agent-based compartmental
681 ABENM: agent-based evolving network modeling
682 ABNM: agent-based network modeling
683 PATH 4.0: Progression and Transmission of HIV, version 4.0
684 ECNA: Evolving Contact Network Algorithm
685 RP: relative prevalence
686 HETF: heterosexual females
687 HETM: heterosexual males

MSM: men who have sex with men (men who have sex with men only and men who have sex with men and women)

DECLARATIONS

- **Competing interests:** CG is also a Guest Researcher at the Division of HIV Prevention, Centers for Disease Control and Prevention. The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. CG, HB, and PJH have no other conflicts of interest.
- **Funding:** Work was supported by the National Science Foundation under NSF 1915481 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI127236. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.
- **Authors' contributions:** CG was involved in conception of study, model development, analyses and interpretation of findings, and manuscript preparation. HB was involved in analyses and interpretation of findings, and manuscript preparation. PJH was involved in analyses and interpretation of findings, and manuscript preparation. All authors read and approved the final manuscript. HB and PJH equally contributed to the work and are listed in alphabetical order.
- **Acknowledgements:** We like to acknowledge Drs. Melanie Taylor, Alexander Viguerie, Paul Farnham, Harrell Chesson, Emily Pollock, Cynthia Lyles, John Beltrami, and Anne-Marie France from the Centers for Disease Control and Prevention for their inputs on the manuscript.
- **Authors' information (optional)**

References

1. Chesson HW, Spicknall IH, Bingham A, Brisson M, Eppink ST, Farnham PG, et al. The Estimated Direct Lifetime Medical Costs of Sexually Transmitted Infections Acquired in the United States in 2018. *Sex Transm Dis.* 2021 Apr;48(4):215–21.
2. Liu HF, Liu M, Xu YL. Analysis of high-risk HPV infection and cervical cytologic screening in HIV positive women. *Zhonghua Fu Chan Ke Za Zhi.* 2016;51(10):734–8.
3. Ferenczy A, Coutlee F, Franco E, Hankins C. Human papillomavirus and HIV coinfection and the risk of neoplasias of the lower genital tract: a review of recent developments. *CMAJ Can Med Assoc J J Assoc Medicale Can.* 2003;169(5):431–4.
4. Liu G, Sharma M, Tan N, Barnabas RV. HIV-positive women have higher risk of human papilloma virus infection, precancerous lesions, and cervical cancer. *AIDS Lond Engl.* 2018 Mar 27;32(6):795–808.
5. Tartaglia E, Falasca K, Vecchiet J, Sabusco GP, Picciano G, Di Marco R, et al. Prevalence of HPV infection among HIV-positive and HIV-negative women in Central/Eastern Italy: Strategies of prevention. *Oncol Lett.* 2017 Dec;14(6):7629–35.
6. Thio CL, Seaberg EC, Skolasky R Jr, Phair J, Visscher B, Munoz A, et al. HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). *Lancet Lond Engl.* 2002;360(9349):1921–6.
7. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex Transm Infect.* 1999 Feb 1;75(1):3–17.
8. Heiligenberg M, Rijnders B, Schim van der Loeff MF, de Vries HJC, van der Meijden WI, Geerlings SE, et al. High prevalence of sexually transmitted infections in HIV-infected men during routine outpatient visits in the Netherlands. *Sex Transm Dis.* 2012 Jan;39(1):8–15.
9. Lin KY, Sun HY, Lee TF, Chuang YC, Wu UI, Liu WC, et al. High prevalence of sexually transmitted coinfections among at-risk people living with HIV. *J Formos Med Assoc Taiwan Yi Zhi.* 2021 Oct;120(10):1876–83.
10. Jones J, Weiss K, Mermin J, Dietz P, Rosenberg ES, Gift TL, et al. Proportion of Incident Human Immunodeficiency Virus Cases Among Men Who Have Sex With Men Attributable to Gonorrhea and Chlamydia: A Modeling Analysis. *Sex Transm Dis.* 2019 Jun;46(6):357–63.
11. Orroth KK, White RG, Korenromp EL, Bakker R, Changalucha J, Habbema JDF, et al. Empirical Observations Underestimate the Proportion of Human Immunodeficiency Virus Infections Attributable to Sexually Transmitted Diseases in the Mwanza and Rakai Sexually Transmitted Disease Treatment Trials: Simulation Results. *Sex Transm Dis.* 2006;33(9):536–44.
12. Adimora AA, Auerbach JD. Structural interventions for HIV prevention in the United States. *J Acquir Immune Defic Syndr.* 2010;55(p S132–S135).

- 748 13. Sipe TA, Barham TL, Johnson WD, Joseph HA, Tungol-Ashmon ML, O'Leary A. Structural
749 Interventions in HIV Prevention: A Taxonomy and Descriptive Systematic Review. *AIDS Behav.*
750 2017;21(12):3366–430.
- 751 14. Blankenship KM, Bray SJ, Merson MH. Structural interventions in public health. *AIDS.* 2000;14(Suppl
752 1:S11-21).
- 753 15. Blankenship KM, Friedman SR, Dworkin S, Mantell JE. Structural interventions: concepts, challenges
754 and opportunities for research. *J Urban Health Bull N Y Acad Med.* 2006;83(1):59–72.
- 755 16. Frieden TR. A framework for public health action: the health impact pyramid. *Am J Public Health.*
756 2010;100(4):590–5.
- 757 17. Craddock J, Barman-Adhikari A, Combs KM, Fulginiti A, Rice E. Individual and Social Network
758 Correlates of Sexual Health Communication Among Youth Experiencing Homelessness. *AIDS Behav.*
759 2020 Jan;24(1):222–32.
- 760 18. Henwood BF, Rhoades H, Redline B, Dzubur E, Wenzel S. Risk behaviour and access to HIV/AIDS
761 prevention services among formerly homeless young adults living in housing programmes. *AIDS*
762 *Care.* 2020 Nov;32(11):1457–61.
- 763 19. Maria DS, Daundasekara SS, Hernandez DC, Zhang W, Narendorf SC. Sexual risk classes among youth
764 experiencing homelessness: Relation to childhood adversities, current mental symptoms, substance
765 use, and HIV testing. *PLOS ONE.* 2020 Jan 3;15(1):e0227331.
- 766 20. Edidin JP, Ganim Z, Hunter SJ, Karnik NS. The mental and physical health of homeless youth: a
767 literature review. *Child Psychiatry Hum Dev.* 2012 Jun;43(3):354–75.
- 768 21. cdc-hiv-surveillance-special-report-number-25.pdf [Internet]. [cited 2021 Nov 20]. Available from:
769 [https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-25.pdf)
770 [number-25.pdf](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-25.pdf)
- 771 22. Huang YL, Frazier EL, Sansom SL, Farnham PG, Shrestha RK, Hutchinson AB, et al. Nearly Half Of US
772 Adults Living With HIV Received Federal Disability Benefits In 2009. *Health Aff Proj Hope.*
773 2015;34(10):1657–65.
- 774 23. Walker N, Tam Y, Friberg IK. Overview of the Lives Saved Tool (LiST). *BMC Public Health.* 2013;13
775 Suppl 3(Journal Article):S1-S1. Epub 2013 Sep 17.
- 776 24. Paul JA, MacDonald L, Hariharan G. Modeling Risk Factors and Disease Conditions to Study
777 Associated Lifetime Medical Costs. *Serv Sci.* 2014;6(1):47.
- 778 25. Damery S, Flanagan S, Combes G. The effectiveness of interventions to achieve co-ordinated
779 multidisciplinary care and reduce hospital use for people with chronic diseases: study protocol for a
780 systematic review of reviews. *Syst Rev.* 2015;4(64).
- 781 26. Jeon YH, Black A, Govett J, Yen L, McRae I. Private health insurance and quality of life: perspectives
782 of older Australians with multiple chronic conditions. *Aust J Prim Health.* 2012;18(3):212–9.

783 27. McPhail SM. Multimorbidity in chronic disease: impact on health care resources and costs. Risk
784 Manag Healthc Policy. 2016;9(Journal Article):143–56.

785 28. Lauer JA, Rohrich K, Wirth H, Charette C, Gribble S, Murray CJ. PopMod: a longitudinal population
786 model with two interacting disease states. Cost Eff Resour Alloc CE. 2003;1(1):6.

787 29. Bershteyn A, Gerardin J, Bridenbecker D, Lorton CW, Bloedow J, Baker RS, et al. Implementation and
788 applications of EMOD, an individual-based multi-disease modeling platform. Pathog Dis. 2018 Jul
789 1;76(5):fty059.

790 30. World Health Organization, OneHealth Tool: Supporting integrated strategic health planning,
791 costing and health impact analysis, World Health Organization, 2013
792 <https://www.who.int/tools/onehealth> (last accessed August 2022).

793 31. Smieszek T, Fiebig L, Scholz RW. Models of epidemics: when contact repetition and clustering should
794 be included. Theor Biol Med Model. 2009;6(Journal Article):11–11.

795 32. Eden M, Castonguay R, Munkhbat B, Balasubramanian H, Gopalappa C. Agent-based evolving
796 network modeling: a new simulation method for modeling low prevalence infectious diseases.
797 Health Care Manag Sci. 2021 Sep;24(3):623–39.

798 33. Singh S, France AM, Chen YH, Farnham PG, Oster AM, Gopalappa C. Progression and transmission of
799 HIV (PATH 4.0)-A new agent-based evolving network simulation for modeling HIV transmission
800 clusters. Math Biosci Eng MBE. 2021 Mar 3;18(3):2150–81.

801 34. Kreisel KM, Spicknall IH, Gargano JW, Lewis FMT, Lewis RM, Markowitz LE, et al. Sexually
802 Transmitted Infections Among US Women and Men: Prevalence and Incidence Estimates, 2018. Sex
803 Transm Dis. 2021 Apr;48(4):208–14.

804 35. Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the United
805 States, 2015–2019. HIV Surveillance Supplemental Report 2021;26(No. 1). <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>. Published May 2021. Accessed [11/20/2021].

807 36. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Åberg Y. The web of human sexual contacts. Nature.
808 2001;411(6840):907–8.

809 37. Glick SN, Morris M, Foxman B, Aral SO, Manhart LE, Holmes KK, et al. A comparison of sexual
810 behavior patterns among men who have sex with men and heterosexual men and women. J Acquir
811 Immune Defic Syndr 1999. 2012 May 1;60(1):83–90.

812 38. Finlayson TJ, Le B, Smith A, Bowles K, Cribbin M, Miles I, et al. HIV risk, prevention, and testing
813 behaviors among men who have sex with men--National HIV Behavioral Surveillance System, 21 U.S.
814 cities, United States, 2008. MMWR Surveill Summ Morb Mortal Wkly Rep Surveill Summ CDC.
815 2011;60(14):1–34.

816 39. Voetsch AC, Lansky A, Drake AJ, MacKellar D, Bingham TA, Oster AM, et al. Comparison of
817 demographic and behavioral characteristics of men who have sex with men by enrollment venue
818 type in the National HIV Behavioral Surveillance System. Sex Transm Dis. 2012;39(3):229–35.

819 40. Friedman MR, Wei C, Klem ML, Silvestre AJ, Markovic N, Stall R. HIV infection and sexual risk among
820 men who have sex with men and women (MSMW): a systematic review and meta-analysis. *PloS*
821 *One*. 2014;9(1):e87139.

822 41. Board AR, Oster AM, Song R, Gant Z, Linley L, Watson M, et al. Geographic Distribution of HIV
823 Transmission Networks in the United States. *J Acquir Immune Defic Syndr* 1999. 2020 Nov
824 1;85(3):e32–40.

825 42. Centers for Disease C, Prevention. Vital signs: HIV testing and diagnosis among adults--United
826 States, 2001-2009. *MMWR Morb Mortal Wkly Rep*. 2010;59(47):1550–5.

827 43. Centers for Disease Control and Prevention. HIV Surveillance Reports,
828 <https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>, [Last Accessed August 2022].

829 44. Broz D, Wejnert C, Pham HT, DiNenno E, Heffelfinger JD, Cribbin M, et al. HIV infection and risk,
830 prevention, and testing behaviors among injecting drug users -- National HIV Behavioral Surveillance
831 System, 20 U.S. cities, 2009. *Morb Mortal Wkly Report Surveillance Summ Wash DC* 2002.
832 2014;63(6):1–51.

833 45. Buchacz K, Armon C, Palella FJ, Baker RK, Tedaldi E, Durham MD, et al. CD4 cell counts at HIV
834 diagnosis among HIV Outpatient Study participants, 2000-2009. *AIDS Res Treat*. 2012;2012:869841.

835 46. Reece M, Herbenick D, Schick V, Sanders SA, Dodge B, Fortenberry JD. Background and
836 considerations on the National Survey of Sexual Health and Behavior (NSSHB) from the
837 investigators. *J Sex Med*. 2010;7 Suppl 5:243–5.

838 47. Chandra A, Mosher WD, Copen C, Sionean C. Sexual behavior, sexual attraction, and sexual identity
839 in the United States: data from the 2006-2008 National Survey of Family Growth. *Natl Health Stat*
840 *Rep*. 2011;2011/05/13(36):1–36.

841 48. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, et al. Estimating the population size
842 of men who have sex with men in the United States to obtain HIV and syphilis rates. *Open AIDS J*.
843 2012;6(Journal Article):98–107.

844 49. Age and sex composition in the United States, United States Census Bureau,
845 <https://www.census.gov/data/tables/2017/demo/age-and-sex/2017-age-sex-composition.html>,
846 [Last accessed August 2022].

847 50. Arias E. United States Life Tables, 2004. *Natl Vital Stat Rep*. 2007;56(9):1–40.

848 51. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, et al. Estimated HIV incidence in the
849 United States, 2006-2009. *PLoS One*. 2011;6(8):e17502.

850 52. Centers for Disease Control and Prevention, Atlas Plus, [https://www.cdc.gov/nchhstp/atlas/about-](https://www.cdc.gov/nchhstp/atlas/about-atlas.html)
851 [atlas.html](https://www.cdc.gov/nchhstp/atlas/about-atlas.html) [Last Accessed August 2022].

852 53. Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the United
853 States, 2010–2015. *HIV Surveill Suppl Rep*. 2018;23(No. 1)(Journal
854 Article):<http://www.cdc.gov/hiv/library/reports/hiv-2018>.

- 855 54. Centers for Disease C, Prevention. HIV surveillance--United States, 1981-2008. *MMWR Morb Mortal*
856 *Rep.* 2011;60(21):689–93.
- 857 55. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care
858 objectives by using HIV surveillance data—United States and 6 dependent areas, 2015. *HIV*
859 *Surveillance Supplemental Report* 2017;22(No. 2).
860 <http://www.cdc.gov/hiv/library/reports/hivsurveillance.html>. Published July 2017. Accessed [August
861 2022]. (Journal Article).
- 862 56. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and
863 care objectives by using HIV surveillance data—United States and 6 dependent areas—2011.
864 *HIV Surveillance Supplemental Report* 2013;18(No. 5). [http://www.cdc.gov/hiv/library/reports/](http://www.cdc.gov/hiv/library/reports/surveillance/)
865 [surveillance/](http://www.cdc.gov/hiv/library/reports/surveillance/). Published October 2013. Accessed August 2022. (Journal Article).
- 866 57. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care
867 objectives by using HIV surveillance data—United States and 6 dependent areas, 2020. *HIV*
868 *Surveillance Supplemental Report* 2022;27(No. 3). [http://www.cdc.gov/hiv/library/reports/hiv-](http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html)
869 [surveillance.html](http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html). Published May 2022. Accessed [August 2022].
- 870 58. Bingham A, Shrestha RK, Khurana N, Jacobson EU, Farnham PG. Estimated Lifetime HIV-Related
871 Medical Costs in the United States. *Sex Transm Dis.* 2021 Apr 1;48(4):299–304.
- 872 59. Bhatele A, Yeom JS, Jain N, Kuhlman CJ, Livnat Y, Bisset KR, et al. Massively Parallel Simulations of
873 Spread of Infectious Diseases over Realistic Social Networks. In: 2017 17th IEEE/ACM International
874 Symposium on Cluster, Cloud and Grid Computing (CCGRID) [Internet]. Madrid, Spain: IEEE; 2017
875 [cited 2022 Dec 10]. p. 689–94. Available from: <http://ieeexplore.ieee.org/document/7973759/>
- 876 60. Grefenstette JJ, Brown ST, Rosenfeld R, DePasse J, Stone NT, Cooley PC, et al. FRED (a Framework
877 for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious
878 diseases and control strategies using census-based populations. *BMC Public Health.* 2013;13(Journal
879 Article):940–940.
- 880 61. Stelzle D, Tanaka LF, Lee KK, Ibrahim Khalil A, Baussano I, Shah ASV, et al. Estimates of the global
881 burden of cervical cancer associated with HIV. *Lancet Glob Health.* 2021 Feb;9(2):e161–9.
- 882 62. Pathela P, Braunstein SL, Blank S, Schillinger JA. HIV incidence among men with and those without
883 sexually transmitted rectal infections: estimates from matching against an HIV case registry. *Clin*
884 *Infect Dis Off Publ Infect Dis Soc Am.* 2013 Oct;57(8):1203–9.
- 885 63. Chesson HW, Kidd S, Bernstein KT, Fanfair RN, Gift TL. The Cost-Effectiveness of Syphilis Screening
886 Among Men Who Have Sex With Men: An Exploratory Modeling Analysis. *Sex Transm Dis.* 2016
887 Jul;43(7):429–32.
- 888 64. Friedman EE, Dean HD, Duffus WA. Incorporation of Social Determinants of Health in the Peer-
889 Reviewed Literature: A Systematic Review of Articles Authored by the National Center for HIV/AIDS,
890 Viral Hepatitis, STD, and TB Prevention. *Public Health Rep Wash DC 1974.* 2018 Aug;133(4):392–412.

65. Health Policy, Healthy People 2030, Department of Health and Human Services, <https://health.gov/healthypeople/objectives-and-data/browse-objectives/health-policy> [last accessed August 2022].
66. Tsuboi M, Evans J, Davies EP, Rowley J, Korenromp EL, Clayton T, et al. Prevalence of syphilis among men who have sex with men: a global systematic review and meta-analysis from 2000-20. *Lancet Glob Health*. 2021 Aug;9(8):e1110–8.
67. Chan PA, Robinette A, Montgomery M, Almonte A, Cu-Uvin S, Lonks JR, et al. Extragenital Infections Caused by *Chlamydia trachomatis* and *Neisseria gonorrhoeae*: A Review of the Literature. *Infect Dis Obstet Gynecol*. 2016;2016:5758387.
68. Werner RN, Gaskins M, Nast A, Dressler C. Incidence of sexually transmitted infections in men who have sex with men and who are at substantial risk of HIV infection – A meta-analysis of data from trials and observational studies of HIV pre-exposure prophylaxis. Mugo PM, editor. *PLOS ONE*. 2018 Dec 3;13(12):e0208107.
69. Ong JJ, Baggaley RC, Wi TE, Tucker JD, Fu H, Smith MK, et al. Global Epidemiologic Characteristics of Sexually Transmitted Infections Among Individuals Using Preexposure Prophylaxis for the Prevention of HIV Infection: A Systematic Review and Meta-analysis. *JAMA Netw Open*. 2019 Dec 11;2(12):e1917134.

919 Table 1: Overview of Disease 1 (HIV) scenarios

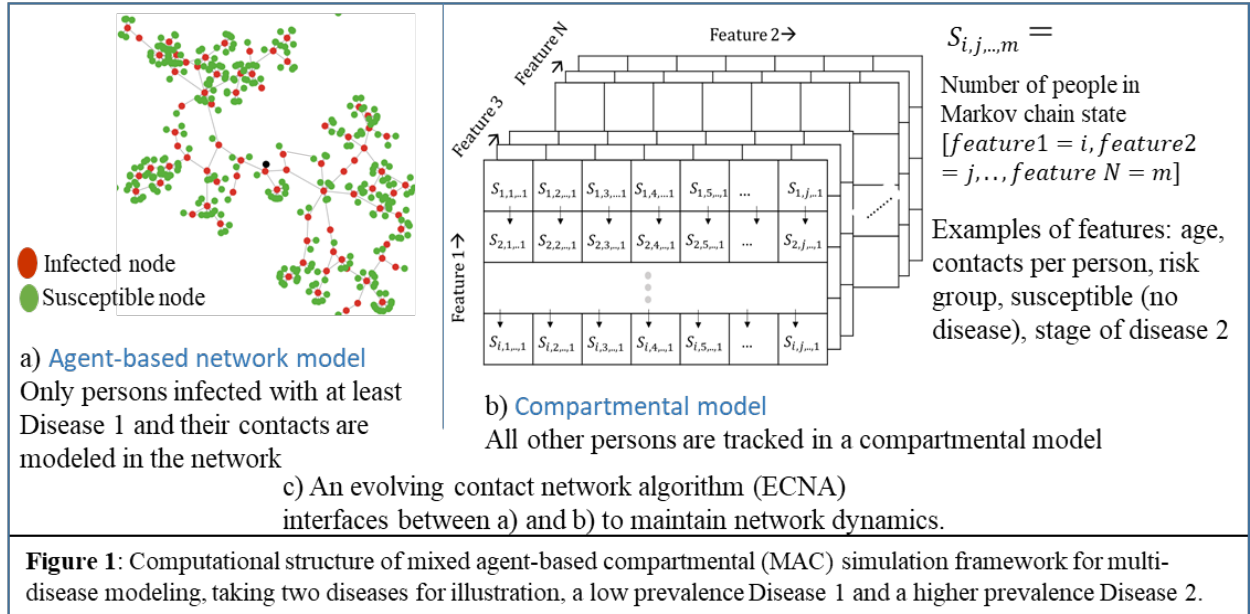
HIV (Disease 1) scenarios	Descriptive name	HIV prevalence assumptions	HIV care assumptions
Scenario 1	status-quo	Initiated the model with a HIV prevalence of 0.1%, 0.05%, and 6% for HETF, HETM, and MSM, respectively. Values correspond to HIV prevalence in the United States in 2006.	Care continuum distributions scaled-up as per values in the United States over the period 2006 – 2017. The corresponding values for the proportion of PWH on treatment with viral suppression was about 20%, 18%, and 19%, in 2006, and 56%, 50%, and 57%, in 2017, for HETF, HETM, and MSM, respectively.
Scenario 2	low prevalence, low care	Initiated the model with a HIV prevalence of 0.07%, 0.04%, and 3.2% for HETF, HETM, and MSM, respectively. Values correspond to HIV prevalence in the United States in 1990s.	Care continuum distributions as per values in 2006, and kept constant for all 12 years of simulation. The corresponding values for the proportion of PWH on treatment with viral suppression was about 20%, 18%, and 19%, for HETF, HETM, and MSM, respectively.
Scenario 3	low prevalence, high care	Initiated the model with a HIV prevalence of 0.07%, 0.04%, and 3.2% for HETF, HETM, and MSM, respectively. Values correspond to HIV prevalence in the United States in 1990s.	Care continuum distributions as per values in 2017, and kept constant for all 12 years of simulation. The corresponding values for the proportion of PWH on treatment with viral suppression was about 56%, 50%, and 57%, for HETF, HETM, and MSM, respectively.
Scenario 4	equal and low prevalence for all transmission groups	Initiated the model with a prevalence of 0.07% for HETF, HETM, and MSM. Values correspond to HETF HIV prevalence in the United States in 1990s.	Care continuum distributions as per values in 2006, and kept constant for all 12 years of simulation. The corresponding values for the proportion of PWH on treatment with viral suppression was about 20%, 18%, and 19%, for HETF, HETM, and MSM, respectively.

920 Note: HIV care assumptions are an input to the simulation. HIV prevalence are an input only for
921 initialization of the model in the first year of simulation. HIV prevalence over time are an outcome of the
922 model. For Scenario 1 (status-quo), the model was validated to match the U.S. epidemic on multiple
923 metrics, including prevalence, for the period 2006 to 2017. Scenarios 2 to 4 are hypothetical.

Table 2: Computational time of the mixed-agent based compartmental model for the two-disease example

Number of nodes in the network			Number of persons in the compartmental model	Total simulation population size	*Computation time per run (minutes) Average (range)
Infected	Susceptible partners	Total			
4,100	17,947	22,047	1,025,050	1,047,097	12(10-16)
8,216	35,740	43,956	2,053,900	2,097,856	31(24-41)
12,020	50,612	62,632	3,005,000	3,067,632	56(46-71)

*Using single thread on Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz 3.70 GHz 64-bit operating system, x64-based processor. Average and range of 10 runs.



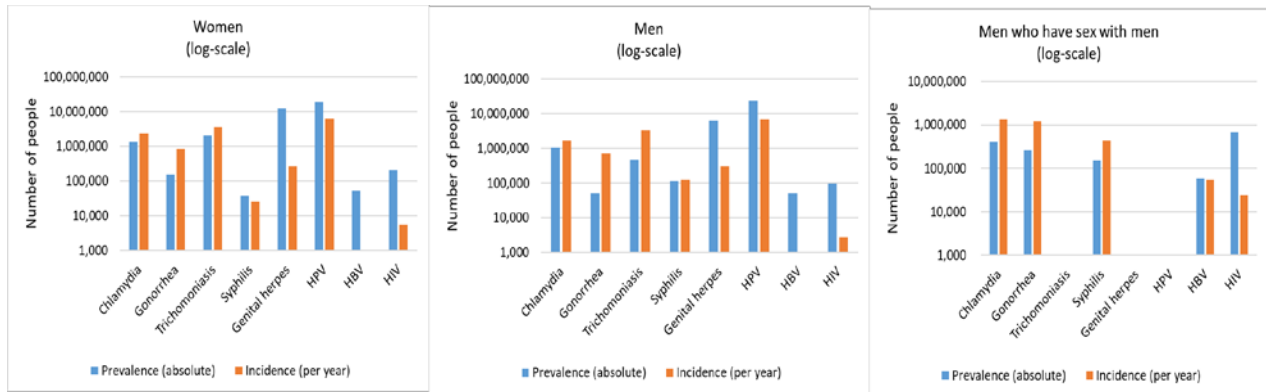


Figure 2a: Prevalence (absolute) and incidence (per year) of STIs in the US [Source: (34,48,66–69)]. *Note: Except for HIV, all other STI incidence in MSM (men who have sex with men) were among MSM at high-risk of STI from pooled global estimates; Not all STI are presented for MSM due to data unavailability.*

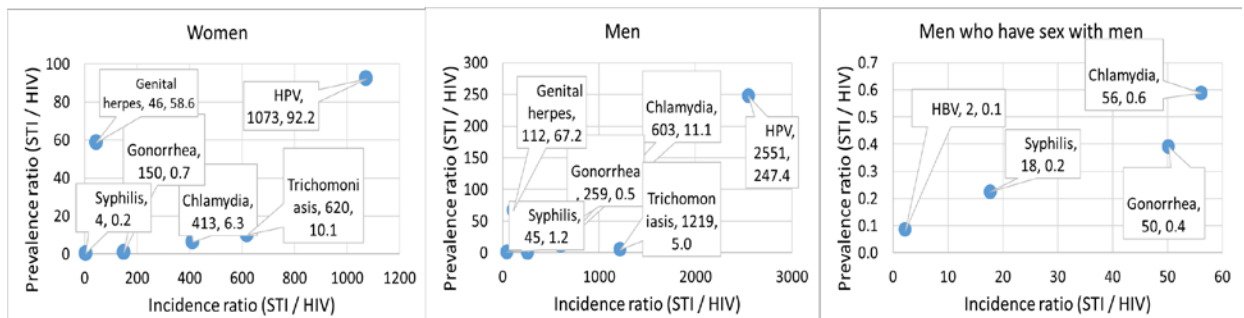


Figure 2b: STI to HIV incidence ratios (x-axis) and prevalence ratios (y-axis) for common STIs in the U.S. in 2018 [Source: (34,66–69)]. *Note: Except for HIV, all other STI incidence in MSM (men who have sex with men) were among MSM at high-risk of STI from pooled global estimates; Not all STI are presented for MSM due to data unavailability.*

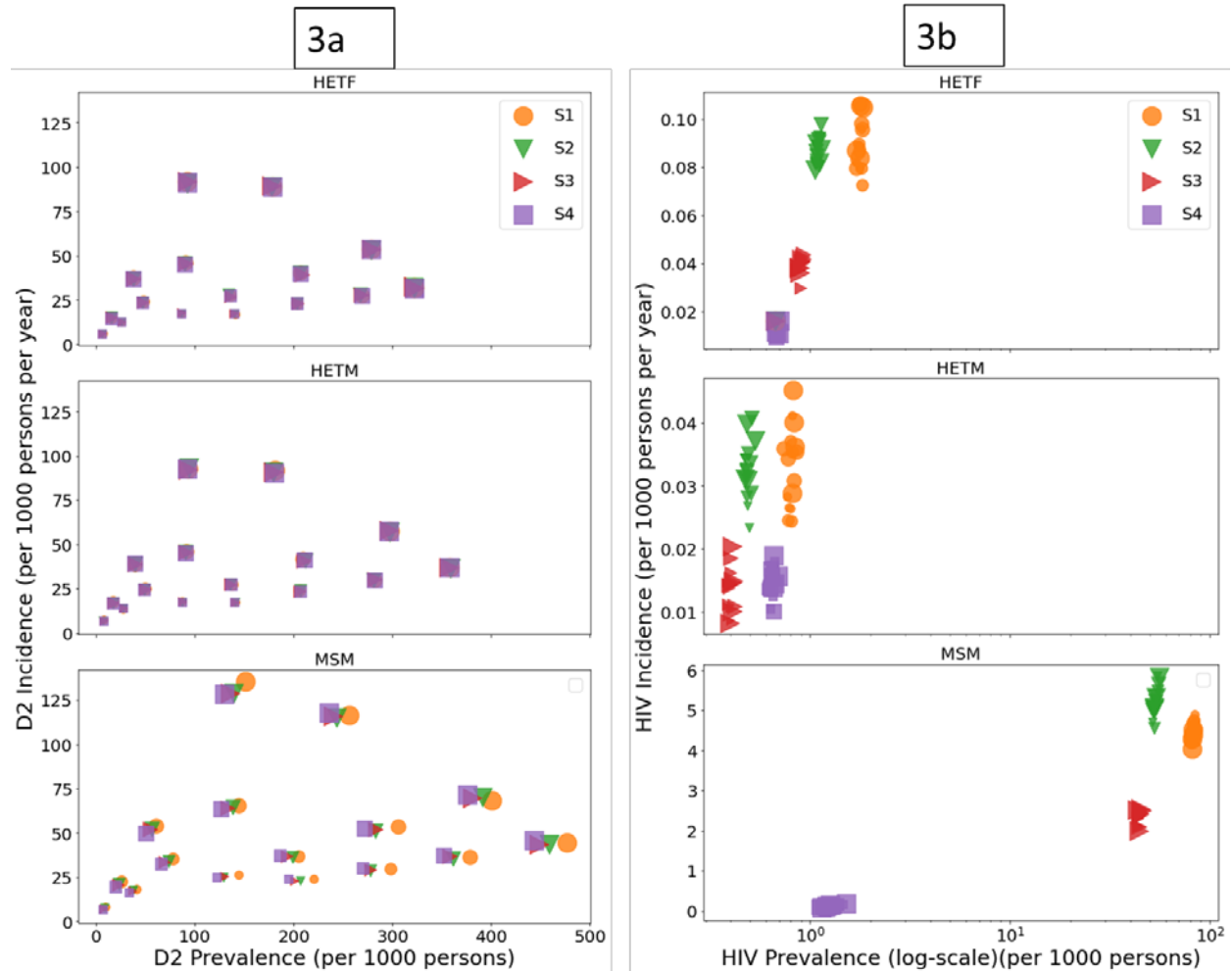


Figure 3: a) Varying transmission rates and recovery rates creates varying burdens of hypothetical Disease 2 (D2).
b) Varying assumptions for HIV across Scenarios 1 to 4 (S1 to S4), creates varying HIV disease burdens (right).
HETF-heterosexual female; HETM-heterosexual male; MSM-men who have sex with men (men only and men and women). Note: For each HIV scenario (S1 to S4), there are 16 data points corresponding to each of the sixteen Disease 2 scenarios (larger the marker size higher the value of D2 transmission rate); For HETF and HETM, data points for S1 to S4 mostly overlap; Results are from last year of 12-year-long simulation (for S1 it corresponds to calendar year 2017).

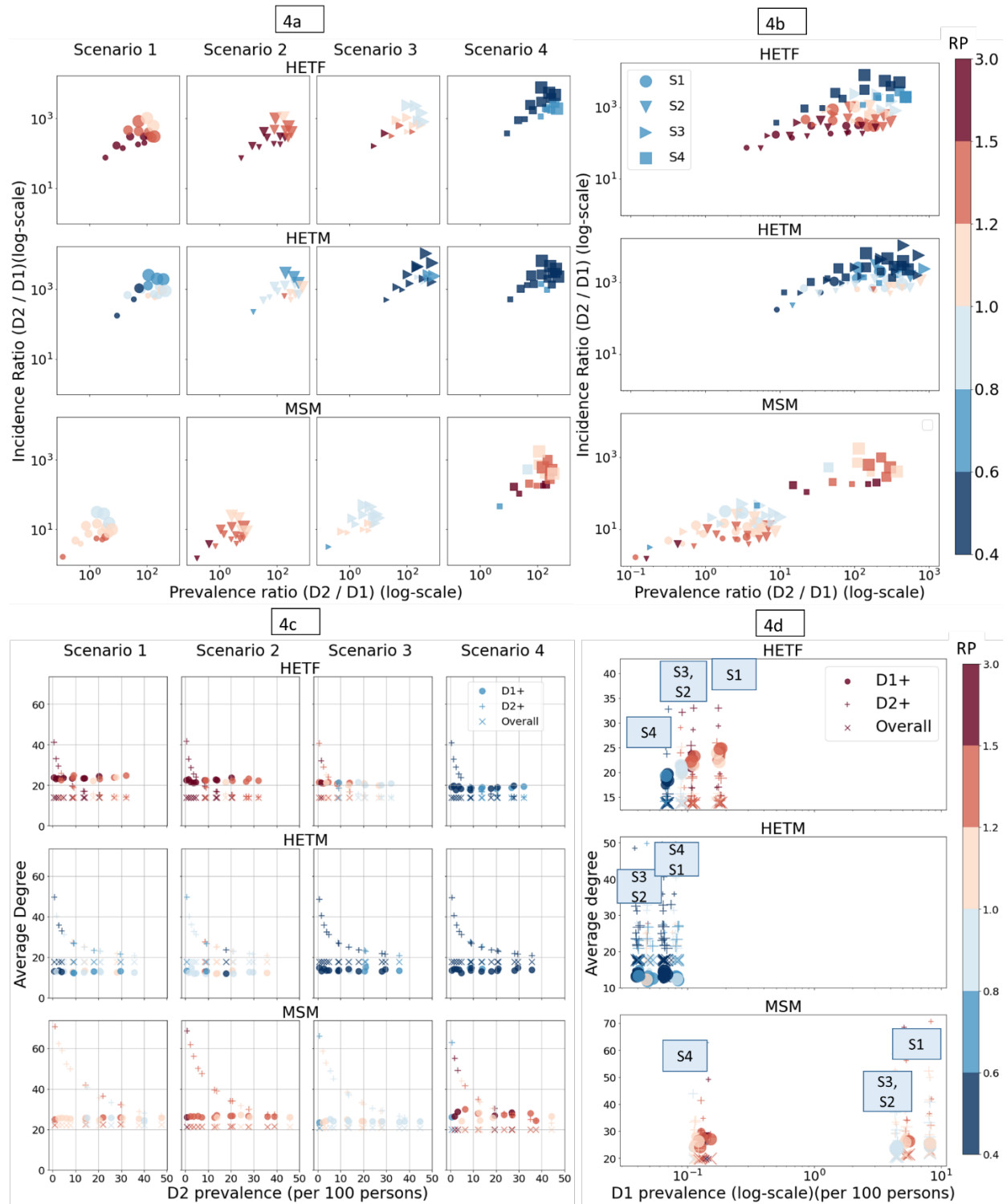


Figure 4: a) Relative prevalence ($RP_{D2|D1}$) (color gradient) presented as a function of $D2/D1$ incidence ratio (y-axis) and prevalence ratio (x-axis) for Scenarios 1 to 4 in each column. b) Same as a) but all scenarios combined into one column. c) Corresponding average degrees (y-axis) among persons with at least D1 (D1+), at least D2 (D2+), and overall, plotted against D2 prevalence (x-axis) for Scenarios 1 to 4. d) The same average degrees (y-axis) but plotted against D1 prevalence;

$RP_{D2|D1}$ is the prevalence of D2 among persons with D1 compared to persons without D1; D1 is HIV; D2 is hypothetical Disease 2; HETF-heterosexual female; HETM-heterosexual male; MSM-men who have sex with men (men only and men and women). Note: For each HIV scenario (S1 to S4), there are 16 data points corresponding to each of the sixteen Disease 2 scenarios (larger the marker size higher the value of D2 transmission rate); Results are from last year of 12-year-long simulation (for S1 it corresponds to calendar year 2017).

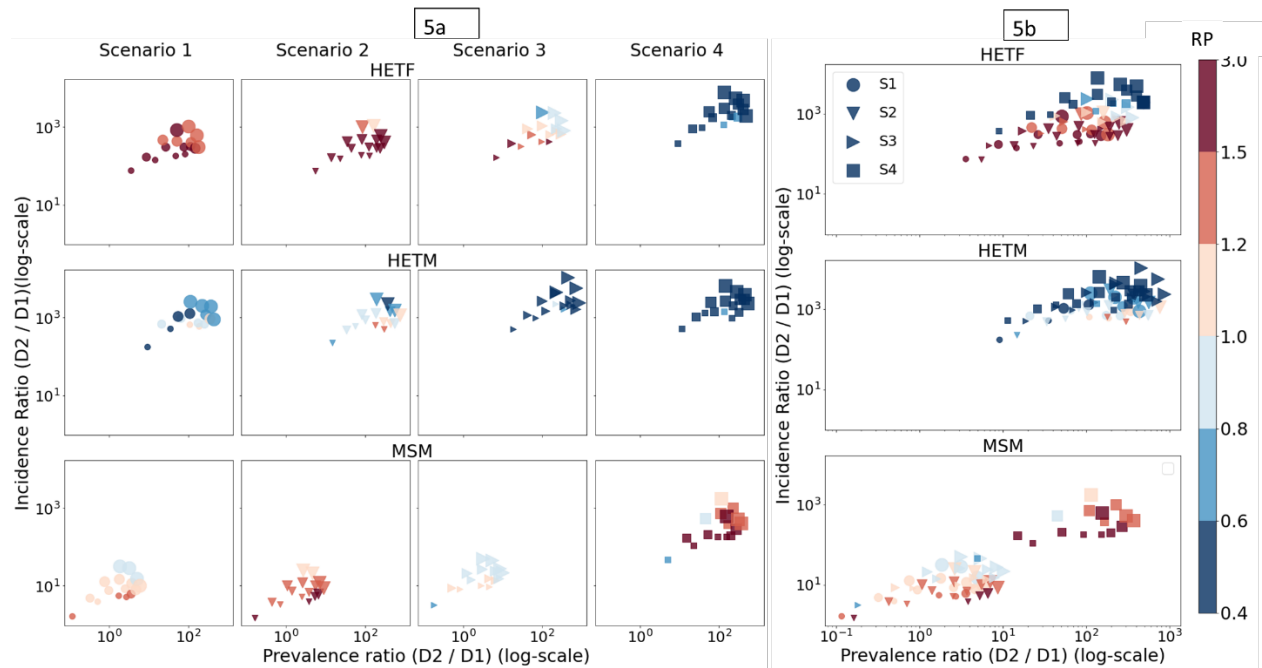


Figure 5: a) Relative prevalence ($RP_{D1|D2}$)(color gradient) as a function of D2/D1 incidence ratio (y-axis) and prevalence ratio (x-axis) for Scenarios 1 to 4 (S1 to S4) in each column. b) Same as a) but all scenarios combined into one column;

$RP_{D1|D2}$ is the prevalence of D1 among persons with D2 compared to persons without D2; D1 is HIV; D2 is hypothetical Disease 2. HETF-heterosexual female; HETM-heterosexual male; MSM-men who have sex with men (men only and men and women). Note: For each HIV scenario (S1 to S4), there are 16 data points corresponding to each of the sixteen Disease 2 scenarios (larger the marker size higher the value of D2 transmission rate); Results are from last year of 12-year-long simulation (for S1 it corresponds to calendar year 2017).

**APPENDIX: A new mixed agent-based network and compartmental simulation framework
for joint modeling of related infectious diseases- Application to sexually transmitted
infections**

Chaitra Gopalappa, PhD, *Hari Balasubramanian, PhD, *Peter J. Haas, PhD

University of Massachusetts Amherst, 160 Governors Drive, Amherst, MA 01003

*HB and PJH equally contributed to the work and are listed in alphabetical order.

Corresponding author: Chaitra Gopalappa, chaitrag@umass.edu, Associate Professor, University
of Massachusetts Amherst, 160 Governors Drive, Amherst, MA 01003

Keywords: joint-modeling diseases, multi-disease modeling, simulation modeling, HIV and
STIs, diseases and social determinants, SDOH

TABLE OF CONTENTS

S1.	Overview of MAC.....	46
S1.1	Overview of MAC using low-prevalence disease 1 and high-prevalence disease 2.....	46
S1.2	Computational structure of MAC	48
S2.	Four main modules of MAC.....	51
S2.1	Compartmental module for simulating high-prevalence Disease 2	51
S2.2	Transmission module for simulating new Disease 1 infections in network (HIV-infection)	55
S2.3	ECNA network generation module for generating partnerships of new HIV-infected nodes	56
S2.3.1	ECNA for determining degree of susceptible partner nodes	57
S2.3.2	Two-step Markov process for determining the partnership distribution by age	58
S2.3.3	Assignment model with heuristic solution for partnership initiation and termination.....	61
S2.4	Disease progression module for simulating Disease 1 progression for nodes in the network.....	64
S3.	General data structure and data inputs used in ECNA module development	65
S3.1	Age structure.....	65
S3.2	Birth cohort for evolving network	65
S3.3	Degree-bin structure	66
S3.4	Risk-group categorization.....	67
S3.5	Data for ECNA module methods: heterosexuals	67
S3.6	Data for ECNA module methods: MSM.....	68
S3.7	Additional age-group assumptions	69
S4.	Model initialization and dry run	69
S5.	Overview of simulation modeling steps	70

The mixed agent-based compartmental (MAC) simulation framework was developed for co-modeling of diseases spread over a common contact network but have varying levels of prevalence and incidence that neither agent-based or compartmental model alone are sufficient. The computational framework for the hybrid agent-based compartmental simulation was enabled through use of a recently developed agent-based evolving network modeling technique (ABENM)(1), applied to the development of the Progression and Transmission of HIV (PATH 4.0) model in the United States and validated against data from the National HIV Surveillance Systems (NHSS) (2). The main concept of ABENM is to simulate only persons infected with at least one low-prevalence disease and their immediate contacts at the individual level as agents of the simulation, and to model all other persons including those with high-prevalence diseases using a compartmental modeling structure. Immediate contacts are defined as all partners a person will have over their lifetime who at the current time-step may be infected or susceptible. As these contacts in the network become infected with a low-prevalence disease, their immediate contacts are added as agents to the network (transitioning from the compartmental portion of the model to the network portion of the model), thus *evolving the contact network*. An Evolving Contact Network Algorithm (ECNA) maintains the network dynamics. Here we provide an overview of the MAC simulation framework, without loss of generality, using a two-disease example, low-prevalence Disease 1 and high-prevalence Disease 2. In the analyses presented in the main paper, we modeled HIV as Disease 1, adopting the validated PATH 4.0 model (2) and a hypothetical Disease 2. Without loss of generality, we can model Disease 2 as a single stage disease, as the process would be similar for multi stages, but for the analyses presented in the main paper, we experimented with varying rates of transmission and recovery to make it representative of a range of diseases. We believe this framework can be generalized to any number of diseases, the computational complexity and relevance informing decisions for modeling it in the ABENM or in the compartmental.

1.1 Overview of MAC using low-prevalence disease 1 and high-prevalence disease 2

We present the framework using HIV as Disease 1 and a hypothetical Disease 2. We track HIV-infected persons and immediate contacts using a dynamic graph $G_t(\mathcal{N}, \mathcal{E})$, with the number of nodes in the graph $Q_t = |\mathcal{N}(G_t)|$ and the number of edges $|\mathcal{E}(G_t)|$ dynamically changing over

time t as persons become newly infected with HIV and their immediate contacts are added to the network. Each person in the network has attributes age, transmission-group (heterosexual female (HETF), heterosexual male (HETM), men who have sex with men (MSM)), degree (number of lifetime partnerships), geographic jurisdiction, and health status ($\{\text{stage of Disease 1, stage of Disease 2}\}$, including $\{0,0\}$ to indicate uninfected with either disease).

We track all other persons using an array S_t of size $A \times R \times D \times G \times H$, where,

A is the number of age-groups,

R is the number of risk-groups,

D is the number of degree-bins (degree is the number of contacts per person, degrees are grouped into bins analogous to age grouped into age-groups),

G is the number of geographic jurisdictions (equal to 1 here corresponding to national), and

H is the number of health states (two in numerical analyses: $\{0,0\}$ and $\{0,1\}$ corresponding to $\{\text{HIV stage, Disease 2 stage}\}$; note: by design persons who are HIV positive will not be in the compartmental model).

Therefore, each element of the array ($S_t[\bar{a}, r, \bar{d}, g, h]$) is the number of people in that specific category, and thus, $Q_t + \sum_{i \in [\bar{a}, r, \bar{d}, g, h]} S_{t,i}$, would be the total number of people in the population at time t .

As all HIV infected persons and exposed partners are in the network, HIV transmissions and HIV disease progression are modeled at the individual level (using HIV transmission and HIV disease progression modules in Appendix S2.1). Disease 2 transmission and progression are modeled using differential equations (as typically done in compartmental modeling technique) but with the consideration that people in both the network $G_t(\mathcal{N}, \mathcal{E})$ and compartmental array S_t can be infected with Disease 2 (see compartmental module in Appendix S2.1).

The mathematical challenges to address for this method to work is to maintain the dynamics between the network $G_t(\mathcal{N}, \mathcal{E})$ and compartmental array S_t , including, transitioning people from S_t to $G_t(\mathcal{N}, \mathcal{E})$. Specifically, determining ‘who’, i.e., the degree, transmission group, age, and geographical location of the persons, are to be added as the immediate contacts of the node newly infected with HIV, as these characteristics of infected persons and their contacts are known to be correlated [12]. Upon determining and adding all lifetime partners of nodes newly infected with

HIV, we need to determine when the partnerships would initiate and dissolve, including the age of both partners at that time, such that the overall dynamics of age-mixing and transmission-group-mixing of the resulting network match that of the U.S. population over time. As network generation techniques in commonly used agent-based models are designed to generate the network between the full population (susceptible and infected persons), they cannot be adopted here. We developed an evolving contact network algorithm (ECNA) for modeling the transitions from S_t to $G_t(\mathcal{N}, \mathcal{E})$ (see ECNA module in Appendix S2.3). As HIV is a chronic infection, once persons enter the network $G_t(\mathcal{N}, \mathcal{E})$ they do not transition back to S_t . However, for low-prevalence diseases that are not chronic, they can be added to the compartments corresponding to their group (stability of the network dynamics for varying epidemiology profiles are discussed elsewhere (1)).

We discuss the computational structure of MAC in Appendix S1.2 and the four simulation modules in Appendix S2. A visual representation of the computational structure is presented in Figure 1 (main manuscript). All notations used in the model are also summarized in Table S1.

1.2 Computational structure of MAC

As noted above, following the compartmental modeling structure, we use a five-dimensional array S_t to keep track of the number of susceptible persons (who are not contacts of persons with HIV infection), i.e., $S_t[\bar{a}, r, \bar{d}, g, h]$ is the number of susceptible persons in age-group $\bar{a}$, transmission-group $\bar{r}$, degree-bin $\bar{d}$, pseudo-geographic jurisdiction g , and health status h at time t .

As noted above, following the ABNEM structure, we use a dynamic graph $G_t(\mathcal{N}, \mathcal{E})$ to track HIV-infected persons and their immediate contacts, where $\mathcal{N}$ is a set of nodes, each node representing an HIV-infected person or a susceptible sexual partner (they may or may not have Disease 2), and $\mathcal{E}(G_t)$ is a set of undirected edges, an edge $\{i, j\}$ representing a sexual partnership between nodes i and j .

The graph $G_t(\mathcal{N}, \mathcal{E})$ has the following features:

- Static adjacency matrix:** C_t of time-variant size $Q_t \times Q_t$, with static elements $C_t[i, j] = 1$ if i and j are sexual partners anytime during their lifetime and $C_t[i, j] = 0$ otherwise, and
- Dynamic adjacency matrix:** V_t of time-variant size $Q_t \times Q_t$, with element $V_t[i, j] = 1$ if i and j are in a partnership during month t and $V_t[i, j] = 0$ otherwise.

1181 Each edge $\{i, j\} \in \mathcal{E}$ has the following features (similar to nodes having features of say age, sex, etc.,
1182 edges can also have features):

1183 **Partnership initiation time:** $\bar{t}(\{i, j\})$ representing the simulation month for when the
1184 partnership initiated,

1185 **Partnership termination time:** $\underline{t}(\{i, j\})$ representing the simulation month when the
1186 partnership terminated,

1187 **Partnership initiation age:** $\{\bar{a}_i, \bar{a}_j\}$ representing the age of nodes i and j , at the time of
1188 partnership initiation, and

1189 **Partnership termination age:** $\{\underline{a}_i, \underline{a}_j\}$ representing the age of nodes i and j , at the time
1190 of partnership termination.

1191 Each node $j \in \mathcal{N}(G_t)$ has the following features:

1192 **Actual degree:** d_j representing the actual number of lifetime sexual partners of node j ,

1193 **Current degree:** $\hat{d}_{t,j}$ representing the number of lifetime sexual partners of person j who
1194 are already added as nodes in $G_t(\mathcal{N}, \mathcal{E})$; if node j is HIV-infected $\hat{d}_{t,j} = d_j$, if node j is
1195 HIV-susceptible $\hat{d}_{t,j} \leq d_j$, and thus dynamically changing with time t ,

1196 **Partnership distribution matrix:** $L_{t,j}$ of size $A \times 2$, where A is the number of age-groups,
1197 $L_{t,j}[\bar{a}, 1]$ is the number of partnerships that node j initiates in age group $\bar{a}$, and $L_{t,j}[\bar{a}, 2]$ is
1198 the number of partnerships that are yet to be assigned; the sub-script t are to indicate that
1199 the values of column 2 of $L_{t,j}$ can change over time, specifically, $L_{t,j}[2]$ is a column of
1200 zeros if the node is HIV-infected as all their partnerships are already assigned, and greater
1201 than or equal to zero if the node is HIV-susceptible (when the HIV-susceptible person is
1202 added as a contact of a different HIV-infected person one of the rows is decremented, and
1203 when the HIV-susceptible person becomes HIV-infected all rows of column 2 are
1204 decremented to zero as their partners are found and added - the HIV-ECNA was
1205 specifically developed for determining when and how to assign these partnerships, and thus
1206 generating the network, which is discussed in Appendix S2.3),

Health status: $h_{t,j} = \{\text{Disease 1 status}, \text{Disease 2 status}\}$, e.g., $h_{t,j} = \{1,0\}$ if node j is stage 1 HIV and no Disease 2 at time t ,

Deceased status: $m_{t,j} = 1$ if node j is alive and 0 otherwise,

Age: $a_{t,j}$ taking an integer value representative of the age of node j ,

Geographic jurisdiction: g_j taking an integer value representative of the geographic location of node j ,

Transmission-group: r_j taking one of the following values, representative of transmission-group of node j , $r_j \in \{\text{heterosexual female}, \text{heterosexual male}, \text{MSM}\}$, and

HIV care continuum and disease stage: $s_{t,j}$ taking one of the following values, 0 (not infected), 1 (infected, acute HIV stage, and undiagnosed), 2 (non-acute HIV, and undiagnosed), 3 (diagnosed and not in care), 4 (in care not on antiretroviral therapy (ART) treatment), 5 (on ART no viral load suppression (VLS)), or 6 (on ART with VLS).

The main relationships between different components of the graph $G_t(\mathcal{N}, \mathcal{E})$ are the following.

Between partnership initiation $\bar{t}(\{i,j\})$ and termination $\underline{t}(\{i,j\})$ times and static and dynamic adjacency matrices (C_t and V_t):

$C_t[i,j] = \begin{cases} 1 & \text{if } \{i,j\} \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$, i.e., if $\{i,j\}$ are partners at some point during their life, this will have a value of 1,

$V_t[i,j] = \begin{cases} 1 & \text{if } \bar{t}(\{i,j\}) \leq t \leq \underline{t}(\{i,j\}) \\ 0 & \text{otherwise} \end{cases}$, i.e., if $\{i,j\}$ are partners at time t this will have a value of 1, and thus,

$C_t[i,j] \geq V_t[i,j]$.

Between actual degree d_j , current degrees $\hat{d}_{t,j}$, and static adjacency matrix C_t :

$\hat{d}_{t,j} \begin{cases} = d_j & \text{if node } j \text{ is infected} \\ \leq d_j & \text{if node } j \text{ is susceptible} \end{cases}$, i.e., if a node is infected, they are linked to all partners they will have (actual degree) over their lifetime and thus $d_j = \hat{d}_{t,j}$, and if a node is susceptible, they are only linked to their infected partners and thus $d_j \leq \hat{d}_{t,j}$, and

$\hat{d}_{t,j} = \sum_{i=1:Q_t} C_t[i,j]$, i.e., C_t keeps track of their current degree.

Between actual degree d_j and partnership distribution matrix $L_{t,j}$:

$\sum_{\bar{a}=1:A} L_{t,j}[\bar{a}, 1] = d_j$, at any t , i.e., as $L_{t,j}[\bar{a}, 1]$ tracks number of partnerships that initiate at age-group $\bar{a}$, when summed over all $\bar{a}$ it should add to the actual degree d_j for all nodes whether infected or susceptible, and

$\sum_{\bar{a}=1:A} L_{t,j}[\bar{a}, 2] = \begin{cases} d_j - \hat{d}_{t,j} & \text{if node } j \text{ is susceptible} \\ 0 & \text{if node } j \text{ is infected} \end{cases}$, i.e., as $L_{t,j}[\bar{a}, 2]$ tracks number of

partnerships that initiate at age-group $\bar{a}$ and are yet to be generated, $\sum_{\bar{a}=1:A} L_{t,j}[\bar{a}, 2]$ would be zero if the node is infected because all partnerships of an infected node are already connected in the network, and would be equal to the number of partners yet to be assigned if the node is susceptible. (Assigning partnerships and all other features related to the network are part of the newly developed HIV-ECNA network generation algorithm, discussed later).

This section presented the computational structure of the model, specifically the compartmental modeling structure, the network structure, and the features of the nodes and edges in the network. A visual representation of the computational structure is presented in Figure 1. The next section describes the methods (modules) used in simulating these features, Appendix S3 discusses the data inputs, and Appendix S4 discusses model initialization, and Appendix S5 provides an overview of the steps of the simulation.

2 Four main modules of MAC

We present the overall MAC framework through four modules, that are run at every time-step (monthly) of the simulation: a compartmental module for simulating Disease 2, a Bernoulli transmission module for simulating new infections, the ECNA network generation module for generating partnership networks of new HIV-infected persons, and a disease progression module for simulating HIV-related events for HIV-infected persons.

2.1 Compartmental module for simulating high-prevalence Disease 2

This module updates the demographic features (births, aging, and deaths) and Disease 2 features (transmission and progression) of persons tracked through the array S_t , using difference equations as follows.

$$S_{t+\Delta t} = S_t + \frac{dS_t}{dt} \Delta t$$

$\frac{dS_t}{dt}$ is the rate of change in S_t , with the general equation for its calculation only varying in element corresponding to health status h , as follows using a simple two stage Disease 2 as in our numerical analyses,

$$\frac{dS_t}{dt} [\bar{a}, r, \bar{d}, g, h = \{0,0\}]$$

$$= -D2 \text{ new infections} + D2 \text{ recovery} - \text{aging_out} + \text{aging_in} - \text{deaths}$$

$$\frac{dS_t}{dt} [\bar{a}, r, \bar{d}, g, h = \{0,1\}]$$

$$= D2 \text{ new infections} - D2 \text{ recovery} - \text{aging_out} + \text{aging_in} - \text{deaths}$$

Each element on the right-hand side can be calculated as follows,

$$D2 \text{ new infections} = \theta_{[\bar{a}, r, \bar{d}, g]} S_t [\bar{a}, r, \bar{d}, g, h = \{0,0\}],$$

$$D2 \text{ recovery} = \gamma S_t [\bar{a}, r, \bar{d}, g, h = \{0,1\}],$$

$$\text{aging_out} = \left(\frac{1}{|\bar{a}|} \right) S_t [\bar{a}, r, \bar{d}, g, h = \{0,0\}],$$

$$\text{aging_in} = \left(\frac{1}{|\bar{a}-1|} \right) S_t [\bar{a}-1, r, \bar{d}, g, h = \{0,0\}], \text{ and}$$

$$\text{deaths} = \delta_{\bar{a}} S_t [\bar{a}, r, \bar{d}, g, h],$$

where,

$\theta_{[\bar{a}, r, \bar{d}, g]}$ is the infection rate for persons in age-group $\bar{a}$, transmission-group r , degree-bin $\bar{d}$, pseudo-geographic jurisdiction g at time t ,

γ is the recovery rate (in this hypothetical analyses we assumed it is static, but could be varied as a function of $\bar{a}, r, g$, to represent epidemiological differences by demography or as a function of time t to represent changes in care or treatment over time,

$|\bar{a}|$ is the age-interval of age-group $\bar{a}$, and

$\delta_{\bar{a}}$ is mortality rate as a function of age-group $\bar{a}$ (here we only modeled all-cause mortality, but it could be varied as a function of health status h).

We calculated infection rate $\theta_{[\bar{a}, r, \bar{d}, g]}$ as follows, (as $G = 1$ here, for clarity of notation, we ignore g in below notations)

$$\theta_{[\bar{a},r,\bar{d}]} = \beta_{[\bar{a},r,\bar{d}]} c_{[\bar{a},r,\bar{d}]} \mathbf{M}_r(\bar{a}, \bar{d})$$

$$\mathbf{M}_i = \sum_{j \in \{HETF, HETM, MSM\}} r_{i,j} (\mathbf{P} \mathbf{A})^T \mathbf{D}$$

Where,

$\beta_{[\bar{a},r,\bar{d}]}$ is the per contact transmission probability among persons in age-group $\bar{a}$, transmission-group r , degree-bin $\bar{d}$,

$c_{[\bar{a},r,\bar{d}]}$ is the number of contacts among persons in age-group $\bar{a}$, transmission-group r , degree-bin $\bar{d}$,

$\mathbf{M}_i$ is a matrix of size $A \times D$, with each element $M_i(\bar{a}, \bar{d})$ the probability that the contacts of a person in transmission-group i and with degree $\bar{d}$ and age-group $\bar{a}$ is infected,

$\mathbf{P}$ is a matrix of size $D \times A$, with each element $P(\bar{d}, \bar{a})$ the prevalence ($\frac{I_{\bar{a},\bar{r},\bar{d},g}}{N_{\bar{a},\bar{r},\bar{d},g}}$) among persons in degree-bin $\bar{d}$ and age-group $\bar{a}$, ($I_{\bar{a},\bar{r},\bar{d},g}$ is the number of persons infected with Disease 2 and $N_{\bar{a},\bar{r},\bar{d},g}$ is the number of persons in that group),

$\mathbf{A}$ is an age-mixing matrix of size $A \times A$, with each element $A(m, n)$ the probability persons in age-group m mix with persons in age-group n ,

$\mathbf{D}$ is a degree-mixing matrix of size $D \times D$, with each element $D(m, n)$ the probability persons in degree-bin m mix with persons in degree-bin n ,

$r_{i,j}$ is the probability persons in transmission-group i mix with persons in transmission-group j , and

T represents matrix transpose.

$$I_{\bar{a},\bar{r},\bar{d},g} = S_t[\bar{a}, r, \bar{d}, g, h = \{0,1\}] + |\{\mathcal{N}_{[\bar{a},r,\bar{d},g,h=\{.,1\}]}|\},$$

$$N_{\bar{a},\bar{r},\bar{d},g} = S_t[\bar{a}, r, \bar{d}, g, h = \{0,0\}] + S_t[\bar{a}, r, \bar{d}, g, h = \{0,1\}] + |\{\mathcal{N}_{[\bar{a},r,\bar{d},g,h=\{.,0\}]}|\} +$$

$$|\{\mathcal{N}_{[\bar{a},r,\bar{d},g,h=\{.,1\}]}|\}$$

i.e., I_j would be the sum of Disease 2 infected persons in group j in the compartmental model

($S_t[\bar{a}, r, \bar{d}, g, h = \{0,1\}]$) and the number of nodes in the network in group j ($|\{\mathcal{N}_{[\bar{a},r,\bar{d},g,h=\{.,1\}]}|\}$),

and similarly, N_j would be the sum of the total number of persons in group j in the compartmental

model ($S_t[\bar{a}, r, \bar{d}, g, h = \{0,0\}] + S_t[\bar{a}, r, \bar{d}, g, h = \{0,1\}]$) and the total number of nodes in group j in the network ($|\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,0\}]}| + |\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,1\}]}|$). Note that, while the first element of h is always zero in notations related to S_t indicating that all persons in the compartmental model are HIV negative, we use a dot in notations related to $\mathcal{N}$ to denote that it sums over all nodes in any stage of Disease 1.

Simulating Disease 2 among HIV infected persons in the network $G_t(\mathcal{N}, \mathcal{E})$

We use the above transition rates (infection rate $\theta_{[\bar{a}, r, \bar{d}, g]}$ and progression rate γ) to simulate Disease 2 among nodes in the network. We first convert rates to probability, as *probability* = $1 - \exp^{-rate \cdot 1}$ (assuming sojourn times follow exponential distribution). We then use a binomial distribution to determine number of nodes to transition, specifically,

$$\text{number of nodes newly infected with Disease 2} = i_{D2} \sim \text{Binomial}(|\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,0\}]}|, 1 - \exp^{-\theta_{[\bar{a}, r, \bar{d}, g]} \cdot 1})$$

$$\text{number of nodes newly recovered from Disease 2} = r_{D2} \sim \text{Binomial}(|\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,1\}]}|, 1 - \exp^{-\gamma \cdot 1})$$

We then randomly choose i_{D2} from node-set $\{\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,0\}]}|\}$ and set their health status $h = \{.,1\}$, and randomly choose r_{D2} from node-set $\{\mathcal{N}_{[\bar{a}, r, \bar{d}, g, h = \{.,1\}]}|\}$ and set their health 2 status $h = \{.,0\}$. Note that the dot in the first element of h indicates that Disease 2 transitions in our numerical are independent of Disease 1 stage, however, to model biological risks of coinfections this assumption would be modified.

Note that, given the simulation time-step is chosen such that the rates are sufficiently small, as the sample size increases, the number of persons to transition drawn from a Binomial distribution will be similar to the number determined by directly using the difference equations, expect that the former is stochastic and generates an integer number and the latter is deterministic and generates a floating point number. To equalize the effects of randomness, the model in the numerical analyses was setup to draw from the Binomial distribution even for the compartmental model. For example, the number of D2 new infections in the compartmental model (Section S2.1) would change from $\theta_{[\bar{a}, r, \bar{d}, g]} S_t[\bar{a}, r, \bar{d}, g, h = \{0,0\}]$ to $\text{Binomial}(|S_t[\bar{a}, r, \bar{d}, g, h = \{0,0\}]|, 1 - \exp^{-\theta_{[\bar{a}, r, \bar{d}, g]} \cdot 1})$.

2.2 Transmission module for simulating new Disease 1 infections in network (HIV-infection)

This module determines if a HIV-susceptible node l in the network $G_t(\mathcal{N}, \mathcal{E})$ becomes infected using a Bernoulli transmission equation and individual-level sexual behaviors and transmission risk factors, each factor modeled as functions of demographic features, and HIV testing and treatment status of infected contacts. Note, as persons in the compartmental model array S_t are not connected to an HIV-infected person, their chance of infection is zero. However, note, persons can move from S_t to $G_t(\mathcal{N}, \mathcal{E})$ upon becoming partners of an HIV-infected person, modeled using the HIV-ECNA algorithm discussed in the next section, which would then expose them to HIV infection. Specifically, every time-step (monthly) of the simulation, this module determines if a HIV-susceptible node l in graph $G_t(\mathcal{N}, \mathcal{E})$ becomes HIV-infected i.e., for nodes with health status $\mathcal{h}_{t-1,l} = \{0, \cdot\}$ or $\mathcal{h}_{t-1,l}[0] = 0$, it estimates its updated value $\mathcal{h}_{t,l}$ as follows.

$$\mathcal{h}_{t,l}[0] = F^{-1}\left(1 - \prod_{j=1}^{Q_t} (1 - \alpha_j \varepsilon)^{s_{t,j} \cdot c_{t,j}} (1 - \alpha_j)^{s_{t,j} \cdot (1 - c_{t,j})}\right), \text{ where,}$$

$\alpha_j = V_t[l, j] \cdot m_{t,j} \cdot p_{t,j}$, where $V_t[l, j]$ and $m_{t,j}$ are the elements of the graph described in Appendix S1.2, and $p_{t,j}$ is the probability of transmission per act modeled as a function of health state $\mathcal{h}_{t,j}$ and transmission-group r_j of the infected node j ; we will have a value of $\alpha_j = p_{t,j}$ if j is a contact of l (i.e., $V_t[l, j] = 1$) and is alive (i.e., $m_{t,j} = 1$), and $\alpha_j = 0$ otherwise; further $p_{t,j} = 0$ if $\mathcal{h}_{t-1,j} = \{0, 0\}$, and $p_{t,j} \geq 0$ for all other values of $\mathcal{h}_{t-1,j}$, i.e., $p_{t,j}$ can be a function of care and disease stage of Disease 1 only or a function of care and disease stage of Disease 1 and Disease 2 if the presence of one infection biologically increases the risk of another,

$\varepsilon = 1$ - condom effectiveness,

$s_{t,j}$ = number of sex acts per month with node j , modeled as a function of age, transmission-group, and number of partners of node j ,

$c_{t,j}$ = proportion of acts with node j that is condom protected, modeled as a function of age, transmission-group, and number of partners of node j ,

$F^{-1}(u)$ = an inverse Bernoulli distribution that takes a value of 1 with probability u and value of 0 with probability $1 - u$.

If node l becomes infected, then the above equation will yield $\mathcal{H}_{t,l}[0] = 1$ i.e., transition to first stage of Disease 1.

Every time-step t , this module also determines and updates any changes in sexual partnerships of HIV-infected nodes. Specifically, for every partnership (k, j) , it updates its active/inactive status as $V_t[k, j] = \begin{cases} 1 & \text{if } \bar{t}(\{k, j\}) \leq t \leq \underline{t}(\{k, j\}), \text{ indicating it is active} \\ 0 & \text{otherwise, indicating it is inactive} \end{cases}$.

2.3 ECNA network generation module for generating partnerships of new HIV-infected nodes

This module controls the overall network dynamics of partnerships between nodes. The main functionality is to generate the contact network for each new HIV-infected node l . The steps of the ECNA are as follows.

For every new HIV-infected node l ,

1. Determine the number of new partnerships (edges) to generate as actual degree minus current degree. Note, these new partnerships would all be with HIV-susceptible persons as any partnerships with HIV-infected were already added when networks of those HIV-infected persons were created.
2. For each new HIV-susceptible partner node, determine node features: number of lifetime partners using *ECNA (discussed below)*, transmission-group, current age-group, and pseudo-geographic jurisdiction using conditional probability distributions, and partnership distribution using a *two-step Markov process (discussed below)*.
3. For each partnership, determine age of both partners and simulation times at partnership initiation and termination, by formulating and solving as *assignment optimization model (discussed below)*.
4. Determine who each new partner is by a uniform random draw from all who are eligible, i.e., all persons who are eligible have an equal chance of selection. All HIV-susceptible nodes in the network $G_t(\mathcal{N}, \mathcal{E})$ and HIV-susceptible non-agents in the compartmental model array S_t with features matching that in steps 2 and 3 above, can be eligible.
5. For each new partner and partnership (determined in previous steps) update their corresponding features in the network $G_t(\mathcal{N}, \mathcal{E})$ and compartmental array S_t .

S2.3.1 ECNA for determining degree of susceptible partner nodes

We assumed that the contact network of sexual partnerships follows a scale-free network, where the distribution of the number of contacts per person follows a power-law distribution (3). A key feature of scale-free networks is that, for a node l , the degree-bin of a partner $\bar{d}_k$ is not independent of its degree-bin $\bar{d}_l$ because of degree correlations between node neighbors (4). That is, $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l) \neq \Pr(\bar{D}_k = \bar{d}_k)$; where $\bar{D}_k$ is the random variable for degree-bin of node k , and thus, $\bar{d}_k$ cannot be directly drawn from the power-law probability mass function. While the literature presents an analytical method for estimation of $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l)$ for general static scale-free networks (4), this method is not suitable for simulating an epidemic in a dynamically evolving contagion network(1). Specifically, in general static scale-free networks, the full network is available so the degree of all node neighbors are available, and thus $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l)$ is an expectation over all possible values of $\bar{d}_l$, i.e., an average over “all” node neighbors. However, as we only simulate HIV-infected nodes and their immediate contacts in the network, in our context, only values corresponding to HIV-infected node neighbors’ are used. As it is more likely that nodes with higher degree get infected first, the value of $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l)$ when $l =$ ‘all node neighbors’ is different compared to when $l =$ ‘HIV-infected node neighbors’. And further, $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l)$ is likely to change over time as the HIV epidemic spreads and the percent of the population that is HIV-infected changes. We developed a neural network model for the prediction of $\Pr(\bar{D}_k = \bar{d}_k | \bar{D}_l = \bar{d}_l)$ using as independent variables, $\bar{d}_l$, $\bar{d}_k$, the minimum degree of the network m , the percent of the population that is infected (p), and the scale-free network parameter λ_{r_l} corresponding to transmission-group of node l (r_l). Details of this algorithm are presented elsewhere (1), and summarized below.

The neural network was trained on data generated by multiple simulations of hypothetical diseases, characterized by different values of probability of transmission, on scale-free networks of different minimum degree and size. Specifically, using the barabasi.game in the R software multiple scale-free network of varying size (N) (1000, 5000, and 10000), and minimum network degree (m) (1, 2, 3, 4, and 5) were generated. Then, hypothetical diseases were simulated on each network, and at every time-step, for every newly infected node, the following data were collected, the independent variables $\bar{d}_l$, $\bar{d}_k$, m , N , p , and the conditional probability as $\Pr(\bar{d}_k | \bar{d}_l) = \frac{z_{k,l}}{\sum_l z_{k,l}}$,

where, $z_{k,l}$ was a counter in the simulation that kept track of the total number of susceptible contacts with degree $d_l \in \bar{d}_l$ for newly infected persons with $d_k \in \bar{d}_k$ for every $\bar{d}_k, \bar{d}_l$ combination. The value of the transmission probability itself were not relevant, however the prevalence (p) was relevant, and therefore we used small values of transmission probability to capture a range of prevalence over time as the infection spread. Although the scale-free network generation package in R (barabasi.game) uses m and N as inputs, the scale-free network power-law degree distribution decay-coefficient λ is more relevant as it is more generalizable. Therefore, we recoded m and N into λ , by calculating λ from the corresponding networks generated, as follows. For a scale-free network following a power law degree distribution, the probability that a node has degree k (represented as $Pr(k)$) can be written as

$$Pr(k) = \frac{k^{-\lambda}}{\sum_{d=m:\bar{m}} d^{-\lambda}}$$

where, the decay-coefficient $\lambda = -\Delta \log(n_k) / \Delta \log(k)$, n_k = number of nodes with degree k , Δ is the gradient, m is the minimum degree, and $\bar{m}$ is the maximum degree (network size influences maximum degree and thus served as a proxy).

The trained neural network was used in the ECNA module for determining the value of $Pr(\bar{d}_k | \bar{d}_l)$, corresponding to the degree of the newly infected node ($\bar{d}_l$), HIV prevalence at the time, the scale-network parameter for the transmission-group (λ_{r_l}), and the minimum degree of the network (m). Data used for the estimation of λ_{r_l} are discussed in Appendix S3.5 and S3.6. We assumed minimum degree of 2 and maximum degree of 128, and thus, for every $\bar{d}_l$, values of $Pr(\bar{d}_k | \bar{d}_l)$ were normalized to add to 1 to keep the range of $\bar{d}_k$ from 2^1 to 2^7 . However, the computational structure can be set to take any degree range (see Appendix S3.3).

S2.3.2 Two-step Markov process for determining the partnership distribution by age

Suppose d_k is the actual degree (number of lifetime partners) of a newly added susceptible node k . We need to determine at what age of k will each partnership initiate. Specifically, suppose there is a matrix $\bar{L}$ of size $A \times D$, with A the number of age-groups and D the number of degree-bins, and element $\bar{L}[\bar{a}, \bar{d}]$ representing the proportion of partnerships that initiate at age-group $\bar{a}$ for persons in degree-bin $\bar{d}$, with each column of $\bar{L}$ adding to 1, for all $\bar{d} \in \{1, 2, \dots, D\}$. Then, for any

node k with actual degree $d_k \in \bar{d}$, we can calculate the partnership distribution matrix, i.e., the number of partnerships that initiate at age $\bar{a}$, as $L_k[\bar{a}, 1] = \bar{L}[\bar{a}, \bar{d}] \cdot d_k$. Direct data for $\bar{L}$ would be a longitudinal survey over the duration of life of an individual, where the individual reports the number of partnerships they initiated at every age points of their life. Such surveys, however, are unavailable. Therefore, we estimated $\bar{L}$ using survey data on the reported number of partners up to that time by persons of different age-groups (see S2.2 and S2.3). Note that, these survey data only represent the number of partners up to the current age of the surveyed individual. Thus, the degree-bin $\bar{d}$ each person would belong to is unknown as $\bar{d}$ represents the number of partners the person would eventually have over their full lifetime. The age at which each partnership initiated is also unknown.

We developed a two-step Markov process and simulation method for estimation of $\bar{L}$ using the survey data. Details of this method were presented previously (2), we discuss it below for completeness.

Step 1: Formulation and solution method for the probabilities of initiating a new partnership-

Let $X_{\bar{a}}$ be the degree-bin corresponding to the number of partners a node has up to the age of age-group $\bar{a}$, $\bar{a} \in \{1, 2, \dots, A\}$. Then, $\{X_{\bar{a}_t}; \Omega, \text{superdiag}(P_1, P_2, \dots, P_{A-1})\}_{t=1}^{\infty}$ is a discrete-time Markov chain with state space $\Omega = \{1_{\bar{a}=1}, 2_{\bar{a}=1}, \dots, D_{\bar{a}=1}, 1_{\bar{a}=2}, 2_{\bar{a}=2}, \dots, D_{\bar{a}=2}, \dots, 1_{\bar{a}=A}, 2_{\bar{a}=A}, \dots, D_{\bar{a}=A}\}$, where each discrete timestep corresponds to the length of time equal to the width of age-group $\bar{a}$, D is the number of degree-bins, A is the number of age-groups, $\text{superdiag}(P_1, P_2, \dots, P_{A-1})$ is a block matrix with $\{P_1, P_2, \dots, P_{A-1}\}$ on the superdiagonal and all other elements equal to zero¹, and $P_{\bar{a}}$ is a $D \times D$ upper triangular matrix, with $P_{\bar{a}}[x, y]$ equal to the probability that a person transitions from state $x \in \{1_{\bar{a}}, 2_{\bar{a}}, \dots, D_{\bar{a}}\}$ in age-group $\bar{a}$ to state $y \in \{1_{\bar{a}+1}, 2_{\bar{a}+1}, \dots, D_{\bar{a}+1}\}$ in age-group $\bar{a} + 1$ (i.e., goes from degree-bin x to degree-bin y in one age-group increment). By assuming that there are no generational changes in partnership behavior and that births are equal to deaths, we can rewrite the above Markov process as a regular Markov chain that is in steady state, i.e., the state distribution is stationary over time. Using the components of each matrix $P_{\bar{a}}$, and defining a row

¹ Alternatively, "... is a block upper bidiagonal matrix with $\{P_1, P_2, \dots, P_{A-1}\}$ on the superdiagonal and the elements of the diagonal blocks equal to zero."

vector $\pi_{\bar{a}}$ with $\pi_{\bar{a}}[x]$ equal to the proportion of people (among those in age-group $\bar{a}$) with number of partners up to age $\bar{a}$ in degree-bin x , we can write

$$\pi_{\bar{a}+1} = \pi_{\bar{a}} P_{\bar{a}}.$$

Note that, for each $\bar{a}$, the elements of $\pi_{\bar{a}}$ add to 1 and $\bar{a}$ is a normalized sub-vector of the steady state distribution of the above Markov chain.

Data for $\pi_{\bar{a}}$ are available from the behavioral survey studies (see Appendix S3.5 and S3.6). We solved for the values in $P_{\bar{a}}$, for each $\bar{a}$, by formulating and solving a linear least-squares optimization problem as follows.

For each $\bar{a}$, we formulated an optimization model as

$$\text{Objective Function: } \min_z \frac{1}{2} \|Cz - \pi_{\bar{a}+1}^T\|_2^2$$

$$\text{Subject to: } Bz = b$$

$$0 \leq z \leq 1$$

where the objective function is equivalent to $\min_{P_{\bar{a}}} \frac{1}{2} \|\pi_{\bar{a}} P_{\bar{a}} - \pi_{\bar{a}+1}\|_2^2$, i.e., equivalent to minimizing the sum of squared errors between the left- and right-hand sides of the equation $\pi_{\bar{a}+1} = \pi_{\bar{a}} P_{\bar{a}}$ defined above ($\|\cdot\|_2$ is the ℓ_2 -norm), and is obtained by reformulating as below:

z is a column vector of length D^2 obtained by stacking the columns of $P_{\bar{a}}$; i.e., $z = \text{vec}(P_{\bar{a}})$,

C is a $D \times D^2$ matrix where, for each row $i = 1, 2, \dots, D$, $C[i, (i-1)D + 1:iD] = \pi_{\bar{a}}$ and all other elements are equal to zero,

$\pi_{\bar{a}+1}^T$ is the transpose of $\pi_{\bar{a}+1}$,

B is a $D \times D^2$ matrix consisting of D side-by-side $D \times D$ identity matrices, and

b is a column vector of length D , all of whose elements are equal to 1.

The reformulation converts the objective function into a linear least squares function, which is easier to solve using standard linear least-squares solvers. The constraint $Bz = b$ ensures that each row of $P_{\bar{a}}$ adds to 1, a necessary Markov chain property. We used the least-squares solver in

MATLAB to solve for the optimal value of elements of vector z , and thus obtained elements of $P_{\bar{a}}$.

Step 2: Simulation- We simulated a hypothetical population of 10,000 persons. At the start of the simulation, all persons are assigned the lowest age group ($\bar{a} = 1$) and lifetime partners are assigned according to the degree distribution of the lowest age group (taking mid-value of the degree-bin). Every time-unit, of length equal to the width of the age-group interval, each person transitions to the next age-group and is assigned additional partners as per the transition probabilities in $P_{\bar{a}}$ estimated in Step 1. This is repeated until all persons reach the last age-group. Taking all persons with lifetime partners in degree-bin $\bar{d}$ at the last age-group, we estimated $\bar{L}[\bar{a}, \bar{d}]$ as the average of the proportion of partnerships that initiated when that person was at age-group $\bar{a}$.

S2.3.3 Assignment model with heuristic solution for partnership initiation and termination

We formulated the problem of assigning age at partnership initiation as a variant of an assignment optimization model. Suppose $n_k[i]$ is the age of partnership initiation for i^{th} partner of node k ($|n_k| = d_k$ the degree of node k). Then the objective is to assign each element (and all elements) of n_l of a newly infected node l to one (and-only-one) element of n_k for each partner k , with constraints to maintain the probability distribution for age-mixing between partners, and partnership distribution matrices ($L_{t,l}$ and $L_{t,k}, \forall k$) of the newly infected node l and each partner k . However, solving it using standard solvers for each infected person is computationally expensive, and therefore, we developed a heuristic solution algorithm. Solving for partnership initiation age for each new partner sets the focal point for assigning the other three parameters partnership initiation time, partnership termination age, and partnership termination time. Details of this method were presented previously (2), and are discussed below for completeness.

The task is to estimate partnership initiation age $\{\bar{a}_l, \bar{a}_k\}$, termination age $\{\underline{a}_l, \underline{a}_k\}$, initiation time $\bar{t}(\{l, k\})$, and termination time $\underline{t}(\{l, k\})$ between nodes l and k . The optimal values are those that ensures age-mixing between partners is maintained, the distribution of partnership age-initiation for newly infected node l (i.e., $L_{t,l}[1]$) is maintained, and the distribution of partnership age-initiation for each of its partners k is maintained. We can formulate this problem as an optimization model as follows. Let,

M be an age-mixing matrix of size $A \times A$ with element $M[i, j]$ the probability that, given a person is in age-group j , his or her partner is in age-group i (here, element j corresponds to the newly infected node, and i the yet to be assigned partners); M will vary by risk-group, but we do not include risk group in the notation for clarity,

$P = M \cdot \text{diag}(L_{t,l}[2])$, where $\text{diag}(v)$ is a diagonal matrix with diagonal elements equal to those of the vector v , be a matrix of size $A \times A$ with element $P[i, j]$ representing the number of partners of age in age-group i yet to be assigned, when the newly infected node l is in age-group j ,

n be a vector of size A with $n[i] = \sum_{j=1:A} P[i, j]$, $\forall i \in \{1, \dots, A\}$ denoting the number of partnerships to initiate when the partner is in age-group i ,

N be a binary matrix of size $A \times D_{t,l}$, where, $D_{t,l} = d_l - \hat{d}_{t,l}$, d_l is the degree of the newly-infected node l and $d_l - \hat{d}_{t,l}$ is the number of partners to newly add, with element $N[\bar{a}, k] = \begin{cases} 1 & \text{if } L_{t,k}[\bar{a}, 1] > 0 \\ 0 & \text{otherwise} \end{cases}$, i.e., $N[\bar{a}, k] = 1$ if partner k is eligible to initiate the partnership at age-group $\bar{a}$, and

R be a binary matrix of size $A \times D_{t,l}$, with element $R[\bar{a}, k] = 1$ if the partnership with k would occur when partner k is in age-group $\bar{a}$.

Then, the problem is to solve for R using the following formulation of the optimization model

$$\text{Objective Function: } \min_R \sum_{i=1:A} \left[\sum_{j=1:d_l - \hat{d}_{t,l}} R[i, j] - \sum_{q=1:A} P[i, q] \right]^2$$

$$\text{Subject to: } R[i, j] \leq N[i, j]; \quad \forall i \in \{1, \dots, A\}, j \in \{1, \dots, d_l - \hat{d}_{t,l}\}$$

$$\sum_{i=1:A} R[i, j] = 1; \quad \forall j \in \{1, \dots, d_l - \hat{d}_{t,l}\}$$

The objective function seeks to minimize the sum of squared error between the number of contacts initiating at a particular age-group and the expected number of contacts to initiate at that age-group (here the age references to the partner's age). The first constraint ensures that the newly infected node does not initiate a partnership in the age-group where the partner does not have an expected partner initiation. The second constraint ensures that any partnership initiates only one time. The above model can be considered a variant of an unbalanced assignment problem, a

category of problems that deal with assigning n jobs (here partners) to m machines (here age-groups). The first variant being the addition of the first constraint (which is similar to a machine assignment problem constraint where not all jobs are eligible on all machines). The second variant being the modification of the objective function, from the typical form $\min_R \sum_{i=1:A} \sum_{j=1:d_l-\hat{a}_{t,l}} R[i, j] c_{ij}$, by setting $c_{ij} = 1$ (the cost of assigning job i to machine j for all i, j combination), and putting a tight constraint that the maximum capacity of the machine should be met ($n[i]$). Notice that, if there exists a solution, there could be more than one solution to this problem. A solution will not exist if for any age-group $\bar{a}$, $n[\bar{a}] > \sum_{j=1:D_{t,l}} N[\bar{a}, j]$, i.e., the number of required partnerships at age-group $\bar{a}$ is greater than that available.

Instead of applying a standard assignment problem optimization solver, which was computationally expensive to apply for every newly infected node, we developed a simple heuristic solution algorithm as follows.

Let e_k be a vector of size A with $e_k[i] = n[i]N[i, k]$, for every new partner k . Then, $e_k[i] > 0$ if k is eligible to initiate a partnership at age-group i and there is a need for a partnership at that age-group. For each new partner k , we search through all elements of e_k by starting at the last element, we select the first occurrence of i with $e_k[i] > 0$ as the solution, i.e., set $R[\bar{a} = i, k] = 1$, and update $n[\bar{a}] = n[\bar{a}] - 1$. This process of starting at the last element of e_k leads to an optimal solution, provided a feasible solution exists, because of the following property assumptions related to the distribution of lifetime partners by age-group of partnership initiation.

Property 1: For any k , if $N[i, k] = 1$ then $N[j, k] = 1 \forall 1 \leq j \leq i$ but the opposite is not necessarily true.

Property 2: Let $\mathcal{S}_{\bar{a}}$ be a set of partners eligible to initiate partnership at age-group $\bar{a}$, i.e., if $N[\bar{a}, j] > 1$, then $j \in \mathcal{S}_{\bar{a}}$. Then, $\mathcal{S}_j \subseteq \mathcal{S}_i \forall j > i$, i.e., $|\mathcal{S}_j| \leq |\mathcal{S}_i|$, where $|\cdot|$ is the size of the set.

Properties 1 and 2 suggest that partners who are eligible to initiate partnership at an older age-group are also eligible to initiate partnership at a younger age-group but not necessarily vice-versa. Property 2 further suggests that the number of partners feasible to initiate a partnership at a specific age-group decreases with age, with the oldest age-group having the least number. Therefore, the heuristic method is equivalent to starting at the oldest age-group, randomly picking

from those eligible, removing them from all sets $\mathcal{S}_{\bar{a}}$, and iterating to the next oldest age-group to repeat the process. We apply the heuristic algorithm even for cases where a solution does not exist (infeasible solution), i.e., when there exists at least one age-group, say $\bar{a}$, where $n[\bar{a}] > \sum_{j=1:D_{t,l}} N[\bar{a}, j]$ indicating that the number of required partnerships at age-group $\bar{a}$ is greater than that available. And at the end of the algorithm, any unassigned partner k , i.e., for k with $\sum R[\cdot, k] = 0$, are assigned to initiate partnership when they are in age-group $\bar{a}$, i.e., set $R[\bar{a}, k] = 1$, and if there are more than one such occurrence, they are randomly selected. This infeasibility in solution occurred 9% on average. Considering that the infeasibility is caused by trying to match various types of data, age-mixing data with partnership initiation derived from number of partners, this margin of error could be expected.

Upon solving for R , for every partner k , using P and R together we can identify the age-groups of l and k at which the partnership will initiate. We select a random age within those age-groups to set as $\{\bar{a}_l, \bar{a}_k\}$. We then set $\bar{t}(\{l, k\}) = t + \bar{a}_l - a_{t,l}$ and $a_{t,k} = \bar{a}_k - (\bar{t}(\{l, k\}) - t)$. If this partnership initiated in the past, we would then have $\bar{a}_l - a_{t,l} < 0$, $\bar{t}(\{l, k\}) < t$, and $a_{t,k} > \bar{a}_k$. If this partnership would initiate in the future, we would then have $\bar{a}_l - a_{t,l} > 0$, $\bar{t}(\{l, k\}) > t$, and $a_{t,k} < \bar{a}_k$. If this partnership would initiate at current time-step, we would then have $\bar{a}_l = a_{t,l}$, $\bar{t}(\{l, k\}) = t$, and $a_{t,k} = \bar{a}_k$ (note: this is of significance in HIV as the susceptible person is then exposed to the acute phase of infection where the transmission is high). Finally, partnership termination time $\underline{t}(\{l, k\})$ is set to the time the next partnership of node l initiates.

Data inputs for this section follows from the estimations from the previous sections, and additionally uses age-mixing matrix discussed in Appendix S3.5 and S3.6.

2.4 Disease progression module for simulating Disease 1 progression for nodes in the network

The disease progression module updates the individual-level demographic and disease dynamics for every HIV-infected person in the network just like an agent-based model. This includes aging, Disease 1-related and natural mortality, progression through Disease 1 disease and care stages. For the analyses in the main paper we adopted the disease progression module from PATH (5).

3 General data structure and data inputs used in ECNA module development

3.1 Age structure

For the numerical analyses, the compartmental model used the following age-groups: {13 – 17, 18 – 24, 25 – 29, 30 – 34, 35 – 39, 40 – 44, 45 – 65}. The agent-based model used individual age and tracks persons until death, though for the numerical analyses we included only persons aged 13-65, to focus on transmissions. For future work, the compartmental model can be easily extended to include age 65+ for purposes of tracking disease progression, such as progression of HPV to cervical or other forms of cancer. Also note that, the age-structure in compartmental model is flexible to change and should be informed by age-groups used in data inputs, to sufficiently capture data heterogeneity (see S3.7 for data that informed choice of age-structure used here). Once an age-structure is chosen, all data common to both compartmental and network should be converted to this age-distribution. Further, the distributions should be kept consistent between the network and compartmental, e.g., though agent-based uses individual age, the age-group for partnership mixing should be drawn from the same distribution as that used in compartmental model. Data specific to agents can use any age-structure and have parameters with varying age-structures. For example, the overall population (network +compartmental) was initialized to the U.S. census data using the age-structure noted above, but the distribution of people with HIV (agents) were distributed using data in Table S2. Some of the sexual behavioral data from the literature are presented in Table S4, which has a different age-structure than in the compartmental model. As these data are common to persons in both compartmental and network they were converted to compartmental age-structure.

3.2 Birth cohort for evolving network

Individuals can only enter the population by aging into the lowest sexually active age-group modeled (13 to 17) and are susceptible upon entering the population, i.e., added to the compartmental model. We assume a constant number of births per year, calculated for each risk group using the overall birth rate in the U.S. in 2015 times the population in the respective risk group.

As the evolving network tracks all life-time partnerships of persons with HIV (Disease 1), when a person becomes newly infected, the current age of one or more of their susceptible contacts could

be outside the ‘alive’ susceptible population, either the susceptible contact has already aged-out, i.e., was a partner in the past, or not yet aged-in, i.e., is a partner in the future. Persons who have aged-out will have an age value greater than the sexually-active age (or dead) and thus, computationally will not influence any part of the model. Persons who have not aged-in will have an age value less than 13, including negative. Computationally, this would not cause an issue in the network because the partnership activation algorithm (discussed in S2.3.3) enforces the activation age to be within the sexually active age-groups modeled here. However, when that person is moved from the susceptible compartment to the network, as the model is set to decrement the number in the corresponding compartment by one, but the first age-group in the compartmental array (S_t) is 13-17, it would create an error. To overcome this, the model maintains a birth-cohort, an array of dimension 100×9 (corresponding to age X degree-bin dimensions), initialized to the number of births per year distributed by degree-bin (degree-bin distribution discussed below), i.e., each row sums to constant number of births per year. If a person of age less than 13 years (say i) with degree-bin j moves from compartmental to network (G_t), then the corresponding element of the birth-cohort ($100 + i - 13, j$) is decremented by 1. Every year, the values in the 100th row of the birth-cohort array are added to the compartmental array (S_t) corresponding to the first age-group (13-17), every row $i \in \{2, \dots, 100\}$ of the birth-cohort is set equal to the value in row $i - 1$, and row 0 is initialized to add to the number of births. For the assumptions used here, a birth-cohort dimension of $65 - 13 \times 9$ would be sufficient, however, we set it to 100×9 for the computational structure to be generalizable to any disease assuming maximum age is 100 years, as its contribution to computational complexity is minimal.

3.3 Degree-bin structure

For the numerical analyses, the computational structure of compartmental array uses a degree-bin distribution of dimension 9 as follows: $\{0, 1, 2, 3 - 4, 5 - 8, 9 - 16, 17 - 32, 33 - 64, 65 - 128\}$. However, in the model we assumed minimum and maximum degree (lifetime number of partnerships) as 2 and 128, respectively, as the probabilities outside of this were low. However, we kept the computational structure to include degree 0 and 1 to keep it generalized and flexible to changes, or for modeling other networks such as needle sharing. The computational structure is also flexible to change the maximum degree. In the methods in Appendix S2.3, degree of 0 and 1 could also be included in parameter estimations (see Appendix S3.5).

3.4 Risk-group categorization

The distribution of persons by HIV transmission-group category are presented in Table S3. We only modeled the first three groups, i.e., only simulated sexual transmissions of HIV, and excluded persons who inject drugs (PWID). Data initialization was specific to these categorizations, where available (Table S2). Sexual behavioral data (Tables S3 to S10) were obtained from surveys of sexual behavior, which may not necessarily have excluded PWID. However, as PWID are a much smaller fraction of the population, we believe it would not have an influence on the overall distributions used here. Data for HIV care continuum distribution over time were also specific to the transmission-group. Our validation metrics for HIV similarly included transmission-group specific parameters, including distribution of new infections across the three groups, incidence divided by prevalence within each group, and distribution of infections by age-group within each transmission-group. Model outputs matched surveillance data in most cases (2).

3.5 Data for ECNA module methods: heterosexuals

The National Survey for Family Growth (NSFG), that surveys sexual behavior among persons aged 15–44 years, presents survey results for the distribution of men (and women) by the reported number of female (and male) sexual partners they have had to this point (age-group) in their lives (Table S5.1 for men and Table S5.2 for women). Note that this data represents $\pi_{\bar{a}}[x]$ (in the Appendix S2.3.2 method), i.e., the proportion of people (among those in age-group $\bar{a}$) with reported number of partners (upto age $\bar{a}$) in degree-bin x . Therefore, we can directly use this data in Appendix S2.3.2 method to infer both the distribution of lifetime number of partners, and for each degree-bin, infer the proportion of partnerships initiating in each age-group. However, that would restrict us to the reported degree-bin $[0, 1, 2, 3-6, 7-14, 15+]$. Therefore, we converted this data to probabilities in degree-bin of interest ($[0, 1, 2, 3-4, 5-8, 9-16, 17-32, 33-64, 65-128]$) as follows. By assuming that the number of partners up to age-group $\bar{a}$, $\forall \bar{a}$, also follows a power-law distribution, we estimated the power-law exponent $\lambda_{\bar{a}}$ for each age-group $\bar{a}$ by fitting to the data in each age-group, and using $\lambda_{\bar{a}}$, determined the $Pr_{\bar{a}}(k) \forall k \in \{0, 1, 2, 3-4, 5-8, 9-16, 17-32, 33-64, 65-128\}$. Note here that we included degree 0 and 1 because for younger age-groups, $Pr_{\bar{a}}(k=0)$ and $Pr_{\bar{a}}(k=1)$ are not small values. We then applied the method in Appendix S2.3.2. Results for the distribution of persons by degree-bin. i.e., $Pr(\bar{d}) \forall \bar{d} \in \{0, 1, 2, 3-4, 5-8, 9-16, 17-32, 33-64, 65-128\}$ are presented in Table S6. The power-

law exponent corresponding to this distribution is used in the simulation as input to the neural network model (Appendix S2.3.1) to predict neighbor's degree.

The proportion of partnerships initiated by age-group $\bar{a}$ for persons in degree-bin $\bar{d}$, i.e., $\sum_{i=1}^{\bar{a}} \bar{L}[i, \bar{d}]$ for each $\bar{d}$, estimated using the method in Appendix S2.3.2, are presented in Tables S7.1 and S7.2 for heterosexual male and female transmission-groups, respectively. These data along with data for partnership age-mixing (Table S10) were used in the simulation as inputs to the method in Appendix S2.3.3, and was applied to every newly infected node, to infer current age of each partner and the age-group of both persons at time of their partnership activation.

3.6 Data for ECNA module methods: MSM

Similar steps as heterosexuals was applied to MSM, expect that the data set used were different. There are no national surveys for MSM that report the distribution of MSM by number of partnerships upto age-group (as that for heterosexuals in Tables S5.1 and S5.2). However, there are smaller surveys, that only report the median and the range of the partners up to the persons current age-group, for both heterosexuals and MSM (Table S8). By assuming that the number of partners up to any given age also follow a power-law distribution, we estimated $\lambda_{\bar{a}}$ specific to each age-group $\bar{a}$ by applying the equation for the median of the power-law distribution, as

$median = m_{\bar{a}} 2^{\frac{1}{\lambda_{\bar{a}} - 1}}$, where $m_{\bar{a}} = 1$ is the minimum degree in age-group $\bar{a}$, and using the data from the smaller surveys (Table S8) as the *median*. With the estimation of these exponents ($\lambda_{\bar{a}}, \forall \{\bar{a}\}$), the same steps as heterosexuals were followed. Corresponding results for the distribution of persons by degree-bin are presented in Table S6. The proportion of partnerships initiated by age-group $\bar{a}$ for persons in degree-bin $\bar{d}$, are presented in Tables S9. Data for partnership age-mixing are presented in Table S10.

Note that the data in Table S8 were based on a small survey compared to the nationally representative NSFG survey used for heterosexuals. Therefore, to compare the differences in the two data sources, and thus the influence of minimal data for MSM, we estimated the proportion of partnerships initiated by age-group $\bar{a}$ for heterosexual male and female using the data from Table S8 and compared it with that estimated from the NSFG survey (Table S5), the comparisons are presented in Figure S1. The youngest age-groups had the most difference, which though improved

with older age-groups, continued to have some differences. The data were closer for heterosexual male than female. Thus, as more data become available these estimations should be updated.

3.7 Additional age-group assumptions

Note that the use of age-group 13-17, 18-24, 25-29, 30-34, 35-39, 40-44, and 45-65 for $\bar{L}[\bar{a}, \bar{d}]$, though the original source starts from 18-24 for MSM (Table S7) and ends at 40-44 in both sources, MSM (Table S7) and heterosexuals (Table S4), was to keep consistent with the age mixing matrix (M) (presented in Table S9), necessary for the methods in Appendix S2.3.3. To do so, we assumed half of partnerships in 18-24 initiate in age group 13-17. We also assumed that the number of partnerships initiating in age group 45-64 is the same as the number initiating in age-group 40-44. Note that, the use of compartmental age-structure was driven by this data, however, as noted in S3.1, it could be modified based on model application and corresponding data.

4 Model initialization and dry run

We can initialize the model to be representative of people in a population in a specific year. For the analyses in the main simulation, we initialized age, transmission-group, degree distributions as per persons in the U.S in 2006 in both compartmental model and network. For distribution by Disease 2 health status in the compartmental model and network, we randomly selected 10% of persons in each category as infected with Disease 2, and dry running the simulations so that the state distributions reach a steady state. We initialized the network for Disease 1 to be representative of HIV in the US in the year 2006 through using two dry runs to ensure network dynamics are generated in addition to epidemic and demographic distributions.

Dry run is a technique used for initialization of the model. It involves running the simulation for a certain period, but the data generated over that period of run is not representative of an actual epidemic projection and thus referred to as a dry run. For a pure agent-based model (without networks), we could initialize the model with a few agents and assign parameters to match surveillance data for the demographic, disease stage, and care-continuum stage distributions. But the agents would be lacking the ‘history’, e.g., the age at infection, and the age and stage of ART initiation, relevant for modeling future events. Therefore, we can do a dry run, which means starting with some number of people, assigning them data according to a specific year, say 2006

surveillance distributions as done here for HIV, running the simulation for several years while maintaining the distributions to match that of 2006. As persons become newly infected, their history is being generated, and the initial persons from day 0 age-out of the model. The above method is also sufficient for diseases modeled in the compartmental model (Disease 2 here). However, for diseases modeled in ABENM (Disease 1 here), the network adds another layer of complexity as static data is infeasible (and moreover, rarely available) for determining the contact network structure, including the links of HIV-infected persons to each of their lifetime partners, the current age of partners, the initiation and termination times of the links, the initiation and termination age of nodes at both ends of a link, the transmission-group of the partner, and the infection status of partner. Therefore, for Disease 1, we first do a dry run to allow for the network to dynamically grow over time (Dry run #1), and then apply the data from surveillance to set the demographic, and disease and care-continuum stages (Dry run #2). Dry runs are explained in more detail in (2) as applied to HIV and validated against NHSS, a similar method can be applied for other diseases modeled.

5 Overview of simulation modeling steps

We provide an overview of the full simulation model in this section. All notations used in the model are also summarized in Table S1.

Step 1: Initialization of the compartmental model array $S_{t=0}$ and graph $G_{t=0}(\mathcal{N}, \mathcal{E})$ at time-step $t = 0$

Step 1a. We initialize the compartmental model array to be representative of the United States population by age, transmission-group, and degree distribution in 2006. We distributed the total population in the U.S. into degree bins by using life-time partner distribution data for MSM and heterosexuals [43]. Within each degree-bin, we distributed the population into seven age-groups, ranging from age 13 to 65, using US census data for distribution by age, and further by risk-group (heterosexual male, heterosexual female, or MSM) using population size estimates for MSM from [35,42].

Step 1b. We initiate a random graph of 1500 nodes and zero edges. For each node, we set their HIV status as newly infected, allocate a degree by randomly drawing from the overall degree distribution, and assign Disease 1-related care continuum and disease stages, age,

1794 and transmission-group through random draws from the corresponding distributions of
1795 HIV in the US in 2006.

1796 Step 2: Dry run- to make the network and disease state-distribution representative of the population
1797 of interest

1798 Dry run the model for a certain number of time-steps (here we chose 200 monthly time-steps) by
1799 running the following steps sequentially for every time-step.

1800 Step 2a. Run the ECNA module (Appendix S2.3) to generate the network of contacts for
1801 every new HIV-infected node in the network.

1802 Step 2b. Run the compartmental module (Appendix S2.1) to update Disease 2 parameters
1803 in the compartmental model and network .

1804 Step 2c. Run the transmission module (Appendix S2.2) to generate new Disease 1
1805 infections in the network.

1806 Step 2d. Run the disease progression module (Appendix S2.4) to update demographics and
1807 Disease 1 progression and care parameters for nodes in the network.

1808 Step 3: Main simulation run.

1809 Repeat Steps 2a to 2d, for the required number of months. For the analyses in the paper, we
1810 simulated years 2006 to 2017 in monthly time-steps.

1811

Table S1: Table of Notations

Notation	Description
t	Simulation time-step.
A	The number of age-groups.
R	The number of transmission-groups.
D	The number of degree bins.
G	The number of geographic jurisdictions (1 representing national in this analyses).
$\bar{a}; r; \bar{d}; g; h$	Used when referring to an age-group, transmission-group, degree-bin, geographic jurisdiction, and health status, respectively, in the compartmental model. We use a dash for age-group and degree-bin to indicate that they are grouped intervals in the compartmental model.
$a_{t,j}; r_j; d_j; \phi_j; h_{t,j}$	Used when referring to the age, transmission-group, degree, geographic jurisdiction, and health status, respectively, of node j in the network at time t . Notations with no t in subscript are variables whose values do not change over time.
$S_t[\bar{a}, r, \bar{d}, g, h]$	An array of size $A \times R \times D \times G$ representing the number of susceptible persons in the model, in age-group $\bar{a}$, transmission-group r , degree-bin $\bar{d}$, pseudo-geographic jurisdiction g , and health status, at time t .
$\mathcal{N}$	A set of nodes, each representing an infected person or a susceptible sexual partner.
$\mathcal{E}$	A set of edges representing sexual partnerships between nodes.
$G_t(\mathcal{N}, \mathcal{E})$	A dynamic graph with $\mathcal{N}$ a set of nodes and $\mathcal{E}$ a set of edges, at time t .
Q_t	The number of nodes in graph G , at time t .
$C_t[i, j]$	A static adjacency matrix of size $Q_t \times Q_t$, with static element $C_t[i, j] = 1$ if i and j are sexual partners anytime during their lifetime and $C_t[i, j] = 0$ otherwise.
$V_t[i, j]$	A dynamic adjacency matrix of size $Q_t \times Q_t$, with element $V_t[i, j] = 1$ if i and j are sexual partners during month t and $V_t[i, j] = 0$ otherwise.
$e = \{i, j\}$	An edge in graph G_t representing a sexual partnership between nodes i and j .
$\bar{t}(\{i, j\})$	The partnership initiation time; represents the simulation month for partnership initiation.
$\underline{t}(\{i, j\})$	The partnership termination time; represents the simulation month for partnership termination.
$\{\bar{a}_i, \bar{a}_j\}$	The age of nodes i and j at the time of their partnership initiation.
$\{\underline{a}_i, \underline{a}_j\}$	The age of nodes i and j at the time of their partnership termination.
$\bar{a}_{t,j}$	Age-group of node j at time t .
$a_{t,j}$	Age of node j at time t .
$\bar{d}_j$	Degree-bin corresponding to the number of lifetime partners of node j .
d_j	The actual number of lifetime sexual partners of node j .

$\hat{d}_{t,j}$	The number of lifetime sexual partners of person j who are already added as nodes in graph G at time t . For infected nodes $d_j = \hat{d}_{t,j}$ for susceptible nodes in G , $d_j \geq \hat{d}_{t,j}$
$L_{t,j}$	A partnership distribution matrix of size $A \times 2$, where $L_{t,j}[\bar{a}, 1]$ is the number of partnerships that initiate at age-group $\bar{a}$, and $L_{t,j}[\bar{a}, 2]$ is the number of partnerships that are yet to be assigned. For infected nodes, $L_{t,j}[\bar{a}, 2] = 0, \forall \bar{a}$
$\hbar_{t,j}$	Infection status of node j at time t . It is an array of size equal to the number of diseases modeled.
$m_{t,j}$	Deceased status of node j at time t .
r_j	Transmission-group of person j .
$p_{t,j}$	Infectiousness or risk of transmission per act for person j at time t .
ε	1-condom effectiveness.
$s_{t,j}$	The number of sex acts per month for person j at time t .
$c_{t,j}$	The proportion of acts condom protected of person j at time t .
$F^{-1}(u)$	The inverse Bernoulli distribution that takes values 1 with probability u and 0 with probability $1 - u$.
D_k	Random variable for degree of node k .
$\Pr(D_k = d_k D_l = d_l)$	Conditional probability distribution for D_k .
$\Pr(D_k = d_k)$	Marginal probability distribution for D_k .
m	Minimum degree of the network.
λ_{r_l}	Scale-free network parameter corresponding to the transmission-group of node l .
$\bar{L}[\bar{a}, \bar{d}]$	A matrix of size $A \times D$, representing the proportion of partnerships that initiate at age-group $\bar{a}$ for persons in degree-bin $\bar{d}$.
$X_{\bar{a}}$	Random variable representing the number of lifetime partners at age-group $\bar{a}$.
$P_{\bar{a}}$	The transition probability matrix of size $D \times D$ for age-group $\bar{a}$.
$\pi_{\bar{a}}$	The steady state distribution for lifetime-partners in age-group $\bar{a}$.
z	A vector of size $D^2 \times 1$, converted from matrix $P_{\bar{a}}$.
C	A matrix of size $D \times D^2$, recreated with values of $\pi_{\bar{a}}$.
B	A binary matrix of size $D \times D^2$.
b	A vector of ones of size $D \times 1$.
$M[i, j]$	An age-mixing matrix of size $A \times A$, which represents the probability that, given a person is in age-group i , his or her partner is in age-group j . Varies by transmission-group.
$P[i, j]$	A matrix of size $A \times A$, representing the expected number of contacts the newly infected node l should have with person in age-group i , when node l is in age-group j .
$N[\bar{a}, k]$	A binary matrix of size $A \times (d_l - \hat{d}_{t,l})$, which represents whether partner k would have a newly initiating partnership at age-group $\bar{a}$.

$R[\bar{a}, k]$	A binary matrix of size $A \times (d_l - \hat{d}_{t,l})$, which represents whether the partnership with partner k would occur when partner k is in age-group $\bar{a}$.
$n[i]$	A vector of size A , denoting the number of partnerships that should initiate when the partner is in age-group i .
$e_k[i]$	A vector of size A , representing node k to be eligible to initiate a partnership at age-group i .

Table S2: Data for initialization people with HIV (PWH) in 2006

	Heterosexual Female	Heterosexual Male	MSM	Source
Distribution of PWH in year 2006 by stage				(6–8)
Acute-unaware	0.16%	0.07%	0.43%	
NonAcute-unaware	5%	3%	14%	
NonAcute Aware- No care	9%	5%	24%	
NonAcute In care- No ART	4%	2%	10%	
NonAcute-On ART- No VLS	1%	1%	3%	
NonAcute-On ART-VLS	5%	2%	12%	
Total	24%	12%	64%	
Age distribution of PWH in year 2006	(same for heterosexuals and MSM)			(7,9)
13-14		0.20%		
15-19		0.90%		
20-24		3%		
25-29		6%		
30-34		9%		
35-39		15%		
40-44		21%		
45-49		19%		
50-54		13%		
55-59		7%		
60-64		3%		
>=65		3%		
Age distribution of new HIV infections in 2006				(7,9)
13-14	0.10%	0.10%	0.10%	
15-19	4%	4%	4%	
20-24	17%	17%	21%	
25-29	15%	15%	19%	
30-34	15%	15%	15%	
35-39	12%	12%	12%	
40-44	11%	11%	12%	
45-49	11%	11%	9%	
50-54	10%	10%	7%	
55-59	5%	5%	2%	
60-64	0%	0%	0%	
>=65	0%	0%	0%	
Distribution of CD4 cell count (cells/ μ L) at diagnosis for those aware of infection by year 2006(10–12)				
<50	10%	10%	10%	
50-200	14%	14%	14%	
200-500	66%	66%	51%	
>500	10%	10%	25%	

Table S3: Distribution of the U.S. population by transmission-category (13–15)

HIV Transmission Category	Percentage
Heterosexual female	47.25%
Heterosexual male	47.25%
Men who have sex with men (MSM)(men only and women)	2.33%
People who inject drugs (PWID)-female	1.20%
PWID- heterosexual male	1.80%
PWID-MSM	0.18%

Table S4: Behavioral parameters related to transmission

Age - group	13-14	15-17	18-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-70	Source
Number of sex acts per year ^a														(16-19)
HET Female	20-41	20-41	73-127	73-127	62-108	51-93	51-93	48-86	48-86	40-73	32-73	35-62	35-62	
HET male	30-60	30-60	68-119	68-119	63-110	59-104	59-104	52-95	39-95	36-73	36-73	24-67	24-67	
MSM	30-60	30-60	68-119	68-119	63-110	59-104	59-104	52-95	39-95	36-73	36-73	24-67	24-67	
Proportion of acts that are anal														(16-19)
HET Female	0.07	0.07	0.07	0.07	0.08	0.06	0.06	0.04	0.04	0.02	0.02	0.04	0.04	
HET Male	0.06	0.06	0.06	0.06	0.11	0.07	0.07	0.09	0.09	0.06	0.06	0.05	0.05	
MSM	1	1	1	1	1	1	1	1	1	1	1	1	1	
Proportion of acts condom protected (main partners) ^b														(20)
HET Female	0.58	0.58	0.39	0.39	0.27	0.18	0.18	0.14	0.14	0.11	0.11	0.09	0.09	
HET male	0.79	0.79	0.45	0.45	0.28	0.26	0.26	0.21	0.21	0.1	0.1	0.06	0.06	
MSM	0.79	0.79	0.45	0.45	0.28	0.26	0.26	0.21	0.21	0.1	0.1	0.06	0.06	
Proportion of acts condom protected (casual partners) (all risk groups)	0.57	0.57	0.57	0.57	0.54	0.54	0.53	0.53	0.53	0.52	0.52	0.52	0.52	(21)
Other parameters													Source	
Proportion of anal sex acts insertive (receptive) (among MSM with other MSM)									50% (50%)				Assumption	
Proportion of HIV-infected MSM who have sex with women (MSMW)									21%				(21-24)	
Proportion of partnerships with female for MSMW									80%				Calibrated in (5)	
Proportion of anal acts with female for MSMW									50%				Calibrated (5)	

^a Number of sex acts were estimated as the average of the reported number of sex acts weighted by the proportion reporting under each category of number of partners/sex acts among those sexually active. For MSM, we used the heterosexual male data as age-distributed data were not available for MSM. Moreover, the median of 1 partner for MSM (24) matched the heterosexual male data. Number of sex acts are uniformly distributed in the given range

^b We applied the heterosexual male data to MSM. These data are for the general population (heterosexual and MSM) unaware of their HIV status, and they apply to their main partners.

Note: All data in the table relate to probability distributions of the parameters and for each person random samples are drawn from these distributions as follows. If proportions are for true or false outcomes, we draw a random number $u \sim \text{Uniform float}[0,1]$, if $u \leq \text{proportion}$ then it is true else false, e.g., determining if MSM is MSMW. If proportions are for behavior of a specific individual then they are directly used as point estimates, e.g., among all sex acts among MSMW, 80% are assigned to women. If they are from probability distributions such as Uniform (e.g., sex acts), or Geometric (e.g., partnership duration), samples are drawn from this distribution.

Table S5.1: For each age group, the distribution of the number of sex partners accrued-to-date for men in that age group (each row adds to 1) (25)

Age Group of Males	Total in Age Group	Number of female sexual partners					
		0	1	2	3-6	7-14	15 or more
15-19	10,208	0.385	0.23	0.092	0.207	0.062	0.025
20-24	9,883	0.09	0.159	0.117	0.335	0.141	0.159
25-29	9,226	0.049	0.1	0.088	0.294	0.232	0.238
30-34	10,138	0.028	0.107	0.069	0.285	0.219	0.292
35-39	10,557	0.02	0.089	0.07	0.28	0.255	0.288
40-44	11,135	0.019	0.088	0.054	0.256	0.242	0.342

Table S5.2: For each age group, the distribution of the number of sex partners accrued-to-date for women in that age group (each row adds to 1) (25)

Age Group of Females	Total in Age Group	Number of male sexual partners					
		0	1	2	3-6	7-14	15 or more
15-19	9,834	0.378	0.272	0.09	0.191	0.05	0.019
20-24	9,840	0.089	0.246	0.13	0.322	0.144	0.069
25-29	9,249	0.025	0.225	0.117	0.313	0.201	0.119
30-34	10,272	0.019	0.205	0.094	0.388	0.18	0.113
35-39	10,853	0.011	0.202	0.112	0.358	0.205	0.112
40-44	11,512	0.014	0.204	0.105	0.374	0.191	0.112

Table S6: Scale-free degree distribution, stratified by risk group

Degree	0	1	2	3-4	5-8	9-16	17-32	33-64	65-128
HET-female	0	0	0.261	0.229	0.179	0.131	0.093	0.064	0.044
HET-male	0	0	0.211	0.202	0.175	0.143	0.113	0.088	0.068
MSM	0	0	0.190	0.189	0.171	0.146	0.122	0.100	0.081

Table S7.1: Within each degree-bin, the estimated proportion of lifetime partnerships that are initiated by the time a person leaves an age-group, for heterosexual men with number of lifetime partnerships in that degree-bin

Degree-bin (degree range)-->	1 (2)	2 (3-4)	3 (5-8)	4 (9-16)	5 (17-32)	6 (33-64)	7 (65-128)
Age-group							
13-17	0.6	0.38	0.22	0.12	0.06	0.03	0.02
18-24	0.67	0.57	0.48	0.39	0.29	0.23	0.17
25-29	0.75	0.7	0.65	0.59	0.49	0.44	0.37
30-34	0.84	0.82	0.79	0.76	0.69	0.66	0.61
35-39	0.84	0.82	0.79	0.76	0.69	0.66	0.61
40-44	0.92	0.91	0.9	0.88	0.85	0.83	0.81
45-65	1	1	1	1	1	1	1

Table S7.2: Within each degree-bin, the estimated proportion of lifetime partnerships that are initiated by the time a person leaves an age-group, for heterosexual women with number of lifetime partnerships in that degree-bin

Degree-bin (degree range)-->	1 (2)	2 (3-4)	3 (5-8)	4 (9-16)	5 (17-32)	6 (33-64)	7 (65-128)
Age-group							
13-17	0.81	0.47	0.26	0.14	0.07	0.04	0.02
18-24	0.89	0.72	0.55	0.45	0.32	0.26	0.26
25-29	1.00	0.98	0.97	0.98	0.96	0.93	0.89
30-34	1.00	0.98	0.97	0.98	0.96	0.95	0.94
35-39	1	1	1	1	1	1	1
40-44	1	1	1	1	1	1	1
45-65	1	1	1	1	1	1	1

Table S8: Median and range of lifetime number of partners accrued-to-date in that age group by transmission-group (26)

Age Group	MSM		Heterosexual men		Heterosexual women	
	Median	Range	Median	Range	Median	Range
18-24	15	(1–3100)	4	(1–99)	4	(1–25)
25-29	30	(1–2562)	8	(1–99)	6	(1–60)
30-34	55	(1–7000)	9	(1–99)	7	(1–40)
35-39	67	(0–9005)	12	(1–99)	7	(1–99)

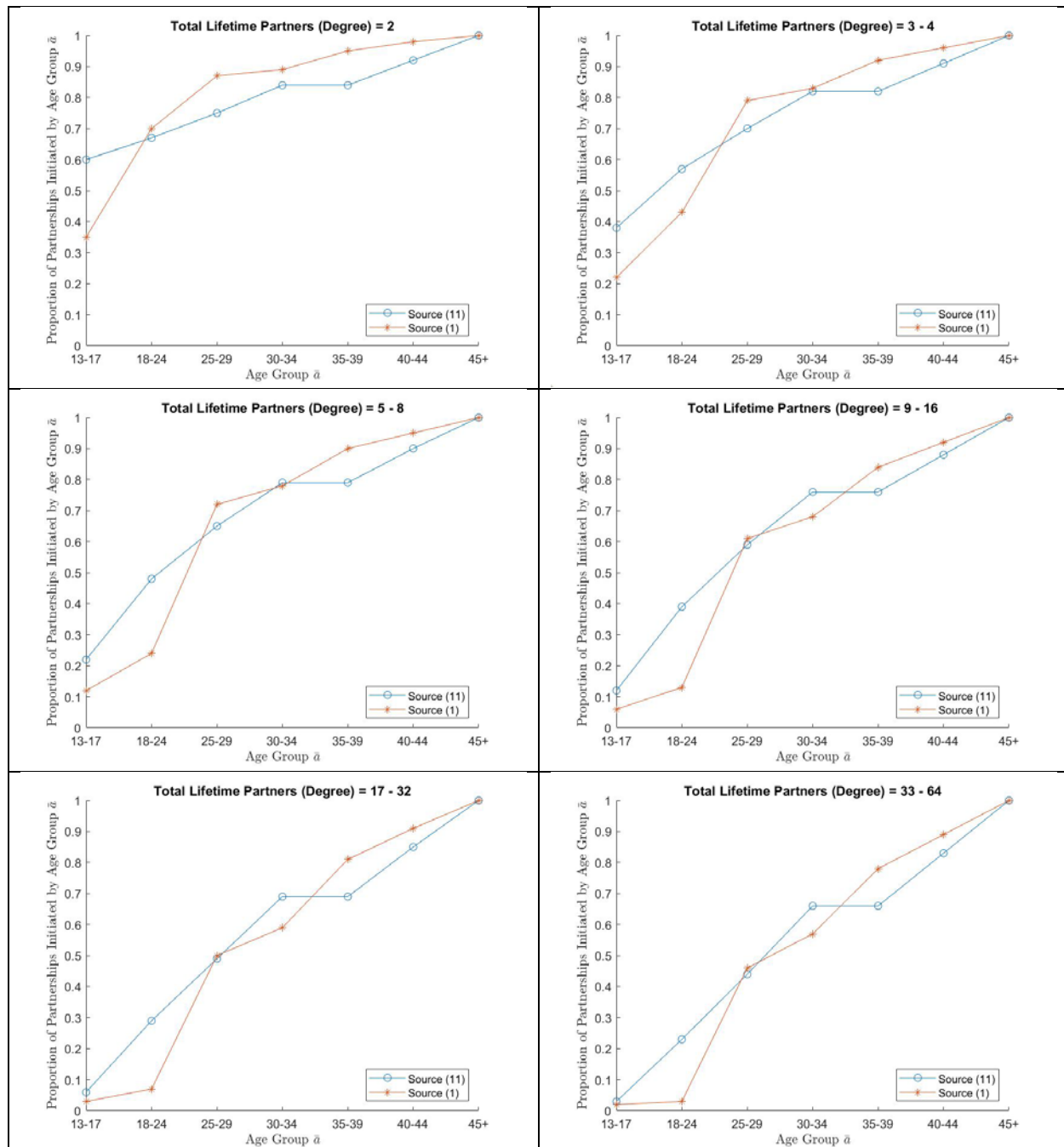
Table S9: Estimated proportion of partnerships that initiated by age-group for persons with lifetime partners in degree-bin for MSM

Degree-bin (degree) -->	1 (2)	2 (3-4)	3 (5-8)	4 (9-16)	5 (17-32)	6 (33-64)	7 (65-128)
Age-group							
13-17	0.35	0.21	0.12	0.06	0.03	0.01	0.01
18-24	0.71	0.42	0.24	0.12	0.06	0.03	0.02
25-29	0.83	0.73	0.61	0.52	0.43	0.35	0.27
30-34	0.92	0.90	0.82	0.79	0.76	0.70	0.64
35-39	0.96	0.94	0.89	0.87	0.85	0.80	0.75
40-44	0.98	0.97	0.94	0.93	0.93	0.90	0.88
45-65	1	1	1	1	1	1	1

Table S10: Age mixing: proportion of partners within own age-group (26)

Age Group	Heterosexual female	Heterosexual male	Men who have sex with men
13–17	91.1	91.05	91.1
18–24	90.0	92.1	48
25–29	57.7	82	55.9
30–34	54.5	54.5	46.3
35–39	81.8	76.2	55.2
40–44	81.8	76.2	55.2
45–65	81.8	76.2	55.2

Figure S1.1: Comparing the proportion of partnerships initiated by age-group estimated using data from source 1 (presented in Table S8) with those estimated using data from source 11 (presented in Tables S5.1 and S5.2), for heterosexual men.



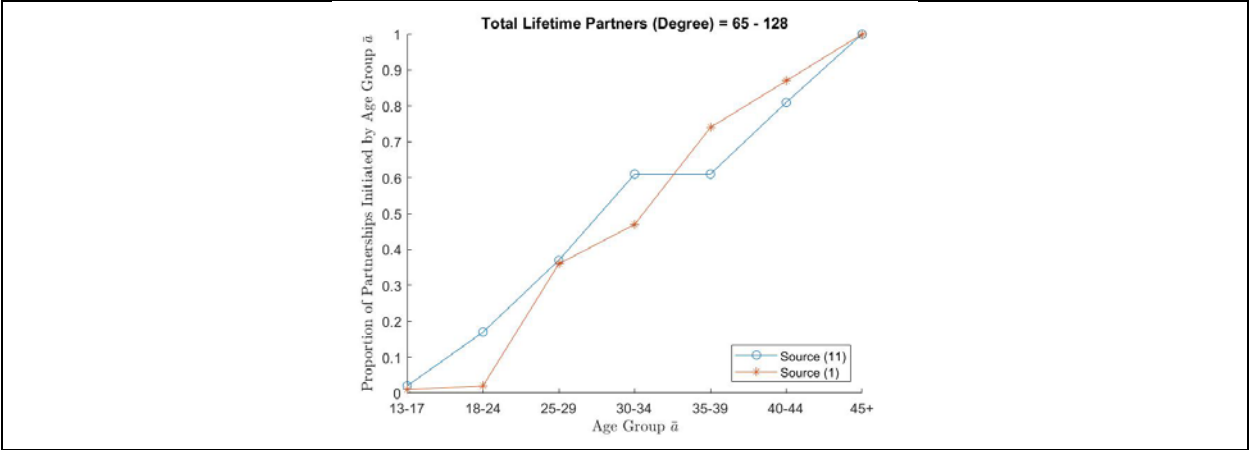
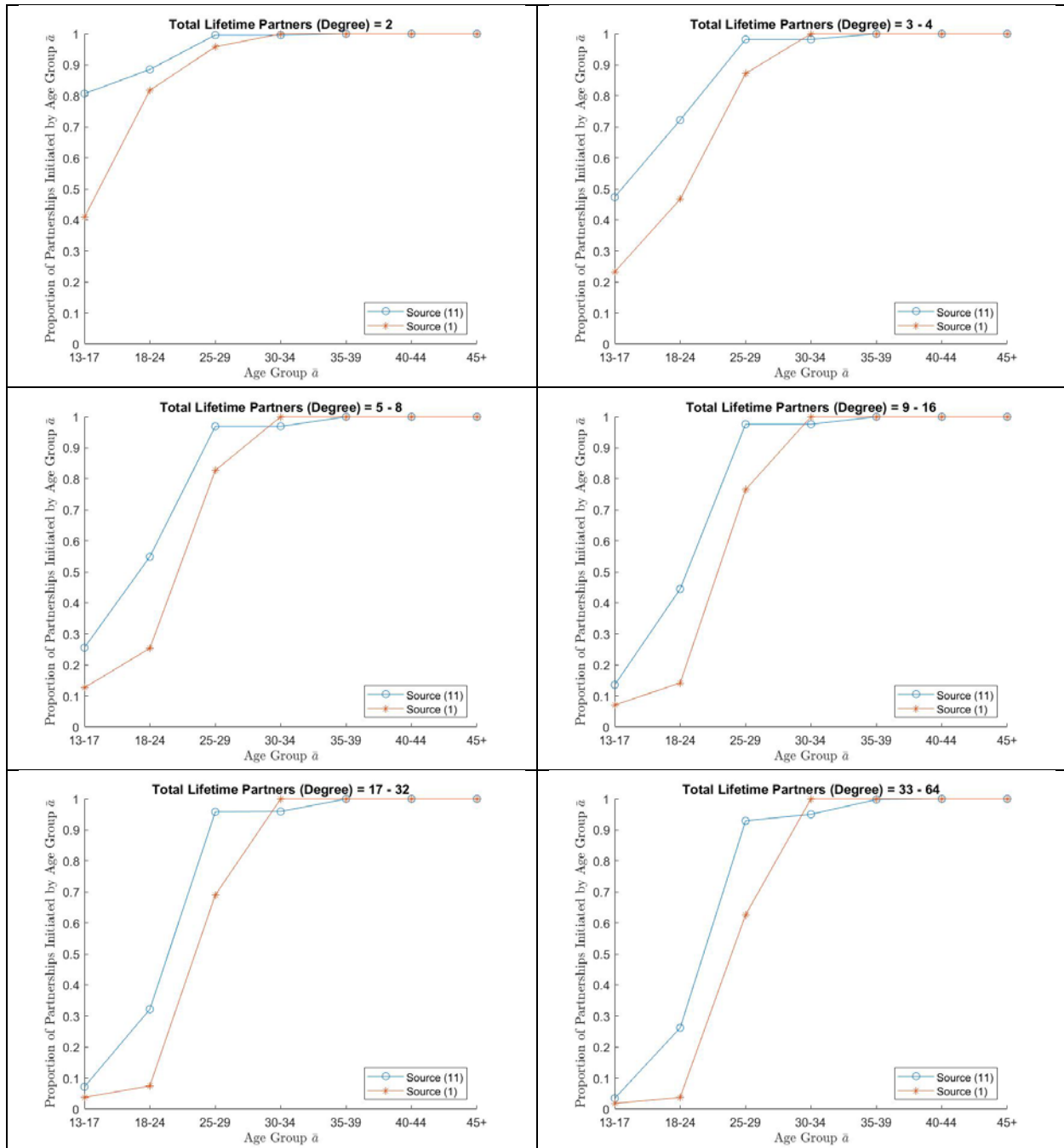
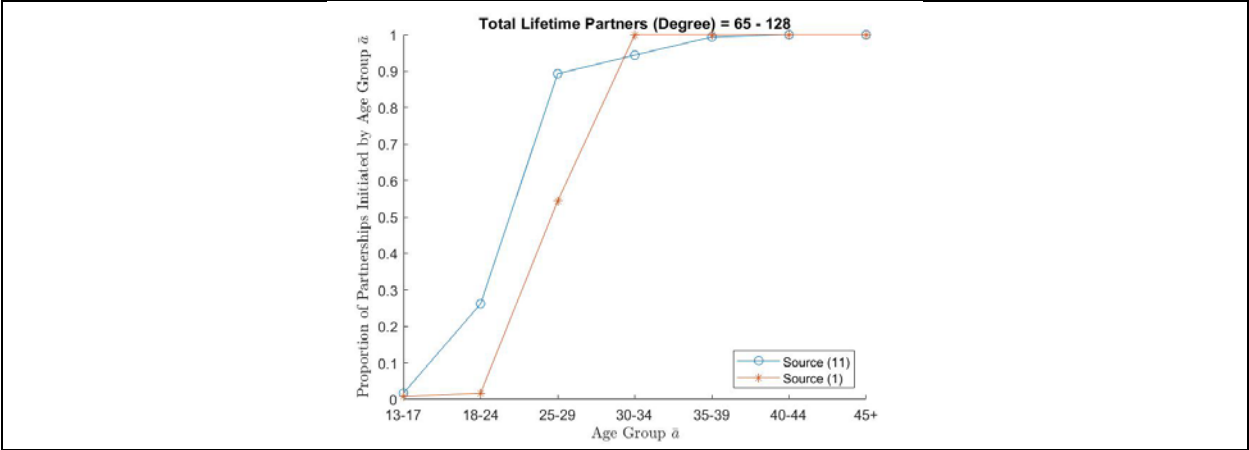


Figure S1.2: Comparing the proportion of partnerships initiated by age-group estimated using data from source 1 (presented in Table S8) with those estimated using data from source 11 (presented in Tables S5.1 and S5.2), for heterosexual female.





Bibliography

1. Eden M, Castonguay R, Munkhbat B, Balasubramanian H, Gopalappa C. Agent-based evolving network modeling: a new simulation method for modeling low prevalence infectious diseases. *Health Care Manag Sci*. 2021 Sep;24(3):623–39.
2. Singh S, France AM, Chen YH, Farnham PG, Oster AM, Gopalappa C. Progression and transmission of HIV (PATH 4.0)-A new agent-based evolving network simulation for modeling HIV transmission clusters. *Math Biosci Eng MBE*. 2021 Mar 3;18(3):2150–81.
3. Barabasi AL. Network Science- The Scale-free Property. *Creat Commons CC -NC-SA 20 PDF V53 09092014*. (Journal Article).
4. Fotouhi B, Rabbat MG. Degree Correlation in Scale-Free Graphs. *Eur Phys J B*. 2013;86(510).
5. Gopalappa C, Farnham PG, Chen YH, Sansom SL. Progression and Transmission of HIV/AIDS (PATH 2.0). *Med Decis Making*. 2017;37(2):224–33.
6. Campsmith ML, Rhodes PH, Hall HI, Green TA. Undiagnosed HIV prevalence among adults and adolescents in the United States at the end of 2006. *J Acquir Immune Defic Syndr*. 2010;53(5):619–24.
7. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, et al. Estimated HIV incidence in the United States, 2006-2009. *PLoS One*. 2011;6(8):e17502.
8. Gardner EM, McLees MP, Steiner JF, Del Rio C, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin Infect Dis*. 2011;52(6):793–800.
9. Centers for Disease C, Prevention. HIV surveillance--United States, 1981-2008. *MMWR Morb Mortal Rep*. 2011;60(21):689–93.
10. Hall HI, McDavid K, Ling Q, Sloggett A. Determinants of progression to AIDS or death after HIV diagnosis, United States, 1996 to 2001. *Ann Epidemiol*. 2006;16(11):824–33.
11. Buchacz K, Armon C, Palella FJ, Baker RK, Tedaldi E, Durham MD, et al. CD4 cell counts at HIV diagnosis among HIV Outpatient Study participants, 2000-2009. *AIDS Res Treat*. 2012;2012:869841.
12. Althoff KN, Gange SJ, Klein MB, Brooks JT, Hogg RS, Bosch RJ, et al. Late presentation for human immunodeficiency virus care in the United States and Canada. *Clin Infect Dis*. 2010;50(11):1512–20.
13. Chandra A, Mosher WD, Copen C, Sionean C. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006-2008 National Survey of Family Growth. *Natl Health Stat Rep*. 2011;2011/05/13(36):1–36.

14. Lansky A, Finlayson T, Johnson C, Holtzman D, Wejnert C, Mitsch A, et al. Estimating the number of persons who inject drugs in the United States by meta-analysis to calculate national rates of HIV and hepatitis C virus infections. *PloS One*. 2014;9(5):e97596.
15. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, et al. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. *Open AIDS J*. 2012;6(Journal Article):98–107.
16. Herbenick D, Reece M, Schick V, Sanders SA, Dodge B, Fortenberry JD. Sexual behavior in the United States: results from a national probability sample of men and women ages 14-94. *J Sex Med*. 2010;7 Suppl 5(Journal Article):255–65.
17. Reece M, Herbenick D, Schick V, Sanders SA, Dodge B, Fortenberry JD. Sexual behaviors, relationships, and perceived health among adult men in the United States: Results from a national probability sample. *J Sex Med*. 2010;7(Journal Article):291–304.
18. Herbenick D, Reece M, Schick V, Sanders SA, Dodge B, Fortenberry JD. Sexual behaviors, relationships, and perceived health status among adult women in the United States: Results from a national probability sample. *J Sex Med*. 2010;7(Journal Article):277–90.
19. Reece M, Herbenick D, Schick V, Sanders SA, Dodge B, Fortenberry JD. Condom use rates in a national probability sample of males and females ages 14 to 94 in the United States. *J Sex Med*. 2010;7(Journal Article):266–76.
20. Rosenberger JG, Reece M, Schick V, Herbenick D, Novak DS, Van Der Pol B, et al. Condom Use during Most Recent Anal Intercourse Event among a U.S. Sample of Men Who Have Sex with Men. *J Sex Med*. 2012 Apr;9(4):1037–47.
21. Finlayson TJ, Le B, Smith A, Bowles K, Cribbin M, Miles I, et al. HIV risk, prevention, and testing behaviors among men who have sex with men--National HIV Behavioral Surveillance System, 21 U.S. cities, United States, 2008. *MMWR Surveill Summ Morb Mortal Wkly Rep Surveill Summ CDC*. 2011;60(14):1–34.
22. Voetsch AC, Lansky A, Drake AJ, MacKellar D, Bingham TA, Oster AM, et al. Comparison of demographic and behavioral characteristics of men who have sex with men by enrollment venue type in the National HIV Behavioral Surveillance System. *Sex Transm Dis*. 2012;39(3):229–35.
23. Friedman MR, Wei C, Klem ML, Silvestre AJ, Markovic N, Stall R. HIV infection and sexual risk among men who have sex with men and women (MSMW): a systematic review and meta-analysis. *PloS One*. 2014;9(1):e87139.
24. Rosenberg ES, Sullivan PS, Dinunno EA, Salazar LF, Sanchez TH. Number of casual male sexual partners and associated factors among men who have sex with men: results from the National HIV Behavioral Surveillance system. *Bmc Public Health*. 2011;11(Journal Article):189.

25. Mosher WD, Chandra A, Jones J. Sexual behavior and selected health measures: men and women 15-44 years of age, United States, 2002. *Adv Data*. 2005 Sep 15;(362):1–55.
26. Glick SN, Morris M, Foxman B, Aral SO, Manhart LE, Holmes KK, et al. A comparison of sexual behavior patterns among men who have sex with men and heterosexual men and women. *J Acquir Immune Defic Syndr* 1999. 2012 May 1;60(1):83–90.