Modelling Epidemic Risk Group Dynamics

Abstract

This is the abstract.

Keywords: TBD, TBD, TBD

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 risks of acquisition and onward transmission 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, levels of sexual mixing between subgroups, and partnership types (Mishra et al., 2012). 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 (Stigum et al., 1994; Boily and Mâsse, 1997). Models with two or more risk groups are increasingly relevant in applied epidemic modelling, such as to help prioritize specific interventions for particular risk groups (Mishra et al., 2012).

Rarely discussed or included in STI transmission models is the movement of individuals between risk groups, which we refer to as "turnover". For example, consider a high risk group representing sex work, for which there are a larger number of sexual partners as paid clients, and other STI-associated vulnerabilities (Watts et al., 2010). Individuals may retire from sex work but continue to be sexually active at a lower level of risk, or enter into sex work following a period of lower risk (Boily et al., 2015). Similar transitions are also possible for individuals engaging in multiple partnerships, and so on.

Several authors have implicated turnover as an important factor in model outputs, such as: the predicted equilibrium prevalence (Stigum et al., 1994; Eaton and Hallett, 2014); the fraction of transmissions occurring during acute infection (Zhang et al., 2012); the basic reproductive number R_0 (Henry and Koopman, 2015); and the level of universal treatment required to achieve epidemic control (Henry and Koopman, 2015). Yet, implementations of risk groups and turnover in recent models vary widely. Some authors only model turnover in the direction of high to low risk, controlled by a single parameter (Stigum et al., 1994; Eaton and Hallett, 2014). In works by Koopman et al., rates of movement between two risk groups are balanced analytically based on the size of the groups, while Boily et al. (2015) use a 100-year burn-in period to equilibrate a complex system of turnover transitions. These various approaches to turnover are contrasted with implementations of model features like assortative sexual mixing, which typically follow "standard" methods, such as that proposed by Nold (1980).

Perhaps widespread and consistent implementation of turnover dynamics in STI models is made difficult by the following two challenges. First, it is not obvious how epidemiological data can be used to inform

rates of turnover among risk groups. Second, as shown by Boily et al. (2015), naive selection of turnover rates can result in imbalanced flows, causing risk groups to initially change size over time; it is not clear how to avoid this problem in all cases.

Therefore, we propose a unified framework for parameterizing risk group dynamics, including risk heterogeneity, population growth, and turnover. We develop this framework in Section 2. Additionally, while some authors have shown the influence of turnover on overall incidence and prevalence (Stigum et al., 1994; Zhang et al., 2012; Henry and Koopman, 2015), the mechanisms and influence of turnover on group-specific incidence and prevalence remain unclear. In Sections 3 and 4, we illustrate these mechanisms using a representative SIR model. Finally, epidemic models are often used to help prioritize interventions following calibration to a specific context. Since turnover can affect model calibration, we also explore some potential implications of failing to model turnover dynamics when they truly exist in Sections 3 and 4. We discuss the results of these experiments in Section 5.

2. System

In this section we introduce a system of risk groups, flows between them, and equations which can be used to describe risk group dynamics in deterministic compartmental epidemic models.

2.1. Notation

Consider a population divided into G risk groups. We denote the number of individuals in risk group $i \in [1, \ldots, G]$ as x_i and the set of all risk groups as $\boldsymbol{x} = \{x_1, \ldots, x_G\}$. The total population size is $N = \sum_i x_i$,[1] and the proportion of the total population that each group represents is denoted as $\hat{x}_i = x_i/N$. We do not consider stratification of age groups. Individuals enter the model at a rate ν per year, and exit at a rate μ per year. However, the proportion of individuals entering into group i from outside the model may be different from the proportion of individuals currently in group i in the model ($\hat{x}_i$). Therefore, we distinguish these proportions and denote the proportion entering into group i as $\hat{e}_i$. For example, a higher proportion of youth (individuals entering the model, $\hat{e}_i$) may engage in high-risk sexual behaviour, as compared to the population overall ($\hat{x}_i$). It will later be shown how rates of turnover can maintain such a system at equilibrium.

Turnover transitions may occur between any two groups, in either direction; therefore we denote the turnover rates as a $G \times G$ matrix ϕ. The element ϕ_{ij} corresponds to the proportion of individuals in group x_i who move from group x_i to group x_j each year. An example matrix is given in Eq. (1), where we write

[1] Here, as in many models, "total population" actually represents a subset of the population with a given duration in the model – e.g. an age-constrained range.

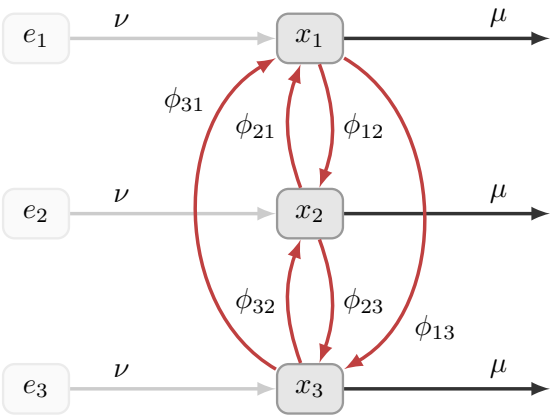


Figure 1: System of states and flows between them for $G = 3$

the diagonal elements as $*$ since they represent transitions from a group to itself, which are inconsequential.

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

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

2.2. Parameterization

Next, we consider the goal of constructing a system like the one introduced above which reflects the risk group dynamics observed in a specific context. We assume that the relative sizes of the risk groups in the model ($\hat{\boldsymbol{x}}$) are already known, and should remain constant over time. Thus, what remains is to estimate the values of the parameters: ν, μ, $\hat{\boldsymbol{e}}$, and ϕ, using commonly available sources of data.

2.2.1. Total Population Size

The total population size $N(t)$ is related to the rates of population entry $\nu(t)$ and exit $\mu(t)$. We consider variation in rate of entry across risk groups via $\hat{\boldsymbol{e}}$, and we do not stratify the rate of exit by activity group (disease-attributable death is assumed to be negligible for now). Therefore, we can assume that ν and μ do not vary across risk groups, allowing these rates to be estimated independent of the risk group distributions $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{e}}$, and turnover ϕ.

The difference between entry and exit rates defines the rate of population growth:

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

The total population may then be defined using an initial population size N_0 as:

$$N(t) = N_0 \exp\left(\int_0^t \log\big(1 + \mathcal{G}(\tau)\big) d\tau\right) \tag{3}$$

which, for constant growth, simplifies to the familiar expression:

$$N(t) = N_0(1 + \mathcal{G})^t \tag{4}$$

In most cases, demographic data will be available on the total size of the population over time $N(t)$, allowing Eqs. (3) and (4) to be used to estimate $\mathcal{G}(t)$.

If we assume that the population size is constant, then $\mathcal{G}(t) = 0$ and $\nu(t) = \mu(t)$. However, this is generally not a good assumption, as it does not reflect the positive population growth observed in many contexts. We will also later show that this assumption may impact epidemic dynamics. Another approach is to fix $\mathcal{G}$ at a constant value. This value can be estimated via Eq. (4) using any two values of $N(t)$, separated by a time interval τ:

$$\mathcal{G}_\tau = \frac{N(t+\tau)^{\frac{1}{\tau}}}{N(t)} - 1 \tag{5}$$

For contexts with variability in $\mathcal{G}$ over time, this process can be repeated for consecutive time intervals, and the complete function $\mathcal{G}(t)$ approximated piecewise by constant values. This approach is generally more feasible than exact solutions using Eq. (3), and can reproduce $N(t)$ accurately for small enough intervals τ, such as one year.

Now, a given value of $\mathcal{G}(t)$ does not imply any particular values of $\nu(t)$ or $\mu(t)$, since any choice of $\mu(t)$ can be compensated by an appropriate choice of $\nu(t)$, and vice versa, as in Eq. (2). However, it is common to assume a constant duration of individuals in the model $\delta(t)$, which is related to the rate of exit by:

$$\delta(t) = \mu^{-1}(t) \tag{6}$$

Such a duration should reflect the demographic data used to define $N(t)$, as well as the epidemiological data used to inform other model parameters. For example, an age range of 15 to 50 years implies a duration in the model of $\delta = 35$ years. The duration may also vary with time to reflect contextual changes to all-cause mortality. In any case, $\mu(t)$ can then be defined as $\delta^{-t}(t)$ following Eq. (6), and $\nu(t)$ defined as $\mathcal{G}(t) - \mu(t)$ following Eq. (2).

2.2.2. Turnover

Next, we present methods for resolving the distribution of individuals entering the risk model $\hat{\boldsymbol{e}}(t)$ and the rates of turnover $\phi(t)$, assuming that entry and exit rates $\nu(t)$ and $\mu(t)$ are known. Similar to above, we first formulate the problem as a system of equations. Then, we explore the data and assumptions which can be leveraged to solve for the values of parameters in the system. The (t) notation is omitted throughout this

section for clarity, though time-varying parameters can be estimated by repeating the necessary calculations for each t.

Depending on the number of risk groups G, there are G unknown elements in $\hat{\boldsymbol{e}}$ and $G(G-1)$ unknown elements in ϕ. We collect these unknowns in the vector $\boldsymbol{\theta} = [\hat{\boldsymbol{e}}, \boldsymbol{y}]$, where $\boldsymbol{y} = \text{vec}_{i \neq j}(\phi)$. For example, for $G = 3$, the vector $\boldsymbol{\theta}$ is defined as:

$$\boldsymbol{\theta} = \begin{bmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 & \phi_{12} & \phi_{13} & \phi_{21} & \phi_{23} & \phi_{31} & \phi_{32} \end{bmatrix} \tag{7}$$

The crux of this framework is then to define a linear system of equations which uniquely determine the elements of $\boldsymbol{\theta}$:

$$\boldsymbol{b} = A\,\boldsymbol{\theta} \tag{8}$$

where A is a $M \times G^2$ matrix and $\boldsymbol{b}$ is a M-length vector. That is, each row in A and $\boldsymbol{b}$ specifies an assumed relationship involving elements of $\hat{\boldsymbol{e}}$ and ϕ. Given a sufficient number $M = G^2$ of unique constraints, the parameters can be calculated algebraically using $\boldsymbol{\theta} = A^{-1}\boldsymbol{b}$. Some examples of constraints are explored below.

Constant Group Size. First, we define the "conservation of mass" equation for group x_i, wherein the rate of change of the group is defined as the sum of flows in / out of the group:

$$\frac{d}{dt}x_i = \nu\, e_i + \sum_j \phi_{ji}\, x_j - \mu\, x_i - \sum_j \phi_{ij}\, x_i \tag{9}$$

While Eq. (9) 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 can be more useful, since N need not be known. If we assume that the average proportions of each group $\hat{x}_i$ is constant over time, then 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. (9), and simplifying yields:

$$\nu\, x_i = \nu\, e_i + \sum_j \phi_{ji}\, x_j - \sum_j \phi_{ij}\, x_i \tag{10}$$

Factoring the left and right hand sides in terms of $\hat{\boldsymbol{e}}$ and ϕ, we obtain G unique constraints. For $G = 3$, this yields the following 3 rows as the basis of $\boldsymbol{b}$ and A:

$$\boldsymbol{b} = \begin{bmatrix} \nu x_1 \\ \nu x_2 \\ \nu x_3 \end{bmatrix}; \qquad A = \begin{bmatrix} \nu & \cdot & \cdot & -x_1 & -x_1 & x_2 & \cdot & x_3 & \cdot \\ \cdot & \nu & \cdot & x_1 & \cdot & -x_2 & -x_2 & \cdot & x_3 \\ \cdot & \cdot & \nu & \cdot & x_1 & \cdot & x_2 & -x_3 & -x_3 \end{bmatrix} \tag{11}$$

These G constraints are necessary to ensure risk groups do not change size over time. However, to obtain a unique solution, we still need an additional $G(G-1)$ constraints. For $G = 3$, this corresponds to 6 additional constraints.

Specified Elements. The simplest type of additional constraint is to directly specify individual elements in $\hat{\boldsymbol{e}}$ or ϕ. This constraint may be appended to $\boldsymbol{b}$ and A using indicator encoding: with b_k as the specified value and $A_k = [0, \ldots, 1, \ldots, 0]$ as the indicator vector. For example, for $G = 3$, if it is known that 20% of individuals enter directly into risk group x_1 upon entry into the model ($\hat{e}_1 = 0.20$), then $\boldsymbol{b}$ and A can be augmented with:

$$\boldsymbol{b}' = \begin{bmatrix} 0.20 \end{bmatrix}; \qquad A' = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{12}$$

If we do not want any turnover from group i to group j, then this approach can also be used to set $\phi_{ij} = 0$. Note that the elements of $\hat{\boldsymbol{e}}$ must sum to one. Therefore, specifying all elements in $\hat{\boldsymbol{e}}$ will only provide $G - 1$ additional constraints, as the last element is redundant. This relationship is implicit in Eq. (11), so it is not necessary to supply a constraint $1 = \sum_i \hat{e}_i$. Similar redundancies or inconsistencies can emerge for constraints on ϕ, as noted below.

Group Duration. Another useful constraint can be derived from the average duration of individuals in a risk group. This duration δ_i is defined as the inverse of all efferent flow rates:

$$\delta_i = \left(\mu + \sum_j \phi_{ij} \right)^{-1} \tag{13}$$

Average durations could be derived from survey data, including for key populations, or they could be assumed. These values can be used to define constraints on ϕ by rearranging Eq. (13) to yield: ${\delta_i}^{-1} - \mu = \sum_j \phi_{ij}$. For example, if for $G = 3$, the average duration in group x_1 is known to be $\delta_1 = 5$ years, then $\boldsymbol{b}$ and A can be augmented with:

$$\boldsymbol{b}' = \begin{bmatrix} 5^{-1} - \mu \end{bmatrix}; \qquad A' = \begin{bmatrix} \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{14}$$

As noted above, redundancies can emerge when constraining turnover rates ϕ via duration δ_i. Namely, specifying all efferent flow rates ϕ_{ij} for one group i will fully determine the duration in the group δ_i. In general, each constraint which is not redundant will increase the rank of A by one, and A must be full rank (G^2) to uniquely determine the parameters in $\boldsymbol{\theta}$.

Turnover Balance. We present one final constraint here. We can assume that the absolute number of individuals moving between two risk groups is related by a ratio $\frac{r_i}{r_j}$ so that: $\phi_{ij} x_i r_i = \phi_{ji} x_j r_j$. This helps constrain ϕ_{ij} and ϕ_{ji}. For example, for $G = 3$, if we assume that the number of individuals moving between groups x_1 and x_2 is related by $\frac{3}{2}$, then $\boldsymbol{b}$ and A can be augmented with:

$$\boldsymbol{b}' = \begin{bmatrix} 0 \end{bmatrix}; \qquad A' = \begin{bmatrix} \cdot & \cdot & \cdot & 3x_1 & \cdot & -2x_2 & \cdot & \cdot & \cdot \end{bmatrix} \tag{15}$$

If we assume that the absolute number of individuals moving between the groups is equal, then simply $\frac{r_i}{r_j} = 1$. Again, care should be taken to avoid redundancies and inconsistencies with other constraints.

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:** $\phi = 0$; Individuals do not move between risk groups.
   (b) **Constant:** $\phi > 0$; Individuals move between risk groups.
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.

Minimization. Lastly, we can avoid specifying a complete set of constraints on $\boldsymbol{\theta}$ (A can be less than full rank) if the problem is posed as a minimization problem, namely:

$$\boldsymbol{\theta}^* = \arg\min f(\boldsymbol{\theta}), \quad \text{subject to: } \boldsymbol{b} = A\,\boldsymbol{\theta};\ \ \boldsymbol{\theta} \geq 0 \tag{16}$$

where f is a function which globally constrains $\boldsymbol{\theta}$, such as: $||\cdot||_2$. That is, when a sufficient number of assumptions cannot be made about $\hat{\boldsymbol{e}}$ and ϕ to yield a unique solution, this approach can be used to find the smallest values of $\hat{\boldsymbol{e}}$ and ϕ which satisfy the given constraints. Numerical solutions to such problems are widely available, such as the Non-Negative Lease Squares solver (Lawson and Hanson, 1995).[2]

2.3. Previous Approaches

Under construction!

. . . reference to Box 1.

[2] The NNLS solver is available in Python: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.nnls.html.

Figure 2: Modelled health states. $\mathcal{S}$: susceptible; $\mathcal{I}$: infected; $\mathcal{R}$: recovered.

3. Experiment

To examine the influence of risk group dynamics on epidemics, we first compared equilibrium prevalence and incidence using models with and without: heterogeneity in risk, population growth, and risk group turnover. We then examined the influence of different rates of turnover on overall and group-specific equilibrium prevalence and incidence. Finally, we examine the influence of turnover on the contribution of the highest risk group to the overall epidemic – as measured by the transmission population attributable fraction (TPAF).

3.1. Model & Simulations

We developed a deterministic 1-sex SIR model which simulates transmission in a population with heterogeneity in risk. The model is not representative of a specific infection but includes balancing contacts as per sexually transmitted infections (Garnett and Anderson, 1994). The model includes three health states: susceptible $\mathcal{S}$, infectious $\mathcal{I}$, and recovered $\mathcal{R}$ (Figure 2), and $G = 3$ levels of risk: high H, medium M, and low L. Risk strata are defined by different number of contacts per unit time so that individuals in risk group i are assumed to form contacts at a rate C_i. The probability ρ_{ik} of contact formation between individuals in group i with individuals in risk group k is assumed to be proportionate to the total number of available contacts within each group:

$$\rho_{ik} = \frac{C_k x_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} x_{\mathrm{k}}} \tag{17}$$

The biological probability of transmission is defined as β per contact. Individuals transition from the infectious $\mathcal{I}$ to susceptible $\mathcal{S}$ health-state via a force of infection λ per year, per susceptible in risk group i:

$$\lambda_i = C_i \sum_k \rho_{ik}\, \beta\, \frac{\mathcal{I}_k}{x_k} \tag{18}$$

Individuals are assumed to transition from the infectious $\mathcal{I}$ to recovered $\mathcal{R}$ health-state at a rate τ per year, reflecting diagnosis and treatment. Individuals in the $\mathcal{R}$ health-state are not infectious and nor susceptible. That is, in this SIR model, individuals cannot become re-infected.

As described in Section 2, individuals enter the model at a rate ν, exit the model at a rate μ, and transition from risk group i to group j at a rate ϕ_{ij}. The turnover rates ϕ and distribution of individuals entering the model by risk group $\hat{\boldsymbol{e}}$ are computed using the methods outlined in Section 2.2.2.

Table 1: Full 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 risk	[0.05 0.20 0.75]
$\hat{\boldsymbol{e}}$	proportion of entering individuals: high, medium, low risk	[0.05 0.20 0.75]
$\boldsymbol{\delta}$	average duration spent in: high, medium, low risk groups	[5 15 25]
C	rate of contact formation among individuals: high, medium, low risk	[25 5 1]
ν	rate of population entry	0.05
μ	rate of population exit	0.03

Our experiments were then carried out under the following assumptions. First, we assumed that the proportion of individuals entering each risk group $\hat{\boldsymbol{e}}$ was equal to the proportion of individuals across risk groups in the model $\hat{\boldsymbol{x}}$. Second, we assumed that the average duration spent in each risk group $\boldsymbol{\delta}$ is known. Third, we assumed that the absolute number of individuals moving between two risk groups in either direction is balanced. The system of equations which results from these assumptions is given in Appendix A.2. To meet all three conditions, there is only one possible value for each element in ϕ and $\hat{\boldsymbol{e}}$ – i.e. A is full rank. In other words, by specifying these three conditions we ensure that a unique set of ϕ and $\hat{\boldsymbol{e}}$ are computed.

Using the above three assumptions, we need only specify the values of $\hat{\boldsymbol{x}}$, $\boldsymbol{\delta}$, ν, and μ. Such parameters could be derived from data; however, in this experiment, we use the illustrative values summarized in Table 1. After resolving the system of equations, $\hat{\boldsymbol{e}}$ is equal to $\hat{\boldsymbol{x}}$ (assumed), and ϕ is:

$$\phi = \begin{bmatrix} * & 0.0833 & 0.0867 \\ 0.0208 & * & 0.0158 \\ 0.0058 & 0.0042 & * \end{bmatrix} \tag{19}$$

The full system of model equations is given in Appendix A.1.

We then simulated epidemics in the base model, and model variants (described below), using these parameters. The model was initialized with $N_0 = 1000$ individuals who are distributed across risk groups according to $\hat{\boldsymbol{x}}$. We seeded the epidemic with one infectious individual in each risk group at $t = 0$. There were no recovered individuals at the start of the epidemic, and so all individuals except the 3 infectious

individuals were susceptible. We numerically solved the system of ordinary differential equations in Python[3] using Euler's method with a time step of $dt = 0.1$ years. All comparative analyses are then conducted at equilibrium, defined as a steady state with <1% difference in incidence.

3.2. Model Variants

To examine the influence of population growth, heterogeneity, and turnover, we simplified the full model accordingly to include or remove each feature. These simplified model variants, and their parameters are shown in Figure 3, and Table 2, respectively.

Experiment 1.1: Influence of risk heterogeneity. V1 assumed no heterogeneity in risk and the V1 contact rate C for all individuals was the weighted average of the Full model's risk-stratified C_i. V1 did not have turnover because there is only one risk group. All other conditions, including population growth rate, remained the same between the Full model and V1 (Table 2).

Experiment 1.2: Influence of population growth. V2 assumed no population growth. The exit rate μ remained fixed and the same as in the Full model, but the entry rate ν was reduced to equal μ. This ensured that the average time spent in the modeled population μ^{-1} remained the same between the Full model and V2.

Experiment 1.3: Influence of turnover. V3 assumed there is no turnover, with all turnover rates $\phi = 0$. Following Eq. (13), this means that in V3, the time spent within all risk groups was equal to the average duration in the modelled population μ^{-1}. Since the proportion of individuals who enter each risk group $\hat{\boldsymbol{e}}$ is assumed to be equal to the proportion of individuals across risk groups in the model $\hat{\boldsymbol{x}}$, as in the Full model, no other modifications are required.

3.3. Influence of Rates of Turnover

To further examine the influence of turnover on equilibrium prevalence and incidence, we varied the rates of turnover in the Full model across a range of values. We then conducted a sensitivity analysis to examine the influence of turnover at different treatment rates (durations of infectiousness).

Experiment 2.1: Influence of the rates of turnover on equilibrium prevalence and incidence. In this experiment, we first considered the influence of turnover at a fixed duration of infectiousness (treatment rate). As in similar experiments (Zhang et al., 2012; Henry and Koopman, 2015), the rates of turnover were scaled by a single parameter. However, because the model used here has $G = 3$ risk groups, multiplying by a set of base rates ϕ by a scalar factor would have resulted in changes to the relative population size of risk

[3] Code for all aspects of the project is available at: https://github.com/c-uhs/turnover

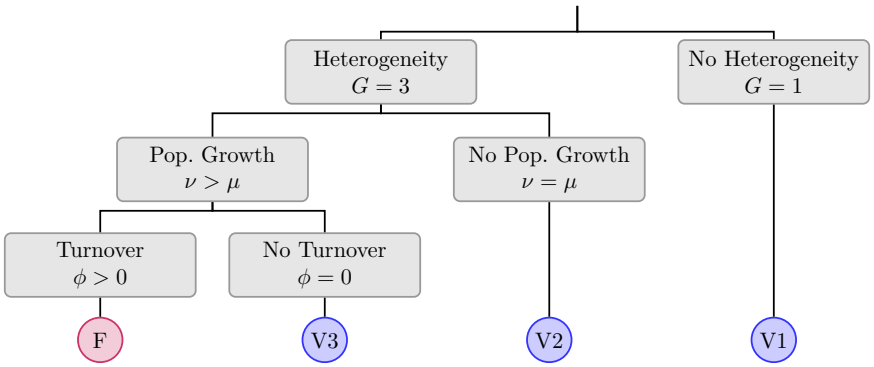


Figure 3: Summary of the Full model (F) and three model variants (V) with respect to heterogeneity in risk, population growth (pop. growth), and turnover. Relative population size of each risk group is the same in all variants. G: number of risk groups, ν: rate of population entry, μ: rate of population exit, ϕ: rates of population turnover.

groups $\hat{\boldsymbol{x}}$. Thus, we controlled the rates of turnover by adjusting the duration of individuals in the high risk group δ_H, such that a shorter period of time spent engaged in highest risk corresponded to higher rates of turnover among all groups. As in Experiment 1, we assumed that: the proportion of individuals who enter each risk group $\hat{\boldsymbol{e}}$ is equal to the proportion of individuals across risk groups in the model $\hat{\boldsymbol{x}}$; and that the absolute number of individuals moving between two groups in either direction is balanced. The duration of individuals in the medium risk group δ_M was then defined as a value between δ_H and the maximum duration μ^{-1} which scales with δ_H following the equation: $\delta_M = \delta_H + \kappa\left(\mu^{-1} - \delta_H\right)$, with $\kappa = 0.3$. Finally, the duration of individuals in the low risk group δ_L similarly scaled with δ_H, but the value was not required to calculate ϕ; it can be determined from ϕ afterwards using Eq. (13). In this way, each value of δ_H was used to define a set of turnover rates ϕ whose elements all scaled inversely with the duration in the high risk group δ_H. The value of δ_H is then varied from 3 to 33 years to examine the influence of turnover.

Experiment 2.2: Influence of the rates of turnover at various treatment rates. Next, we conducted a 2-way sensitivity analyses to examine how the influence of turnover might vary at different treatment rates. The treatment rate controls the duration of infectiousness $\delta_\mathcal{I}$ as in $\delta_\mathcal{I} = \tau^{-1}$. Treatment rate τ was varied from 1 to 0.05, implying a duration of infectiousness of 1 to 20 years. The duration of time spent in the high risk group δ_H was varied from 3 to 33 years as in Experiment 2.1. We examine the influence of the rates of turnover on equilibrium prevalence and incidence across the range of treatment rates using multiple 1D plots and 2D surface plots.

3.4. Influence of turnover on the contribution of highest risk group to overall transmission

Finally, we examined the influence of turnover when fitting models to observed data on disease prevalence across risk groups. We use two models: the Full model with turnover, and model variant V3 without

Table 2: Parameters for model variants.

Parameter	Full	V1	V2	V3
$\hat{\boldsymbol{x}}$	[0.05 0.20 0.75]	—	[0.05 0.20 0.75]	[0.05 0.20 0.75]
$\hat{\boldsymbol{e}}$	[0.05 0.20 0.75]	—	[0.05 0.20 0.75]	[0.05 0.20 0.75]
C	[25 5 1]	[**3**][a]	[25 5 1]	[25 5 1]
δ	[5 15 25]	[**33.3**][b]	[5 15 25]	[**33.3 33.3 33.3**][b]
ν	0.05	0.05	**0.03**[c]	0.05
μ	0.03	0.03	0.03	0.03

$\hat{\boldsymbol{x}}$: proportion of individuals in the model by risk group (high, medium, low); $\hat{\boldsymbol{e}}$: proportion of individuals entering the model by risk group; C: rate of contact formation by risk group (per year); δ: average duration in each risk group (years); ν: rate of population entry (per year); μ: rate of exit (per year).

[a] Weighted average of risk-stratified C

[b] Without turnover, duration in all groups must be equal to the inverse of the exit rate, μ^{-1}

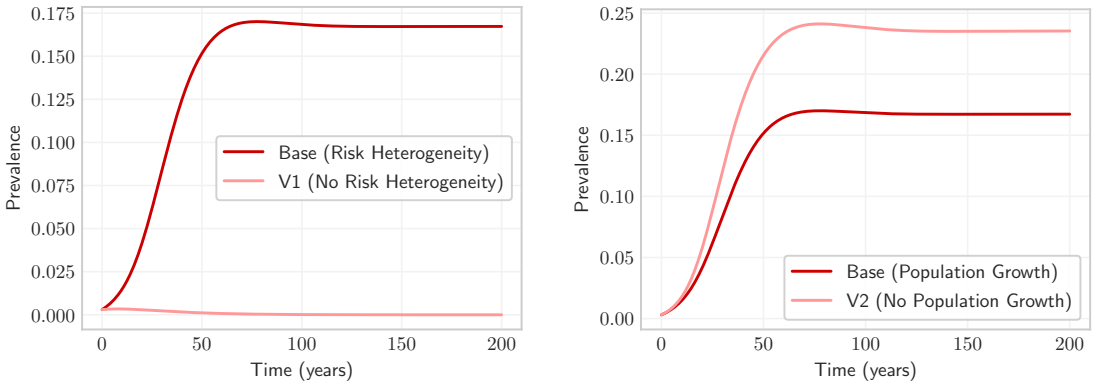
[c] Adjusting the entry rate, versus exit rate, does not affect average duration in the model

turnover (Figure 3: F and V3). Both models include population growth and 3 risk groups. We then fit the following parameter in both models: the contact rates among the highest and lowest risk groups, C_H and C_L. Specifically, we fit the models to equilibrium disease prevalence in the highest and and lowest risk groups. We then estimate the contribution of the highest risk group to overall transmission from both models. The contribution is estimated as the transmission population attributable fraction (TPAF) (Mishra et al., 2016) of the highest risk group.

Experiment 3.1: Influence of turnover on inferred risk heterogeneity. The Full model (turnover) and model variant V3 (no turnover) are calibrated to 25% prevalence in the high risk group, and 5% prevalence in the low risk group at equilibrium. The fitted parameters are the contact rates of the high and low risk groups: C_H and C_L.[4] The ratio of fitted (or posterior) contact rates C_H / C_L represents the degree of risk heterogeneity in the population, after fixing all other parameters, which produce the group-specific observed disease prevalence.

Experiment 3.2: Influence of turnover on the predicted TPAF of the highest risk group. We then calculate and compare the TPAF of the highest risk group across the model variants with and without turnover. We repeat this analysis using the default model parameters (before model fitting), and again using the inferred values of C_H and C_L (after model fitting).

[4] The fitted parameters are estimated by minimizing the negative log-likelihood of each predicted prevalence versus the target (assuming a sample size of 1000) using the Sequential Least SQuares Programming (SLSQP) method (Kraft, 1988) from the `scipy.optimize.minimize` Python package.



(a) Risk heterogeneity (Base vs V1)

(b) Population growth (Base vs V2)

Figure 4: Comparison of model projections with and without risk heterogeneity and with and without population growth

4. Results

In this section we summarize the results of the experiments described in Section 3.

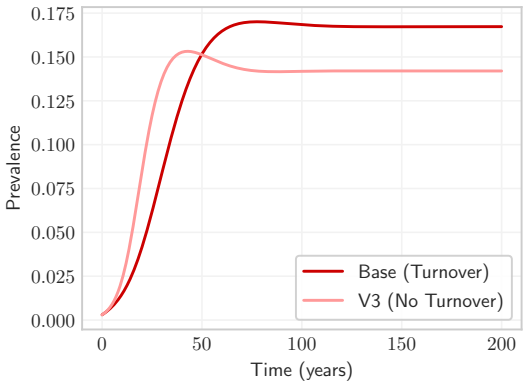
4.1. Model Variants

First, the comparisons of models with and without key features of risk group dynamics are presented.

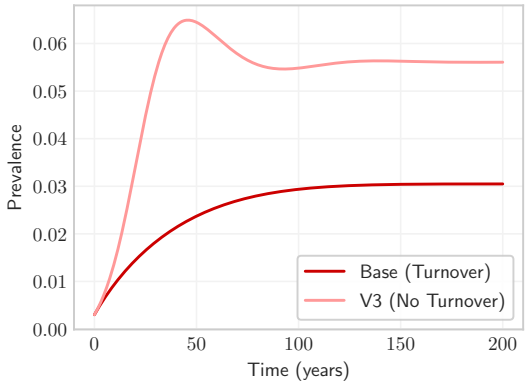
Experiment 1.1: Heterogeneity in Risk. Figure 4a shows the modelled prevalence with and without heterogeneity in risk (Base vs V1). As previously noted in discussions of core group theory (Yorke et al., 1978; Stigum et al., 1994), epidemic dynamics and endemic equilibrium are influenced by the presence of heterogeneity in risk within a population. For this model and parameters (Table 1), failure to model heterogeneity in risk (V1) results in a basic reproduction number $R_0 < 1$, and no epidemic, while the model with heterogeneity (Base) predicts a nonzero endemic equilibrium.

Experiment 1.2: Population Growth. Figure 4b shows the modelled prevalence with and without population growth (Base vs V2). Inclusion of population growth in the model results in lower equilibrium prevalence. This result can be explained following the results of Hadeler and Ngoma (1990), who observed that under exponential population growth, and without disease-attributable mortality, the rate of growth of susceptible individuals exceeds the rate of growth of infected individuals. Thus equilibrium prevalence declines relative to a model with constant population size. For a simplified version of the model used here, the relationship between equilibrium prevalence and model entry rate ν is also given in Appendix A.4.

Experiment 1.3: Turnover. Finally, the influence of including risk group turnover in an epidemic model on equilibrium prevalence is considered. Under the default treatment rate $\tau = 0.1$, overall projected equilibrium



(a) Treatment rate $\tau = 0.2$



(b) Treatment rate $\tau = 0.2$

Figure 5: Comparison of overall projected prevalence with and without risk group turnover (Base vs V3 model variants), under two different treatment rates τ.

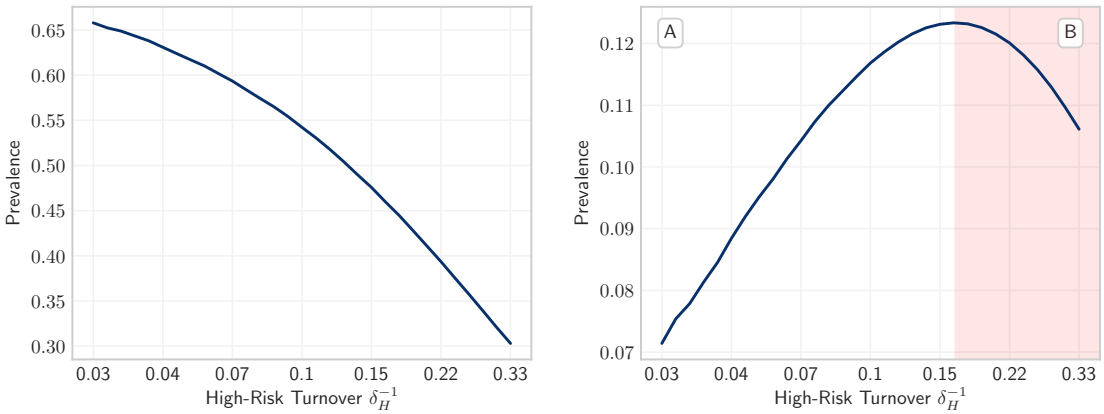
prevalence is slightly higher with turnover than without (Figure 5a). However, if the treatment rate is increased to $\tau = 0.2$, the model with turnover projects a lower equilibrium prevalence than the model without turnover (Figure 5b). Thus, inclusion of risk group turnover influences the equilibrium prevalence. The nature of this influence, however, depends on several factors, including treatment rate and risk group parameters. The next section aims to clarify and explain this influence through exploration of group-specific incidence and prevalence at equilibrium under different rates of turnover ϕ and treatment τ.

4.2. Influence of Turnover

This section presents trends in the influence turnover on incidence and prevalence, with focus on systems at equilibrium.

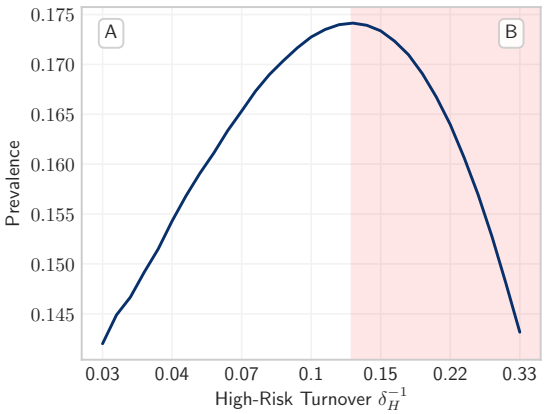
Experiment 2.1: Turnover Magnitude. Figure 6 illustrates trends in equilibrium prevalence versus turnover among the high and low risk groups, as well as overall. As turnover increases, prevalence among the highest risk group decreases (Figure 6a). This is because the proportion of individuals exiting the highest risk group via turnover who are infected is higher than the proportion of individuals entering the highest risk group via turnover who are infected, since prevalence is highest in the highest risk group. That is, the highest risk group experiences a net reduction in the number of infected individuals (illustrated in Figure 7). For low to moderate rates of turnover, this exchange also increases prevalence among the lowest risk group (Figure 6b, region A), and the population overall (Figure 6c, region A).[5] However, at high rates of turnover, turnover

[5] In this model, the lowest risk group usually dominates trends in overall prevalence because this risk group represents 75% of the population (Table 1: $\hat{\boldsymbol{x}}$).



(a) High risk

(b) Low risk



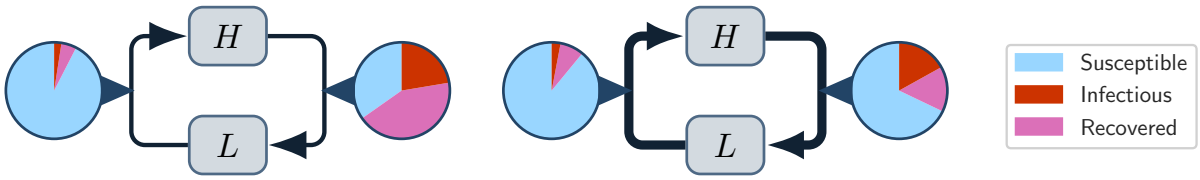
(c) Overall

Figure 6: Equilibrium prevalence among risk groups versus turnover, as controlled by the duration in the high risk group δ_H. Turnover shown in log scale.

decreases the prevalence among the lowest risk group and overall (Figure 6b and 6c, region B), This peak and decline is due to the influence of turnover on incidence, which has not yet been considered.

To understand the influence of turnover on incidence in this model, consider the force of infection equation, Eq (18). As shown in Appendix A.3, the driving component in this expression is the proportion of available partnerships which are offered by infectious individuals, denoted as $\boldsymbol{C}_{\mathcal{I}}$. This component can be further broken down into the following two factors: 1) the average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$, and 2) the proportion of the population who are infectious $\hat{\mathcal{I}}$. Thus the influence of turnover on overall incidence can be understood through the influence of turnover on these two factors.

As shown in Figure 8a, turnover decreases $\hat{C}_{\mathcal{I}}$ the average contact rate among infectious individuals. This is because turnover results in a net movement of infected individuals from high to low risk (Figure 7) (Henry and Koopman, 2015). However, at low to moderate rates of turnover, turnover increases $\hat{\mathcal{I}}$ the



(a) Low turnover: region A

(b) High turnover: region B

Figure 7: Illustrative schematic showing average health states of individuals moving between high H and low L risk groups due to different rates of turnover. Turnover acts to homogenize the distribution of health states across risk groups.

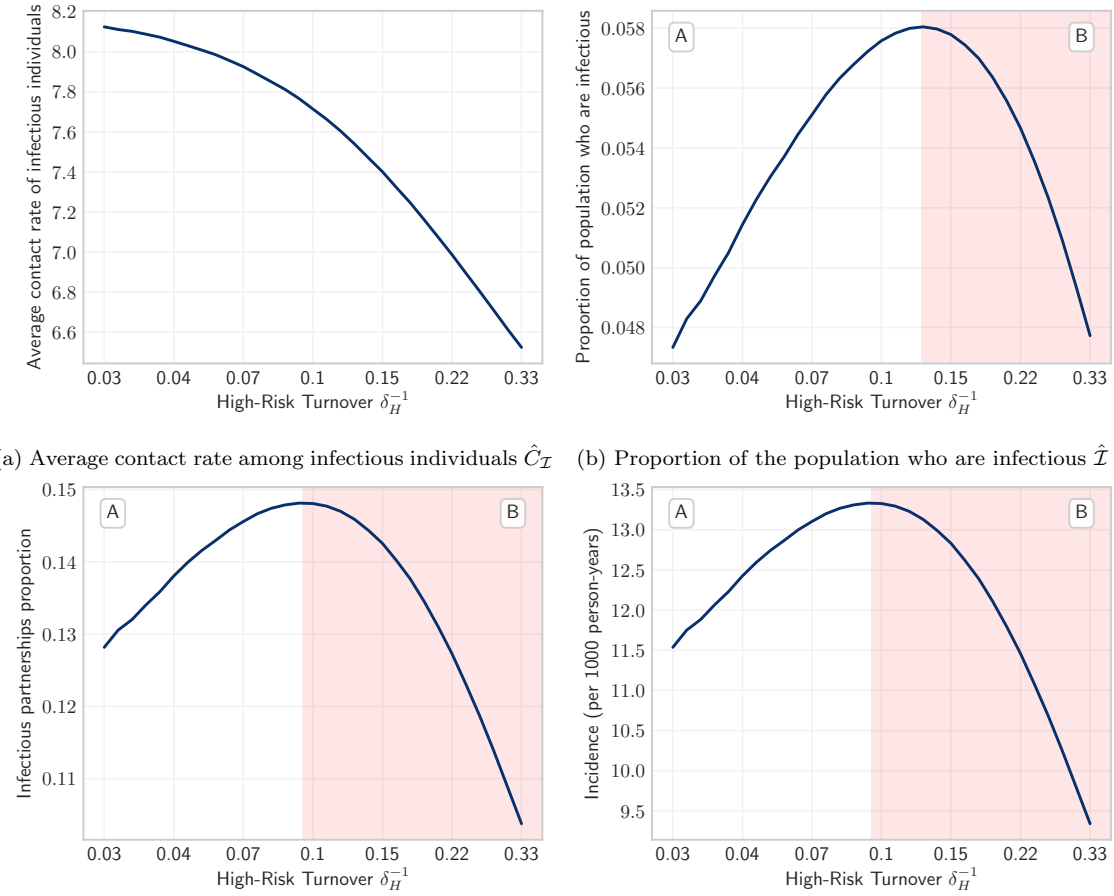
proportion of the population who are infectious (Figure 8b, region A). This is related to the overall increase in prevalence with turnover shown in region A of Figure 6c (Zhang et al., 2012). Under the conditions shown in region A of Figure 8, the proportion of the population who are infectious $\hat{\mathcal{I}}$ increases faster with turnover than the average contact rate of infectious people $\hat{C}_{\mathcal{I}}$ decreases. Thus the overall $\boldsymbol{C}_{\mathcal{I}}$ increases with turnover in region A (Figure 8c) and incidence increases proportionally (Figure 8d).

It therefore follows that the peak in incidence, and the transition between regions A and B in Figure 8c occurs when the dominating factor between $\hat{\mathcal{I}}$ and $\hat{C}_{\mathcal{I}}$ reverses. That is, the transition occurs when the average contact rate of infectious people $\hat{C}_{\mathcal{I}}$ decreases due to turnover more than the proportion of the population who are infectious $\hat{\mathcal{I}}$ increases due to turnover. As rates of turnover continue to increase, declining incidence then reverses the upward trend in $\hat{\mathcal{I}}$, and incidence and prevalence decrease across all groups in a snowball effect. This mechanism then explains the observations shown in region B throughout.

Finally, it is worth noting that the rates of turnover which maximize incidence (Figure 8d) are lower than the rates of turnover which maximize prevalence among the lowest risk group, as well as prevalence overall (Figure 6b and 6c) That is, incidence "peaks" first with respect to turnover. This is because, after peak incidence, increasing turnover may reduce the total number of infections, but it still moves infected individuals from high to low risk (Figure 7).

Experiment 2.2: Turnover and Treatment Rate. So far, the influence of turnover on equilibrium incidence and prevalence has been explored for a single treatment rate. This section explores additional treatment rates τ. First, the factors of incidence are considered in Figure 9. Increasing the treatment rate τ actually increases the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$ (Figure 9a). This is because increasing treatment concentrates infections in the highest risk group [*JK: need to find citation!*], so that $\hat{C}_{\mathcal{I}}$, on average, increases. However, increasing treatment reduces the proportion of the population who are infectious $\hat{\mathcal{I}}$ (Figure 9b), as the rate of transition between $\mathcal{I}$ and $\mathcal{R}$ increases. The dominant effect is that of $\hat{\mathcal{I}}$, which tends towards zero faster than $\hat{C}_{\mathcal{I}}$ tends towards infinity. Thus, incidence declines with treatment (Figure 9d).

Figure 9d also shows that the rate of turnover which maximizes incidence decreases with increasing treatment. That is, as treatment rates increase, turnover is more likely to decrease incidence than it is to



(a) Average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$

(b) Proportion of the population who are infectious $\hat{\mathcal{I}}$

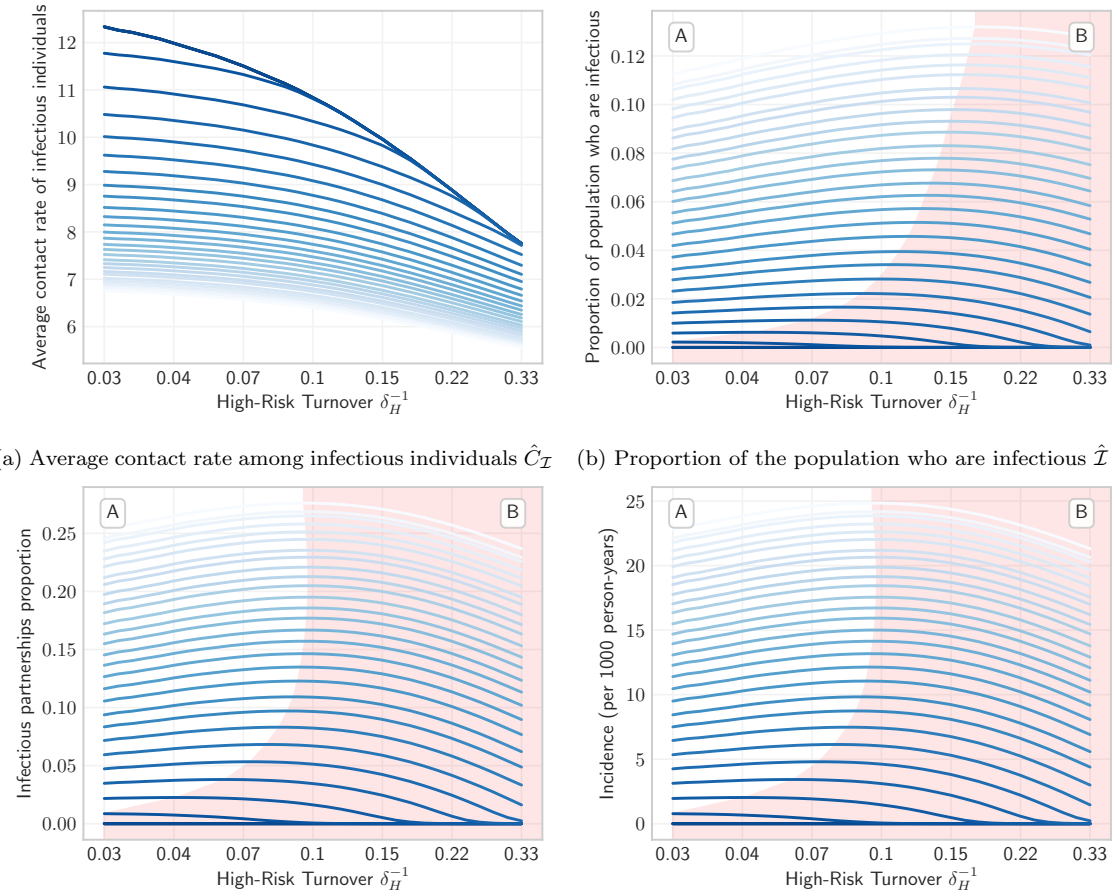
(c) Proportion of available partnerships which are offered by infectious individuals $\boldsymbol{C}_{\mathcal{I}}$

(d) Overall incidence λ

Figure 8: Incidence and the dynamic factors of incidence versus turnover. The product of components (a) and (b) is proportional to (c) the proportion of total available contacts which are with infectious individuals and (d) overall incidence.

increase incidence (region B grows). This effect can be explained as follows. Recall that the mechanism by which turnover decreases incidence is through reduction of the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$, due to net movement of infectious individuals from high to low risk (Figure 7). If treatment increases the concentration of infections in the high risk group, then the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$ will be more sensitive to redistribution of those infectious individuals via turnover. In Figure 9a, this is shown as the larger downward slope of $\hat{C}_{\mathcal{I}}$ versus turnover at higher treatment rates (darker blue). Therefore, at higher treatment rates, turnover is more likely to decrease incidence than increase incidence because movement of infectious individuals from high to low risk has a larger impact on the average contact rate of infectious individuals.

Finally, Figure 10 summarizes trends in overall equilibrium incidence and group-specific prevalence with



(a) Average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$

(b) Proportion of the population who are infectious $\hat{\mathcal{I}}$

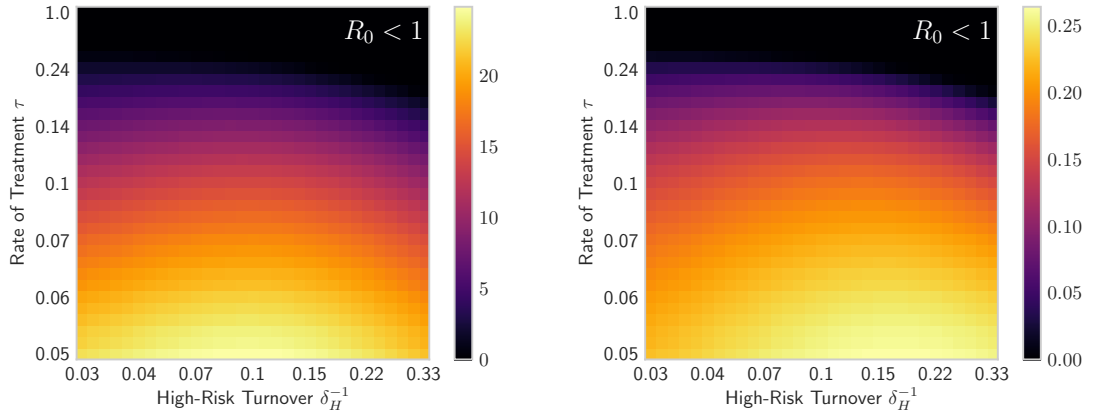
(c) Proportion of available partnerships which are offered by infectious individuals $\boldsymbol{C}_{\mathcal{I}}$

(d) Overall incidence λ

Figure 9: Incidence and the dynamic factors of incidence versus turnover, for a range of treatment rates. Darker blue indicates higher treatment rate. The product of components (a) and (b) is proportional to (c) the proportion of total available contacts which are with infectious individuals and (d) overall incidence.

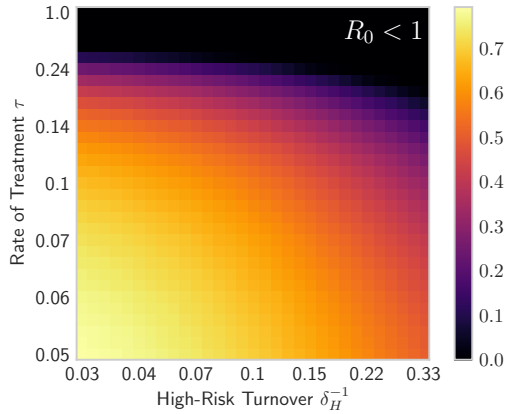
respect to both turnover ϕ and treatment rate τ.[6] Treatment consistently decreases equilibrium incidence (as noted above), as well as prevalence, at all rates of turnover. As suggested in Experiment 2.1, prevalence among the highest risk group (Figure 10c) also decreases with turnover for any rate of treatment. Similarly, prevalence among the low risk group (Figure 10d) increases with turnover for moderate rates of turnover and treatment. However, as turnover increases past the point which maximizes incidence, prevalence among the low risk group peaks, and then declines. As shown in Figure 9, the rate of turnover at which this occurs decreases with treatment rate. Once again, trends in overall prevalence roughly reflect those of the lowest risk group. Finally, for high rates of treatment and/or turnover, the product of the average contact rate of

[6] Figure 10a is the surface projection of the profiles shown in Figure 9d.

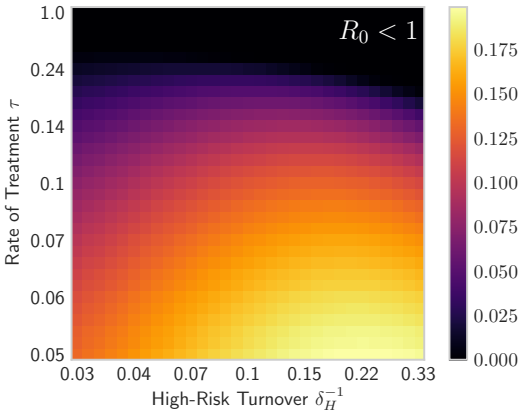


(a) Overall incidence

(b) Overall prevalence



(c) High risk prevalence



(d) Low risk prevalence

Figure 10: Equilibrium prevalence and incidence for different rates of turnover ϕ (log scale) and treatment τ (log scale).

infectious individuals $\hat{C}_{\mathcal{I}}$ and the proportion of the population who are infected $\hat{\mathcal{I}}$ is too low to sustain the epidemic. That is, the basic reproductive number R_0 declines to less than one, and no epidemic is observed.[7]

4.3. Fitted Models with Turnover

Finally, the influence of turnover on fitted models is explored. For reference, the pre-calibration equilibrium prevalence predicted among the and low high risk groups are shown in Figure B.2. The prevalence ratios are 3.5 with turnover, and 9.4 without.

Experiment 3.1: Inferred Risk Heterogeneity. Following calibration of contact rates C_H and C_L, both models predict an equilibrium prevalence 25% and 5%, as desired. However, the fitted contact rates required to

[7] In fact, it can be shown that for extreme rates of turnover, a heterogeneous system (e.g. Base model) will converge on a homogeneous system (e.g. model V1). This result is shown in Figure B.1.

Table 3: Equilibrium contact rates C and prevalence P among the high H and low L risk groups predicted by the Base model (turnover) and V3 (no turnover) before and after model fitting.

Model	C_H	C_L	C_H / C_L	P_H	P_L	P_H / P_L
Base	25.0	1.0	25	42%	12%	**3.5**
V3	25.0	1.0	25	66%	7%	**9.4**
Base [fit]	16.9	0.28	**60**	25%	5%	5.0
V3 [fit]	15.8	2.49	**6.3**	25%	5%	5.0

yield these outputs are different with and without turnover (Table 3); the ratio of C_H / C_L with turnover (60) is much higher than the ratio without turnover (6.3). Thus, for the model and conditions explored here, the inferred heterogeneity in risk is higher in the model with turnover than in the model without turnover. This is because turnover acts to counteract the concentration of risk. That is, in order to observe the same prevalence ratio in a system with turnover, the "risk homogenizing effects" of turnover must be overcome by greater heterogeneity in risk, as compared to a system without turnover.

Experiment 3.2: TPAF of the High Risk Group. Figure 11 shows the estimated TPAF of the high risk group with and without turnover, and before and after model fitting. The TPAF approaches 1.0 for all models over a 100 year time horizon, indicating that unmet treatment needs of the high risk group are central to epidemic persistence in all models. Additionally, no TPAFs intersect during this period, so relative differences between TPAFs by model can be described irrespective of time horizon.

Before model fitting (Figure 11, solid lines), the model without turnover estimates a larger TPAF of the high risk group than the model with turnover. This can be attributed to the larger equilibrium prevalence ratio before model fitting (Table 3, Figure B.2), which results in more onward transmission from the high prevalence high risk group. However, after fitting contact rates C_H and C_L to high and low risk prevalence targets as described above, the model with turnover estimates a higher TPAF of the high risk group than the model without turnover (Figure 11, dashed lines). This reversal in which model predicts a higher TPAF can be explained by two factors. First, the equilibrium prevalence ratios predicted by the models with and without turnover are equalized through model fitting to the same targets. As a result, higher prevalence among the high risk group in the model without turnover no longer contributes to an increased TPAF estimate by this model. Second, as shown in Experiment 3.1, the ratio of fitted contact rates C_H / C_L in the model with turnover are higher than in the model without. This affords a higher risk of onward transmission to the high risk group in the model with turnover, and thus an increased TPAF. This result then implies that models which fail to capture turnover dynamics which are truly present in reality may underestimate the TPAF of high risk groups. Consequently, the importance of prioritizing high risk groups

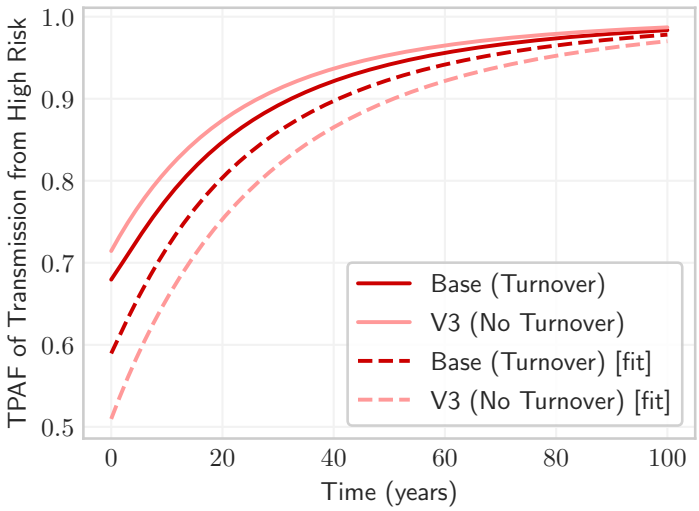


Figure 11: Transmission population attributable fraction (TPAF-from) of the high risk group with and without turnover, and with and without fitted contact rates to group-specific prevalence.

to achieve epidemic control may be similarly underestimated by such models.

5. Discussion

Under construction!

6. Conclusion

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.

Garnett, G. P. and Anderson, R. M. (1994). Balancing sexual partnership in an age and activity stratified model of HIV transmission in heterosexual populations. *Mathematical Medicine and Biology*, 11(3):161–192.

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.

Hadeler, K. and Ngoma, K. (1990). Homogeneous models for sexually transmitted diseases. *Rocky Mountain Journal of Mathematics*, 20(4):967–986.

Henry, C. J. and Koopman, J. S. (2015). Strong influence of behavioral dynamics on the ability of testing and treating HIV to stop transmission. *Scientific Reports*, 5(1):9467.

Koopman, J. S., Jacquez, J. A., Welch, G. W., Simon, C. P., Foxman, B., Pollock, S. M., Barth-Jones, D., Adams, A. L., and Lange, K. (1997). The role of early HIV infection in the spread of HIV through populations. *Journal of Acquired Immune Deficiency Syndromes*, 14(3):249–58.

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.

Lawson, C. L. and Hanson, R. J. (1995). *Solving least squares problems*, volume 15. SIAM.

Mishra, S., Boily, M. C., Schwartz, S., Beyrer, C., Blanchard, J. F., Moses, S., Castor, D., Phaswana-Mafuya, N., Vickerman, P., Drame, F., Alary, M., and Baral, S. D. (2016). Data and methods to characterize the role of sex work and to inform sex work programs in generalized HIV epidemics: evidence to challenge assumptions. *Annals of Epidemiology*, 26(8):557–569.

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.

Nold, A. (1980). Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, 52(3-4):227–240.

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.

Watts, C., Zimmerman, C., Foss, A. M., Hossain, M., Cox, A., and Vickerman, P. (2010). Remodelling core group theory: the role of sustaining populations in HIV transmission. *Sexually Transmitted Infections*, 86(Suppl 3):iii85–iii92.

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.

Zhang, X., Zhong, L., Romero-Severson, E., Alam, S. J., Henry, C. J., Volz, E. M., and Koopman, J. S. (2012). Episodic HIV Risk Behavior Can Greatly Amplify HIV Prevalence and the Fraction of Transmissions from Acute HIV Infection. *Statistical Communications in Infectious Diseases*, 4(1).

Appendix A. Supplemental Equations

Appendix A.1. Model Equations

For each risk group i:

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

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

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

Appendix A.2. Complete Example Turnover System

$$\begin{array}{r} \left.\begin{array}{r} \\ \\ \\ \end{array}\right\} \text{conservation of mass} \\ \left.\begin{array}{r} \\ \\ \\ \end{array}\right\} \text{entering distribution} \\ \left.\begin{array}{r} \\ \\ \\ \end{array}\right\} \text{group duration} \\ \left.\begin{array}{r} \\ \\ \\ \end{array}\right\} \text{balance turnover} \end{array} \left\{ \begin{bmatrix} \nu x_1 \\ \nu x_2 \\ \nu x_3 \\ e_1^* \\ e_2^* \\ e_3^* \\ {\delta_1}^{-1} - \mu \\ {\delta_2}^{-1} - \mu \\ {\delta_3}^{-1} - \mu \\ 0 \\ 0 \\ 0 \end{bmatrix} \right. = \begin{bmatrix} \nu & \cdot & \cdot & -x_1 & -x_1 & x_2 & \cdot & x_3 & \cdot \\ \cdot & \nu & \cdot & x_1 & \cdot & -x_2 & -x_2 & \cdot & x_3 \\ \cdot & \cdot & \nu & \cdot & x_1 & \cdot & x_2 & -x_3 & -x_3 \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & \cdot & x_1 & \cdot & -x_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & x_1 & \cdot & \cdot & -x_3 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & x_2 & \cdot & -x_3 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \phi_{12} \\ \phi_{13} \\ \phi_{21} \\ \phi_{23} \\ \phi_{31} \\ \phi_{32} \end{bmatrix} \tag{A.4}$$

Appendix A.3. Factors of Incidence

Rearranging the force of infection λ_i to isolate the dynamic (not constant) component ($*$):

$$\begin{aligned} \lambda_i &= C_i \sum_k \rho_{ik}\beta \frac{\mathcal{I}_k(t)}{N_k} \\ &= C_i\beta \sum_k \frac{C_k N_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} N_{\mathrm{k}}} \frac{\mathcal{I}_k(t)}{N_k} \\ &= C_i\beta \underbrace{\frac{\sum_k C_k \mathcal{I}_k(t)}{\sum_{\mathrm{k}} C_{\mathrm{k}} N_{\mathrm{k}}}}_{*} \end{aligned} \tag{A.5}$$

This component $(*)$ is: $\boldsymbol{C}_{\mathcal{I}}$ *the proportion of available partnerships which are offered by infectious individuals.* As the only dynamic component, only this component can be affected by turnover.

Now consider that $\boldsymbol{C}_{\mathcal{I}}$ can be written in terms of the following three factors:

- The average contact rate among infectious individuals $\hat{C}_{\mathcal{I}} = \frac{\sum_k C_k \mathcal{I}_k}{\sum_k \mathcal{I}_k}$
- The proportion of the population who are infectious $\hat{\mathcal{I}} = \frac{\sum_k \mathcal{I}_k}{\sum_k N_k}$
- The average contact rate among all individuals (constant) $\hat{C} = \frac{\sum_k C_k N_k}{\sum_k N_k}$

$$\begin{aligned} \boldsymbol{C}_{\mathcal{I}} &= \hat{C}_{\mathcal{I}} \times \hat{\mathcal{I}} \times \hat{C}^{-1} \\ &= \frac{\sum_k C_k \mathcal{I}_k}{\sum_k \mathcal{I}_k} \times \frac{\sum_k \mathcal{I}_k}{\sum_k N_k} \times \frac{\sum_k N_k}{\sum_k C_k N_k} \\ &= \frac{\sum_k C_k \mathcal{I}_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} N_{\mathrm{k}}} \end{aligned} \tag{A.6}$$

Therefore, actually only two dynamic factors control the force of infection: 1) the average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$, and 2) the proportion of the population who are infectious $\hat{\mathcal{I}}$; and the product of these factors (scaled by $\hat{C}^{-1}$) gives $\boldsymbol{C}_{\mathcal{I}}$. Overall incidence is then directly proportional to $\boldsymbol{C}_{\mathcal{I}}$, following Eq. (A.5). In fact, the incidence in each group individually is proportional to $\boldsymbol{C}_{\mathcal{I}}$, as C_i is only factor depending on i.

Appendix A.4. Simplified Equilibrium Prevalence vs Growth Rate

The aim of this derivation is to relate the equilibrium prevalence P to the rate of population entry ν for a simplified epidemic model. A homogeneous ($G = 1$) SI model is assumed. The expression for prevalence is given by:

$$P = \frac{\mathcal{I}}{\mathcal{S} + \mathcal{I}} \tag{A.7}$$

At equilibrium, prevalence is unchanging. The expression for $\frac{d}{dt}P$ can be obtained by the quotient rule:

$$\frac{d}{dt}P = \frac{\left(\frac{d}{dt}\mathcal{I}\right)\left(\mathcal{S} + \mathcal{I}\right) - \left(\mathcal{I}\right)\left(\frac{d}{dt}(\mathcal{S} + \mathcal{I})\right)}{\left(\mathcal{S} + \mathcal{I}\right)^2} \tag{A.8}$$

which simplifies to:

$$\frac{d}{dt}P = \frac{\mathcal{S}\lambda - \mathcal{I}\nu}{\mathcal{S} + \mathcal{I}} \tag{A.9}$$

Separating terms in the numerator, this can be further re-written as follows:

$$\frac{d}{dt}P = \lambda(1 - P) - \nu P \tag{A.10}$$

Moreover, considering the simplified system, the force of infection can be written as $\lambda = \beta \frac{\mathcal{I}}{\mathcal{S}+\mathcal{I}} = \beta P$. Thus, the rate of change of prevalence is given by:

$$\frac{d}{dt}P = \beta P(1-P) - \nu P \tag{A.11}$$

Now, at equilibrium, $\frac{d}{dt}P = 0$, and so $\beta P_{eq}(1-P_{eq}) = \nu P_{eq}$, which simplifies to:

$$\frac{\nu}{\beta} = (1-P_{eq}) \tag{A.12}$$

Therefore, as the population growth rate ν increases, the equilibrium prevalence P_{eq} must decrease.

Appendix B. Supplemental Figures

Appendix B.1. Homogenizing Effect of Turnover

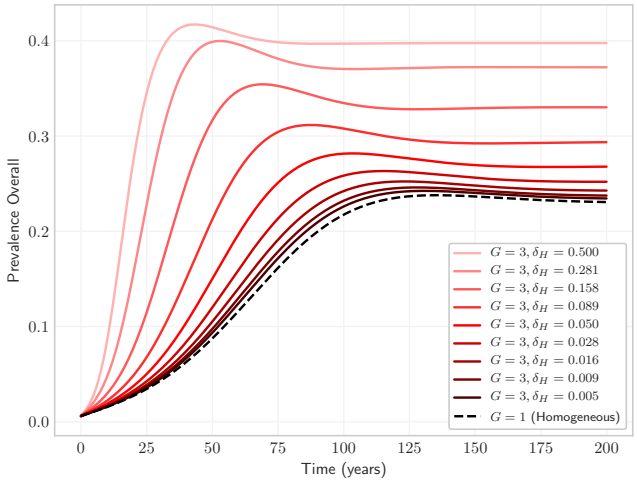
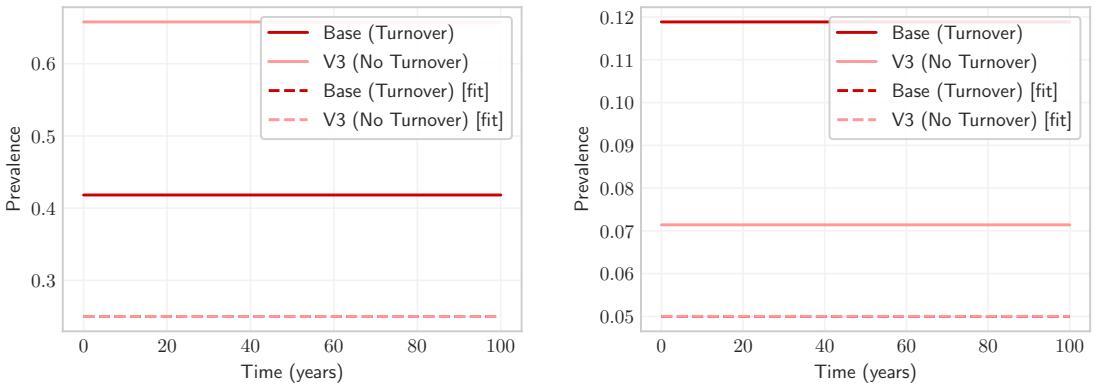


Figure B.1: Overall prevalence predicted by a heterogeneous system for a range of high turnover rates. Note how the heterogeneous model ($G = 3$) converges on a homogeneous model ($G = 1$) with very high turnover rates. Compared to the Base model, transmission probability is increased to $\beta = [TBD]$ in order to yield non-zero equilibrium prevalence in the homogeneous model.

Appendix B.2. Equilibrium Prevalence Before and After Model Calibration



(a) High risk

(b) Low risk

Figure B.2: Equilibrium prevalence among high and low risk groups with and without turnover, and with and without fitted C_i to group-specific prevalence.