Modelling Epidemic Risk Group Dynamics

Jesse Knight[a,*], Linwei Wang[a], Huiting Ma[a], Katherine Young[b], Harry Hausler[b], Sheree Schwartz[c], Stefan Baral[c], Sharmistha Mishra[a]

[a] *MAP Centre for Urban Health Solutions, Unity Health Toronto*
[b] *TB HIV Care, South Africa*
[c] *Dept. Epidemiology, Johns Hopkins Bloomberg School of Public Health*

Abstract

This is the abstract.

Keywords: TBD, TBD, TBD

* Corresponding author (knightje@smh.ca)

Abbreviations: HIV: human immunodeficiency virus, TPAF: transmission population attributable fraction

1. Introduction

Core group theory has long underpinned the study of epidemics of sexually transmitted infections (STI). The theory posits that heterogeneity in acquisition and transmission risk are sometimes necessary and sometimes sufficient for an STI epidemic to emerge and persist. This heterogeneity is often demarcated by identifying potential cores, comprised of sub-populations or geographies, where onward transmission risks are the highest, such that the core's unmet STI prevention and treatment needs sustain local epidemics (Yorke et al., 1978; Gesink et al., 2011).

Mathematical models of STI transmission include heterogeneity in risk by stratifying the modelled population by features such as the partner change rate, and levels of sexual mixing between subgroups, and partnership types . The implications of including heterogeneity, as compared to assumptions of homogeneity, include higher basic reproductive ratios R_0, and lower overall STI prevalence (provided the latter still results in $R_0 > 1$) (Boily and Mâsse, 1997). R_0 and overall STI prevalence are further influenced by mixing between subgroups (Boily and Mâsse, 1997). Thus, models with more than two risk groups are increasingly relevant for exploring epidemic nuance and for aligning model outputs with programmatic decision support – i.e. prioritization of specific interventions for specific risk groups .

Less often discussed or included in STI transmission models is the influence of movement of individuals between risk groups, which we herein refer to as "turnover". For example, a period of higher risk could represent the average duration in sex work, which is often associated with larger number of sexual partners as paid clients, and other STI-associated vulnerabilities . It could also represent periods of higher partner change outside the formal sex work. . Stigum et al. modelled movement between risk groups as a form of "migration", and showed that and thus, had nearly as large an influence on overall STI prevalence as sexual mixing between sub-groups (Stigum et al., 1994). It has also been shown that rates of movement between risk groups can play an important role during estimation of intervention impact, following model fitting to calibration targets, since (Eaton and Hallett, 2014).

Yet, implementation of risk groups and turnover in recent models vary widely, from no modelled risk groups to seven risk groups with highly context-specific turnover (Boily et al., 2015). A common challenge in structuring and parameterizing STI models are considerations on how to best incorporate turnover and duration of periods of risk using available data. Models require parameters on transition rates between risk groups, but must also content with considerations such as stability in the relative size of risk groups over time.

First, estimating the rates of movement between groups directly from cross-sectional survey data is difficult, and typically requires strong assumptions. Second, ensuring the relative sizes of risk groups do not vary dramatically over time requires careful selection of rates of turnover among groups, or other compensatory parameters.

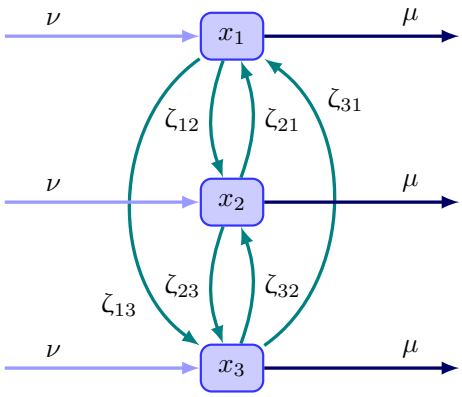


Figure 1: System of states and flows between them for $G = 3$

2. System

This section introduces a system of states, flows between them, and equations which can be used to describe risk group dynamics in deterministic compartmental epidemic models.

We denote the size of risk group $i \in [1, \ldots, G]$ as x_i and the vector of all x_i as $\boldsymbol{x}$. The total population size is denoted $N = \sum_i x_i$,[1] and the proportion that each group represents by $\hat{x}_i = x_i/N$. The rate of population entry overall is denoted ν, and the rate of exit by μ. We do not consider disease-attributable death, which may vary by group, though this should be the subject of future work. All rates have units *per year*. The proportion of the entering population who will enter group i is denoted $\hat{e}_i$. Since the rate of entry ν is typically expressed as a proportion of the total size N, we model the theoretical entering population $\boldsymbol{e}$ as also having size N, so that $e_i = \hat{e}_i N$.

Turnover transitions can occur between any two groups, in either direction; therefore we denote the turnover rates as a $G \times G$ matrix ζ, where ζ_{ij} corresponds to the transition $x_i \to x_j$. An explicit definition is given in Eq. (1), where the diagonal elements are written $*$ since they represent transitions from a group to itself, which are inconsequential.

$$\zeta = \begin{bmatrix} * & x_1 \to x_2 & \cdots & x_1 \to x_G \\ x_2 \to x_1 & * & \cdots & x_2 \to x_G \\ \vdots & \vdots & \ddots & \vdots \\ x_G \to x_1 & x_G \to x_2 & \cdots & * \end{bmatrix} \tag{1}$$

These transition flows and the associated rates are summarized for $G = 3$ in Figure 1.

[1] Here, as in many models, "total population" actually refers to an age-constrained range.

2.1. Parameterization

Next, we explore methods for estimating the values of parameters in the system described above (ν, μ, $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{e}}$, and ζ) directly from some commonly available sources of data.

2.1.1. Total Population Size

The model of total population size over time is defined by entry and exit rates, ν and μ, as in:

$$N(t) = N_0\big(1 + \mathcal{G}(t)\big)^t \tag{2a}$$

$$\mathcal{G}(t) = \nu(t) - \mu(t) \tag{2b}$$

and we note that the average duration of an individual in the model at time t is given by:

$$\delta(t) = \mu^{-1}(t) \tag{3}$$

Variation in rate of entry across risk groups is captured in $\hat{\boldsymbol{e}}$, and we generally do not stratify rate of exit by activity group (besides disease-attributable death); therefore, we can assume that ν and μ do not vary across risk groups, which allows us to derive them with $N(t)$, independent of the population proportions $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{e}}$, and turnover ζ.

The simplest approach assumes a constant population size $N(t) = N_0$, or a growth rate of zero, yielding $\nu = \mu$. However, this does not reflect the true positive population growth of most contexts, and may result in underestimated incidence, due to the relative reduction in inflow of susceptible individuals.

Another approach is to fix $\mathcal{G}(t)$ as some constant. When using this approach, extra care should be taken to ensure the resulting $N(t)$ matches any available population size estimates to a reasonable degree.

Typically, data will be available for the total size of the population over time $N(t)$, so the growth rate for each time interval t_i can be derived by rearranging Eq. (2a):

$$\mathcal{G}(t_i) = \left(\frac{N(t_{i+1})}{N(t_i)}\right)^{-(t_{i+1}-t_i)} - 1 \tag{4}$$

All of these approaches help define $\mathcal{G}(t)$, but leave one degree of freedom, since any choice of $\mu(t)$ can be compensated by $\nu(t)$ to yield the desired $\mathcal{G}(t)$. However, we can usually leverage the known duration of individuals in the model $\delta(t)$ to choose $\mu(t)$ as in Eq. (3). This can come from an assumed duration of sexual activity, or a constant, predefined age range relevant to parameterization. Then, we can solve for $\nu(t)$ using Eq. (2b).

2.1.2. Turnover

Next, we assume that $\nu(t)$ and $\mu(t)$ are known, and we focus on resolving $\hat{\boldsymbol{e}}(t)$ and $\zeta(t)$. Similar to above, we will first formulate the problem as a system of equations; then we will consider which data and assumptions we can leverage to solve the system.

We begin by defining the "conservation of mass" equation for a given group x_i, where that the rate of change of the group is simply the sum of flows in / out of the group:

$$\frac{d}{dt}x_i = \nu\, e_i + \sum_j \zeta_{ji}\, x_j - \mu\, x_i - \sum_j \zeta_{ij}\, x_i \tag{5}$$

While Eq. (5) is written in terms of absolute population sizes $\boldsymbol{x}$ and $\boldsymbol{e}$, it is equivalent to divide through by N, yielding a system in terms of proportions $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{e}}$, which is often more useful, since N need not be known.

We further assume that the average proportions of each group $\hat{x}_i$ do not change over time. Therefore, the desired rate of change for risk group i will be equal to the growth of the risk group, $\mathcal{G}x_i$. Substituting this into Eq. (5), and simplifying, we have:

$$\nu\, x_i = \nu\, e_i + \sum_j \zeta_{ji}\, x_j - \sum_j \zeta_{ij}\, x_i \tag{6}$$

Now, depending on the number of risk groups, we have G and $G(G-1)$ unknowns in $\boldsymbol{e}$ and ζ, totalling G^2 variables to resolve. We denote these variables as the vector $\boldsymbol{\theta} = [\boldsymbol{e}, \boldsymbol{z}]$, where $\boldsymbol{z} = \text{vec}_{i\neq j}(\zeta)$; this allows us to define a system of linear equations of the form:

$$\boldsymbol{b} = A\,\boldsymbol{\theta} \tag{7}$$

where A is a $G \times G^2$ matrix and $\boldsymbol{b}$ is a G-length vector, representing the right-hand side and left-hand side of Eq. (6), respectively. In this form, we can use $A^{-1}\boldsymbol{b} = \boldsymbol{\theta}$ to solve for $\boldsymbol{\theta}$.

Unfortunately, for any $G > 1$, the system is underdetermined by a factor of $G(G-1)$, meaning there are many combinations of $\boldsymbol{e}$ and ζ which satisfy Eq. (6). Therefore, we now resume our task of leveraging data and assumptions to define a unique solution.

Our first tool is another equation. We note that the duration of time spent in a particular group δ_i is the inverse of all efferent flow rates:

$$\delta_i = \left(\mu + \sum_j \zeta_{ij}\right)^{-1} \tag{8}$$

These durations could be derived from survey data, including for key populations, or they could be assumed. Rearranging Eq. (8), we obtain ${\delta_i}^{-1} - \mu = \sum_j \zeta_{ij}$, which yields an additional G equations in our linear system – i.e. rows of A and $\boldsymbol{b}$. For $G = 2$, this provides enough constraints to fully determine the system, as shown in Appendix B.2 Example Turnover System Equations, but for larger G, still more constraints are needed.

The simplest additional constraints can be elements in $\boldsymbol{\theta}$ which are directly specified – i.e. elements of of $\boldsymbol{e}$ or ζ. For example, the proportion of individuals who move from one risk group to another each year (ζ_{ij}) may be assumed or derived from data. Similarly, the distribution of individuals across risk groups in the entering population $\hat{\boldsymbol{e}}$ may be approximated using the proportions among the lowest age group for which

Box 1: Common assumptions regarding the dynamics of risk groups

1. **Risk Groups:** Populations are stratified by risk of infection acquisition.
   (a) **No:** $G = 1$; Populations are homogeneous in risk of infection acquisition.
   (b) **Yes:** $G > 1$; Heterogeneity in risk of infection acquisition within populations is considered.
2. **Turnover:** Individuals may move between risk groups.
   (a) **No:** $\zeta = 0$; Individuals do not move between risk groups.
   (b) **Constant:** $\zeta > 0$; Individuals move between risk groups at a constant rate.
3. **Population Growth:** Increase in the total N over time.
   (a) **No:** $\nu = \mu$; Population size N is constant.
   (b) **Yes:** $\nu > \mu$; Population size N increases, at some constant or data-driven rate.

data are available. In each case, the value specified is appended to $\boldsymbol{b}$, and a row appended to A of the form: $[0, \ldots, 1, \ldots, 0]$, with 1 in the position of the element in $\boldsymbol{\theta}$.

There are, however, two notable caveats to this approach. First, not all combinations of specified elements will add an equal number of constraints. Specifying all elements of $\boldsymbol{e}$ will only add $G-1$ (not G) constraints, since $\sum \hat{\boldsymbol{e}} = 1$, so the final element adds no new information. Similarly, specifying all elements of ζ_{ij} for a given i as well as the duration for the group δ_i will only add $G-1$ (not G) constraints, since Eq. (8) must hold. Second, not all combinations of specified values will yield a valid solution,[2] and it is unfortunately difficult to anticipate problematic combinations.

Finally, we note that additional constraints may be avoided altogether if we pose the problem as an optimization problem, namely:

$$\boldsymbol{\theta}^* = \arg\min f(\boldsymbol{\theta}), \quad \text{subject to: } \boldsymbol{b} = A\,\boldsymbol{\theta};\ \ \boldsymbol{\theta} \geq 0 \tag{9}$$

where f is a function such as $||\cdot||_2$. However, the choice of f implies a prior on the values of $\boldsymbol{\theta}$, and so introduces bias in the solution.

2.2. Previous Approaches

INP

[2] Even rank-deficient systems be inconsistent.

Figure 2: Modelled health states

3. Experiment

We have described an approach to parameterizing models of risk group dynamics which hopefully highlights the *feasibility* of including such model components based on available data. Next, we will explore the *importance* of including these components through comparison of projected model outputs across different implementations of risk group dynamics. Specifically, we will first compare structural variants involving differences in population growth, number of risk groups, and turnover. Next, we will explore the impact of different rates of turnover on overall and group-specific prevalence and incidence, and on implications for model fitting.

3.1. Model & Simulations

We start with a simple deterministic SIR model of transmission in a heterogeneous population, which is meant to be representative of an arbitrary epidemic, rather than any specific context or disease. The model includes three health states: susceptible $\mathcal{S}$, infected $\mathcal{I}$, and recovered $\mathcal{R}$, as shown in Figure 2, and $G = 3$ levels of risk: high H, medium M, and low L, stratified by different rates of contact formation across groups. Gender is not modelled. Individuals in risk group i are assumed to form contacts at a rate C_i, and the probability of contact formation between these individuals with partners in risk group i is assumed to be proportionate with the total number of available contacts:

$$\rho_{ii} = \frac{C_{\mathrm{i}} x_{\mathrm{i}}}{\sum_i C_i x_i} \tag{10}$$

Transmission of the infection from infected $\mathcal{I}$ to susceptible $\mathcal{S}$ individuals is assumed to occur with probability β per contact. Recovered individuals $\mathcal{R}$ are not considered infectious, but do not return to the susceptible state. The force of infection for susceptible individuals in risk group i is therefore modelled using the following equation:

$$\lambda_i = C_i \sum_{\mathrm{i}} \rho_{i\mathrm{i}} \, \beta \frac{\mathcal{I}_{\mathrm{i}}}{N_{\mathrm{i}}} \tag{11}$$

Infected individuals are assumed to be diagnosed and begin treatment at a rate τ (per year). As described in Section 2, individuals enter the model at a rate ν, exit at a rate μ, and transition from risk group i to group j at a rate ζ_{ij}. The default parameters for this base model are summarized in Table 1, and the full system of model equations is given in Appendix B.1 Model Equations.

Table 1: Base model parameters. All rates have units year^{-1} and durations are in years.

Symbol	Description	Value
$\boldsymbol{\beta}$	transmission probability per contact	0.03
τ	rate of treatment initiation among infected	0.1
N_0	initial population size	1000
$\hat{\boldsymbol{x}}$	proportion of system individuals: high, medium, low activity	[0.04 0.20 0.76]
$\hat{\boldsymbol{e}}$	proportion of entering individuals: high, medium, low activity	[0.04 0.20 0.76]
$\boldsymbol{\delta}$	average duration spent in: high, medium, and low activity groups	[5 15 25]
C	rate of contact formation among individuals: high, medium, low activity	[25 5 1]
ν	rate of population entry	0.05
μ	rate of population exit	0.03

Using this model (and variants, described below), simulated epidemics are initialized at $t = 0$ with $N_0 = 1000$ individuals, distributed according to $\hat{\boldsymbol{x}}$. Among these individuals, three are infected, and the remaining are susceptible; for $G = 3$, this corresponds to one infected person in each risk group, while for $G = 1$, this corresponds to simply three infected people. Simulated epidemics are solved numerically using Euler's method with a time step of $dt = 0.1$ years.

3.2. Model Variants

Drawing on the most common assumptions outlined in Box 1, we define a series of four structural model variants (V1 – V4) for investigation. These variants, summarized in Figure 3, include homogeneous versus heterogeneous risk ($G = 1$ vs $G = 3$), zero versus nonzero population growth ($\nu = \mu$ vs $\nu > \mu$), and zero versus nonzero turnover among risk groups ($\zeta = 0$ vs $\zeta > 0$).

In order to facilitate fair comparisons across model variants with respect to parameter values, we start from the base model describe above above (V1), and aim to make simplifications which have only a singular impact on the system. For example, when moving from $\nu > \mu$ to $\nu = \mu$, we ensure the average duration of individuals in the model μ^{-1} is unchanged by fixing μ and reducing ν to match (V4). Similarly, when collapsing the stratification of risk groups from $G = 3$ to $G = 1$ (V3), we define the contact rate C for all individuals as the weighted average of previously risk-stratified C. Finally, considering rates of turnover ζ (which are only applicable to V1 and V3) we fully determine the system, as outlined in Section 2.1.2, by specifying the average duration of individuals in each group $\boldsymbol{\delta}$, and balancing the number of individuals moving between two groups in both directions. The resulting parameter values for each scenario are summarized in Table 2. For each model variant, we project the simulated epidemic using these fixed parameter

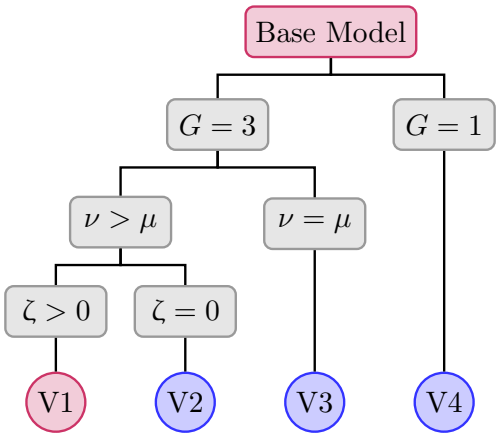


Figure 3: Summary of four structural model variants with respect to simulated risk group dynamics. ν: rate of population entry, μ: rate of population exit, G: number of risk groups, ζ: rates of population turnover

Table 2: Model parameters for structural variants. All rates have units year^{-1} and durations are in years.

Parameter	V1	V2	V3	V4
$\hat{\boldsymbol{x}}$	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[0.04 0.20 0.76]	—
$\hat{\boldsymbol{e}}$	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[0.04 0.20 0.76]	—
C	[25 5 1]	[25 5 1]	[25 5 1]	[**2.76**]
δ	[5 15 25]	[**33 33 33**]	[5 15 25]	[**33**]
ν	0.05	0.05	**0.03**	0.05
μ	0.03	0.03	0.03	0.03

values, and calculate the incidence and prevalence for each risk group, as well as overall. Comparing the results, we highlight trends in the projected outputs across the structural variants.

3.3. Impact of Turnover

The overall impact of risk group turnover in an epidemic is not straightforward. On one hand, movement of infected individuals from high to low risk groups increases disease penetration in the lower risk groups; however, turnover also decreases the duration of risk exposure among individuals in the higher risk groups. In order to clarify which effect dominates when, we additionally explore a wide range of turnover magnitudes with model variant V1. Moreover, since the average exposure experienced by each group is directly affected by the duration of infectiousness δ_I, we also explore the impact of treatment rate $\tau = {\delta_I}^{-1}$ on this result.

Since particular turnover rates ζ_{ij} are difficult to conceptualize, we leverage the methods outlined in Section 2.1.2 to derive the necessary transition matrix given specified durations δ_i in each risk group, which are inversely correlated to turnover rates. That is, we actually vary $\boldsymbol{\delta}$, and calculate ζ to ensure risk group

proportions $\hat{\boldsymbol{x}}$ are maintained over time.[3] Furthermore, we balance the absolute number of individuals moving between two groups via turnover as in Eq. (**??**), define the duration in the medium risk group as a linear function of δ_H like: $\delta_M = \delta_H + 0.3\left(\mu^{-1} - \delta_H\right)$, and the leave the duration in the high risk group as a free parameter to allow a consistent solution. Using this approach, a single parameter δ_H controls the magnitude of turnover for the entire system,[4] and so the values of $\hat{\boldsymbol{e}}$ and ζ required to maintain group proportions $\hat{\boldsymbol{x}}$ can be resolved using Eq. (7). To explore the impact of turnover on the model, the duration in the high risk group δ_H is then varied from 3 to 33 years, and the duation of infectiousness δ_I from 1 to 20 years (using the rate treatment τ).

3.4. Implications for Model Fitting

Finally, since almost all context-specific applications of epidemic models entail fitting uncertain model parameters to data-driven calibration targets, we explore some potential implications of omitting turnover from a fitted model. In particular, we calibrate model variants V1 (includes turnover) and V2 (no turnover) to 25% prevalence in the high risk group, and 5% prevalence in the low risk group, at quasi-equilibrium, by fitting the contact rate of groups C_H and C_L.[5] We then compare the fitted contact rates, and estimate the transmission population attributable fraction (TPAF) (Mishra et al., 2012) of the high risk group, using both the pre-calibration and post-calibration models (4 total variants). TPAF estimates the proportion of cumulative new infections which are attributable to prevention gaps among a specific population. In this case, we toggle transmission "from" the high risk group (not "to"). We aim to highlight potential differences in estimation of TPAF due to omission of turnover in models, before and after model fitting.

4. Results

The aim of the experiments was to explore the impact of different implementations of risk group dynamics on model outputs in a generalized epidemic model. We compare the projected incidence and prevalence across four model variants and range of turnover magnitudes, as well as the estimated TPAF of the high risk group, with and without turnover.

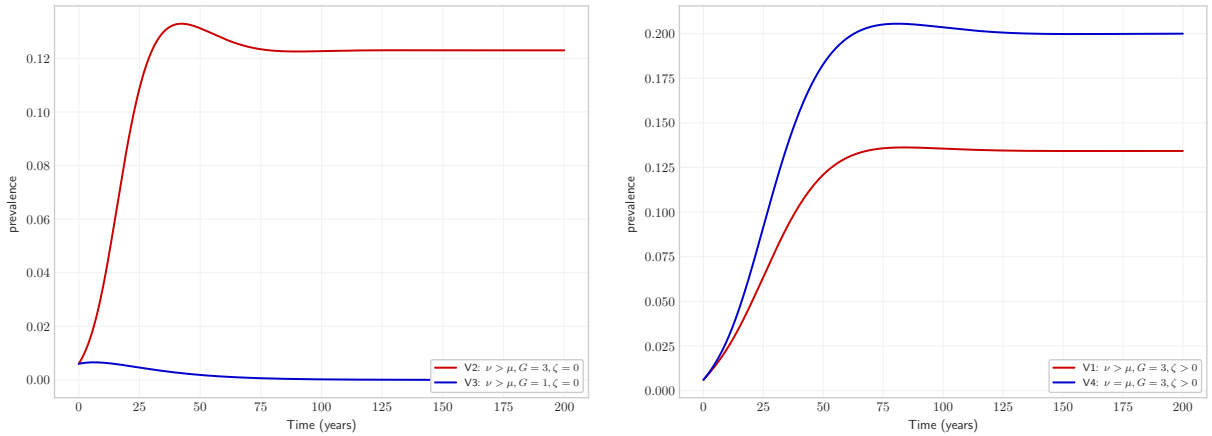
4.1. Model Variants

Figure 4a shows the prevalence predicted by the model with and without population heterogeneity (V1 vs V3). As previously noted in numerous discussions of core group theory (Yorke et al., 1978; Stigum

[3] Throughout this work, we define duration δ_i as a "single average pass through the group" which does not consider reentrance after exiting to another group.

[4] Note that from Eq. (8), we can see that no risk-group duration can exceed the overall duration in the model μ^{-1}.

[5] The negative log-likelihood of each predicted prevalence (assuming a sample size of 1000) was minimized using the Sequential Least SQuares Programming (SLSQP) method (Kraft, 1988) from the `scipy.optimize.minimize` Python package.



(a) Risk heterogeneity (V1 vs V3)

(b) Population growth (V1 vs V4)

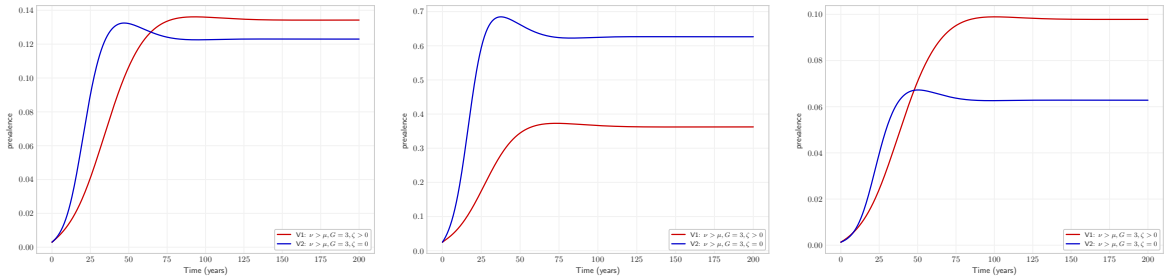
Figure 4: Comparison of model projections with and without structural model features

et al., 1994), epidemic characteristics are dramatically impacted by the presence of heterogeneity in risk within a population. In this case, failure to model heterogeneity in risk (V3) results in a basic reproduction number $R_0 < 1$, and no epidemic, while the model with heterogeneity (V1) predicts an endemic prevalence of over 12%.

Figure 4b shows the prevalence predicted by the model with and without population growth (V1 vs V4). Since all model entrants are assumed to be susceptible, simulated population growth reduces the proportion of possible contacts who are infected, and so the relative epidemic spread is reduced, both in terms of incidence and equilibrium prevalence.

Next, we consider the impact of risk group turnover. From Figure 5a, we can see that the impact on overall equilibrium prevalence is not obvious, and that the effects on high versus low risk groups are opposite for this system (Figures 5b and 5c). We can observe, however, that initial epidemic spread is slowed for all groups, as the concentration of the epidemic in a core group is eroded by decreasing duration of exposure in the high risk group. In general, we might summarize this effect as a "risk-homogenizing" effect, or the opposite of increased risk heterogeneity.[6] In order to characterize additional trends, we look to the next section, where a wider range of epidemic contexts are explored, in terms of durations in each risk group and durations of disease infectiousness.

[6] In fact, it can be shown that for sufficiently high rates of risk-group turnover, model outputs converge on those outputs predicted by a model without stratification of risk groups at all. This result is shown in Figure A.1.



(a) Overall (b) High risk (c) Low risk

Figure 5: Comparison of projected prevalence with and without risk group turnover (V1 vs V2)

4.2. Impact of Turnover

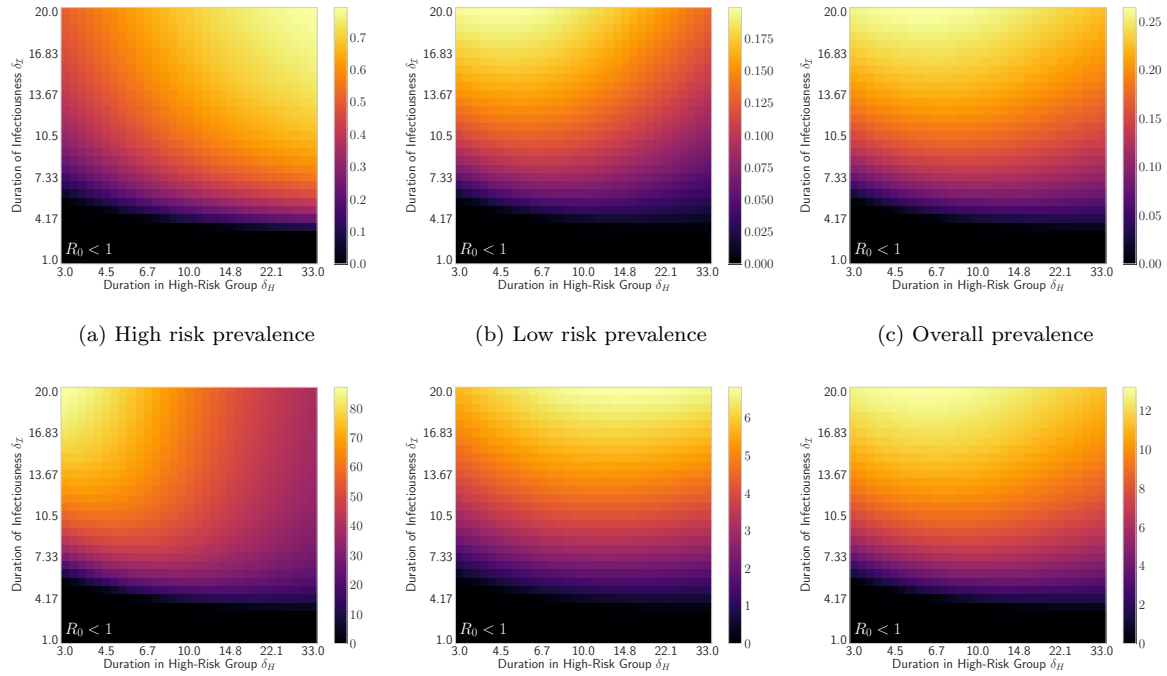
Figure 6 shows the equilibrium prevalence and incidence across a wide range of turnover rates and durations of infectiousness, stratified by high and low risk groups, as well as overall. We note that increasing the duration of infectiousness increases prevalence and incidence across all groups, regardless of turnover. However, the impact of increasing rates of turnover can depend on the duration of infectiousness, making it difficult to anticipate the effect of including or omitting turnover in a model.

One exception is that prevalence among the high risk group always decreases with turnover (Figure 6a), since the duration of exposure in this group decreases. Turnover also increases movement of infected individuals into low risk groups from the high risk group, and movement of mainly susceptible individuals from low to high risk groups. When prevalence in the high risk group is high, the additional susceptible individuals are quickly infected, increasing incidence in the high risk group (Figure 6d, top left). Movement of infected individuals from high to low risk groups also contributes a considerable proportion of prevalent infections in the low risk group under these conditions (Figure 6b, top left, and Figure A.2). However, if prevalence in the high risk group is sufficiently low, due to short duration of infectiousness or very high turnover, then the replacement of infected individuals with mainly susceptible individuals through turnover decreases, rather than increases, incidence in the group (Figure 6d, bottom left).[7] As a result, the ability of the "core-group" to maintain the epidemic declines, and both incidence and prevalence decrease across all groups as turnover increases (Figure 6, all panels, bottom left). Thus, for short durations of infectiousness, or for very high turnover, no epidemic is observed, and we infer that $R_0 < 1$.

4.3. Fitted Models with Turnover

Model variants V1 (turnover) and V2 (no turnover) were both calibrated to 25% and 5% prevalence in the high and low risk groups, respectively, by fitting the contact rates C_i. The ratio between fitted

[7] For simple SI* models, incidence is maximized when $\mathcal{S} = \mathcal{I}$.



(a) High risk prevalence (b) Low risk prevalence (c) Overall prevalence

(d) High risk incidence (e) Low risk incidence (f) Overall incidence

Figure 6: Steady-state prevalence and incidence for different rates of turnover (based on δ_H; log scale; turnover increases right to left) and duration of infectiousness δ_I (reflecting treatment/disease course; linear scale; treatment increases top to bottom).

high and low contact rates was much higher with turnover ($C_H/C_L = 16.92/0.28 = 59.63$) than without ($C_H/C_L = 15.79/2.49 = 6.33$). This implies that, if turnover dynamics present in reality are not captured in the model, the relative risk experienced by the high risk group may be underestimated by calibrated models. This is because, in order to observe the same prevalence ratios between the high and low risk groups in a system which has turnover, the difference in risks experienced by each group would have to be larger, since turnover acts to homogenize the average risk experienced by each individual throughout their life.

These dynamics also have implications when estimating the TPAF of particular risk groups. For instance, before model calibration, a system without turnover estimates the TPAF of the high risk group to be greater than a system with turnover (Figure 7, solid lines). In this case, we think that the high prevalence among the high risk group in the absence of turnover maximizes incident infections *from* the group, yielding a higher TPAF. However, after calibration of group-specific contact rates C_i, the trend reverses, and the system with turnover predicts a higher TPAF for the high risk group (Figure 7, solid lines). We hypothesize that this reversal is mainly attributable to the larger ratio of fitted contact rates under turnover, which dramatically increases the proportion of contacts (and therefore infections) which occur with the high risk group. Therefore, if real-world turnover processes are omitted from a calibrated epidemic model, then the relative importance of reaching the high risk group with interventions may be underestimated.

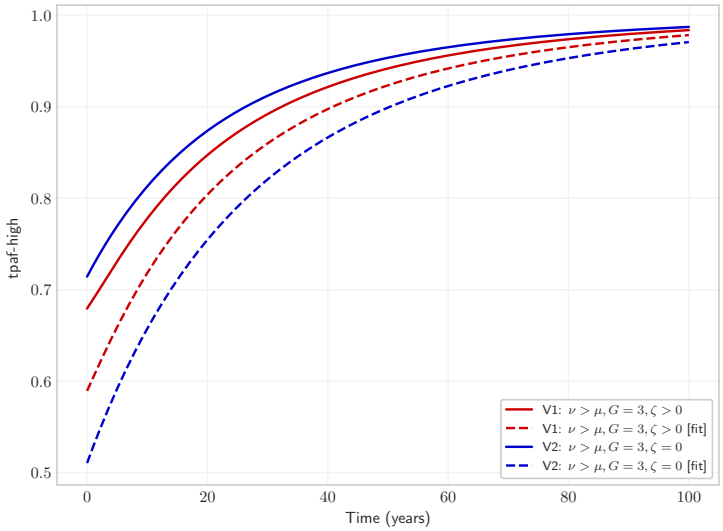


Figure 7: Transmission population attributable fraction (TPAF-from) of the high risk group with and without turnover, and with and without fitted C_i to group-specific prevalence.

5. Discussion

6. Conclusion

7. References

Boily, M. C. and Mâsse, B. (1997). Mathematical models of disease transmission: A precious tool for the study of sexually transmitted diseases. *Canadian Journal of Public Health*, 88(4):255–265.

Boily, M. C., Pickles, M., Alary, M., Baral, S., Blanchard, J., Moses, S., Vickerman, P., and Mishra, S. (2015). What really is a concentrated HIV epidemic and what does it mean for West and Central Africa? Insights from mathematical modeling. *Journal of Acquired Immune Deficiency Syndromes*, 68:S74–S82.

Eaton, J. W. and Hallett, T. B. (2014). Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence. *Proceedings of the National Academy of Sciences*, 111(45):16202–16207.

Gesink, D. C., Sullivan, A. B., Miller, W. C., and Bernstein, K. T. (2011). Sexually transmitted disease core theory: Roles of person, place, and time. *American Journal of Epidemiology*, 174(1):81–89.

Kraft, D. (1988). A software package for sequential quadratic programming. Technical Report DFVLR-FB 88-28, DLR German Aerospace Center — Institute for Flight Mechanics, Koln, Germany.

Mishra, S., Steen, R., Gerbase, A., Lo, Y. R., and Boily, M. C. (2012). Impact of High-Risk Sex and Focused Interventions in Heterosexual HIV Epidemics: A Systematic Review of Mathematical Models. *PLoS ONE*, 7(11):e50691.

Stigum, H., Falck, W., and Magnus, P. (1994). The core group revisited: The effect of partner mixing and migration on the spread of gonorrhea, chlamydia, and HIV. *Mathematical Biosciences*, 120(1):1–23.

Yorke, J. A., Hethcote, H. W., and Nold, A. (1978). Dynamics and control of the transmission of gonorrhea. *Sexually Transmitted Diseases*, 5(2):51–56.

Appendix A. Supplemental Figures

Appendix A.1. Homogenizing Effect of Turnover

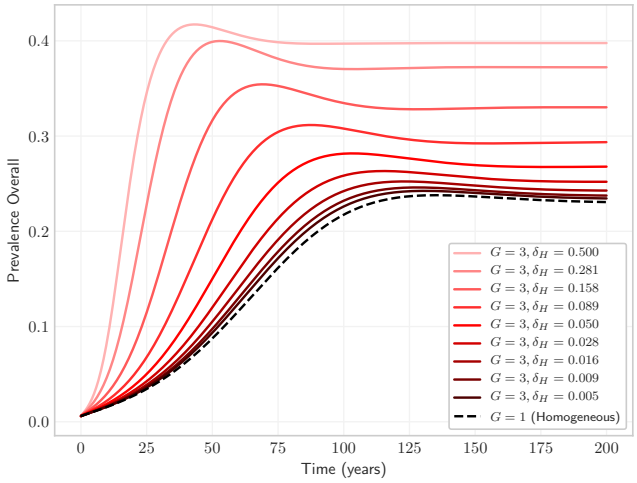


Figure A.1: Overall prevalence predicted by a heterogeneous system under a wide range of high turnover rates. Note how the heterogeneous system ($G = 3$) converges on a homogeneous system ($G = 1$) with very high turnover rates (duration δ_H approaches zero).

Appendix A.2. Infection Convection

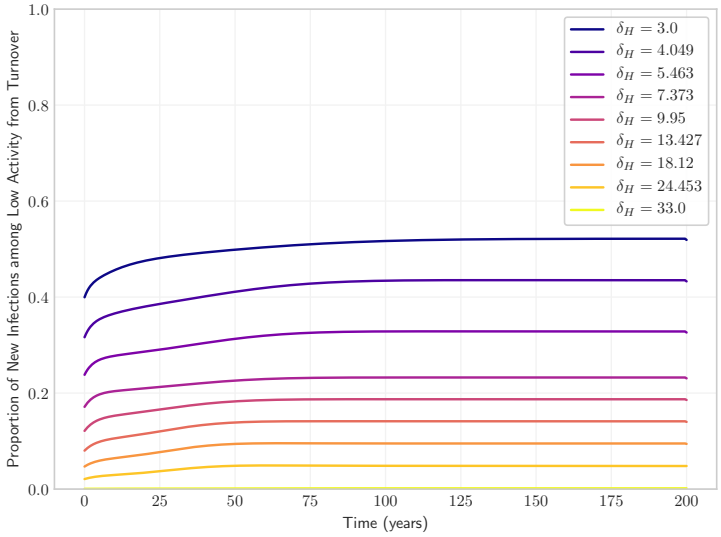


Figure A.2: Proportion of new infections among the low risk group which represent turnover of infected individuals versus infection of individuals in the low risk group ($\delta_I = 10$ years). At high levels of turnover, an notable proportion of prevalence in the low risk group can be attributed to turnover of infected individuals from higher risk groups.

Appendix B. Supplemental Equations

Appendix B.1. Model Equations

$$\frac{d}{dt}\mathcal{S}_i(t) = \sum_j \zeta_{ji}\mathcal{S}_j(t) \; - \sum_j \zeta_{ij}\mathcal{S}_i(t) \; - \mu\mathcal{S}_i(t) \; + \nu\hat{e}_i N(t) - \lambda_i(t)\mathcal{S}_i(t) \tag{B.1}$$

$$\frac{d}{dt}\mathcal{I}_i(t) = \sum_j \zeta_{ji}\mathcal{I}_j(t) \; - \sum_j \zeta_{ij}\mathcal{I}_i(t) \; - \mu\mathcal{I}_i(t) \qquad\qquad + \lambda_i(t)\mathcal{S}_i(t) - \tau\mathcal{I}_i(t) \tag{B.2}$$

$$\frac{d}{dt}\mathcal{R}_i(t) = \sum_j \zeta_{ji}\mathcal{R}_j(t) - \sum_j \zeta_{ij}\mathcal{R}_i(t) - \mu\mathcal{R}_i(t) \qquad\qquad + \tau\mathcal{I}_i(t) \tag{B.3}$$

Appendix B.2. Example Turnover System Equations

Appendix B.2.1. $G = 1$

$$\left[\begin{array}{c} \nu x_1 \end{array}\right] = \left[\begin{array}{c} \nu \end{array}\right] \left[\begin{array}{c} e_1 \end{array}\right] \tag{B.4}$$

Appendix B.2.2. $G = 2$

$$\left[\begin{array}{c} \nu x_1 \\ \nu x_2 \\ {\delta_1}^{-1} - \mu \\ {\delta_2}^{-1} - \mu \end{array}\right] = \left[\begin{array}{cccc} \nu & \cdot & -x_1 & x_2 \\ \cdot & \nu & x_1 & -x_2 \\ \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 \end{array}\right] \left[\begin{array}{c} e_1 \\ e_2 \\ \zeta_{12} \\ \zeta_{21} \end{array}\right] \tag{B.5}$$

Appendix B.2.3. $G = 3$

$$\left[\begin{array}{c} \nu x_1 \\ \nu x_2 \\ \nu x_3 \\ {\delta_1}^{-1} - \mu \\ {\delta_2}^{-1} - \mu \\ {\delta_3}^{-1} - \mu \end{array}\right] = \left[\begin{array}{ccccccccc} \nu & \cdot & \cdot & -x_1 & -x_1 & x_2 & \cdot & x_3 & \cdot \\ \cdot & \nu & \cdot & x_1 & \cdot & -x_2 & -x_2 & \cdot & x_3 \\ \cdot & \cdot & \nu & \cdot & x_1 & \cdot & x_2 & -x_3 & -x_3 \\ \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 \end{array}\right] \left[\begin{array}{c} e_1 \\ e_2 \\ e_3 \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{21} \\ \zeta_{23} \\ \zeta_{31} \\ \zeta_{32} \end{array}\right] \tag{B.6}$$