On Parameterizing Risk Group Dynamics in Simulated STI Epidemics

Jesse Knight, Linwei Wang, Huiting Ma, Sharmistha Mishra

January 22, 2019

Contents

Key Contributions

1. Formalize a mathematical framework for risk group demographics
2. Describe methods for deriving risk group demographic parameters from common data sources
3. Illustrate differences in modelled projections for different implementations of risk group demographics, using an example sexually transmitted infection

1 Introduction

Core group theory has long underpinned modelling of STI epidemics. It generally describes maintenance of an epidemic by high levels of exposure in a small subset of individuals, relative to a larger general population with less exposure. When considering one or more core groups – i.e. a population composed of several risk-stratified groups – the epidemic characteristics are known to depend on the group sizes, relative exposures, and rates of sexual mixing between groups. Less discussed is a similar dependence on movement of individuals between risk groups, which we call "turnover".[1] This turnover of individuals between risk groups has a similar effect to sexual mixing between groups, making it an important feature to include in representative models [1].

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 specific interventions for specific risk groups. Yet, implementations of risk groups and turnover in recent models vary widely, from no modelled risk groups [2, 3] to seven risk groups with highly context-specific turnover [[best Mishra 7-groups citation]]. In fact, many HIV models still do not consider turnover, despite the fact that rates of movement between risk groups can play an important role during estimation of intervention impact following model fitting to calibration targets [4].

Two major challenges exist to implementing turnover, which may help explain its inconsistent usage. First, unlike behavioural parameters, 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 rates of turnover among groups, or other compensatory parameters. Prior works have generally solved this problem *ad hoc*, without providing a generalized approach, while some simply rely on a "burn-in" period, which permits equilibriation of risk group sizes due to turnover dynamics before introduction of the infection.

We therefore expect that a unified framework for defining and parameterizing risk group dynamics would be of great use. We present such a framework here, and draw direct links to modelling assumptions and relevant sources of data. Building on previous work by Stigum et al. (1994) [1], we then leverage this framework to explore the impact of several risk group implementations on model outputs (incidence, prevalence) in a representative model.

2 The System

This section introduces a system of compartments, flows between them, and equations which can be used to describe risk group dynamics.

We denote the variable representing the size of risk group $i \in [1, \dots, G]$ as x_i and the vector of all x_i as $\boldsymbol{x}$. The total population size is denoted $N = \sum_i x_i$,[2] and the proportions represented by each group by $\hat{x}_i = x_i N^{-1}$. The rate of population entry for all groups is denoted by ν, and the rate of exit by μ. We do not consider disease-attributable death, which may vary by group, though this will be the subject of future work. All rates have units *per year* (yr^{-1}). The proportion of the entering population who are in group i, which may not be equal to the proportion of the current population in group i, is denoted $\hat{e}_i$. Since the rate of entry ν is typically expressed as a proportion of the total population 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. (2.1), where the diagonal elements are denoted $*$ since they represent transitions from a group to itself, which is inconsequential.

$$\zeta = \begin{bmatrix} * & x_1 \to x_2 & \cdots & x_1 \to x_G \\ x_2 \to x_1 & * & \cdots & x_2 \to x_G \\ \vdots & \vdots & \ddots & \vdots \\ x_G \to x_1 & x_G \to x_2 & \cdots & * \end{bmatrix} \tag{2.1}$$

[1] In early works, such as [1], movement of individuals between risk groups is often called "migration"; in order to avoid confusion with population entry / exit, we prefer the term "turnover".

[2] In many models, "total population" actually refers to an age-constrained range.

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

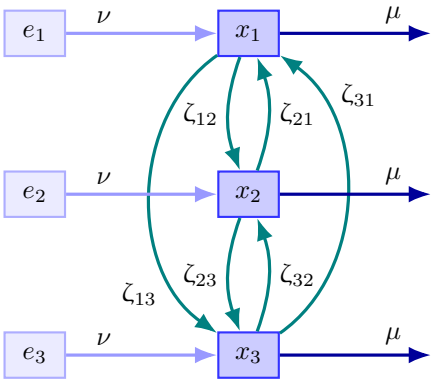


Figure 1: System of compartments and flows between them for $G = 3$

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.

In most cases, there will not be sufficient data to directly estimate all parameters, especially ζ. The next section outlines additional methods to solve for these values.

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:

$$N(t) = N_0\big(1 + \mathcal{G}(t)\big)^t \tag{2.2a}$$

$$\mathcal{G}(t) = \nu(t) - \mu(t) \tag{2.2b}$$

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

$$\delta(t) = \mu^{-1}(t) \tag{2.3}$$

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. (2.2a):

$$\mathcal{G}(t_i) = \left(\frac{N(t_{i+1})}{N(t_i)}\right)^{-(t_{i+1}-t_i)} - 1 \tag{2.4}$$

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. (2.3). 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. (2.2b).

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:

$$\frac{d}{dt}x_i = \nu\, e_i + \sum_j \zeta_{ji}\, x_j - \mu\, x_i - \sum_j \zeta_{ij}\, x_i \tag{2.5}$$

While Eq. (2.5) 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. (2.5), and simplifying, we have:

$$\nu\, x_i = \nu\, e_i + \sum_j \zeta_{ji}\, x_j - \sum_j \zeta_{ij}\, x_i \tag{2.6}$$

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:

$$\boldsymbol{b} = A\,\boldsymbol{\theta} \tag{2.7}$$

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. (2.6), 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. (2.6). 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:

$$\delta_i = \left(\mu + \sum_j \zeta_{ij}\right)^{-1} \tag{2.8}$$

These durations could be derived from survey data, including for key populations, or they could be assumed. Rearranging Eq. (2.8), 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 A Example Systems, 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 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. (2.8) must hold. Second, not all combinations of specified values will yield a valid solution,[3] 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:

$$\boldsymbol{\theta}^* = \arg\min f(\boldsymbol{\theta}), \quad \text{subject to: } \boldsymbol{b} = A\,\boldsymbol{\theta};\ \ \boldsymbol{\theta} \geq 0 \tag{2.9}$$

[3] Even rank-deficient systems be inconsistent.

Box 1: Common assumptions regarding the dynamics of risk groups

1. **Risk Groups:** Major demographic groups are stratified by risk of HIV acquisition.
   1.1. **No:** $G = 1$; Major demographic groups are homogeneous in risk of HIV acquisition.
   1.2. **Yes:** $G > 1$; Heterogeneity in risk of HIV acquisition within major demographic groups is considered.
2. **Turnover:** Individuals may move between risk groups.
   2.1. **No:** $\zeta = 0$; Individuals do not move between risk groups.
   2.2. **Constant:** $\zeta > 0$; Individuals move between risk groups at a constant rate.
3. **Population Growth:** Increase in the total N over time.
   3.1. **No:** $\nu = \mu$; Population size N is constant.
   3.2. **Yes:** $\nu > \mu$; Population size N increases, at some constant or data-driven rate.

Table 1: Summary of prior work with respect to modelled risk group dynamics.

Ref.	Year	Author	Risk Groups G	Mortality ϕ	Dynamic Rebalance RB	Turnover ζ
[5]	2008	Hallett et al.	3	Yes	None	None
[3]	2012	Barnighausen et al.	1	Yes	None	N/A
[2]	2012	Estill et al.	1	Yes	None	N/A
[6]	2013	Cremin et al.	3	Yes	None	None
[4]	2014	Eaton and Hallett	3	Yes	None	Constant

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

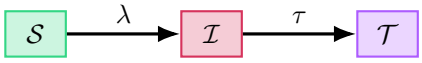


Figure 2: Modelled health states

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 compare structural variants involving differences in population growth, number of risk groups, and turnover.

3.1 Model & Simulations

We start with a simple deterministic model of heterosexual HIV transmission, which is meant to be generally representative of a plausible HIV epidemic, rather than any specific context. The model includes three health states: susceptible $\mathcal{S}$, infected $\mathcal{I}$, and on treatment $\mathcal{T}$, as shown in Figure 2, and $G = 3$ levels of sexual activity (risk groups): high H, medium M, and low L. Individuals of sex k in risk group i are assumed to form sexual partnerships at a rate C_{ki} with individuals of the opposite sex k. The probability of partnership formation with a partner of sex k in risk group i is assumed to follow the formulation proposed by Garnett and Anderson [7]:

$$\rho_{ki\mathrm{ki}} = (1-\epsilon)\psi_{ki\mathrm{ki}} + (\epsilon)\pi_{ki\mathrm{ki}} \tag{3.1}$$

where ϵ is a parameter controlling the dominance of assortative ψ ($\epsilon = 0$) or proportional π ($\epsilon = 1$) mixing.

Transmission of HIV from infected $\mathcal{I}$ to susceptible $\mathcal{S}$ individuals is assumed to occur with probability $\beta_{k\mathrm{k}}$ per partnership. Individuals on treatment $\mathcal{T}$ are not considered infectious. The force of infection for susceptible individuals of sex k in risk group i is therefore modelled using the following equation:

$$\lambda_{ki} = C_{ki} \sum_{\mathrm{ki}} \rho_{ki\mathrm{ki}}\, \beta_{k\mathrm{k}} \frac{\mathcal{I}_{\mathrm{ki}}}{N_{\mathrm{ki}}} \tag{3.2}$$

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 2.

Using this model (and variants, described below), simulated epidemics are initialized at $t = 0$ with $N_0 = 1000$ individuals, distributed proportionally according to $\hat{\boldsymbol{x}}$. Among these individuals, 6 are infected, and the remainder are susceptible; for $G = 3$, this corresponds to one infected person in each sex-risk group, while for $G = 1$, this corresponds to three infected people in each sex. Simulated epidemics are solved numerically using Euler's method with a time step of $dt = 0.1$ years.

Table 2: Base model parameters. All rates have units year^{-1} and durations are in years.

Symbol	Description	Value
$\boldsymbol{\beta}$	transmission probability per partnership: men → women, women → men	[0.04 0.02]
ϵ	mixing parameter where: 1 → proportional and 0 → assortative [7]	1.0
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 partner change among individuals: high, medium, low activity	[25 5 1]
ν	rate of population entry	0.05
μ	rate of population exit	0.03

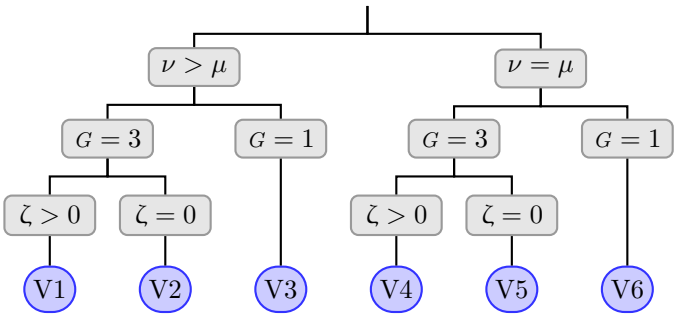


Figure 3: Summary of 6 structural model variants with respect to simulated risk group dynamics. ν: rate of population entry, μ: rate of population exit, G: number of risk groups, ζ: rates of population turnover

Table 3: Model parameters for structural variants. All rates have units year^{-1} and durations are in years.

Parameter	V1	V2	V3	V4	V5	V6
$\hat{\boldsymbol{x}}$	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[1.0]	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[1.0]
$\hat{\boldsymbol{e}}$	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[1.0]	[0.04 0.20 0.76]	[0.04 0.20 0.76]	[1.0]
C	[25 5 1]	[25 5 1]	[2.76]	[25 5 1]	[25 5 1]	[2.76]
δ	[5 15 25]	[33 33 33]	[33]	[5 15 25]	[33 33 33]	[33]
ζ_{HL}	0.10	—	—	0.10	—	—
ν	0.05	0.05	0.05	0.03	0.03	0.03
μ	0.03	0.03	0.03	0.03	0.03	0.03

3.2 Model Variants

Drawing on the most common assumptions outlined in Box 1, we define a series of six structural model variants (V1 – V6) for investigation. These variants, summarized in Figure 3, include zero versus nonzero population growth ($\nu = \mu$ vs $\nu > \mu$), homogeneous versus heterogeneous risk ($G = 1$ vs $G = 3$), 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 described above (V1 in Figure 3), and aim to make simplifications with minimal impact on system characteristics. 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 – V6). Similarly, when collapsing the stratification of risk groups from $G = 3$ to $G = 1$ (V3, V6), we define the homogeneous partner change rate C as the weighted average of the previously risk-stratified C. Finally, considering rates of turnover ζ (which are only applicable for V1, V4) we fully determine the system, as outlined in Section 2.1.2, by specifying the average duration of individuals in each group $\boldsymbol{\delta}$, as well as the distribution of individuals entering the model $\hat{\boldsymbol{e}}$, and the proportion of high activity individuals moving to the low activity group each year ζ_{HL}. The resulting parameter values for each scenario are summarized in Table 3.

For each model variant, we project the simulated epidemic using these fixed parameter 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 effect of including turnover in a model is not straightforward. On one hand, transfer of infected individuals from high to low risk groups increases disease penetration in the lower risk groups; however, turnover also decreases the duration of individuals in the higher risk groups, and in turn, their cumulative exposure to risk. In order to clarify which effect dominates when, we additionally explored 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, we also explored the impact of treatment rate τ on this result.

Particular values of ζ are difficult to conceptualize, so we leveraged the methods outlined in Section 2.1.2 to derive the necessary matrix for specified durations δ_i in each risk group. That is, we actually varied $\boldsymbol{\delta}$, and calculated ζ to ensure risk group proportions $\hat{\boldsymbol{x}}$ were maintained over time.[4] Note that rates of turnover are inversely proportional to the duration in each group. Furthermore, we defined $\delta_M = \min\{5\delta_H, \mu^{-1}\}$ and $\delta_L = \min\{80\delta_H, \mu^{-1}\}$ so that a single parameter δ_H controls the entire system. In order to fully determine the system, the turnover flows from each group were evenly divided among the destination groups using $\zeta_{ij_1} = \zeta_{ij_2} = \frac{1}{2}\left(\delta_i^{-1} - \mu\right)$. The values of $\hat{\boldsymbol{e}}$ were the solved to maintain group proportions $\hat{\boldsymbol{x}}$ as in Eq. (2.7).

The duration in the highest risk group δ_H was then varied from approximately 33 years through 2 days. From Eq. (2.8), we can see that no duration can exceed μ^{-1}, or 33.3 years here. Technically, no lower bound on δ_i exists, though we note that the very small values explored here (< 1 year) are mainly for illustrative purposes. For each δ_H, and for a range of $\tau \in [0, 0.2]$, an epidemic is simulated with these fixed parameters as before.

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 50% HIV prevalence in the High risk group, and 10% HIV prevalence in the Low risk group, both at quasi-equilibrium, 200 years after the epidemic start, fitting the partner change rate of the High and Low risk groups, C_H and C_L.[5] We then estimate the transmission population attributable fraction (TPAF) of Women in the High risk group, who are approximately representative of female sex workers, using the pre-calibration and post-calibration models (4 total variants). TPAF estimates the proportion of onward transmission which is attributable to prevention gaps for a specific population. We aim to highlight potential underestimation of the TPAF of women in the High-risk group due to omission of turnover in models, before and after model fitting.

4 Results & Discussion

The aim of these experiments was to explore the impact of different implementations of risk group demographics on model outputs. We compared projected outputs from 6 model variants and explored the effect of overall turnover magnitude. In this discussion, we generally assume that the first model variant presented (S1) is the *most plausible* variant, since it includes population growth, risk heterogeneity, and risk group turnover.

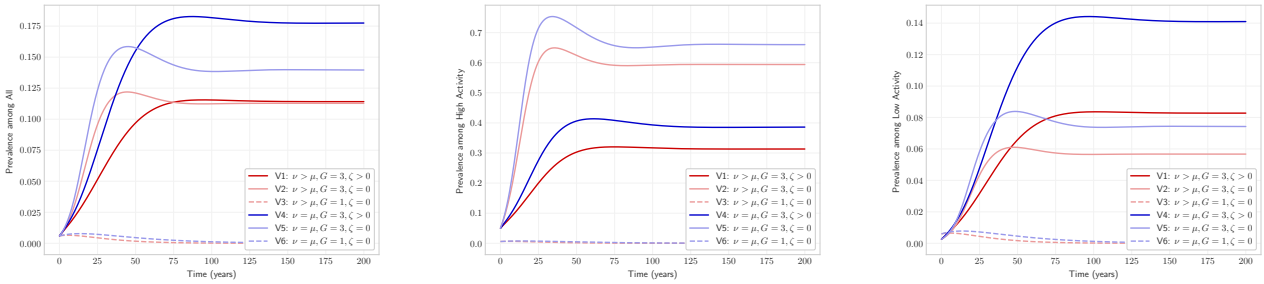
4.1 Model Variants

Figure 4 illustrates the projected HIV prevalence for each model variant for the population overall, and among High and Low activity groups specifically. Omission of risk heterogeneity ($G = 1$ vs $G = 3$) has the most dramatic impact on system dynamics, reducing R_0 below 1 for both variants V3 and V6, resulting in no epidemic. This is supported by long-standing literature on core group theory, and so should come as no surprise. Next, we note that population growth (V1 – V3) universally reduces projected HIV prevalence, especially among the low risk group. This effect is mainly attributable to increases in the denominator N of the prevalence equation with susceptible entrants to the model, since absolute incidence given population growth remains higher than with constant population size – just not high enough to keep pace with the additional population. Again, these results should not be surprising, but we take this opportunity to reiterate the importance of stratifying modelled populations by risk, and reasonably approximating population growth, considering the impacts demonstrated here.

Finally, we turn to risk group turnover. Generally, turnover decreases HIV prevalence in the high-risk group through replacement of exiting infectious individuals with mainly susceptible individuals from lower risk groups (Figure 4b),

[4] 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.

[5] The absolute difference between the target and predicted prevalences were minimized using the Sequential Least SQuares Programming (SLSQP) method [8] from the `scipy.optimize.minimize` Python package: https://docs.scipy.org/doc/scipy/reference/optimize.minimize-slsqp.html



(a) Prevalence Overall (b) High activity prevalence (c) Low activity prevalence

Figure 4: Projected prevalence under each of the 6 model structural variants.

and increases prevalence in the low-risk group, through "retirement" of infected individuals from high to low risk groups (Figure 4c). This parallels the characteristics of a system under proportionate mixing ($\epsilon = 1$) as compared to assortative mixing ($\epsilon = 0$), since both proportionate mixing and turnover act to erode the concentration of risk among higher risk groups.[6] Therefore, if turnover is omitted from the model ($\zeta = 0$), the initial incidence and prevalence are overestimated relative to when turnover is considered, particularly in the high risk group, but also overall. At steady state, the *overall* impact of turnover is not so straightforward, but the trends in prevalence described above generally persist. These trends are further clarified in the next section, which considers both the impact of turnover magnitude and treatment rate.

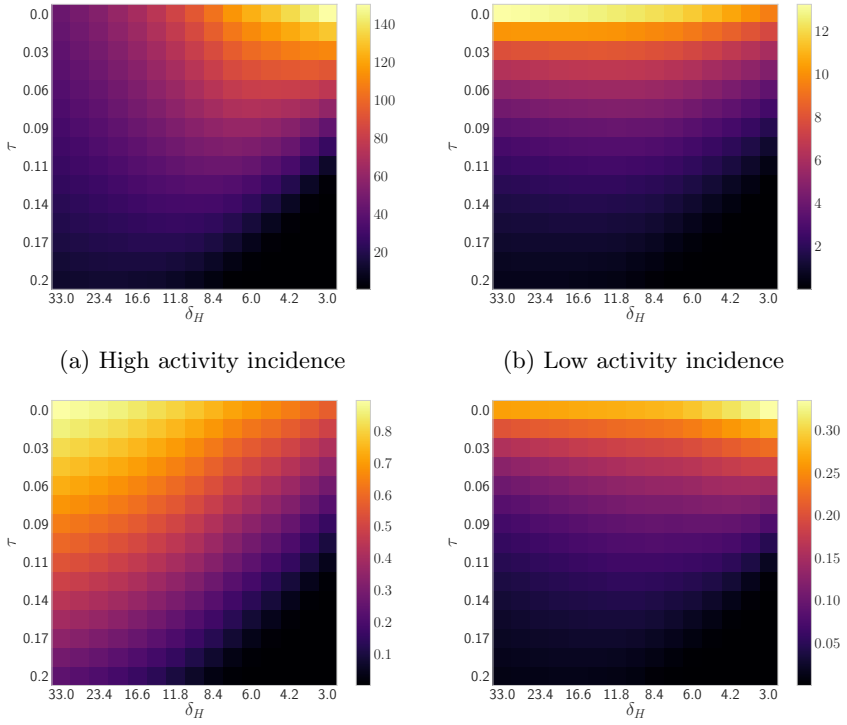
4.2 Impact of Turnover

Figure 5 shows the projected steady-state incidence and prevalence for model V1 under several rates of overall turnover and treatment. As suggested above, increasing turnover (decreasing δ_H) reduces HIV prevalence in the high risk group (Figure 5c), due to reduced cumulative exposure to risk with shorter duration in the group. However, at low treatment rate, turnover actually increases incidence in the high risk group (Figure 5a), since with prevalence over 50%, supply of susceptibles, not infectious individuals, increases the probability of serodiscordant partnerships. Among the low-risk group, still with low treatment rate, the prevalence trend approximately reverses, as increased turnover contributes additional infected individuals to the low-risk group. Interestingly, incidence again exhibits the opposite trend of prevalence; we hypothesize that this time, it is due to decreased prevalence among the high-risk group, representing the main source of serodiscordant partnerships under the proportional mixing assumptions here.

As treatment rates increases, incidence and prevalence predictably decline among all groups. However, since differences across τ tend to dominate the magnitude of prevalence in Figure 5, we additionally normalize these plots by the turnover-free magnitude ($\delta_H = 33$) for each τ. From the resulting plots, we can see that increasing turnover continues to erode prevalence among the high-risk group (Figure 6b), while prevalence among the low-risk group increases with turnover up to a τ-dependent threshold (Figure 6c). After this point – for high rates of treatment and turnover – wash-out of the core group rapidly reduces steady-state incidence and prevalence. At the extreme, even the core group R_0 is reduced to < 1, yielding no epidemic. Since the overall prevalence in this system is dominated by the large low-risk group, the effect of turnover on low-risk prevalence approximates the effect on overall prevalence, yielding the complex relationship with turnover and treatment rate shown in Figure 6a. Given this result, turnover cannot be said to simply increase or decrease overall estimated prevalence; indeed, even the trends in Figure 6a may subject to the specific parameters and assumptions employed here.

Finally, we note that the homogenizing effect of turnover can be illustrated empirically. Specifically, we show that model variant V1 will actually converge on V3 ($G = 1$) as the rates of turnover increase, though well beyond plausible values, since individuals spend so little time in each risk group that their overall risk behaviour and exposure becomes approximately uniform. After increasing $\boldsymbol{\beta}$ by a factor of 2.5 so that V3 obtains an epidemic,

[6] A useful analogy is to describe proportionate partnership mixing as risk "conduction" (direct contact) between groups, and turnover as risk "convection" (movement of individuals).



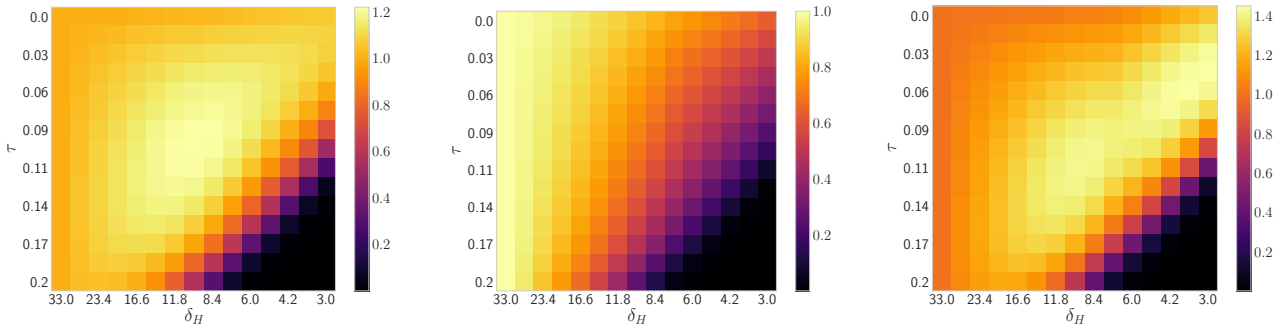
(a) High activity incidence

(b) Low activity incidence

(c) High activity prevalence

(d) Low activity prevalence

Figure 5: Steady-state incidence and prevalence for different rates of turnover (based on δ_H, log scale) and treatment rate τ (linear scale).



(a) Overall

(b) High activity

(c) Low activity

Figure 6: Steady-state prevalence for different rates of turnover (based on δ_H, log scale) and treatment rate τ (linear scale), after normalizing by prevalence at $\delta_H = 33$ for each τ, in order to highlight the impact of turnover for a given τ.

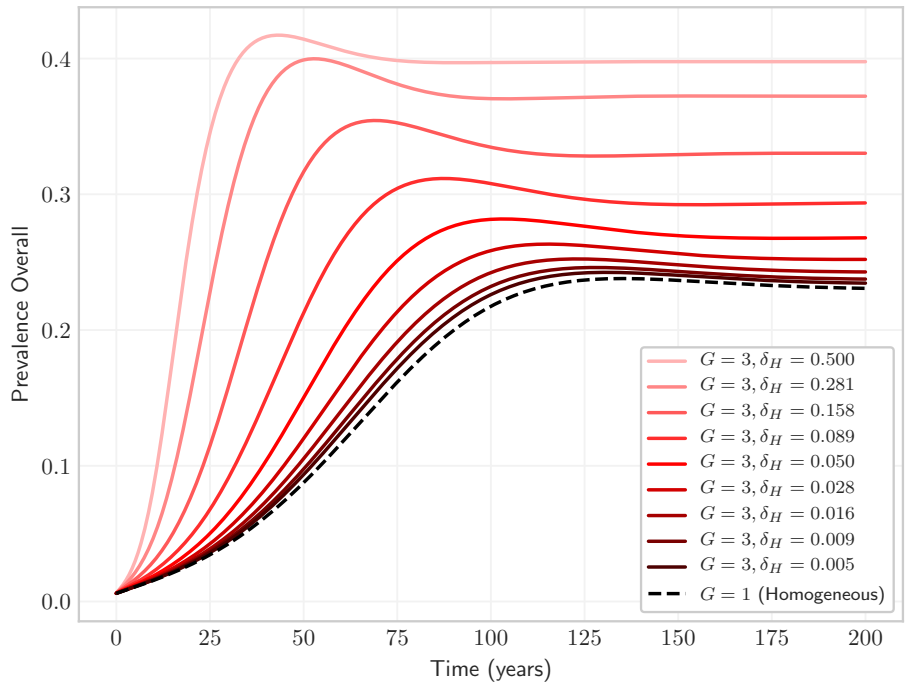


Figure 7: Overall prevalence predicted by a heterogeneous system under a wide range of high turnover rates. Note how the heterogeneous system ($G = 3$) converges on a homogeneous system ($G = 1$) with very high turnover rates (duration δ approaches zero).

Figure 7 demonstrates the convergence of overall projected prevalence from model V1 with high turnover on that from model V3.

4.3 Fitted Models with Turnover

Finally, we return to model fitting, where parameters (typically behavioural) are adjusted so that the model projections match historically observed data, such as prevalence. As shown in Figures 6b and 6c, omission of turnover promises to overestimate prevalence in the high-risk group and (for low treatment rate) underestimate prevalence in the low-risk group. Therefore, when fitting to consistent group-wise prevalence targets without turnover, the risk behaviour of high-risk groups will be underestimated, while that of low-risk groups will likely be overestimated. Following the experiment described in Section 3.4, we fit C_H and C_L to prevalence targets for the high and low risk groups, with and without turnover. With turnover, the ratio of fitted C_H/C_L was 70.6, while without turnover, the ratio was only 9.4.

The TPAF of high-risk women with turnover and fixed parameters is already larger than that without turnover for all time horizons (Figure 8, solid lines), implying that models lacking turnover are liable to underestimate the importance of reaching key populations like female sex workers with interventions. However, after fitting, with the adjustments to behavioural parameters described above, the gap in estimated TPAF grows (Figure 8, dotted lines), implying that *fitted* models lacking turnover underestimate the importance of reaching core groups even more.

4.4 Implications

4.5 Limitations

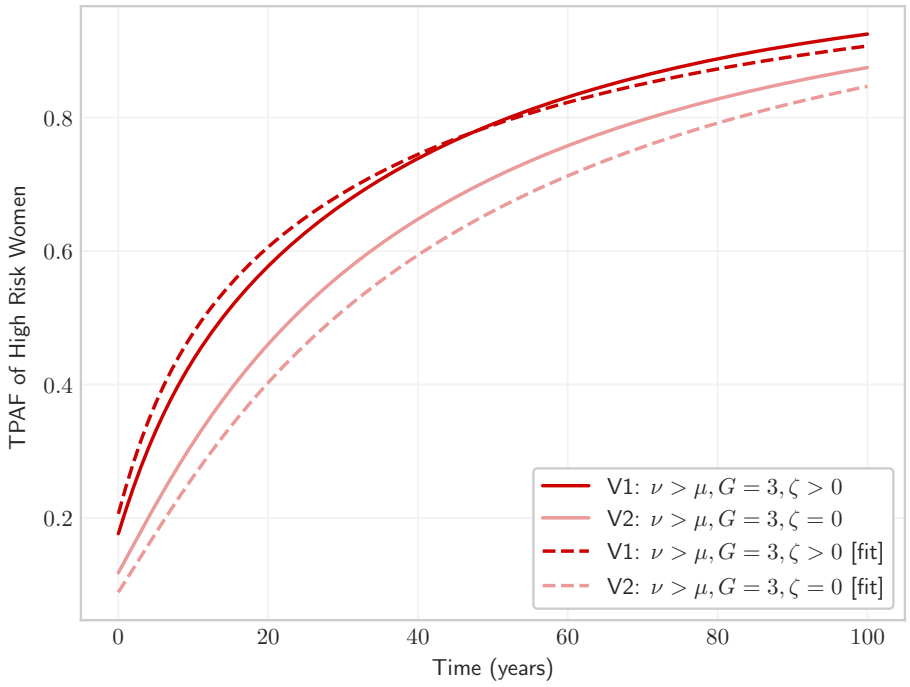


Figure 8: Estimated transmission population attributable fraction (TPAF) of high-risk women with and without turnover, and with and without fitting C to group-specific prevalence. Turnover increases the estimated TPAF of high risk groups for fixed parameters. Moreover, the difference in estimated TPAF with and without turnover increases after fitting C to group-specific prevalence.

5 References

[1] Hein Stigum, W. Falck, and P. Magnus. "The core group revisited: The effect of partner mixing and migration on the spread of gonorrhea, chlamydia, and HIV". In: *Mathematical Biosciences* 120.1 (1994), pp. 1–23. DOI: 10.1016/0025-5564(94)90036-1.

[2] Janne Estill et al. "Viral load monitoring of antiretroviral therapy, cohort viral load and HIV transmission in Southern Africa: A mathematical modelling analysis". In: *AIDS* 26.11 (2012), pp. 1403–1413. DOI: 10.1097/QAD.0b013e3283536988.

[3] T. Barnighausen, D. E. Bloom, and S. Humair. "Economics of antiretroviral treatment vs. circumcision for HIV prevention". In: *Proceedings of the National Academy of Sciences* 109.52 (2012), pp. 21271–21276. DOI: 10.1073/pnas.1209017110.

[4] Jeffrey W. Eaton and Timothy B. Hallett. "Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence". In: *Proceedings of the National Academy of Sciences* 111.45 (2014), pp. 16202–16207. DOI: 10.1073/pnas.1323007111.

[5] Timothy B. Hallett et al. "Understanding the impact of male circumcision interventions on the spread of HIV in southern Africa". In: *PLoS ONE* 3.5 (2008). Ed. by Atle Fretheim, e2212. DOI: 10.1371/journal.pone.0002212.

[6] Ide Cremin et al. "The new role of antiretrovirals in combination HIV prevention: A mathematical modelling analysis". In: *AIDS* 27.3 (2013), pp. 447–458. DOI: 10.1097/QAD.0b013e32835ca2dd.

[7] Geoffrey P. Garnett and Roy M. Anderson. "Balancing sexual partnership in an age and activity stratified model of HIV transmission in heterosexual populations". In: *Mathematical Medicine and Biology* 11.3 (1994), pp. 161–192. DOI: 10.1093/imammb/11.3.161.

[8] Dieter Kraft. *A software package for sequential quadratic programming*. Tech. rep. DFVLR-FB 88-28. Koln, Germany: DLR German Aerospace Center — Institute for Flight Mechanics, 1988.

A Example Systems

A.1 $G = 1$

$$\begin{bmatrix} \nu x_1 \end{bmatrix} = \begin{bmatrix} \nu \end{bmatrix} \begin{bmatrix} e_1 \end{bmatrix} \tag{A.1}$$

A.2 $G = 2$

$$\begin{bmatrix} \nu x_1 \\ \nu x_2 \\ {\delta_1}^{-1} - \mu \\ {\delta_2}^{-1} - \mu \end{bmatrix} = \begin{bmatrix} \nu & \cdot & -x_1 & x_2 \\ \cdot & \nu & x_1 & -x_2 \\ \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \zeta_{12} \\ \zeta_{21} \end{bmatrix} \tag{A.2}$$

A.3 $G = 3$

$$\begin{bmatrix} \nu x_1 \\ \nu x_2 \\ \nu x_3 \\ {\delta_1}^{-1} - \mu \\ {\delta_2}^{-1} - \mu \\ {\delta_3}^{-1} - \mu \end{bmatrix} = \begin{bmatrix} \nu & \cdot & \cdot & -x_1 & -x_1 & x_2 & \cdot & x_3 & \cdot \\ \cdot & \nu & \cdot & x_1 & \cdot & -x_2 & -x_2 & \cdot & x_3 \\ \cdot & \cdot & \nu & \cdot & x_1 & \cdot & x_2 & -x_3 & -x_3 \\ \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{21} \\ \zeta_{23} \\ \zeta_{31} \\ \zeta_{32} \end{bmatrix} \tag{A.3}$$

B Code

The code for this work is available at:
https://github.com/c-uhs/turnover
which uses the `epi-model` framework from:
https://github.com/c-uhs/turnover