Isolating the sources of racial disparities in HIV prevalence among men who have sex with men (MSM) in Atlanta, GA: A modeling study

Supplementary Technical Appendix

TABLE OF CONTENTS

1 INTRODUCTION

This supplementary technical appendix describes the mathematical model structure, parameterization, and statistical analysis of the accompanying paper in further detail.

1·1 Model Framework

The mathematical models for HIV transmission dynamics presented in this study are agent-based microsimulation models in which uniquely identifiable sexual partnership dyads were simulated and tracked over time. This partnership structure is represented through the use of separable temporal exponential-family random graph models (STERGMs), described in Section 2. On top of this dynamic network simulation, the larger epidemic model represents demography (entries, exits, and aging), interhost epidemiology (disease transmission), intrahost epidemiology (disease progression), and clinical epidemiology (disease diagnosis and treatment). Individual attributes related to these processes are stored and updated in discrete time over the course of each epidemic simulation.

The modeling methods presented here depend upon and extend the *EpiModel* software to incorporate HIV-specific epidemiology. The HIV extensions for men who have sex with men (MSM) were originally developed by Goodreau et al. for use in prior modeling studies of MSM in the United States and South America.[1–3]

In this manuscript, each scenario was simulated 16 times. This number was chosen because it represents the number of cores per node on the Hyak simulation cluster (see below); 16 was deemed sufficient as the main model outcomes showed high stability (e.g. see Figure 3). Models were run until equilibrium prevalence was achieved; consistent with previous work,[4] this was assessed as a <5% change in prevalence over the previous 1000 time steps (weeks), evaluated at each time step consistently for 3000 time steps.

1·2 Model Software

The models in this study were programmed in the R and C++ software languages using the *EpiModel* [http://epimodel.org/] software platform for epidemic modeling. *EpiModel* was developed by the authors for simulating complex network-based mathematical models of infectious diseases, with a primary focus on HIV and other sexually transmitted infections (STIs). *EpiModel* depends on *Statnet* [http://statnet.org/], a suite of software in R for the representation, visualization, and statistical analysis of complex network data.[5]

EpiModel allows for a modular expansion of its built-in modeling tools to address novel research questions. For this current research study, we have developed extension modules into

an add-on software package to *EpiModel* called *Mardham*. This open-source software is available for download along with the scripts used in the execution of these models. The tools and scripts to run these models are contained in a Github software repository:

- [http://github.com/statnet/Mardham] contains the general extension software package. Installing this using the instructions listed at the repository homepage will also load in *EpiModel* and the other dependencies.

Simulations were performed on the Hyak high-performance computing (HPC) system at the University of Washington [https://itconnect.uw.edu/service/shared-scalable-compute-cluster-for-research-hyak/]. This allowed execution of multiple simulations in parallel to reduce the overall computation time.

2 MAIN DATA SOURCES

The various behavioral modules were parameterized using two studies of HIV/STI disparities in Black and White non-Hispanic MSM, conducted from 2010-2014 in Atlanta, GA. The Involvement Study was a prospective HIV incidence cohort of 803 MSM and the MAN Project was a cross-sectional chain-referral sexual networks study of 314 MSM. Both samples were recruited contemporaneously using venue-time-space sampling, using a modified frame from the 2008 MSM round of the National HIV Behavioral Surveillance system. Study participants completed common self-administered computer-based questionnaire modules that assessed demographics, prevention behaviors, and a detailed dyadic (partnership) section that collected demographic, behavioral, and structural (partnership duration and sequence) data.

We first created a combined ego dataset of Black and White non-Hispanic MSM in Atlanta, ages 18-40 from the baseline visit of Involvement (n=803) and network seed-level respondents from the MAN Project (n=196), for a total of 999 egos. We then created a combined dyadic dataset for partnerships among those egos, which included up to 5 most recent sex partners in the previous 6 months per ego for Involvement or 10 partners in 12 months per ego for MAN Project. Only Black and White non-Hispanic male partners were included, and dyads were limited to those in which AI occurred at least once (at last sex or during the 6 or 12 month interval), resulting in a total of 2,626 dyads. We refer to this as the “combined dyadic dataset” below.

3 NETWORKS OF SEXUAL PARTNERSHIPS

We model networks of three interacting types of sexual relations: main partnerships, casual (but persistent) partnerships, and one-time AI contacts. We first describe the methods conceptually, including the parameters used to guide the model and their derivation (Section 3·1), and then present the formal statistical modeling methods (Section 3·2). Consistent with our parameter derivations, all relationships are defined as those in which AI is expected to occur at least once.

3·1 Conceptual Representation of Sexual Networks

Our modeling methods aim to preserve certain features of the cross-sectional and dynamic network structure as reported in behavioral studies, while also allowing for mean relational durations to be targeted to those reported for different groups and relational types. Our methods do so all within the context of changing population size (due to births, deaths, arrivals and departures from the population) and changing composition by attributes such as age.

The network features that we aim to preserve are as follows, with the parameters for each described in turn:

- The proportion of men (by race) in any given combination of main and casual partnerships (for example, in 1 main and 0 casual partnerships) at any time point.
- The expected number of one-time contacts per time step had by men in each main-casual combination (by race).
- Variation across men in the numbers of one-time contacts.
- Race mixing within each of the different relational types.
- Age mixing within each of the different relational types, by the races of the two men involved.
- Prohibitions against partnering for two men who are both exclusively insertive or exclusively receptive.

3·1·1 Number of Ongoing Main and Casual Partnerships

Ongoing partnerships (whether main or casual) were defined from the combined dyadic dataset as those in which sex had already occurred more than once, and in which the respondent anticipated having sex again. Within this set, partnerships were defined as main if the respondent indicated that it was someone they “felt committed to above all others” or that they considered the person their “primary sex partner”; if neither of these conditions held, the partner was defined as casual. This yielded the following proportions of men with a given number of main and casual relationships at a point in time (i.e. the expected *momentary degree distribution*):

	Black				White		
	0 Casual	*1 Casual*	*2 Casual*		*0 Casual*	*1 Casual*	*2 Casual*
0 Main	50·6%	15·1%	5·3%		43·5%	18·4%	9·5%
1 Main	20·7%	6·1%	2·2%		23·3%	3·3%	2·0%

Note that this implies a slightly higher rate of relational concurrency within these two types of relationships for White men (9·5% + 3·3% + 2·0% = 14·8%) than for Black men (5·3% + 6·1% + 2·2% = 13·6%). It also implies a slightly higher proportion of Black men in a main partnership (29·0%) than White men (28·6%), but a higher average number of casual partners for White men (weighted average = 0·447 casual partnerships per person) than Black men (0·362).

3·1·2 Expected Number of One-Time AI Contacts, by Main/Casual Degree

Respondents in the combined dyadic dataset were asked whether they had had sex with each partner once or more than once; the former response led to the contact being defined as one-time. These contacts cannot be analyzed in terms of momentary degree distributions, since none are ongoing at the point of interview, by definition. Instead, we turn the observed frequencies into expected rates of one-time contacts per time step for men under different conditions. One of the sources of heterogeneity in men's propensity for one-time AI contacts is their current relationship status. The expected numbers are given by:

	Black				White		
	0 Casual	*1 Casual*	*2 Casual*		*0 Casual*	*1 Casual*	*2 Casual*
0 Main	0·073	0·091	0·080		0·057	0·084	0·091
1 Main	0·055	0·052	0·052		0·057	0·058	0·058

3·1·3 Heterogeneity in the One-Time Contact Rate

In addition to differences by relational status, men also have underlying fixed heterogeneities in their propensity to engage in one-time AI. The distribution of one-time contacts was divided into quintiles, within which the expected values of one-time AI per time step are:

Quintile	Black	White
Lowest quintile	0·000	0·000
Second quintile	0·010	0·003
Third quintile	0·039	0·036
Fourth quintile	0·074	0·068

Highest quintile	0·212	0·231

Men are assigned a quintile upon entry into the population, which remains fixed. Any individual man's propensity for AI is determined as a combination of their quintile and their current main/casual partnership counts. Our statistical methods (described below) translate both propensities into conditional log-odds, allowing for their combination. Note that the means of the columns in the quintile table equal the means of the values in Section 3·1·2 weighted by the proportions in Section 3·1·1. These reflect the overall expected value across all men within each race for one-time AI acts per time step, which are nearly identical to each other (0·0670 for Black MSM, 0·0676 for White MSM).

3·1·4 Race Mixing

Respondents reported on their perception of the race and ethnicity (Hispanic/non-Hispanic) for each partner. Since this model is limited to young Black and White MSM, we calculated our race mixing proportions based on contacts reported within these two groups. Limiting to these, the self-reports of relations within race were:

	Black	**White**
Main partnerships	94·8%	91·5%
Casual partnerships	90·2%	85·1%
One-time contacts	90·9%	89·4%

These numbers must balance in our model; that is, the reports by Black men and by White men each imply a specific expected number of cross-race relationships, given the overall numbers of relationships for each group. Although the implied numbers are similar, they are not exactly equal. To reconcile this, we calculated the expected cross-race relationships given each number above and all of the other relational statistics in the model, and took the midpoint of the two values. These equaled 93%, 88%, and 90%, for main, casual and one-time, respectively.

3·1·5 Age Mixing

Respondents also reported on the estimated age of each partner. We model age mixing within a given relational type and race combination using a single parameter for each, the expected mean difference in square root of the ages of men in a relationship, consistent with previous work.[1,3,6] For instance, a relationship between a 23-year-old and a 28-year-old would represent $\left|\sqrt{23} - \sqrt{28}\right|$ = 0·496.

	B-B Dyads	B-W Dyads	W-W Dyads
Main partnerships	0·417	0·454	0·520
Casual partnerships	0·498	0·629	0·632
One-time contacts	0·456	0·585	0·590

3·1·6 Mixing by Sexual Role

We assign men a fixed sexual role preference (exclusively insertive, exclusively receptive, versatile). The model then includes an absolute prohibition, such that two exclusively insertive men cannot partner, nor can two exclusively receptive men. Men's roles at last sex for each of the last 5 (Involvement) or 10 (MAN Project) partners were aggregated; those who had engaged in one role across all of those acts were deemed to be exclusively receptive or insertive, and those who had engaged in at least one act of each were deemed to be versatile.

	Black	White
Exclusively insertive	24·2%	22·8%
Versatile	43·7%	54·4%
Exclusively receptive	32·1%	22·8%

3·1·7. Partnership Durations

We model relational dissolution as a memoryless process with a single parameter per relational type and race combination. This implies an exponential distribution for relational durations within each category. As detailed in previous work,[1] for memoryless processes, the expected age of an extant relationship at any moment in time matches the expected uncensored duration of relationships, given the balancing effects of right-censoring and length bias for this distribution. To derive our values, we take the median of the observed distribution and then calculate the mean for the exponential distribution with that median. Duration was calculated as the difference between first and last sex date for each dyad the ego reported sex with more than once in the interval. The resulting expected relational durations were:

	B-B Dyads	B-W Dyads	W-W Dyads
Main partnerships	348 days	372 days	555 days
Casual partnerships	131 days	286 days	144 days

3·2 Statistical Representation of Sexual Networks

Exponential-family random graph models (ERGMs) and their dynamic extension separable temporal ERGMs (STERGMs) provide a foundation for statistically principled simulation of local and global network structure given a set of target statistics from empirical data. Main and casual relationships were modeled using STERGMs,[7] since they persist for multiple time steps. One-time contacts, on the other hand, were modeled using cross-sectional ERGMs.[8] Formally, our statistical models for relational dynamics can be represented as five equations for the conditional log odds (logits) of relational formation and persistence at time *t* (for main and casual relationships) or for relational existence at time *t* (for one-time contacts):

$$logit\left(P\left(Y_{ij,t}=1\middle|\, Y_{ij,t-1}=0,\, Y_{ij,t}^{C}\right)\right) = \theta_m^{+\prime}\partial\left(g_m^{+}(y)\right) \qquad \text{Main partnership formation}$$

$$logit\left(P\left(Y_{ij,t}=1\middle|\, Y_{ij,t-1}=0,\, Y_{ij,t}^{C}\right)\right) = \theta_c^{+\prime}\partial\left(g_c^{+}(y)\right) \qquad \text{Casual partnership formation}$$

$$logit\left(P\left(Y_{ij,t}=1\middle|\, Y_{ij,t-1}=1,\, Y_{ij,t}^{C}\right)\right) = \theta_m^{-\prime}\partial\left(g_m^{-}(y)\right) \qquad \text{Main partnership persistence}$$

$$logit\left(P\left(Y_{ij,t}=1\middle|\, Y_{ij,t-1}=1,\, Y_{ij,t}^{C}\right)\right) = \theta_c^{-\prime}\partial\left(g_c^{-}(y)\right) \qquad \text{Casual partnership persistence}$$

$$logit\left(P\left(Y_{ij,t}=1\middle|\, Y_{ij,t}^{C}\right)\right) = \theta_o{}'\partial\left(g_o(y)\right) \qquad \text{One-time contact existence}$$

where:

- $Y_{ij,t}$ = the relational status of persons *i* and *j* at time *t* (1 = in relationship/contact, 0 = not)
- $Y_{ij,t}^{C}$ = the network complement of *i,j* at time *t*, i.e. all relations in the network other than *i,j*
- $g(y)$ = vector of network statistics in each model
- θ = vector of parameters in the model

For $g(y)$ and θ, the superscript distinguishes the formation model (+), persistence model (-) and existence models (neither). The subscript indicates the main (m), casual (c) and one-time (o) models.

The recursive dependence among the relationships renders the model impossible to evaluate using standard techniques; we use MCMC in order to obtain the maximum likelihood estimates for the $\boldsymbol{\theta}$ vectors given the $\boldsymbol{g}(\boldsymbol{y})$ vectors.

Specific model statistics are listed below. Together these sets allow us to retain all of the network features listed in Section 3·1. it is important to note that, although the statistics are expressed here in terms of number of relationships and enter into the estimation model in this form, the simulation model is then parametrized using the resulting θ coefficients. This means that, as population size and composition changes, it is not the absolute number of relationships

of different kinds that will be preserved, but the relative numbers (e.g. the mean number of relationships per person). Similar conversions hold for the other statistics (e.g. the mean age difference per relationship is preserved, not the sum across all relationships).

Main partner formation model statistics: $\mathrm{g}_{\mathrm{m}}^{+}(\mathrm{y})$ vector:

- $g_{m1}^{+}(y)$ = number of main partnerships
- $g_{m2}^{+}(y)$ = number of Black-White main partnerships
- $g_{m3}^{+}(y)$ = number of White-White main partnerships
- $g_{m4}^{+}(y)$ = number of main partnerships for Black men with 1 casual partner
- $g_{m5}^{+}(y)$ = number of main partnerships for Black men with 2 casual partners
- $g_{m6}^{+}(y)$ = number of main partnerships for White men with 1 casual partner
- $g_{m7}^{+}(y)$ = number of main partnerships for White men with 2 casual partners
- $g_{m8}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across BB main partnerships
- $g_{m9}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across BW main partnerships
- $g_{m10}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across WW main partnerships

The main partner formation model also includes three constraints on the sample space: networks are prohibited unless they contain no edges between two men who are both exclusively insertive; no edges between two men who are both exclusively receptive; and no men with more than 1 main partner simultaneously

Main partner persistence model terms: $\mathrm{g}_{\mathrm{m}}^{-}(\mathrm{y})$ vector:

- $g_{m1}^{-}(y)$ = number of main partnerships
- $g_{m2}^{-}(y)$ = number of Black-White main partnerships
- $g_{m3}^{-}(y)$ = number of White-White main partnerships

Casual partner formation model terms: $\mathrm{g}_{\mathrm{c}}^{+}(\mathrm{y})$ vector:

- $g_{c1}^{+}(y)$ = number of casual partnerships
- $g_{c2}^{+}(y)$ = number of Black-White casual partnerships
- $g_{c3}^{+}(y)$ = number of White-White casual partnerships
- $g_{c4}^{+}(y)$ = number of casual partnerships for Black men with 1 main partner
- $g_{c5}^{+}(y)$ = number of casual partnerships for White men with 1 main partner

- $g_{c6}^{+}(y)$ = number of Black men with 2 casual partners
- $g_{c7}^{+}(y)$ = number of White men with 2 casual partners
- $g_{c8}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across BB casual partnerships
- $g_{c9}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across BW casual partnerships
- $g_{c10}^{+}(y)$ = sum of the absolute difference in the square root of partners' ages across WW casual partnerships

The casual partner formation model also includes three constraints on the sample space: networks are prohibited unless they contain no edges between two men who are both exclusively insertive; no edges between two men who are both exclusively receptive; and no men with more than 2 casual partners simultaneously

Casual partner persistence model terms: $g_c^{-}(y)$ vector:

- $g_{c1}^{-}(y)$ = number of casual partnerships
- $g_{c2}^{-}(y)$ = number of Black-White casual partnerships
- $g_{c3}^{-}(y)$ = number of White-White casual partnerships

One-time contact existence model terms: $g_o(y)$ vector:

- $g_{o1}(y)$ = number of one-time contacts
- $g_{o2}(y)$ = total # of one-time contacts for Black men with 0 main and 1 casual partnership
- $g_{o3}(y)$ = total # of one-time contacts for Black men with 0 main and 2 casual partnerships
- $g_{o4}(y)$ = total # of one-time contacts for Black men with 1 main and 0 casual partnerships
- $g_{o5}(y)$ = total # of one-time contacts for Black men with 1 main and 1 casual partnership
- $g_{o6}(y)$ = total # of one-time contacts for Black men with 1 main and 2 casual partnerships
- $g_{o7}(y)$ = total # of one-time contacts for White men with 0 main and 0 casual partnerships
- $g_{o8}(y)$ = total # of one-time contacts for White men with 0 main and 1 casual partnership
- $g_{o9}(y)$ = total # of one-time contacts for White men with 0 main and 2 casual partnerships

- $g_{o10}(y)$ = total # of one-time contacts for White men with 1 main and 0 casual partnerships
- $g_{o11}(y)$ = total # of one-time contacts for White men with 1 main and 1 casual partnership
- $g_{o12}(y)$ = total # of one-time contacts for White men with 1 main and 2 casual partnerships
- $g_{o13}(y)$ = total # of one-time contacts for Black men in risk quintile 1
- $g_{o14}(y)$ = total # of one-time contacts for Black men in risk quintile 2
- $g_{o15}(y)$ = total # of one-time contacts for Black men in risk quintile 4
- $g_{o16}(y)$ = total # of one-time contacts for Black men in risk quintile 5
- $g_{o17}(y)$ = total # of one-time contacts for White men in risk quintile 1
- $g_{o18}(y)$ = total # of one-time contacts for White men in risk quintile 2
- $g_{o19}(y)$ = total # of one-time contacts for White men in risk quintile 4
- $g_{o20}(y)$ = total # of one-time contacts for White men in risk quintile 5
- $g_{o21}(y)$ = total # of race-homophilous one-time contacts
- $g_{o22}(y)$= sum of the absolute difference in the square root of partners' ages across BB one-time contacts
- $g_{o23}(y)$= sum of the absolute difference in the square root of partners' ages across BW one-time contacts
- $g_{o24}(y)$= sum of the absolute difference in the square root of partners' ages across WW one-time contacts

The one-time existence model also includes two constraints on the sample space: networks are prohibited unless they contain no edges between two men who are both exclusively insertive; and no edges between two men who are both exclusively receptive.

Our method of converting the statistics laid out in Section 3·1 into our fully specified network models consists of the following steps:

1. Construct a cross-sectional network of 10,000 men with no relationships
2. Assign men sexual roles based on prevalences listed in Section 3·1·6, as well as one-time risk quintiles (20% of the men in each race per quintile)
3. Calculate the target statistics (i.e., the expected count of each statistic at any given moment in time) associated with the terms in the formation model (for the main and casual partnerships) and in the existence model (for one-time contacts).

4. Assign each node a place-holder main and casual degree (number of on-going partnerships) that is consistent with the race-specific matrices, and store these numbers as a nodal attribute. (Note: this does not actually require individuals to be paired up into the partnerships represented by those degrees).
5. For the main and casual networks, use the mean relational durations by race combination to calculate the parameters of the persistence model, using closed-form solutions, given that the models are dyadic-independent (each relationship's persistence probability is independent of all others).
6. For the main and casual networks, estimate the coefficients for the formation model that represent the maximum likelihood estimates for the expected cross-sectional network structure.
7. For the one-off network, estimate the coefficients for the existence model that represent the maximum likelihood estimates for the expected cross-sectional network structure.

Steps 5-7 occur within the *Statnet* software, and use the ERGM and STERGM methods therein. They are made most efficient by the use of an approximation in Step 6.[9] During the subsequent model simulation, we use the method of Krivitsky et al.[10] to adjust the coefficient for the first term in each model at each time step, in order to preserve the same expected mean degree (relationships per person) over time in the face of changing network size and nodal composition. At all stages of the project, simulated partnership networks were checked to ensure that they indeed retained the expected cross-sectional structure and relational durations throughout the simulations.

4 BEHAVIOR WITHIN SEXUAL PARTNERSHIPS

We model four phenomena consecutively within relationships at each time step: HIV+ status disclosure, number of anal sex acts, condom use per sex act, and sexual role per sex act. We model these only within the relationships in our network that are HIV-discordant (whether diagnosed or not), given that only they may lead to potential transmissions.

4·1 Disclosure

We model the process by which someone who knows he is HIV-positive discloses this fact to partners of all types. Disclosure affects subsequent decision-making around condom use. We do not explicitly model other forms of serostatus discussion, since our source data do not include these all; our behavioral estimates in the absence of HIV+ disclosure marginalize over those cases in which men disclose as concordant negative and do not discuss at all. Disclosure

may occur at the point of a relation commencing (if HIV+ status is already known) or it may occur at the point of diagnosis, in the case of on-going relationships. In the former case, disclosure of HIV+ status was determined from the combined dyadic dataset using the HIV status of the respondent and their response to the question, “Did you and this partner share both of your HIV statuses before you first had sex?” In the latter case, we did not have data and assumed it to be universal.

Probability of Disclosure of HIV+ Status	**Black**	**White**
to new main partner at outset of relationship	0·685	0·889
to new casual partner at outset of relationship	0·527	0·828
to one-time contact	0·445	0·691
to ongoing partner if diagnosis occurs during relationship	1·000	1·000

4·2 Number of AI Acts

The number of anal sex acts per week for each ongoing relationship is determined from a Poisson draw, with mean specific to the relational type and race combination of the dyad. For one-time contacts, the number is set deterministically to 1 for the time step in which it occurs.

AI Acts/Week/Partnership	**B-B Dyads**	**B-W Dyads**	**W-W Dyads**
Main partnerships	1·19	1·79	1·56
Casual partnerships	0·75	1·13	0·98

These rates were calculated based on the two Atlanta studies, derived from questions asking the number of coital acts per partnership during the recall periods.[11,12] These were then rescaled from the length of the recall period into the weekly rates listed in the table above.

4·3 Condom use

We conducted logistic regressions to identify the significant predictors of condom use within HIV-discordant relationships (whether diagnosed or not) in our data. Respondents were asked if they had had unprotected anal sex with each partner during the recall periods. Predictors included the race combination of the men in the relationship, the type of relationship, the HIV diagnosis status of the HIV+ partner (i.e. whether or not he himself knew that he was HIV+), and the disclosure status of the HIV+ partner (whether he had told his partner he was HIV+).

Predictors that dropped out of the model included sexual position and perceived monogamy of the partnership.

Base model coefficients for the nine race/partnership types were defined as logit(P(condom use|anal intercourse) =

	B-B Dyads	B-W Dyads	W-W Dyads
Main partnership	-0·49	-2·20	-1·73
Casual partnership	-0·45	-2·09	-1·66
One-time contact	-0·04	-1·73	-1·27

Note that for these, the reference category is the case in which the HIV+ man is undiagnosed, hence the relatively low values of condom use. Modifiers for these logit coefficients are:

Condition	Coefficient
HIV+ diagnosis	0·67
HIV+ status disclosure	0·85

Together, these values, in combination with the frequencies with which AI occurs in all of the different types of situations, implies an overall rate of condom use average across all acts of approximately 50%.

4·4 Sexual role

Men are assigned an individual sexual role preference (exclusively insertive, exclusively receptive, or versatile) as described in Section 3·1·6. Relationships between two exclusively insertive or two exclusively receptive men are prohibited via the ERGM and STERGM models. Versatile men are further assigned an insertivity preference drawn from a uniform distribution between 0 and 1. When two versatile men are determined to have an AI act, their sexual positions must be determined (all other combinations have only one feasible combination). One option is for men to engage in intra-event versatility (IEV; i.e. both engage in insertive and receptive AI during the act). The probability of this is specific to the race combination of the men (Black-Black = 42%, Black-White = 56%, White-White = 49%), and is derived from the partner-specific role data described in Section 3·1·6. If IEV does not occur, then each man's probability of being the insertive partner equals his insertivity quotient divided by the sum of the two men's insertivity quotients.

5 DEMOGRAPHY

In this model, there are three demographic processes: entries, exits, and aging. Entries and exits are conceptualized as flows to and from the sexually active population of interest: MSM aged 18 to 40 years old. Entry into this population represents the time at which persons become at risk of infection via male-to-male sexual intercourse, and we model these flows as starting at an age after birth (age 18) and ending at an age potentially before death (age 40).

5·1 Entry at Sexual Onset

All persons enter the network at age 18, which was the lower age boundary of our two main source studies. The number of new entries at each time step is based on a fixed rate (10 per 10,000 persons per weekly time step) that keeps the overall network size in a stable state over the time series of the simulations. The model parameter governing this rate was calibrated iteratively in order to generate simulations with a population size at equilibrium, given the inherent variability in population flows related to background mortality, sexual maturation (i.e., reaching the upper age limit of 40), and disease-induced mortality. At each time step, the exact number of men entering the population was simulated by drawing from a Poisson distribution with the rate parameter.

5·2 Initialization of Attributes

Persons entering the population were assigned attributes, some of which remained fixed by definition (e.g., race), others fixed by assumption (e.g., insertive versus receptive sexual role), and yet others allowed to vary over time (e.g., age and disease status). Here we describe three attributes in the first category:

- For **race/ethnicity**, this model was based on a population composition that was 50% black MSM and 50% white MSM. This 1:1 ratio comes close to that for the Atlanta metropolitan area and also provides analytical clarity.
- **Circumcision** status was randomly assigned to incoming men. Based on empirical data from Atlanta MSM,[11] 89·6% of men were circumcised before sexual onset. Circumcision was associated with a 60% reduction in the per-act probability of infection for HIV- males for insertive anal intercourse only (i.e., circumcision did not lower the *transmission* probability if the HIV+ partner was insertive).[2,13]
- The **CCR5-Δ32 genotype** was modeled by assigning a value of zero, one, or two Δ32 alleles. Compared to men without the Δ32 allele, heterozygous men (those with Δ32 alleles) were 70% less likely to become infected and Δ32 homozygous men (those with two Δ32 alleles) were fully immune from infection.[14,15] The population distribution of

genotypes was differential by race, with 0% of black men and 3·4% of white men expressing as Δ32 homozygous, and 2·1% of black men and 17·6% of white men expressing as Δ32 heterozygous.[14]

5·3 Exits from the Network

All persons exited the network by age 40, either from mortality or by reaching the upper age bound of the MSM target population of interest. This upper limit of 40 was modeled deterministically (probability = 1), but other exits due to mortality were modeled stochastically. Mortality included both natural (non-HIV) and disease-induced mortality causes before age 40. Background mortality rates were based on US all-cause mortality rates specific to age and race from the National Vital Statistics life tables.[16] The following table shows the probability of mortality per year by age and race.

Age	White	Black
18–24	0·00103	0·00159
25–34	0·00133	0·00225
35–39	0·00214	0·00348

Natural mortality was applied to persons within the population at each time step stochastically by drawing from a binomial distribution for each eligible person with a probability parameter corresponding to that person's risk of death tied to his age. Disease-related mortality, in contrast, was modeled based on clinical disease progression, as described in Section 6.

5·4 Aging

The aging process in the population was linear by time step for all active persons. The unit of time step in these simulations was one week, and therefore, persons were aged in weekly steps between the minimum and maximum ages allow (18 and 40 years old). Evolving age impacted background mortality, age-based mixing in forming new partnerships, and other behavioral features of the epidemic model described below. Persons who exited the network were no longer active and their attributes such as age were no longer updated.

6 INTRAHOST EPIDEMIOLOGY

Intrahost epidemiology includes features related to the natural disease progression within HIV+ persons in the absence of clinical intervention. The main component of progression that was explicitly modeled for this study was HIV viral load. In contrast to other modeling studies that

model both CD4 and viral load, our study used viral load progression to control both interhost epidemiology (HIV transmission rates) and disease progression eventually leading to mortality.

Following prior approaches,[1,2] we modeled changes in HIV viral load to account for the heighted viremia during acute-stage infection, viral set point during the long chronic stage of infection, and subsequent rise of VL at clinical AIDS towards disease-related mortality. The HIV viral load has a crucial impact on the rates of HIV transmission within serodiscordant pairs in the model, and this interaction is detailed in Section 8. A starting viral load of 0 copies per mL is assigned to all persons upon infection. From there, the natural viral load curve is fit with the following parameters:

Parameter	Value	Reference
Time to peak viremia in acute stage	45 days	Little[17]
Level of peak viremia	6·886 $\log_{10}$ copies per mL	Little[17]
Time from peak viremia to viral set point	45 days	Little,[17] Leynaert[18]
Level of viral set point	4·5 $\log_{10}$ copies per mL	Little[17]
Duration of chronic stage infection (no ART)	3550 days	Buchbinder,[19] Katz[20]
Duration of AIDS stage	728 days	Buchbinder[19]
Peak viral load during AIDS (at death)	7 $\log_{10}$ copies per mL	Estimated from average duration of AIDS

After infection, it takes 45 days to reach peak viremia, at a level of 6·886 log 10 copies per mL. From peak viremia, it takes another 45 days to reach viral set point, which is set at a level of 4·5 log 10 copies per mL. The total time of acute stage infection is therefore 3 months. The duration of chronic stage infection in the absence of clinical intervention is 3550 days, or 9·7 years. The total duration of pre-AIDS disease from infection is therefore approximately 10 years. At onset of AIDS, HIV viral load rises linearly from 4·5 log 10 copies per mL to 7 log 10 copies per mL, at which point mortality is assumed to occur. The time spent in the AIDS stage is 728 days, or 2 years. This viral load trajectory is for ART-naïve persons only, and the influence of ART on disease progression is detailed in Section 7. These transitions are deterministic for all ART-naïve persons.

7 CLINICAL EPIDEMIOLOGY

Clinical epidemiological processes refer to all steps along the HIV care continuum after initial infection: diagnosis, linkage to care, treatment initiation and adherence, and HIV viral load suppression. In this model, these clinical features have critical interactions with behavioral features detailed above, as well as impacts on the rates of HIV transmission, detailed below. The features of our model's clinical processes generally follow the steps of the HIV care continuum, in which persons transition across states from infection to diagnosis to medical care linkage and ART initiation to HIV viral suppression.[21]

7·1 HIV Diagnostic Testing

Persons in our models were divided into non-testers (through age 40) and regular interval-based testers. Based on empirical data for Atlanta MSM,[11] 6·5% of MSM did not receive HIV testing before age 40. This was calculated based on a survey about never tested prior to the study, which may overestimate the final proportion who would have never tested before age 40. A fixed individual attribute for HIV treatment trajectories that characterized progression through the care continuum was randomly assigned upon entry into the population, with this group of 6·5% of MSM not accessing HIV testing or other forms of post-diagnostic HIV medical services.

The remaining 93·5% who entered the HIV care continuum HIV tested at regular intervals, with the estimated mean time between tests for HIV-negative persons at 301 days for black MSM and 315 days for white MSM.[11,22] This was calculated based on time since last test in the survey, with the assumption that testing was a memoryless process. Diagnostic testing was simulated stochastically using draws from a binomial distribution with probability parameters equal to the reciprocal of this interval. This generated a population-level geometric distribution of times since last test.

We also modeled a 21-day window period after infection during which the tests of the truly HIV+ persons would show as negative to account for the lack of antibody response immediately after infection.[23] HIV+ persons who tested after this window period would be correctly diagnosed with 100% test sensitivity. Individual-level attributes for diagnosis status and time since last HIV test were recorded for all MSM.

7·2 Antiretroviral Therapy (ART) Initiation

Consistent with previous models,[1,2] we simulated the initiation of ART and subsequent clinical outcomes of full or partial HIV viral suppression based on men being in one of three clinical states: never tested, on treatment and partially virally suppressed, and on treatment with full

viral suppression. There was insufficient empirical data to represent the patterns and rates at which individual men switch among these three states over the course of their infection, since the clinical ART landscape is constantly evolving. Therefore, we modeled men as being on one of the three fixed treatment trajectories as an individual-level attribute such that our model matched the population-level data on the prevalence of durable HIV viral suppression and treatment-naïve mortality.[24,25]

Following HIV diagnosis (for the 93·5% of men who ever HIV test before age 40), MSM initiated treatment at a rate of 0·924 per week (Black MSM) or 0·1271 (White MSM). This translates into an average interval between testing and treatment initiation of 10·8 weeks (Black MSM) and 7·9 weeks (White MSM), consistent with empirical data.[22] In the absence of quantitative data, we assumed no gap between treatment entry and ART initiation.

7·3 ART Adherence and Viral Suppression

MSM who initiated ART could cycle on and off treatment, where cycling off treatment resulted in an increase in the VL back up to the assumed set point of 4·5 $\log_{10}$. The slope of changes to VL were calculated such that it took a total of 3 months to transition between the set point and the on-treatment viral loads.[26] Men on treatment could achieve partial or full suppression. Men with partial suppression were assumed to have a viral load of 3·5 $\log_{10}$ copies per mL, compared to 1·5 $\log_{10}$ copies per mL among those who were fully suppressed.[26] The latter corresponds to an absolute viral load below the standard levels of detection (VL = 50 copies per mL).[27]

The patterns of ART adherence leading to partial and full HIV viral suppression were estimated based on an analysis of HIV care patterns among MSM in the United States,[24] which was required in order to obtain parameters that were specific to young MSM by race. Parameterizing our model used three types of inputs: (1) the proportion of those diagnosed who are on ART; (2) the proportion of those diagnosed who are virally suppressed; (3) the level of durable suppression (proportion on ART who have been suppressed for a year). Our source included recent estimates for (1) by race and by age, but not the interaction of the two. We used a weighted average of their 18-29 and 30-39 year old data, and assumed that the overall prevalence ratio by race that they observed for each outcome held within this age group as well. This suggested that 30·2% of young Black MSM who were diagnosed were in care, and 74·0% of those were on ART, for a combined value of 22·3% of young Black MSM who were diagnosed being on ART at any time point. Analogous figures for young White MSM were 46·6%, 84·0% and 39·2%. For (3), we used the same method of deriving estimates specific to young Black MSM (47% of those on ART are durably suppressed) and young White MSM (60%

for the corresponding figure). For (2), we used figures by race from the same paper; however, similar figures by age were not included. Instead, we adjusted by using the relative rates of retention in care and suppression for young adults (25-44) compared to all respondents from an additional analysis of the care continuum for members of all risk groups (not just MSM-specific) in the US.[28] This yielded estimates for the percent of young MSM on ART who are virally suppressed of 62·4% for Blacks and 67·7% for Whites.

None of these three sets of values entered the model directly as inputs. Parameter (3) was converted into a per-time step probability of falling out of suppression, by using the inverse geometric function to calculate the probability consistent with observed levels of durable suppression after 1 year. Our other two input parameters were the proportion of those initiating ART who achieved full suppression, and the per-time step probability of re-achieving suppression after one had previously fallen out. We simulated our full model iteratively until we identified the unique values of these parameters by race that yielded the values estimated for parameters (1) and (2) above. The resulting set of model inputs were:

Parameter	Black MSM	White MSM
Proportion of those initiating ART who achieved full suppression	0·614	0·651
Per-time step probability of falling out of suppression	0·0102	0·0071
Per-time step probability of re-achieving suppression	0·00066	0·00291

7·4 Disease Progression and Mortality after ART Initiation

Mortality after ART initiation was modeled based on the cumulative time on and off ART for persons who were fully or partially suppressed. The maximum time between infection and the start of AIDS was 9·7 years.[19] If a person in either the full or partial suppression categories who spent this much time off ART during the course of infection progressed to AIDS. For the partially suppressed, we assumed a maximum time on ART of 15 years, similar to previous models, to account for treatment failure.[1] For this group, the time to AIDS was an additive function of two ratios: (time on treatment / maximum time on treatment) + (time off treatment / maximum time off treatment). AIDS was simulated to begin when the sum of this score exceeded 1. Persons who had ever initiated ART progressed through AIDS at a similar rate as those who were ART-naïve.

8 INTERHOST EPIDEMIOLOGY

Interhost epidemiological processes represent the HIV-1 disease transmission within the model. Disease transmission occurs between sexual partners who are active on a given time step. This section will describe how the overall rate is calculated as a function of the intrahost epidemiological profile of each member of a partnership, and behavioral features within the dyad.

8·1 Disease-Discordant Dyads

At each time step in the simulation, a list of active dyads was selected based on the current composition of the network. This was called an "edgelist." Given the three types of partnerships detailed above, the full edgelist was a concatenation of the type-specific sublists. The complete edgelist reflects the work of the STERGM- and ERGM-based network simulations, wherein partnerships formed on the basis of nodal attributes and degree distributions (see Section 2). Dyads active were considered active at a specific time step if the terminus of that simulated edge was less than or equal to the current time step (right-censored). From the full edgelist, a disease-discordant subset was created by removing those dyads in which both members were HIV- or both were HIV+. This left dyads that were discordant with respect to HIV status, which was the set of potential partnerships over which infection may be transmitted at that time step.

8·2 Per-Act HIV Transmission Probability

Within disease-discordant dyads, HIV transmission was modeled based on a sexual act-by-act basis, in which multiple acts of varying infectiousness could occur within one partnership within a weekly time step. Determination of the number of acts within each discordant dyad for the time step, as well as condom use and role for each of those acts, was described in Section 3. Transmission by act was then modeled as a stochastic process for each discordant sex act following a binomial distribution with a probability parameter that is a multiplicative function of the following predictors of the HIV- and HIV+ partners within the dyad.

Predictor	Partner	Parameters	References
Sexual role (insertive or receptive)	HIV-	*Receptive:* 0·008938 base probability when HIV+ partner has 4·5 $\log_{10}$ viral load copies per mL	Vittinghoff[29]
		Insertive: 0·003379 base probability when HIV+ partner has 4·5 $\log_{10}$ viral load copies per mL	Vittinghoff[29]
HIV viral load (VL)	HIV+	Multiplier of $2{\cdot}45^{(VL - 4{\cdot}5)}$, where VL is expressed as $\log_{10}$ copies per mL	Wilson[30]

Acute stage	HIV+	Multiplier of 6	Leynaert,[18] Bellan[31]
CCR5 status	HIV-	Δ32 homozygote: multiplier of 0	Marmor[14]
		heterozygote: multiplier of 0·3	Marmor[14]
Condom use	Both	Multiplier of 0·25	Varghese,[32] Weller[33], Smith[34]
Circumcision status	HIV-, insertive	Multiplier of 0·40	Gray[13]

For each act, the overall transmission probability was determined first with a base probability that was a function of whether the HIV- partner was in the receptive or insertive role, with the former at a 2·6-fold infection risk compared to the latter. The HIV+ partner's viral load modifies this base probability in a non-linear formulation, upwards if the VL was above the VL set point during chronic stage infection in the absence of ART, and downwards if it was below the set point. Following others, we modeled an excess transmission risk in the acute stage of infection above that predicted by the heightened VL during that period. Three covariates of the HIV- partner could reduce the risk of infection: the Δ32 allele on the CCRR5 gene, condom use within the act, circumcision status (only if the HIV- partner was insertive in that act).

The final transmission rate per partnership per weekly time step was a function of the per-act probability of transmission in each act and the number of acts per time step. The per-act transmission probability could be heterogeneous within a partnership due to various types of acts in each interval: for example, a HIV- man who is versatile in role may have both insertive and receptive intercourse within a single partnership; some acts within a partnership may be protected by condom use while others are condomless. Transmission was simulated for each act within each serodiscordant dyad, based on draws from a binomial distribution with the probability parameter equal to the per-act transmission probabilities detailed above.

9 ADDITIONAL RESULTS: SUSTAINING VS. GENERATING DISPARITY

One focus of this work is to probe the relationship between *sustaining* disparities and *generating* disparities. That is, if we know that a given disparity by race in some causal factor (e.g. sexual behavior, care continuum) generates a given race disparity in HIV incidence or prevalence, what does that tell us about the ability for that same causal factor to *sustain* a pre-existing disparity in incidence and prevalence over time?

We are motivated to ask this question for two reasons. One is the fact that existing differences in HIV prevalence by race are used in the literature as explanatory factors for differences in HIV incidence. This argument states that, given high levels of race assortative mixing, and higher

existing prevalence among Black MSM than White MSM, HIV-negative Black MSM are more likely than HIV-negative White MSM to have HIV-positive partners, and thus to acquire HIV infection. Indeed, partner race is a strong explanatory factor in incidence studies.[11,35] If a pre-existing disparity plus assortative mixing together can explain *short-term* differences in HIV incidence, then it would likely seem reasonable for many readers of this literature to imagine that the same phenomena might be sufficient to sustain disparities indefinitely and thus provide the explanation for *long-term* disparities.

However, modeling theory suggests that the short-term explanatory power of a pre-existing disparity combined with assortative mixing does not generally translate into long-term explanatory power.[36] This theory involves complex mathematics such as Lyapunov functions, which makes it accessible only to a relatively small audience. Moreover, while it has been proven for more simple models of HIV infection involving two groups, it is worth confirming that the same holds true for our more realistic, and thus more complex, epidemiological model. Thus, our second reason for exploring this topic is to confirm the theory for our model, and to demonstrate the implications to a wider audience.

A basic summary of the theory is as follows. Consider an infection that exhibits the basic features of HIV (e.g. no recovery, multiple stages of infection varying in transmissibility, infection-induced mortality), and a population containing two subgroups that preferentially partner within subgroup. Select a set of parameters guiding each aspect of the transmission system (e.g. the contact rates for each subgroup, the level of assortative mixing, the transmissibility by stage, the death rates by subgroup). Simon and Jacquez (1992)[36] prove that, for any set of parameters chosen, there is a unique pair of values representing HIV prevalence for each subgroup that the system will move towards, regardless of what the current levels of HIV prevalence are in the two subgroups. This unique pair of values is called the “endemic equilibrium.” (The only exception is if both subgroups have no infection; in this case, the epidemic cannot take off, and it instead remains at the “disease-free equilibrium”).

Imagine, then, the case where the two subgroups are exactly equal in all ways, including current HIV prevalence, but they exhibit assortative mixing. This situation should not generate a disparity, since the groups have the same HIV prevalence, and are otherwise identical in every way. This would suggest that the unique equilibrium for this scenario is one in which the two groups have no HIV disparity. The theory tells us, then, that if the only thing we change is the pre-existing disparity – that is, if Black MSM have higher HIV prevalence and White MSM have lower HIV prevalence – the equilibrium is still the same, and the population will move towards

no disparity. Ceteris paribus, assortative mixing *alone* cannot sustain any disparity at all in the long term.

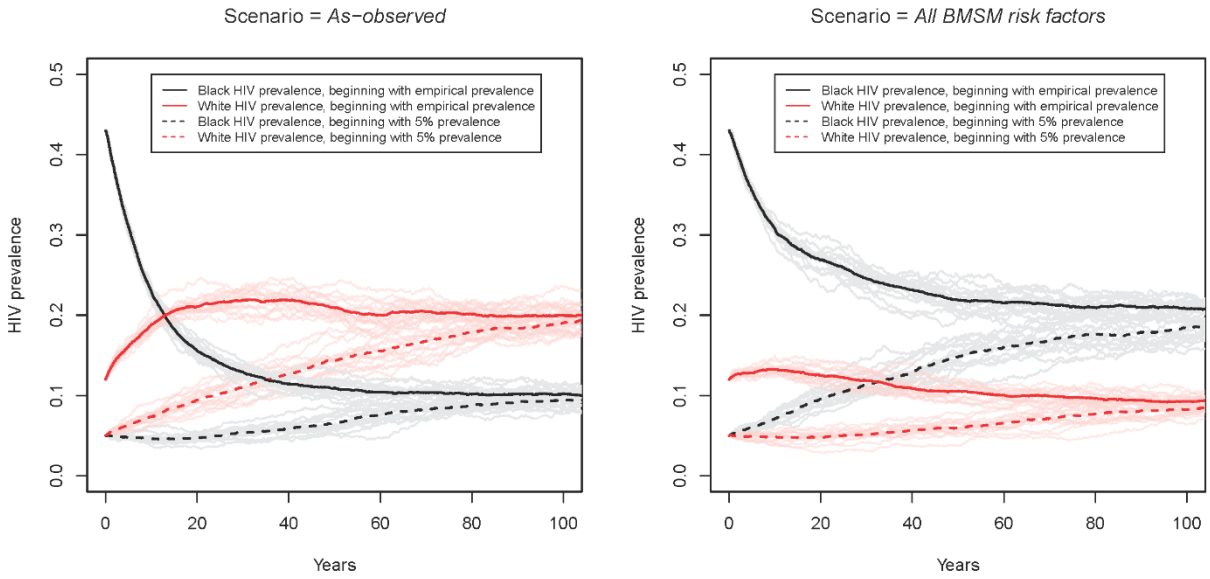
Moreover, any other scenario considered – involving an arbitrary set of differences between Black and White MSM, whether behavioral, clinical, demographic or otherwise, should lead to a specific disparity in HIV burden; the magnitude of the current disparity does not change the ultimate disparity. That is, the ability for any factor to *generate* a disparity is identical to its ability to *sustain* a pre-existing one indefinitely.

As mentioned above, the theory is based on mathematical proofs that make various simplifying assumptions, and the model in our manuscript is much more complicated, so much so that the theory cannot be proven mathematically in this case. However, we can use simulation to demonstrate that the effect is still observed for any given scenario of interest.

Figure 1a in the main text demonstrates that the theory does indeed hold for the case where Black and White MSM are identical, but exhibit strong assortative mixing (~90% of contacts are within-race), and a large pre-existing disparity (43% HIV prevalence for Black MSM and 13% for White MSM). Although in the short term, HIV incidence is higher among Black MSM than White MSM, this is not sustained, and both incidence and prevalence begin to converge noticeably within a few years. This behavior is in contrast to that observed empirically, in that racial disparities in HIV have been measured over decades, and have in general not been shown to be narrowing. Eventually they reach the unique endemic equilibrium for this scenario, in which Black and White HIV prevalence converge and the disparity disappears.

To demonstrate that analogous behavior occurs for other scenarios, we consider our *All observed* model, i.e. the one in which all parameters are set to the race-specific values observed in our data. Figure 1c in the main text shows the time series of HIV prevalence for Black and White MSM for this model, beginning with observed HIV prevalence. In Figure S1 below, we repeat this figure in the left panel (albeit with a different scale on the x-axis), and overlay with it the prevalence trajectories that occur with the same scenario begun at 5% prevalence for each group. We see that Black and White prevalence each head to a specific value over time, which remains the same regardless of their initial values. The theory appears to hold; for our more complex model there remains a single equilibrium defined by two race-specific prevalence values. Thus, the disparity that a given risk factor can generate is the same as the one it can sustain. To provide yet more confirmation, we repeat the exercise with our *All BMSM risk factor* scenario (Figure S1, right panel), and again see the expected behavior.

FIGURE S1



It is worth repeating that, although time to fully achieve equilibrium may be very long (on the order of 100 years, far longer than the HIV epidemic has existed), the measurable convergence of the two prevalence levels in the direction of that equilibrium occurs immediately. Thus, if past, unmeasured factors generated a racial disparity in HIV between Black and White MSM, and current measured factors reflect any kind of change from those past unmeasured ones, we should expect to see HIV prevalence move in the direction sustainable by the new factors rapidly. In practice, however, HIV disparities between Black and White MSM have persisted for decades.

In the main paper, we ran our scenarios for Question 2 beginning with 5% prevalence for each race since we defined our question in terms of generating disparities. The results in this section confirm that our main results are in fact not dependent on this choice; if we had begun the scenarios at observed prevalence, or indeed any other non-zero prevalence, we would have headed to the same disparity.

10 REFERENCES

1. Goodreau SM, Carnegie NB, Vittinghoff E, et al. What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? *PLoS One*. 2012;7(11):e50522.
2. Goodreau SM, Carnegie NB, Vittinghoff E, et al. Can male circumcision have an impact on the HIV epidemic in men who have sex with men? *PLoS One*. 2014;9(7):e102960.
3. Carnegie NB, Goodreau SM, Liu A, et al. Targeting pre-exposure prophylaxis among men who have sex with men in the United States and Peru: partnership types, contact rates,

and sexual role. *J Acquir Immune Defic Syndr*. 2015;69(1):119-125.
4. Khanna AS, Goodreau SM, Gorbach PM, Daar E, Little SJ. Modeling the impact of post-diagnosis behavior change on HIV prevalence in Southern California men who have sex with men (MSM). *AIDS Behav*. 2014;18(8):1523-1531. doi:10.1007/s10461-013-0646-2.
5. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *J Stat Softw*. 2008;24(1):1548-7660.
6. Sullivan PS, Carballo-Diéguez A, Coates T, et al. Successes and challenges of HIV prevention in men who have sex with men. *Lancet*. 2012;380(9839):388-399.
7. Krivitsky PN, Handcock MS. A Separable Model for Dynamic Networks. *J R Stat Soc Ser B Stat Methodol*. 2014;76(1):29-46.
8. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *J Stat Softw*. 2008;24(3):nihpa54860.
9. Carnegie NB, Krivitsky PN, Hunter DR, Goodreau SM. An approximation method for improving dynamic network model fitting. *J Comput Graph Stat*. 24(2):502-519.
10. Krivitsky PN, Handcock MS, Morris M. Adjusting for Network Size and Composition Effects in Exponential-Family Random Graph Models. *Stat Methodol*. 2011;8(4):319-339.
11. Sullivan PS, Rosenberg ES, Sanchez TH, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. *Ann Epidemiol*. 2015;25(6):445-454.
12. Grey JA, Rothenberg RB, Sullivan PS, Rosenberg ES. Disassortative Age-Mixing Does Not Explain Differences in HIV Prevalence between Young White and Black MSM: Findings from Four Studies. *PLoS One*. 2015;10(6):e0129877.
13. Gray RH, Kigozi G, Serwadda D, et al. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet*. 2007;369(9562):657-666.
14. Marmor M, Sheppard HW, Donnell D, et al. Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. *J Acquir Immune Defic Syndr*. 2001;27(5):472-481.
15. Zimmerman PA, Buckler-White A, Alkhatib G, et al. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol Med*. 1997;3(1):23-36.
16. United States Census Bureau. *Mortality Data*.; 2012.
17. Little SJ, McLean AR, Spina CA, Richman DD, Havlir D V. Viral dynamics of acute HIV-1 infection. *J Exp Med*. 1999;190(6):841-850.
18. Leynaert B, Downs AM, de Vincenzi I. Heterosexual transmission of human immunodeficiency virus: variability of infectivity throughout the course of infection. European Study Group on Heterosexual Transmission of HIV. *Am J Epidemiol*. 1998;148(1):88-96.
19. Buchbinder SP, Katz MH, Hessol NA, O'Malley PM, Holmberg SD. Long-term HIV-1 infection without immunologic progression. *AIDS*. 1994;8(8):1123-1128.
20. Katz MH, Hessol NA, Buchbinder SP, Hirozawa A, O'Malley P, Holmberg SD. Temporal trends of opportunistic infections and malignancies in homosexual men with AIDS. *J Inf Dis*. 1994;170(1):198-202.
21. Mugavero MJ, Amico KR, Horn T, Thompson MA. The state of engagement in HIV care in the United States: from cascade to continuum to control. *Clin Infect Dis*. 2013;57(8):1164-1171.
22. Rosenberg ES, Millett GA, Sullivan PS, Del Rio C, Curran JW. Understanding the HIV disparities between black and white men who have sex with men in the USA using the HIV care continuum: a modeling study. *Lancet HIV*. 2014;1(3):e112-e118.

23. Fiebig EW, Wright DJ, Rawal BD, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS*. 2003;17(13):1871-1879.
24. Beer L, Oster AM, Mattson CL, Skarbinski J. Disparities in HIV transmission risk among HIV-infected black and white men who have sex with men, United States, 2009. *AIDS*. 2014;28(1):105-114.
25. Bertolli J, Shouse RL, Beer L, et al. Using HIV surveillance data to monitor missed opportunities for linkage and engagement in HIV medical care. *Open AIDS J*. 2012;6:131-141.
26. Chu H, Gange SJ, Li X, et al. The effect of HAART on HIV RNA trajectory among treatment-naive men and women: a segmental Bernoulli/lognormal random effects model with left censoring. *Epidemiology*. 2010;21 Suppl 4:S25-S34.
27. Chun T-W, Carruth L, Finzi D, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997;387(6629):183-188.
28. Hall HI, Frazier EL, Rhodes P, et al. Differences in human immunodeficiency virus care and treatment among subpopulations in the United States. *JAMA Intern Med*. 2013;173(14):1337-1344.
29. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *Am J Epidemiol*. 1999;150(3):306-311.
30. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM. Relation between HIV viral load and infectiousness: a model-based analysis. *Lancet*. 2008;372(9635):314-320.
31. Bellan SE, Dushoff J, Galvani AP, Meyers LA. Reassessment of HIV-1 acute phase infectivity: accounting for heterogeneity and study design with simulated cohorts. *PLoS Med*. 2015;12(3):e1001801.
32. Varghese B, Maher JE, Peterman TA, Branson BM, Steketee RW. Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use. *Sex Transm Dis*. 2002;29(1):38-43.
33. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. *Cochrane Database Syst Rev*. 2002;(1):CD003255.
34. Smith D, Herbst J, Zhang X. Condom effectiveness for HIV prevention by consistency of use among men who have sex with men in the United States. *JAIDS J Acquir*. 2015. http://journals.lww.com/jaids/Abstract/2015/03010/Condom_Effectiveness_for_HIV_Prevention_by.14.aspx. Accessed August 12, 2016.
35. Garofalo R, Hotton AL, Kuhns LM, Gratzer B, Mustanski B. Incidence of HIV Infection and Sexually Transmitted Infections and Related Risk Factors Among Very Young Men Who Have Sex With Men. *J Acquir Immune Defic Syndr*. 2016;72(1):79-86. doi:10.1097/qai.0000000000000933.
36. Simon CP, Jacquez JA. Reproduction numbers and the stability of equilibria of {SI} models for heterogeneous populations. *SIAM J Appl Math*. 1992;52(2):541-576.