MBE, 18(3): 2150–2181.  
DOI: 10.3934/mbe.2021109  
Received: 23 December 2021  
Accepted: 18 February 2021  
Published: 03 March 2021

http://www.aimspress.com/journal/MBE

***Research article***

**Progression and Transmission of HIV (PATH 4.0) – A new Agent-based Evolving Network Simulation for Modeling HIV Transmission Clusters**

**Sonza Singh1, Anne Marie France2, Yao-Hsuan Chen2, Paul G. Farnham2 Alexandra M. Oster2, and Chaitra Gopalappa1,\***

**1**  University of Massachusetts Amherst, Amherst, MA, United States

**2** Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

**\* Correspondence:** E-mail: chaitrag@umass.edu; Tel: +1-413-545-2306.

**S1 Overview of PATH 4.0**

The Progression and Transmission of HIV (PATH 4.0) is a stochastic dynamic simulation of HIV, constructed using a newly developed agent-based evolving network modeling technique. At every time-step (monthly) of the simulation, the model runs four different modules, a compartmental module for simulating susceptible persons, a Bernoulli transmission module for simulating new infections, a HIV-ECNA (evolving contact network algorithm) network generation module for generating sexual partnership networks of newly infected persons, and a disease progression module for simulating HIV-related events for HIV-infected persons. We discuss each module below. We also discuss implementation of PATH 4.0 for simulating HIV in the US for years 2006 to 2017.

**S2 Compartmental module for simulating susceptible population**

As discussed in the main manuscript, a compartmental modeling structure is used to keep track of the number of susceptible persons in the US, who are not immediate contacts of the infected. It is specifically denoted by an array ( of dimension where is the number of age-groups, is the number of risk groups, and is the number of degree bins, and is the number of pseudo-geographic jurisdictions. As degree i.e., the number of lifetime partners, follows a power-law distribution where a large number of nodes have low degree and a few nodes have high degree, as commonly done, we used log 2-binning for degree instead of uniform-binning, i.e., all persons with degree , are grouped into one group. We assumed minimum degree as 2 and maximum degree as 128, and thus, the endpoints of degree bin ranged from to . We distributed the total population in the US into degree-bins using the probability density function , with , , and λ estimated from lifetime partner distribution data for MSM and heterosexuals[1]. Within each degree bin, we distributed the population across seven age groups, ranging from age 13 to 65, using US census data for distribution by age, and further by risk group. We assumed 2.27% of the US population are MSM [2], 49.03% are heterosexual female, and the remaining 48.70% heterosexual male.

*S2.1 Births*

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. There is a constant number of births per year, calculated for each risk group using the overall birth rate in the US in 2015 and the population in the respective age group [3].

*S2.2 Deaths*

Susceptible individuals can leave the simulation model in two different ways: 1) death by non-HIV/AIDS causes, and 2) ageing out of the system. Deaths by non-HIV/AIDS causes was modeled using the all-cause mortality in the general population (see Table S1). As less than 2% of new diagnoses occur among individuals aged 65 and above, we assume the maximum age for sexual transmissions of HIV is 65 years and ageing out of the population occurs when an individual turns 65 years of age [4].

*S2.3 Aging*

Individuals move from one age group to the next at a constant rate equal to the inverse of the width of the age-group interval. Thus, the number of people moving from one age group to the next every year is given by the following equation.

*S2.4 Age distribution*

We determined the initial age distribution of the susceptible population based on the age distribution of all PWH. As we only consider ages 13 to 65 in our model, the percentages do not sum to 100. Once we determine the age distribution of the infected population, we distribute the remaining susceptible population into age-groups so the total number of people in a specific age-group matches the US census data for that age group, presented in Tables S2.1 and S2.2.

*S2.5 Risk group and geographic distribution*

The risk group distribution of persons in the US are presented in Table S3. We split the population equally into 50 pseudo-geographic jurisdictions to represent geographic heterogeneity in contact mixing, as persons more likely form partnerships with other persons in the same geographical jurisdiction [5]. However, we did not specifically model any geographical features related to HIV-infected persons, as the scope of this work was national level estimations. Future versions of the model could expand this to consider more specific geographical data.

**S3 Transmission module for simulating new infections**

*S3.1 Bernoulli transmission model*

This section is the same as the main manuscript, but present here for completeness. Every time-step (monthly) of the simulation, this module determines if a susceptible node in graph becomes infected using a Bernoulli transmission equation, specifically, for nodes with HIV infection status it estimates its updated value as follows.

, where,

, where are the elements of the graph described in Section 2.1, and is the probability of transmission per act modeled as a function of disease and care stage and risk group of the infected node ; we will have a value of if is a contact of (i.e., , is infected (i.e., ), and is alive (i.e., ), and otherwise,

= condom effectiveness,

= number of sex acts per month with node , modeled as a function of age, risk group, and number of partners of node ,

= proportion of acts with node that is condom protected, modeled as a function of age, risk group, and number of partners of node ,

an inverse Bernoulli distribution that takes a value of 1 with probability and value of 0 with probability .

If node becomes infected, i.e., if the above equation yields , we set its HIV stage .

Every time-step , this module also determines and updates any changes in sexual behavior of infected nodes. Specifically, it updates the following values. For every partnership , its active/inactive status as . For every infected node , the number of sex acts per month as a random draw from age-group and risk-group specific uniform distribution, corresponding to age and risk group of node . For every infected node , the number of sex acts per partner . For every infected node , condom use as a function of number of active partners () and disease stage (specifically, diagnosed status of HIV, with higher condom use if aware, i.e., ). For every infected node , the infectiousness as a function of risk group and stage of node . Data assumptions for the above behavioral parameters are presented in Table S4.

**S4 HIV-ECNA network generation module for simulating sexual partnerships**

*S4.1 Neural network for calculating degree correlation*

As discussed in the main paper, we apply a newly developed evolving contact network algorithm (ECNA) to generate a scale-free contact network for newly infected persons, where the distribution of the number of contacts per person follows a power-law distribution [6]. We assumed that the contact network of sexual partnerships follows a scale-free network [7–9]. Studies that derived network statistics from molecular detection of clusters among persons infected with HIV and in care, concluded that while power-law distributions are a good approximation, Waring’s distribution, which is also a scale-free network, provided the best fit. Waring’s distribution approximates to a power-law distribution when the maximum degree is sufficiently large [10], which is true in our case as we model the lifetime number of partners. The contacts to be generated by the ECNA represent sexual partners of the infected node, and all partners the person would have over his/her lifetime are added. A key aspect of this algorithm is determining who each susceptible partner is, including the degree (the number of lifetime partners) of the susceptible partner, which is correlated with the degree of the infected person. Our previous work [11] showed that this degree correlation is also influenced by the stochastic process of HIV transmissions, and developed a neural network prediction model, a machine learning model, for determining the degree of a susceptible partner conditional on the degree of the newly infected node. 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. Eden et al. [11] further showed that degree correlations were not influenced by variations in the probability of transmission but significantly influenced by the prevalence of disease and scale-free network properties. Therefore, in training the neural network with the conditional probability as the response variable, the following five inputs were set as the independent variables:

* , the degree-bin of the newly infected agent ,
* , the degree-bin of the susceptible neighbor of the newly infected agent
* the minimum degree of the network ,
* the percent of the population that is infected (to account for the changes in epidemic paths over time), and
* the size of the full network.

Specifically, by simulating hypothetical diseases on scale-free networks, the numerical data for conditional probabilities was generated as , where, was a counter in the simulation that kept track of the total number of susceptible contacts with degree for newly infected persons with for every combination

The neural network was trained, in the R software, on data from 15 different scale-free networks. The data set included data from networks of size (𝑁) 1000, 5000, and 10000, and minimum network degree (𝑚) of 1, 2, 3, 4, and 5, which are the inputs for generating the scale-free networks in R using barabasi.game [12], and percent infected at 5%, 13%, 25%, 35%, 50%, 60%, and 90%. The neural net (NN) had one hidden layer and the number of hidden nodes was a hyperparameter.

Tuning of the hyperparameter and validation of the NN was conducted as follows. The scale-free networks were split into test and train networks, all networks except {𝑁 = 1000, 𝑚 = 1}, {𝑁 = 5000, 𝑚 = 2}, and {𝑁 = 10000, 𝑚 = 4}, were set as train networks. Train networks were further split into 60% and 40% train and test data, respectively, through random selection. Only train data of train networks were used in NN prediction. The hyperparameter was iteratively set to values between 5 and 14 and, under each value, the corresponding mean square error (MSE) between predicted and actual were estimated for train networks (for both train and test data) and test networks (all data in this are test data). While the MSE decreased in the train data of train networks as the number of hidden nodes increased, on the test data and test networks, MSE first decreased and then started to increase after 8 hidden nodes [11]. Therefore, we set the NN hyperparameter value at 8 hidden nodes. Comparison of neural network predictions with actual estimates on the test data sets are presented in the Appendix of Eden et. al. [11].

In this work, instead of using the size of the network (𝑁) as an independent variable of the NN, to keep it generalized for application to any network, we recoded the independent variable to use the scale-free network power-law degree distribution decay-coefficient . Specifically, as the scale-free network generation package in R (barabasi.game) [12] uses and 𝑁 as inputs, we calculated from the corresponding networks generated. For a scale-free network following a power law degree distribution, the probability that a node has degree (represented as) can be written as

where, the decay-coefficient , = number of nodes with degree , is the gradient, is the minimum degree, and is the maximum degree (network size influences maximum degree and thus served as a proxy in the previous application).

Empirical fits to sexual and injecting drug use networks have generated values of λ between 2 to 3 [13,14] for annual number of sexual partners. We set the minimum and maximum degree as 2 and 128, respectively, and estimated the value of λ for MSM and heterosexuals by using the distribution of number of lifetime partners, described in section S4.2 [1]. Upon estimation of using the neural network model, the set of conditional probabilities for each was normalized to 1 for use in the PATH model. The scale-free degree distribution used, stratified by risk group, is presented in Table S5.

*S4.2 Methods for estimation of lifetime number of partners*

As discussed in the main manuscript, for every infected node, the model keeps track of all partners over its lifetime, including the initiation and termination age for every partnership. If a partnership is initiated and not yet terminated, the susceptible partner is at risk of infection. If the partnership is not yet initiated, or has been terminated, the risk of infection is zero in the given time-step. As discussed in the main manuscript, the initiation and termination age can be calculated from matrix , where represents the proportion of partnerships that initiate at age-group for persons in degree-bin , each column of adding to 1, for all . Direct data for are unavailable. However, data for the distribution of persons by number of partners up to that age for persons at different age-groups are available through survey studies (see Tables S6.1, S6.2). We developed a two-step Markov process and simulation method for estimation of using the survey data, which we present next.

S4.2.1 Markov process and simulation method for estimation of partnership distribution matrix

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

Let be the degree-bin corresponding to the number of partners a node has up to the age of age-group, . Then, is a discrete-time Markov chain with state space

,

where each discrete timestep corresponds to the length of time equal to the width of age-group , is the number of degree-bins, is the number of age-groups, is a block matrix with {} on the superdiagonal and all other elements equal to zero[[1]](#footnote-1), and is a upper triangular matrix, with equal to the probability that a person transitions from state in age-group to state in age-group (i.e., goes from degree-bin to degree-bin 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 , and defining a row vector with equal to the proportion of people (among those in age-group ) with number of partners up to age in degree-bin , we can write

.

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

Data for are available from the behavioral survey studies (see Tables S6.1, S6.2). We solved for the values in , for each , by formulating and solving a linear least-squares optimization problem as follows.

For each , we formulated an optimization model as

Objective Function:

Subject to:

where the objective function is equivalent to , i.e., equivalent to minimizing the sum of squared errors between the left- and right-hand sides of the equation defined above ( is the -norm), and is obtained by reformulating as below:

is a column vector of length obtained by stacking the columns of ; i.e., ,

is a matrix where, for each row , and all other elements are equal to zero,

is the transpose of ,

is a matrix consisting of *D* side-by-side identity matrices, and

is a column vector of length , 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 ensures that each row of 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 , and thus obtained elements of .

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 () 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 estimated in Step 1. This is repeated until all persons reach the last age-group. Taking all persons with lifetime partners in degree-bin at the last age-group, we estimated as the average of the proportion of partnerships that initiated when that person was at age-group .

### S4.2.2 Data for determining the distribution of lifetime partners by age for heterosexual risk groups

The National Survey for Family Growth (NSFG), that surveys sexual behavior among persons aged 15–44 years [15], 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 S6.1 for men and Table S6.2 for women). The proportion of partnerships initiated by age-group for persons in degree-bin , i.e., for each estimated using the method in S4.2.1, are presented in Tables S7.1 and S7.2 for heterosexual male and female risk-groups, respectively. Note that, before applying the method in S4.2.1, the survey data from Tables S6.1 and 6.2 were modified to be distributed by the degree-bins used here (, by estimating the value of the power-law exponent for each age-group , based on the assumption that the number of partners up to age-group , also follows a power-law distribution, and by applying the probability density function to estimate the distribution by degree-bin.

### S4.2.3 Data for determining the distribution for lifetime partners by age for MSM risk group

There are no national surveys for MSM that report the distribution of MSM by number of partnerships, by age-group, as in Tables S6.1 and S6.2. 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 [1] (Table S8). By assuming that the number of partners up to any given age also follow a power-law distribution, we estimated specific to each age-group by applying the equation for the median of the power-law distribution, as

, where is the minimum degree in age-group , and using the data from the smaller surveys (Table S8) as the .

For each age-group , using the estimated , we calculated the distribution of people by degree-bin . Then, by applying the method described in S.4.2.1, we estimated the proportion of partnerships initiated by age-group , i.e., , for persons with number of lifetime partners in degree-bin , (presented in Table S9).

Note that the data in [1], used in the estimation of , for MSM, were based on a small survey compared to the nationally representative NSFG survey, used in the estimation of , for heterosexual male and female risk groups in S4.2.2. Therefore, to compare the differences in the two data sources, and thus the influence of minimal data for MSM, we estimated ,, for heterosexual male and female risk groups using the data from Table S8 [1] and compared it with that estimated from the NSFG survey, the comparisons are presented in Figure S3. The youngest age-groups had the most difference, which though improved with older age-groups, continued to have some differences. Thus, as more data become available these estimations should be updated.

### S4.2.4 Additional assumptions made

Note that the use of age-group 13-17, 18-24, 25-29, 30-34, 35-39, 40-44, and ≥45 for , though the original source starts from 18-24 for MSM (Table 8) and ends at 40-44 in both sources, MSM (Table 8) and heterosexuals (Table S6), was to keep consistent with the age mixing matrix ()(discussed in S4.3 and presented in Table S10), necessary for the methods in Section S4.3. 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, which was included into the estimation method in S4.2.1 in determination of and calculation of the proportions in .

### *S4.3 Optimization model for simulating partnership activation and deactivation times*

Determining partnership initiation and termination age and times ,, ,

We discuss estimation of partnership initiation age , termination age , intiation time , and termination time between nodes and .The optimal values are those that ensures age-mixing between partners is maintained, the distribution of partnership age-initiation for newly infected node (i.e., ) is maintained, and the distribution of partnership age-initiation for each of its partners is maintained. We can formulate this problem as an optimization model as follows. Let,

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

, where is a diagonal matrix with diagonal elements equal to those of the vector , be a matrix of size with element representing the number of partners of age in age-group yet to be assigned, when the newly infected node is in age-group ,

be a vector of size with , denoting the number of partnerships to initiate when the partner is in age-group ,

be a binary matrix of size , where,, is the degree of the newly-infected node and is the number of partners to newly add, with element , i.e., if partner is eligible to initiate the partnership at age-group , and

be a binary matrix of size , with element if the partnership with would occur when partner is in age-group .

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

*Objective Function:*

*Subject to:*

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 jobs (here partners) to 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 by setting (the cost of assigning job to machine for all combination), and putting a tight constraint that the maximum capacity of the machine should be met (). 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 , , i.e., the number of required partnerships at age-group 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 be a vector of size with , for every new partner . Then, if is eligible to initiate a partnership at age-group and there is a need for a partnership at that age-group. For each new partner , we search through all elements of by starting at the last element, we select the first occurrence of with as the solution, i.e., set , and update . This process of starting at the last element of leads to an optimal solution, provided a feasible solution exists, because of the following properties related to the distribution of lifetime partners by age-group of partnership initiation specific to the data in the US (Tables S6-S9).

Property 1: For any , if then but the opposite is not necessarily true.

Property 2: Let be a set of partners eligible to initiate partnership at age-group , i.e., if , then . Then, , i.e., , where 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 , 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 , where indicating that the number of required partnerships at age-group is greater than that available. And at the end of the algorithm, any unassigned partner , i.e., for with , are assigned to initiate partnership when they are in age-group , i.e., set , 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 , for every partner , using and together we can identify the age-groups of and at which the partnership will initiate. We select a random age within those age-groups to set as . We then setand . If this partnership initiated in the past, we would then have , , and . If this partnership would initiate in the future, we would then have , , and . If this partnership would initiate at current time-step, we would then have , , and (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 is set to the time the next partnership of node initiates.

**S5 Disease progression and continuum care module**

The disease progression module in PATH 4.0 is similar to the version in PATH 2.0 [16]. This module simulates disease progression at the individual level for every HIV infected person, including natural HIV disease progression, HIV progression upon diagnosis, care and treatment, and mortality. It also simulates events of diagnosis, care, and treatment to match corresponding care uptake metrics in the population being simulated. We provide brief descriptions of these below.

### *S5.1 Natural disease progression of HIV*

In the disease progression model, each HIV-infected person transitioned through the following disease stages: acute infection, asymptomatic infection, symptomatic infection/AIDS (acquired immune deficiency syndrome), and death. The acute phase of infection was characterized by high HIV viral loads, ranging between 4.4 and 6.2 log10 copies/ml [17,18], for the first three months of infection (Table S11). We assumed the CD4 count at the time of new infection was between 750 and 900 cells/µL [19]. During the asymptomatic infection phase, we assumed there were no AIDS-related symptoms, but the stage was marked by a steady viral load, corresponding to an HIV viral load set point between 4 and 5 log10 copies/ml [20], and declining CD4 count in the absence of treatment. We used estimates of the rate of CD4 count decline for different ranges of HIV viral load reported by Rodriguez et al. (2006) [21]. Symptomatic HIV infection or AIDS was characterized by the occurrence of an opportunistic infection (OI), determined by different probabilities, or a drop in CD4 count to below 200 cells/µL. We assumed the probability of having an OI increased with a decline in CD4 count, and we modeled six infections [22,23].

### *S5.2 Progression of HIV upon diagnosis, care, and treatment*

We assumed persons could be diagnosed, linked to care, and started on treatment at any time after the acute phase of infection with rates determined by the setting being analyzed. In this paper, we calibrated the transition rates to match the distribution of people with HIV (PWH) along care continuum stages as observed through surveillance data for years 2006 to 2015 and is discussed in Section S5.4 below. We assumed the natural progression of HIV described above was altered upon initiation of treatment with antiretroviral therapy (ART), which is associated with suppressed viral load, higher CD4 cell counts, improved life expectancy, and improved quality of life [24,25]. Once suppressed, HIV viral load was assumed to be maintained between 1.0 and 2.7 log10 copies/ml as long as the regimen was effective [26]. When a particular treatment regimen ceased to be effective, we assumed that the HIV viral load rebounded to between 3.1 and 4.5 log10 copies/ml [27]. In accordance with expert opinion and recent clinical trials [28–32], we assumed the probability of initial viral load suppression when taking an ART regimen depended on the CD4 count at the start of treatment. If initial suppression was obtained, the person continued with the regimen for a duration of time until viral rebound occurred, after which the person started on the next line regimen. The duration of time on each regimen was determined by a random number drawn from a geometric distribution. The mean (or rate) of the geometric distribution varied by CD4 count at the start of ART. The rates were derived by calibrating against expected life-expectancies from the Antiretroviral Therapy Cohort Collaboration population [33,34]. If there was no initial suppression, the person moved to the next line regimen. We considered three lines of suppressive regimens, which were based on the Department of Health and Human Services (DHHS) guidelines [35] and expert opinion, followed by salvage therapy. The rate of treatment change for the 2nd and 3rd lines of regimen was 1.18 times the rate of the previous regimen [36]. The maximum CD4 count that could be achieved during sustained HIV viral load suppression (VLS) depended upon the CD4 count at initiation of the first ART regimen [37]. A summary of the disease progression input parameters is provided in Table S11 and a schematic flow is presented in Figure S1.

### *S5.3 Mortality*

We assumed HIV-infected persons in the model could die from causes either related to HIV/AIDS or other factors. For persons not yet on treatment, we used quarterly rates of death [38] that increased as a person’s CD4 count declined (Table S11). The maximum number of years of life remaining for a person infected with HIV in the PATH model was limited by life expectancy at the age of HIV infection based on the general population as reported in US life tables [39]. At the end of every time step of the simulation, we remove all agents that are dead from the model. This is done for computational efficiency. The agent node is removed from the graph.

### *S5.4 Simulating changes diagnosis, care and treatment*

We calibrated diagnosis, linking to care, and initiation of treatment to match the surveillance estimates for the distribution of PWH by care continuum stage, specifically, to the expected proportions unaware, aware but not in care, and on ART with VLS, which we discuss here. Figure S2 shows the flow diagram for disease incidence and transition along the stages of care continuum, used in the calibration process. Note that this representation of the flow diagram combines the acute and non-acute unaware stages into a single stage, and combines ‘in-care no ART’, ‘on-ART with VLS’, and ‘on ART no VLS’ into a single stage. This is because, transitioning from acute to non-acute, and between VLS and no-VLS, is a function of the disease progression and not care behavior, and transitioning from ‘in-care no ART’ to ‘ART with VLS’ is based on ART guidelines at the time, and thus not part of care behavior.

Let,

be the number of infected persons attime *,*

be the number of new infections at time *,*

be the proportion of infected persons in care-continuum stage *,* such thatwould be the number of people in stageat time; where,unaware;aware but not in care or treatment;and in-care (no ART or on ART),

be the rate of diagnosis*,*

be the proportion of persons linking to care and initiating treatment at the time of diagnosis,

be the rate of entering care and treatment among those not in care,

be the rate of dropping-out of care and treatment, and

be the number of new deaths.

Then, at a sufficiently small incremental time-step(we use monthly increments), we can write the equations for the number of people in each stage by formulating as a system of differential equations,

where, is the rate of change in , i.e., the change in the number of infected persons in stageat .

Specifically, for each stage we can write the compartmental equations as,

Additionally,

Note, these equations are applied for each risk-group separately, but for clarity of notations we do not indicate risk-group in subscripts.

At any time-step , values for and for each are known as they are estimated at the previous time-step of simulation. At the beginning of the simulation this would be set to the care continuum distribution in 2006. We analytically calculate the remaining parameters iteratively as follows:

* Estimate transmissions: They are estimated at the individual level for every susceptible-infected partnership. (Modeling partnerships are discussed in Section S4 and modeling transmissions are discussed in Section S3. Transmissions are modeled as a function of stage of the infected person at time *.* And thus, the number of new infections is a function of *(i.e., )*.
* Estimate deaths: They are also modeled at the individual level using stage- and age- specific rates, as discussed in Section S2.2 and thus, can also be estimated through simulation.
* Estimate number of persons to diagnose: We can write the monthly change in proportion unaware as , where, is the annual change in proportion unaware, taken from surveillance (see Table S12). We setin the compartmental equation for above, and rewrite it to estimate diagnostic rate as

and the corresponding number of persons to diagnose as

* Estimate number of persons linking to care at diagnosis: As *.* Data for are available through surveillance [40–45] and presented in Table S12.
* Estimate number of persons entering or re-entering care as : For we use estimates from studies in the literature [40–45] as discussed in Section S6.5.
* Estimate number of persons to drop-out: We can write the monthly change in proportion V as , where, is the annual change in proportion in care, taken from surveillance (see Table S12). We set in the compartmental equation for above and rewrite it to estimate rate of drop-out as

and the corresponding number of persons dropping out as

i.e., the remaining persons are retained in care.

**S6 Implementation of PATH 4.0 for simulating HIV in the US**

### *S6.1 Model initialization to year 2006*

We initially set up the model to represent a cross-sectional view of persons living with HIV in the US in the year 2006. We did this in two main steps. First, we generated a network that is replicative of the contact structure among HIV infected persons and their immediate contacts in the US. That is, the model was set to initiate with 10,000 persons weighted to represent all persons living with HIV in the US in 2006 by sex, age, degree, and transmission risk group (heterosexual male, heterosexual female, and MSM). Next, we ensured that the distribution of epidemic and demographic features, such as stage of disease (acute versus non-acute), and diagnostic, care, and treatment status (whether the person had diagnosed infection, was in care, was taking an ART regimen, and had suppressed viral load (Tables S12-S13)), are representative of HIV infected persons the US in 2006 [46–50]. Because the age distribution of new infections would likely be different from the age distribution of established infections, we identified the proportion of persons in the model living with HIV in 2006 who were newly infected from years 2004 to 2006 and assigned them the published age distributions of new cases from 2006 to 2008 [47,49]. We used new infections from 2006 to 2008 [47,49], as approximate values for new infections from 2004 to 2006 because we did not have estimates of new infections in 2004 and 2005.

At the start of the simulation, we assumed that, for those already diagnosed of their infection, their CD4 count at diagnosis matched the reported estimates in the US (Table S13) [51–53], 7% of persons in care had linked to care within 3 months of diagnosis [54], and persons on ART had an equal probability of being on one of the 3 suppressive ART regimens or salvage therapy.

We randomly assigned each HIV-infected person a number between 1 and 50 to represent their geographic jurisdiction. In this version of the model, we added geographic heterogeneity only to more realistically model contact networks, as persons from the same jurisdiction are more likely to form partnerships. However, we did not specifically use data from jurisdictions, such as population size or differences in care continuum distributions, as the scope of this work was national level estimations only. Future versions of the model should consider more specific jurisdictional data.

### *S6.2 Dry runs*

Dry run is a technique used for initialization of the model. It involves running the simulation for a certain period of time, 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. In this work, we are simulating the US HIV epidemic starting from 2006. For this, we need to initialize the model with a population that is representative of the persons living with HIV in 2006, including the contact structures between them, their demographic distributions, and the disease and care-continuum stage distributions. 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 do a dry run, which means starting with some number of people (assigning them data according to 2006 surveillance distributions), 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. A network model 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 risk-group of the partner, and the infection status of partner. Therefore, 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).

We set up the model as described in Section 5.1 using 2 dry runs. Dry run #1 of the model is used for generating a HIV network structure replicative of sexual transmission contact structure in the US – specifically, to generate an initial network with age, degree, and risk group correlations between neighboring agents to be representative of actual correlations. Dry run #2 is done to ensure demographics of HIV infected persons, namely age and risk group distributions, match 2006 PWH distributions from NHSS, and to further generate HIV disease and care continuum parameters to match cross-sectional distributions in 2006. Dry runs are explained in more detail in Section S7.

### *S6.3 Age adjustment during initialization*

The age of the initial HIV-infected population was distributed to match the reported age distribution of all people living with HIV in the US in 2006. The age for each PWH was initially determined before the first dry run, using surveillance estimates for age distribution of new infections [47,49]. Before the second dry run, the PWH were redistributed to ensure the age distribution of PWH was maintained. This was done by taking the age-group with the largest HIV-infected population size, calculating the total population based on that age group, then adding or removing people from the infected population, to match the number of PWH required in each age group, to match the distribution.

### *S6.4 Risk group adjustment during initialization*

The risk group of the initial HIV-infected population was distributed to match the proportion of people living with HIV in each risk group in the US in 2006. Before the first dry run, risk group is randomly assigned according to the risk-group distribution in the US in 2006. Before the second dry run, adjustments are made to ensure that the distribution of people by risk group match that in the US in 2006. This was done by first calculating the total population of PWH from the number of people in the largest risk group (MSM), then removing the people from the infected population in the other risk groups that exceeded the proportion required in the other risk groups (namely heterosexual male and heterosexual female). Specifically, we calculated ‘number PWH’ as equal to the ‘number of MSM in the simulation’ divided by the ‘proportion of HIV-infected persons who are MSM in the US in 2006’. Then calculated ‘the required number of heterosexual men (women) in the simulation’ as ‘number PWH’ times the ‘proportion of HIV-infected persons who are heterosexual men (women) in the US in 2006’. Then a ‘number of heterosexual men (women) in the simulation’ minus ‘the required number of heterosexual men (women) in the simulation’ number of heterosexual men were removed from the simulation.

### *S6.5 Risk group mixing and pseudo-geographic mixing*

During the simulation, the risk-group of a partner was determined using a risk-group mixing matrix. We assumed heterosexual men only mix with heterosexual women. We assumed 21% of MSM are bisexual, with 20% of partnerships with other MSM and 80% of partnerships with women, and the remaining 71% of MSM only mix with other MSM [55–58]. We assumed 90% of partnerships are with persons in the same pseudo-geographic jurisdiction and 10% are with a person from one of the other randomly chosen pseudo-geographic jurisdictions, to approximate the observed mixing between persons by their geographic location [5].

### *S6.6 Simulating HIV care continuum for years 2006 to 2017*

During the simulation for years 2006 to 2017, the number of persons with diagnosed infection, linked to care, retained in care, and initiating ART was controlled to match the observed proportions of persons aware of their infection, in-care, and taking an ART regimen (Table S12). That is, we simulated diagnosis, linkage to care and treatment, and dropping out of care and re-entry into care and treatment as using methods described in Section S5. For persons who dropped out, the simulation kept track of their status at the time of drop-out (i.e., either taking or not taking an ART regimen) and assigned a time for their re-entry such that, when that time arrived, the person would be re-initiated into care or treatment based on their status at drop-out. The time for re-entry was assumed to be within 1 to 2 years for 45% of those who dropped out [48]. Due to lack of data on the remaining 55%, we assumed 40% would re-enter when their CD4 count decreased to 200 cells/µL and 15% when their CD4 count decreased to 40 cells/µL.

### *S6.7 Simulation and validation timelines*

We ran the simulation model for years 2006 to 2017. However, we validated against years 2010 to 2017 as the surveillance data for the validation metrics, presented in the main paper, were available for only this timeline on ATLAS, the online CDC surveillance database [59]. Starting the model in 2006, a few years prior to first validation year of 2010, allows for a longer period for the model to generate infections, reducing the influence of the initial population, and thus providing more assurance in the validation. We applied the cluster generation algorithm for years 2015 to 2017 as the molecular clusters, used in the cluster validation in the main paper, correspond to analyses conducted using persons diagnosed during January 1, 2015 to December 31, 2017.

We generated 100 runs of the simulation. To obtain statistically significant results, we must consider the number of simulation runs that are sufficient [60–62]. Byrne [60] suggests the use of confidence interval widths to determine the optimal number of simulation runs and found it can range from 19 to 38416 based on the desired confidence of the result. Cheng [61] developed a method for determining the optimal number of runs for discrete-time simulations. They showed that the probability of selecting a best alternative choice out of many possible choices using a small data sample size is 97% for 100 simulation runs. These methods also acknowledge the tradeoff between the number of simulation runs that can be done and the computational complexity of the simulation model. As PATH 4.0 is computationally complex taking about 30 minutes per run on a standard desktop computer, we chose 100 simulation runs to balance computational time with sufficient number of samples.

### S7 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 S14.

Step 1: Initialization of the compartmental model array and graph at time-step

Step 1a. We initialize the compartmental model array to be representative of the U.S. population by age, risk group, and degree distribution in 2006. We distributed the total population in the U.S. into degree bins by using , assuming minimum degree and maximum degree , and estimating the power-law parameter λ from 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 care continuum stage, disease stage, age, and risk group through random draws from the corresponding distributions of PWH in the US in 2006. Note, this step does not attempt to make it be representative of the 2006 population, that attempt is done in Step 2 through dry runs.

Step 2: Dry run #1- Generate a network that is representative of the sexual partnership network in the US

Dry run #1 of the model is used for generating a HIV network structure replicative of the sexual transmission contact structure in the United States – specifically, to generate an initial network with age, degree, and risk group correlations between neighboring agents to be representative of actual correlations. We repeat the following steps of the simulation model for a random number of time-steps (we chose 50) by calibrating to the care continuum distribution in 2006 for Step 2d. Though dry run #1 calibrates to the 2006 disease and care continuum distributions in Step 2d, the main purpose of dry run #1 is to generate an initial network with age, degree, and risk group correlations between neighboring nodes to be representative of actual correlations by repeated application of the HIV-ECNA in Step 2a.

Step 2a. Generate the network of contacts for every newly infected node by applying the HIV-ECNA module.

Step 2b. Update demographic parameters of susceptible persons by applying the compartmental module.

Step 2c. Generate new infections after updating behavioral parameters by applying the Bernoulli transmission module.

Step 2d. Update demographics and HIV disease progression and care for every HIV infected node by applying the disease progression module, by calibrating to the care continuum distribution in the year corresponding to the time-step of the simulation.

Step 3: Dry run #2- Generate an initial population that is representative of the PWH in the US in 2006

Dry run #2 is done to ensure that the demographics of HIV infected persons, namely their age and risk group distributions, and their disease stage and care continuum stage distributions match that of PWH in 2006 in the United States (as reported in the NHSS).

Step 3a: From among the nodes at the end of the dry run #1, remove or add HIV infected nodes to ensure that the overall age distribution match that of PWH in 2006. Removal is preferred to addition of nodes as it removes nodes while maintaining the degree correlations between neighboring nodes. Therefore, we use the age-group with the largest proportion as reference and scale down the number of nodes in all other age-groups by removing nodes to achieve the required distribution by age. As age distribution of recently infected nodes (those infected in 2004 to 2006 in this case) are typically different from the overall age-distribution of PWH, from every age-group, we randomly select a proportion of nodes equivalent of proportion of new infections in that age group in years 2006 to 2008 (43,45) and set their time of infection to between 2004 and 2006. We used the age-group from those infected in years 2006 to 2008, as data for 2004 to 2006 were unavailable and assumed they will be similar.

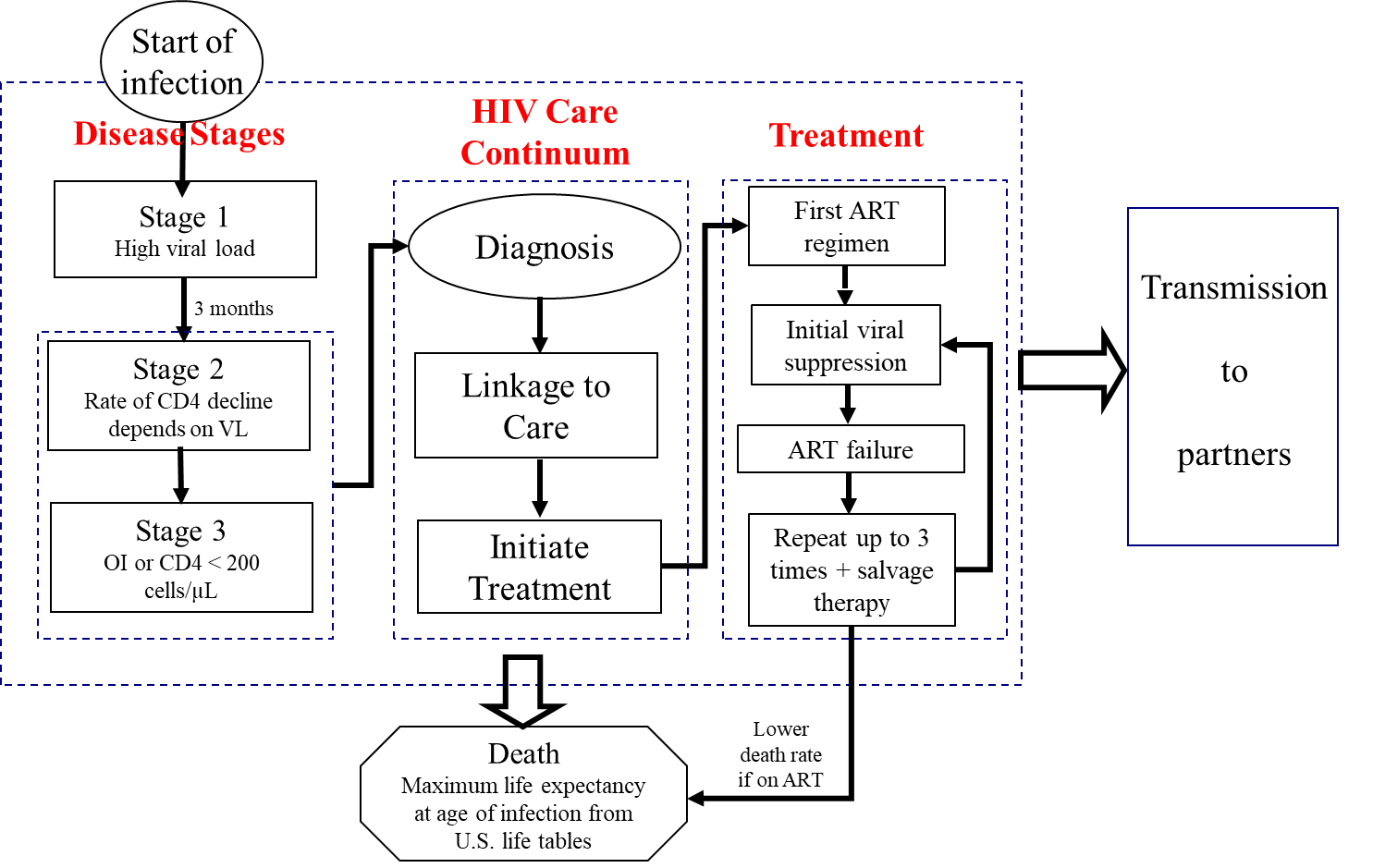
Step 3b: Remove HIV infected nodes to ensure that the distribution by risk-group match that of PWH in 2006. The number of MSM is set as the reference and the number of heterosexual males and females are scaled down through random selection. Though this would alter the age distributions from above, it would only be mild (as nodes are randomly selected) and further rectified during Step 3e.

Step 3c: Reassign HIV disease and care continuum stage to ensure that the distributions among those assigned as new infections in Step 3a match the actual distribution of newly infected PWH in 2006 in the US, and the distributions in all HIV infected persons match those of the actual overall PWH in 2006 in the US. For example, if there are 20% of PWH who are undiagnosed, and currently the model has 25%, it randomly selects 5% of the undiagnosed and sets their status as diagnosed.

Step 3d: Repeat Step 1a. to reinitialize the compartmental model array population to match the US population in 2006.

Step 3e: Repeat the simulation Steps 2a to 2d for a certain number of time-steps (we used 180). For Step 2d, calibrate to the care continuum distribution in year corresponding to simulation start year (2006 in our implementation to the US). This is the main step of the dry-run, a type of burn-in period where the simulation is run several times to ensure that the dynamics between individual-level features, including disease and care continuum stages of PWH, and behavioral features of partnerships, and the population-level features of contact structures and overall distributions are well aligned.

Step 4: Main simulation run. Repeat the steps outlined in 2a to 2d, for the required number of months, we simulated years 2006 to 2017.



**Figure S1**. Schematic overview of the PATH disease progression model.

A picture containing clock

Description automatically generated

**Figure S2.** Flow diagram for disease incidence and transition along the stages of care continuum. S: Population susceptible, U: Population unaware, A: Population aware but not in care, V: Population in care (no ART and on ART), : diagnostic rate, : rate of entering care and treatment among those not in care, and ρ: rate of dropping-out of care, and : rate of diagnosis and linked to care at the time of diagnosis.

|  |  |
| --- | --- |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | |

|  |  |
| --- | --- |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | A close up of a map  Description automatically generated |
| A close up of a map  Description automatically generated | |

**Figure S3.** 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 [15] (presented in Tables S6.1 and S6.2), for heterosexual men. 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 [15] (presented in Tables S6.1 and S6.2), for heterosexual female.

**Table S1**. Rate of natural mortality.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Annual probability of death if HIV uninfected | 13–17 Years | 18–24 Years | 25–34 Years | 35–44 Years | 45–64  Years | Source |
| HET or MSM |  |  |  |  |  |  |
| Female | 0.0002 | 0.0004 | 0.0006 | 0.0013 | 0.0048 | [39] |
| Male | 0.0005 | 0.0013 | 0.0014 | 0.0021 | 0.0079 |

**Table S2.1.** Percentage of PWH in 2006 by age [63] .

|  |  |
| --- | --- |
| Age group | Percentage of PWH |
| 13-24 | 5.1% |
| 25-34 | 16.1% |
| 35-44 | 35.3% |
| 45-54 | 30.4% |
| 55-70 | 13.1% |

**Table S2.2.** The cumulative percentage of population, accumulated along age groups by age in 2006, stratified by newly infected (persons infected in years 2004, 2005, or 2006) and all people with HIV (PWH).

|  |  |  |
| --- | --- | --- |
| Age group | Cumulative percentage of newly infected [47] | Cumulative percentage of all PWH [46] |
| 13-14 | 0.1% | 0.3% |
| 15-19 | 4.1% | 0.7% |
| 20-24 | 21.1% | 1.3% |
| 25-29 | 36.1% | 6.0% |
| 30-34 | 51.1% | 14.0% |
| 35-39 | 66.1% | 29.0% |
| 40-44 | 78.1% | 51.2% |
| 45-49 | 89.1% | 70.9% |
| 50-54 | 95.6% | 85.0% |
| 55-59 | 99.6% | 92.9% |
| 60-64 | 100.0% | 96.8% |
| 65-70 | 100.0% | 100.0% |

**Table S3.** Percentage of U.S. population by sex & risk group in 2006[64–66].

|  |  |
| --- | --- |
| Risk Group | Percentage |
| HET-female | 47.25% |
| HET-male | 47.25% |
| MSM | 2.33% |
| PWID-female | 1.20% |
| PWID-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** | **Source** |
| Number of sex acts per year a |  |  |  |  |  |  |  |  |  |  |  |  | [67–70] |
| 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 |  |
| 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 |  |
| 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 |  |
| Proportion of acts that are anal |  |  |  |  |  |  |  |  |  |  |  |  | [67–70] |
| 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 |  |
| 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 |  |
| MSM | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |
| Proportion of acts condom protected (main partners) b |  |  |  |  |  |  |  |  |  |  |  |  | [71] |
| 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 |  |
| 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 |  |
| 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 |  |
| 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 | [55] |

|  |  |  |
| --- | --- | --- |
| **Other parameters** |  | Source |
| Reduction in unprotected acts when aware (all risk groups) | 53% | [72] |
| 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% | [55–58] |
| Proportion of partnerships with female for MSMW | 80% | [16] |
| Proportion of anal acts with female for MSMW | 50% | [16] |

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 [58] 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~ Uniform float[0,1], if u <= 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.** Scale-free degree distribution, stratified by risk group.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Degree Bin | 0 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 |
| HET-female | 0 | 0 | 0.261179 | 0.22850951 | 0.178619 | 0.130822 | 0.092571 | 0.064362 | 0.043939 |
| HET-male | 0 | 0 | 0.211253 | 0.20189226 | 0.174722 | 0.142791 | 0.113207 | 0.08837 | 0.067764 |
| MSM | 0 | 0 | 0.190383 | 0.18901156 | 0.170899 | 0.146404 | 0.121882 | 0.099992 | 0.081429 |

**Table S6.1.** 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) [15].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Number of female sexual partners | | | | | |
| Age Group of Males | Total in Age Group | 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 S6.2.** 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) [15].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Number of male sexual partners | | | | | |
| Age Group of Females | Total in Age Group | 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 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 (estimation discussed in Section 4.2).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| >44 | 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 (estimation discussed in Section 4.2).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| >44 | 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 risk group [1].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 risk group (estimation discussed in Section 4.2).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |
| >44 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table S10.** Age mixing matrix for MSM: age distribution of sexual partners (columns) for a person in age-group given in row.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age Group | 13–17 | 18–24 | 25–29 | 30-34 | 35–39 | 40-44 | 45-64 | Source |
| 13–17 | 91.1 | 4.2 | 1.1 | 1.1 | 1.1 | 1.1 | 0 | [1] |
| 18–24 | 46.4 | 48 | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 |  |
| 25-29 | 19.8 | 19.8 | 55.9 | 1.1 | 1.1 | 1.1 | 1.1 |
| 30–34 | 18.9 | 18.9 | 12.6 | 46.3 | 1.1 | 1.1 | 1.1 |
| 35-39 | 14.2 | 14.2 | 7.1 | 7.1 | 55.2 | 1.1 | 1.1 |
| 40–44 | 12.5 | 12.5 | 6.2 | 6.2 | 6.2 | 55.2 | 1.1 |
| 45–64 | 11.2 | 11.2 | 5.6 | 5.6 | 5.6 | 5.6 | 55.2 |

**Table S11.** Disease progression parameters.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Mean Value** | **Range** | **Source** |
| **Natural Disease Progression** |  |  |  |
| CD4 cell count when infected (cells/µL) | 870 | 750 - 900 | [19] |
| Acute phase HIV viral load (log10 copies/ml) | 5.3 | 4.4 – 6.2 | [17,18] |
| HIV viral load set point (log10 copies/ml) | 4.5 | 4.0 – 5.0 | [20,73] |
| Natural rate of CD4 cell count decline (cells/µL/quarter) as a function of HIV viral load stratum (log10 copies/ml) |  |  | [21] |
| ≤ 2.7 | 5.1 | 2.4 – 7.8 |  |
| 2.7 – 3.3 | 9.9 | 7.2 – 12.3 |  |
| 3.3 – 4.0 | 12.0 | 9.9 – 13.8 |  |
| 4.0 – 4.6 | 14.1 | 11.7 – 16.2 |  |
| ≥ 4.6 | 19.5 | 17.1 – 21.9 |  |
| **Quarterly Rate of Developing an Opportunistic Infection (OI) (%)** |  |  | [22,23] |
| *Pneumocystis pneumonia* (PCP) | 0.1 – 10.7a |  |  |
| *Mycobacterium avium* complex | 0.0 – 3.6 |  |  |
| Toxoplasmosis | 0.0 – 0.8 |  |  |
| Cytomegalovirus infection | 0.0 – 5.5 |  |  |
| Fungal infection | 0.0 – 3.3 |  |  |
| Other | 0.1 – 11.4 |  |  |
| Cumulative probability for all OIs | 0.3 – 35.3 |  |  |
| **Quarterly Rates of Death for Antiretroviral Therapy (ART)-Naïve Individuals (%)** |  |  |  |
| CD4 cell count (cells/µL) |  |  | [38] |
| ≥ 650 | 0.043 |  |  |
| 500 – 649 | 0.05 |  |  |
| 350 – 499 | 0.08 |  |  |
| 200 – 349 | 0.145 |  |  |
| 50 – 199 | 0.767 |  |  |
| < 50 | 4.9 |  |  |
| **ART Regimens** |  |  |  |
| Suppressed HIV viral load level (log10 copies/ml) | 1.3 | 1.0 – 2.7 | [26] |
| Rebound HIV viral load level (log10 copies/ml) | 3.7 | 3.1 – 4.5 | [27] |
| Maximum number of ART regimens and regimen drug composition | 3 + Salvage Therapy (I. EFV/TDF/FTC;  II. ATV/r+ABC/3TC; III. RAL+TDF/FTC) |  | b |
| Probability of initial virologic suppression in ART regimens: CD4 cell count (cells/µL) at ART initiation |  |  | [28,29] |
| >200 | 0.84 |  |  |
| 50 - 200 | 0.79 |  |  |
| < 50 | 0.774 |  |  |
| Rate of HIV viral load rebound (% experiencing rebound after one year on first regimen) by CD4 cell count (cells/µL) at ART initiation |  |  | [33,34] (estimated by calibrated to match life-expectancy in cited reference) |
| <100 | 5 |  |  |
| 100-200 | 2.45 |  |  |
| 200-300 | 1.83 |  |  |
| 300-400 | 1.46 |  |  |
| 400-500 | 1.45 |  |  |
| Percent increase in rate of HIV viral load rebound for each successive regimen compared to its previous regimen | 18 |  | [36] |
| HIV viral load above set-point during salvage therapy (log10 copies/ml) | 0.8 | 0.0 – 1.5 | [74] |
| HIV viral load above set-point during salvage therapy after onset of AIDS (log10 copies/ml) | 1 | 0.0 – 2.0 | [74] |
| Quarterly increase in CD4 cell count during HIV viral load suppression (cells/µL/quarter) |  |  | [37] |
| Quarters 1 – 2 | 68 |  |  |
| Quarters 3 – 12 | 40 |  |  |
| Quarters 12+ | 0 |  |  |
| Maximum CD4 cell count achieved based on CD4 cell count at initiation of ART (cells/µL) |  |  | [37] |
| < 50 | 410 |  |  |
| 50 – 200 | 548 |  |  |
| 201 – 350 | 660 |  |  |
| 351 - 500 | 780 |  |  |
| > 500 | 870 |  |  |
| **Quarterly Rate of Death After Initiation of ART (%)** |  |  | [24,25] |
| No AIDS symptoms |  |  |  |
| Age 16 – 29 years | 0.09 – 0.26c |  |  |
| Age 30 – 39 years | 0.12 – 0.32 |  |  |
| Age 40 – 49 years | 0.15 – 0.43 |  |  |
| Age ≥ 50 years | 0.29 – 0.81 |  |  |
| Clinical symptoms of AIDS |  |  |  |
| Age 16 – 29 years | 0.19 – 0.53 |  |  |
| Age 30 – 39 years | 0.25 – 0.69 |  |  |
| Age 40 – 49 years | 0.32 – 0.93 |  |  |
| Age ≥ 50 years | 0.64 – 1.77 |  |  |

aThe lower and upper bounds for various types of OIs reflect probabilities for CD4 cell counts of > 500 cells/µL and 0 – 50 cells/µL respectively. Probabilities of an OI at intermediate CD4 cell counts lie within these bounds.

bExpert opinion (2009); EFV/TDF/FTC = efavirenz/tenofovir/emtricitabine, ATV/r = atazanavir/ritonavir, ABC/3TC = abacavir/lamivudine, RAL = raltegravir

cThe lower and upper bounds reflect the probability of death for CD4 cell counts ≥ 350 cells/µL and < 25 cells/µL, respectively. Probabilities of death at intermediate CD4 cell counts lie within these bounds.

**Table S12.** Annual proportion of PWH in continuum care stages.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | Unaware of infection [41,44,45,75,76] | Aware of infection but not in care [41–45,75,76] | In care but not on ART [42,77] | Virally suppressed [40–45,75] | On ART – Not virally suppressed | Linked to care within 3 months of diagnosis [40–43,45,76] |
| 2006 | 21.0% | 40.0% | 16.0% | 19.0% | 4.0% | 77.0% |
| 2007 | 18.3% | 43.0% | 12.0% | 22.0% | 4.7% | 78.0% |
| 2008 | 17.8% | 45.0% | 8.0% | 24.0% | 5.2% | 78.0% |
| 2009 | 17.3% | 47.6% | 3.7% | 26.5% | 4.9% | 79.0% |
| 2010 | 20.1% | 28.4% | 3.5% | 37.0% | 11.0% | 79.0% |
| 2011 | 19.4% | 26.6% | 3.1% | 40.5% | 10.4% | 81.0% |
| 2012 | 18.7% | 24.7% | 2.9% | 43.3% | 10.4% | 81.0% |
| 2013 | 18.0% | 22.6% | 0.0% | 47.5% | 11.9% | 85.0% |
| 2014 | 17.3% | 21.4% | 0.0% | 50.6% | 10.7% | 85.0% |
| 2015 | 16.7% | 20.9% | 0.0% | 52.5% | 9.9% | 84.0% |
| 2016 | 16.4% | 24.2% | 0.0% | 54.4% | 5.0% | 85.6% |
| 2017 | 16.0% | 22.7% | 0.0% | 55.2% | 6.10% | 86.8% |

**Table S13.** Distribution of parameters of people living with HIV in the United States in 2006.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Female | Male | MSM | Source |
| Distribution of PWH in year 2006 by stage | | | | [46–48] |
| 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) | | | [47,49] |
| 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 infections each year | | | | [47,49] |
| 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/μL) at diagnosis for those aware of infection by year 2006 | | | | |
| <50 | 10% | 10% | 10% | [51] |
| 50-200 | 14% | 14% | 14% | [51] |
| 200-500 | 66% | 66% | 51% | [52,53] |
| >500 | 10% | 10% | 25% | [52,53] |

**Table S14.** Table of Notations.

|  |  |
| --- | --- |
| **Notation** | **Description** |
|  | Simulation time-step. |
|  | The number of age-groups. |
|  | The number of risk-groups. |
|  | The number of degree bins. |
|  | The number of pseudo-geographic jurisdictions. |
| ; ; ; | Used when referring to an age-group, risk-group, degree-bin, and pseudo-geographic jurisdiction, respectively. |
| , , , | An array of size representing the number of susceptible persons in the model, in age-group , risk group , degree-bin , and pseudo-geographic jurisdiction , at time t. |
|  | A set of nodes, each representing an infected person or a susceptible sexual partner. |
|  | A set of edges representing sexual partnerships between nodes. |
|  | A dynamic graph with a set of nodes and a set of edges, at time t. |
|  | The number of nodes in graph *G*, at time *t*. |
|  | A static adjacency matrix of size , with static element if and are sexual partners anytime during their lifetime and 0 otherwise. |
|  | A dynamic adjacency matrix of size , with element if and are sexual partners during month and otherwise. |
|  | An edge in graph representing a sexual partnership between and |
|  | The partnership initiation time; represents the simulation month for partnership initiation. |
|  | The partnership termination time; represents the simulation month for partnership termination. |
|  | The age of nodes and at the time of their partnership initiation. |
|  | The age of nodes and at the time of their partnership termination. |
|  | Age-group of node at time . |
|  | Age of node *j* at time *t.* |
|  | Degree-bin corresponding to the number of lifetime partners of node . |
|  | The actual number of lifetime sexual partners of node . |
|  | The number of lifetime sexual partners of person who are already added as nodes in graph *G* at time *t*. For infected nodes for susceptible nodes in *G,* |
|  | A partnership distribution matrix of size , where is the number of partnerships that initiate at age-group , and is the number of partnerships that are yet to be assigned. For infected nodes, |
|  | Infection status of node *j* at time *t*. |
|  | Deceased status of node *j* at time *t*. |
|  | Risk-group of person *j*. |
|  | Care continuum or disease stage of person *j* at time *t*. |
|  | Infectiousness or risk of transmission per act for person *j* at time *t*. |
|  | Condom effectiveness. |
|  | The number of sex acts per month for person *j* at time *t*. |
|  | The proportion of acts condom protected of person *j* at time *t*. |
|  | The inverse Bernoulli distribution that takes values 1 with probability and 0 with probability . |
|  | Random variable for degree of node . |
|  | Conditional probability distribution for . |
|  | Marginal probability distribution for . |
| *m* | Minimum degree of the network. |
|  | Scale-free network parameter corresponding to the risk-group of node *l*. |
|  | A matrix of size , representing the proportion of partnerships that initiate at age-group for persons in degree-bin . |
|  | Random variable representing the number of lifetime partners at age-group. |
|  | The transition probability matrix of size for age-group . |
|  | The steady state distribution for lifetime-partners in age-group. |
|  | A vector of size , converted from matrix . |
|  | A matrix of size , recreated with values of . |
|  | A binary matrix of size . |
|  | A vector of ones of size . |
|  | An age-mixing matrix of size , which represents the probability that, given a person is in age-group, his or her partner is in age-group . Varies by risk-group. |
|  | A matrix of size , representing the expected number of contacts the newly infected node should have with person in age-group , when node is in age-group . |
|  | A binary matrix of size , which represents whether partner would have a newly initiating partnership at age-group . |
|  | A binary matrix of size , which represents whether the partnership with partner would occur when partner is in age-group . |
|  | A vector of size *A*, denoting the number of partnerships that should initiate when the partner is in age-group *i*. |
|  | A vector of size *A*, representing node *k* to be eligible to initiate a partnership at age-group *i*. |

**S8 References**

1. S. N. Glick, M. Morris, B. Foxman, S. O. Aral, L. E Manhart, K. K. Holmes, 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*., **60** (2012), 83–90. doi: 10.1097/QAI.0b013e318247925e
2. J. A Grey, K. T. Bernstein, P. S. Sullivan, D. W. Purcell, H. W. Chesson, T. L. Gift, et al. Estimating the Population Sizes of Men Who Have Sex With Men in US States and Counties Using Data From the American Community Survey. *JMIR Public Health Surveill.,* **2(1)** (2016), e14.
3. J. A. Martin, B. E. Hamilton, M. J. K. Osterman, A. K. Driscoll, T. J. Mathews. Births: Final data for 2015. *Natl. Vital Stat. Rep*., **66(1)** (2017), 1.
4. *Centers for Disease Control and Prevention. “HIV Among People Aged 50 and Over.”* Centers for Disease Control and Prevention, 12 Nov. 2019, Available from: www.cdc.gov/hiv/group/age/olderamericans/index.html.
5. A. R. Board, L. Linley, A. M. Oster, et al. Geographic distribution of HIV transmission networks in the United States. *J. Acquir. Immune Defic. Syndr*., (2020). In Press.
6. A-L. Barabasi, R. Albert. Emergence of scaling in random networks. *Science (New York, NY).,* **286** (1999), 509–512. doi: 10.1126/science.286.5439.509
7. A. J. Leigh Brown, S. J. Lycett, L. Weinert, G. J. Hughes, E. Fearnhill, D. T. Dunn; UK HIV Drug Resistance Collaboration. Transmission Network Parameters Estimated From HIV Sequences for a Nationwide Epidemic. *J. Infect. Dis.,* **204(9)** (2011), 1463-9. doi: 10.1093/infdis/jir550
8. J. O Wertheim, A. J. Leigh Brown, N. L. Hepler, S. R. Mehta, D. D. Richman, D. M. Smith, et al. The Global Transmission Network of HIV-1. *J. Infect. Dis.,* **209(2)** (2014), 304–313. doi: 10.1093/infdis/jit524
9. J. O. Wertheim, S. L. Kosakovsky Pond, L. A. Forgione, S. R. Mehta, B. Murrell, S. Shah, et al. Social and Genetic Networks of HIV-1 Transmission in New York City. *PLoS Pathog.,* **13(1)** (2017), e1006000. doi: 10.1371/journal.ppat.1006000
10. F. Brauer, P. van den Driessche, J. Wu, editors. Mathematical Epidemiology. *Springer.,* (2000).
11. M. Eden, R. Castonguay, B. Munkhbat, et al. Agent-based evolving network modeling: a new simulation method for modeling diseases with low prevalence. *Health Care Manag. Sci.*, (2019). Accepted.
12. “Igraph v0.1.1” Function | R Documentation, www.rdocumentation.org/packages/igraph/versions/0.1.1/topics/barabasi.game.
13. K. Dombrowski, R. Curtis, S. Friedman, B. Khan. Topological and Historical Considerations for Infectious Disease Transmission among Injecting Drug Users in Bushwick, Brooklyn (USA). *World J. AIDS.,* **3(1)** (2013), 1-9. doi: 10.4236/wja.2013.31001
14. F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, Y. Åberg. The web of human sexual contacts. *Nature*., **411** (2001), 907–908. doi: 10.1038/35082140
15. W. D. Mosher, A. Chandra, J. Jones. Sexual behavior and selected health measures: men and women 15-44 years of age, United States, 2002. *Adv. Data*., **(362)** (2005); 1–55.
16. C. Gopalappa, P. G. Farnham, Y. H. Chen, S. L. Sansom. Progression and Transmission of HIV/AIDS (PATH 2.0). *Med. Decis. Making.*, **37(2)** (2017), 224-233. doi: 10.1177/0272989X16668509
17. T. W. Schacker, J. P. Hughes, T. Shea, R. W. Coombs, L. Corey. Biological and virologic characteristics of primary HIV infection. *Ann. Intern. Med*., **128(8)** (1998), 613–620. doi: 10.7326/0003-4819-128-8-199804150-00001
18. T. Schacker, A. C. Collier, J. Hughes, T. Shea, L. Corey. Clinical and epidemiologic features of primary HIV infection. *Ann. Intern. Med*., **125(4)** (1996), 257–264. doi: 10.7326/0003-4819-125-4-199608150-00001
19. B. J. Turner, F. M. Hecht, R. B. Ismail. CD4+ T-lymphocyte measures in the treatment of individuals infected with human immunodeficiency virus type 1. A review for clinical practitioners. *Arch. Intern. Med*., **154(14)** (1994), 1561–1573.
20. J. T. Herbeck, G. S. Gottlieb, X. Li, Z. Hu, R. Detels, J. Phair, et al. Lack of evidence for changing virulence of HIV-1 in North America. *PLOS One*., **3(2)** (2008), e1525. doi: 10.1371/journal.pone.0001525
21. B. Rodríguez, A. K. Sethi, V. K. Cheruvu, W. Mackay, R. J. Bosch, M. Kitahata, et al. Predictive value of plasma HIV RNA level on rate of CD4 T-cell decline in untreated HIV infection. *JAMA*., **296(12)** (2006), 1498–1506.
22. M. C. Weinstein, S. J. Goldie, E. Losina, C. J. Cohen, J. D. Baxter, H. Zhang, et al. Use of genotypic resistance testing to guide hiv therapy: clinical impact and cost-effectiveness. *Ann. Intern. Med*., **134** (2001), 440–450. doi: 10.7326/0003-4819-134-6-200103200-00008
23. A. D. Paltiel, M. C. Weinstein, A. D. Kimmel, G. R. Seage 3rd, E. Losina, H. Zhang, et al. Expanded screening for HIV in the United States–an analysis of cost-effectiveness. *N. Engl. J. Med.,* **352(6)** (2005), 586–595. doi: 10.1056/NEJMsa042088
24. Antiretroviral Therapy Cohort Collaboration. Importance of baseline prognostic factors with increasing time since initiation of highly active antiretroviral therapy: collaborative analysis of cohorts of HIV-1-infected patients. *J. Acquir. Immune Defic. Syndr.,* **46(5)** (2007), 607–615. doi: 10.1097/QAI.0b013e31815b7dba
25. M. May, J. A. C. Sterne, C. Sabin, D. Costagliola, A. C. Justice, R. Thiébaut, et al. Prognosis of HIV-1-infected patients up to 5 years after initiation of HAART: collaborative analysis of prospective studies. *AIDS.,* **21(9)** (2007), 1185–1197. doi: 10.1097/QAD.0b013e328133f285
26. J. M. Raboud, J. S. G. Montaner, B. Conway, S. Rae, P. Reiss, S. Vella, et al. Suppression of plasma viral load below 20 copies/ml is required to achieve a long-term response to therapy. *AIDS*., **12(13)** (1998), 1619–1624.
27. United Kingdom Collaborative HIV Cohort Study. K. J. Lee, D. Dunn, R. Gilson, K. Porter, L. Bansi, T. Hill, et al. Treatment switches after viral rebound in HIV-infected adults starting antiretroviral therapy: multicentre cohort study. *AIDS.,* **22(15)** (2008), 1943–1950. doi: 10.1097/QAD.0b013e32830e4cf3
28. J. L. Lennox, E. Dejesus, D. S. Berger, A. Lazzarin, R. B. Pollard, J. V. Ramalho Madruga, et al. Raltegravir versus Efavirenz regimens in treatment-naive HIV-1-infected patients: 96-week efficacy, durability, subgroup, safety, and metabolic analyses. *J. Acquir. Immune Defic. Syndr*., **55(1)** (2010), 39–48. doi: 10.1097/QAI.0b013e3181da1287
29. J. L. Lennox, E. DeJesus, A. Lazzarin, R. B. Pollard, J. V. Madruga, D. S. Berger, et al. Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naive patients with HIV-1 infection: a multicentre, double-blind randomised controlled trial. *Lancet.,* **374** (2009), 796–806. doi: 10.1016/S0140-6736(09)60918-1
30. A. M. Mills, M. Nelson, D. Jayaweera, K. Ruxrungtham, I. Cassetti, P. M. Girard, et al. Once-daily darunavir/ritonavir vs. lopinavir/ritonavir in treatment-naive, HIV-1-infected patients: 96-week analysis. *AIDS.,* **23(13)** (2009), 1679–1688. doi: 10.1097/QAD.0b013e32832d7350
31. J-M. Molina, J. Andrade-Villanueva, J. Echevarria, P. Chetchotisakd, J. Corral, N. David, et al. Once-daily atazanavir/ritonavir compared with twice-daily lopinavir/ritonavir, each in combination with tenofovir and emtricitabine, for management of antiretroviral-naive HIV-1-infected patients: 96-week efficacy and safety results of the CASTLE study. *J. Acquir. Immune Defic. Syndr.*, **53(3)** 2010, 323–332. doi: 10.1097/QAI.0b013e3181c990bf
32. J-M. Molina, J. Andrade-Villanueva, J. Echevarria, P. Chetchotisakd, J. Corral, N. David, et al. Once-daily atazanavir/ritonavir versus twice-daily lopinavir/ritonavir, each in combination with tenofovir and emtricitabine, for management of antiretroviral-naive HIV-1-infected patients: 48 week efficacy and safety results of the CASTLE study. *Lancet.,* **372** (2008), 646–655. doi: 10.1016/S0140-6736(08)61081-8
33. Antiretroviral Therapy Cohort Collaboration. Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet.,* **372(9635)** (2008), 293–299. oi: 10.1016/S0140-6736(08)61113-7
34. When To Start Consortium, J. A. C. Sterne, M. May, D. Costagliola, F. de Wolf, A. N. Phillips, R. Harris, et al. Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies. *Lancet.,* **373(9672)** (2009), 1352–1363. doi: 10.1016/S0140-6736(09)60612-7
35. Centers for Disease Control and Prevention. Report of the NIH Panel to Define Principles of Therapy of HIV Infection and Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents. *MMWR Morb. Mortal Wkly. Rep.*, **47(No. RR-5)** (2012), 1–239.
36. C. J. Smith, A. N. Phillips, B. Dauer, M. A. Johnson, F. C. Lampe, M. S. Youle, et al. Factors associated with viral rebound among highly treatment-experienced HIV-positive patients who have achieved viral suppression. *HIV Med*., **10(1)** (2009), 19–27. doi: 10.1111/j.1468-1293.2008.00650.x
37. L. Gras, A. M. Kesselring, J. T. Griffin, A. I. van Sighem, C. Fraser, A. C. Ghani, et al. ATHENA, Netherlands National Observational Cohort Study. CD4 cell counts of 800 cells/mm3 or greater after 7 years of highly active antiretroviral therapy are feasible in most patients starting with 350 cells/mm3 or greater. *J. Acquir. Immune Defic. Syndr*., **45(2)** (2007), 183–192.
38. A. N. Phillips, B. Gazzard, R. Gilson, P. Easterbrook, M. Johnson, J. Walsh, et al. The United Kingdom Collaborative HIV Cohort (CHIC) Study. Rate of AIDS diseases or death in HIV-infected antiretroviral therapy-naive individuals with high CD4 cell count. *AIDS*. **21(13)** (2007); 1717–1721.
39. E. Arias. United States life tables, 2004. National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics. *Natl. Vital Stat. Rep*., **56** (2007), 1–39.
40. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2011*. HIV Surveillance Supplemental Report 2013; 18(No. 5). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-18-5.pdf Published October 2013. Accessed May 2020.
41. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2012.* HIV Surveillance Supplemental Report 2014; 19(No. 3). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-19-3.pdf Published November 2014. Accessed May 2020.
42. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2013*. HIV Surveillance Supplemental Report 2015; 20(No. 2). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-20-2.pdf Published July 2015. Accessed May 2020.
43. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2014*. HIV Surveillance Supplemental Report 2016; 21(No. 4). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-21-4.pdf Published July 2016. Accessed May 2020.
44. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2015*. HIV Surveillance Supplemental Report 2017; 22(No. 2). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-22-2.pdf Published July 2017. Accessed May 2020.
45. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2016*. HIV Surveillance Supplemental Report 2018; 23(No. 4). Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-23-4.pdf Published June 2018. Accessed May 2020.
46. M. L. Campsmith, P. H. Rhodes, H. I. Hall, T. A. Green. Undiagnosed HIV prevalence among adults and adolescents in the United States at the end of 2006. *J. Acquir. Immune Defic. Syndr.,* **53(5)**(2010), 619–624.
47. J. Prejean, R. Song, A. Hernandez, R. Ziebell, T. Green, F. Walker, et al. Estimated HIV incidence in the United States, 2006-2009. *PloS One*., **6(8)** (2011), e17502. doi: 10.1371/journal.pone.0017502
48. E. M. Gardner, M. P. McLees, J. F. Steiner, C. Del Rio, W. J. Burman. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin. Infect. Dis.,* **52(6)** (2011), 793–800. doi: 10.1093/cid/ciq243
49. *Centers for Disease Control and Prevention*. HIV Surveillance Report, 2008; vol. 20. Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2008-vol-20.pdf Published June 2010. Accessed May 2020.
50. Centers for Disease Control and Prevention (CDC). HIV surveillance--United States, 1981-2008. *MMWR Morb. Mortal Wkly. Rep.,* **60(21)** (2011), 689–693.
51. H. I. Hall, K. McDavid, Q. Ling, A. Sloggett. Determinants of progression to AIDS or death after HIV diagnosis, United States, 1996 to 2001. *Ann. Epidemiol.* **16(11)** (2006), 824–833.
52. K. Buchacz, C. Armon, F. J. Palella, R. K. Baker, E. Tedaldi, M. D. Durham, et al. CD4 Cell Counts at HIV Diagnosis among HIV Outpatient Study Participants, 2000-2009. *AIDS Res. Treat*., **2012** (2012), 869841.
53. K. N. Althoff, S. J. Gange, M. B. Klein, J. T. Brooks, R. S. Hogg, R. J. Bosch, et al. Late presentation for human immunodeficiency virus care in the United States and Canada. *Clin. Infect. Dis*., **50(11)** (2010), 1512–1520.
54. *Centers for Disease Control and Prevention.* Vital Signs: HIV Prevention Through Care and Treatment — United States [Internet]. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6047a4.htm.
55. T. J. Finlayson, B. Le, A. Smith, K. Bowles, M. Cribbin, I. Miles, et al; Centers for Disease Control and Prevention (CDC). 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.,* **60(14)** (2011), 1-34.
56. A. C. Voetsch, A. Lansky, A. J. Drake, D. MacKellar, T. A. Bingham, A. M. Oster, 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.,* **39(3)** (2012), 229-235.
57. M. R. Friedman, C. Wei, M. L. Klem, A. J. Silvestre, N. Markovic, R. Stall. HIV infection and sexual risk among men who have sex with men and women (MSMW): a systematic review and meta-analysis. *PLoS One.,* **9(1)** (2014), e87139. doi: 10.1371/journal.pone.0087139
58. E. S. Rosenberg, P. S. Sullivan, E. A. Dinenno, L. F. Salazar, T. H. Sanchez. 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.,* **11** (2011), 189. doi: 10.1186/1471-2458-11-189
59. *Centers for Disease Control and Prevention.* NCHHSTP AtlasPlus. Updated 2019. Available from: https://gis.cdc.gov/grasp/nchhstpatlas/tables.html. Accessed July 2020.
60. M. Byrne. How many times should a stochastic model be run? An approach based on confidence intervals. In: Proceedings of the 12th international conference on cognitive modeling, Ottawa. (2013), 445–450.
61. R. Cheng. Determining efficient simulation run lengths for real time decision making. In Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come (WSC ’07) *IEEE Press*., (2007), 340–345.
62. J. Yao. Minimum number of simulation runs for reliable building energy and peak load prediction at different building scales: a study on stochastic shade adjustment. *J. Asian Archit. Build*., **19** (2020).
63. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 U.S. dependent areas—2010*. HIV Surveillance Supplemental Report 2012; 17(No. 3, part A). Available from: https://www.cdc.gov/hiv/library/reports/hiv-surveillance/hiv-surveillance-supplement-17-3.html Published June 2012. Accessed May 2020.
64. A. Chandra, W. D. Mosher, C. Copen, C. Sionean. Sexual behavior, sexual attraction, and sexual identity in the United States: data from the 2006-2008 National Survey of Family Growth. *Natl. Health Stat. Report*., **(36)** (2011); 1–36.
65. A. Lansky, T. Finlayson, C. Johnson, D. Holtzman, C. Wejnert, A. Mitsch, 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*., **9(5)** (2014), e97596. doi: 10.1371/journal.pone.0097596
66. D. W. Purcell, C. H. Johnson, A. Lansky, J. Prejean, R. Stein, P. Denning, 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*., **6** (2012), 98–107. doi: 10.2174/1874613601206010098
67. D. Herbenick, M. Reece, V. Schick, S. A. Sanders, B. Dodge, J. D. Fortenberry. Sexual behavior in the United States: results from a national probability sample of men and women ages 14-94. *J. Sex. Med*., **7 Suppl 5** (2010), 255–265. doi: 10.1111/j.1743-6109.2010.02012.x
68. M. Reece, D. Herbenick, V. Schick, S. A. Sanders, B. Dodge, J. D. Fortenberry. Background and considerations on the National Survey of Sexual Health and Behavior (NSSHB) from the investigators. *J. Sex. Med*., **7 Suppl 5** (2010), 243–245. doi: 10.1111/j.1743-6109.2010.02038.x
69. M. Reece, D. Herbenick, V. Schick, S. A. Sanders, B. Dodge, J. D. Fortenberry. Sexual behaviors, relationships, and perceived health among adult men in the United States: results from a national probability sample. *J. Sex. Med*., **7 Suppl 5** (2010), 291–304. doi: 10.1111/j.1743-6109.2010.02009.x
70. D. Herbenick, M. Reece, V. Schick, S. A. Sanders, B. Dodge, J. D. Fortenberry. Sexual behaviors, relationships, and perceived health status among adult women in the United States: results from a national probability sample. *J. Sex. Med.,* **7 Suppl 5** (2010), 277–290. doi: 10.1111/j.1743-6109.2010.02010.x
71. J. G. Rosenberger, M. Reece, V. Schick, D. Herbenick, D. S. Novak, B. Van Der Pol, 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*., **9(4)** (2012), 1027–1048.
72. G. Marks, N. Crepaz, R. S. Janssen. Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. *AIDS.,* **20(10)** (2006), 1447-1450. doi: 10.1097/01.aids.0000233579.79714.8d
73. T. T. N. Vo, B. Ledergerber, O. Keiser, B. Hirschel, H. Furrer, M. Battegay, et al. Durability and outcome of initial antiretroviral treatments received during 2000-2005 by patients in the Swiss HIV Cohort Study. *J. Infect. Dis.,* **197(12)** (2008), 1685–1694.
74. G. D. Sanders, A. M. Bayoumi, V. Sundaram, S. P. Bilir, C. P. Neukermans, C. E. Rydzak, et al. Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. *N. Engl. J. Med.,* **352(6)** (2005), 570–585.
75. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2018.* HIV Surveillance Supplemental Report 2020; 25(No. 2). Available from: http://www.cdc.gov/hiv/library/reports/ hiv-surveillance.html. Published May 2020. Accessed July 2020.
76. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2017*. HIV Surveillance Supplemental Report 2019; 24(No. 3). Available from: http://www.cdc.gov/hiv/library/reports/ hiv-surveillance.html. Published June 2019. Accessed July 2020.
77. *Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 U.S. dependent areas—2010*. HIV Surveillance Supplemental Report 2013; 18(No. 2, part B). Available from: https://www.cdc.gov/hiv/library/reports/hiv-surveillance/hiv-surveillance-supplement-18-2.html Published January 2013. Accessed May 2020.

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)

1. Alternatively, “… is a block upper bidiagonal matrix with {} on the superdiagonal and the elements of the diagonal blocks equal to zero.” [↑](#footnote-ref-1)