Modelling Epidemic Risk Group Dynamics

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

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

Abstract

This is the abstract.

Keywords: TBD, TBD, TBD

* Corresponding author (knightje@smh.ca)

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

1. Introduction

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

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

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

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

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

2. System

This section introduces 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. The number of individuals in risk group $i \in [1, \ldots, G]$ is denoted as x_i and the set of all risk groups is denoted 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 $\hat{x}_i = x_i/N$. 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 (model entrants, $\hat{e}_i$) 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 can occur between any two groups, in either direction; therefore the turnover rates are denoted 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 the diagonal elements are written $*$ 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, consider the goal of constructing a system like the one described above, which reflects the risk group dynamics observed in a real-world context. It will be assumed that the number of risk groups to model is already determined, and also that the relative sizes of these groups ($\hat{x}_i$) should remain constant over time. Thus, what remains is to estimate the values of the following parameters: ν, μ, $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{e}}$, and ϕ. Methods to do so, which leverage some commonly available sources of data, are presented next.

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

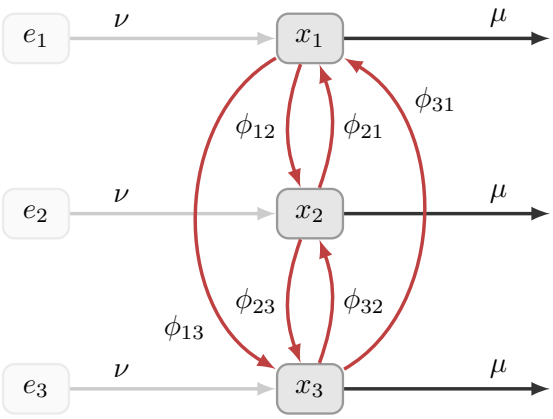


Figure 1: System of states and flows between them for $G = 3$

2.2.1. Total Population Size

The difference between the entry rate $\nu(t)$ and exit rate $\mu(t)$ for a population, gives a net growth rate $\mathcal{G}(t)$. The total population $N(t)$ is then defined by the exponential growth of the original population N_0 according to $\mathcal{G}(t)$,

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

$$\tag{2b}$$

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

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

Variation in rate of entry across risk groups is captured in $\hat{\boldsymbol{e}}$, and rate of exit is generally not stratified by activity group (besides disease-attributable death, which we do not consider here). Therefore, it can be assumed that ν and μ do not vary across risk groups, allowing these rates to be derived independent of the population proportions $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{e}}$, and turnover ϕ.

The simplest approach to obtaining ν and μ assumes a constant population size $N(t) = N_0$. This implies a growth rate of zero, yielding $\nu = \mu$. However, this does not reflect the positive population growth of most real contexts. Another approach is to fix $\mathcal{G}(t)$ as some constant. When using this approach, care should be taken to ensure the resulting $N(t)$ matches 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. (**??**):

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

All of these approaches help define $\mathcal{G}(t)$, but leave one degree of freedom, since many choices of $\mu(t)$ can be compensated by an appropriate choice of $\nu(t)$ to yield the desired $\mathcal{G}(t)$. However, an assumed duration of

individuals in the model $\delta(t)$ can usually be leveraged to choose $\mu(t)$ as in Eq. (3). This duration can be assumed, or be chosen to reflect the age range in data sources for other model parameters. Then, $\nu(t)$ can be resolved using the growth rate in Eq. (2a).

2.2.2. Turnover

Next, methods for resolving $\hat{\boldsymbol{e}}(t)$ and $\phi(t)$ are presented, assuming that $\nu(t)$ and $\mu(t)$ are known. Similar to above, the problem is first formulated as a system of equations. then the data and assumptions which can be leveraged to solve the system will be considered.

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:

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

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

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

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

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

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

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

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

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

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

These durations could be derived from survey data, including for key populations, or they could be assumed. Rearranging Eq. (8), we obtain ${\delta_i}^{-1} - \mu = \sum_j \phi_{ij}$, 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.2 Complete Example Turnover System, 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 ϕ. For example, the proportion of individuals who move from one risk group to another each year (ϕ_{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 ϕ_{ij} for a given i as well as the duration for the group δ_i will only add $G - 1$ (not G) constraints, since Eq. (8) must hold. Second, not all combinations of specified values will yield a valid solution,[2] and it is unfortunately difficult to anticipate problematic combinations.

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

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

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

2.3. Previous Approaches

INP

[2] Even rank-deficient systems be inconsistent.

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.

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

3. Experiment

In the previous section, we have described an approach to parameterizing models of risk group dynamics which highlights the *feasibility* of including such model components based on available data. In this section, we explore the *importance* of including these components. The aim of these experiments is to explore the influence of different risk group dynamics, and especially turnover, on epidemic model outputs. To do so, we compare prevalence and incidence at equilibrium, with and without: population growth, heterogeneity in risk, and risk group turnover. We also explore the influence of different rates of turnover on overall and group-specific equilibrium prevalence and incidence. Finally, we highlight the importance of turnover for calibrated models, especially when estimating the transmission population attributable fraction (TPAF) of risk groups.

3.1. Model & Simulations

We start with a deterministic 1-sex SIR model of transmission in a population with heterogeneity in risk. The model is not representative of a specific infection but requires contact matching as per a sexually transmitted infection. The model includes three health states: susceptible $\mathcal{S}$, infected $\mathcal{I}$, and recovered $\mathcal{R}$ (Figure 2), and $G = 3$ levels of risk: high H, medium M, and low L. Risk levels are stratified by different rates of contact formation across groups, reflecting sexual mixing, so that individuals in risk group i are assumed to form contacts at a rate C_i. The probability ρ of contact formation between individuals in group i with partners in risk group k is assumed to be proportionate with the total number of available contacts:

$$\rho_{ik} = \frac{C_k x_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} x_{\mathrm{k}}} \tag{10}$$

Transmission of the infection from infected $\mathcal{I}$ to susceptible $\mathcal{S}$ individuals is assumed to occur with probability β per contact. The force of infection for susceptible individuals in risk group i is therefore modelled using the following equation:

$$\lambda_i = C_i \sum_k \rho_{ik}\, \beta\, \frac{\mathcal{I}_k}{x_k} \tag{11}$$

Infected individuals $\mathcal{I}$ are assumed to be diagnosed, treated and recover at a rate τ (per year). Recovered individuals $\mathcal{R}$ are not considered infectious, but do not return to the susceptible state.

As described in the previous section, 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 model

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

Symbol	Description	Value
$\boldsymbol{\beta}$	transmission probability per contact	0.03
τ	rate of treatment initiation among infected	0.1
N_0	initial population size	1000
$\hat{\boldsymbol{x}}$	proportion of system individuals: high, medium, low 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

entrants by risk group $\hat{\boldsymbol{e}}$ are resolved using the methods outlined in Section 2.2.2. To this end, the following three assumptions ensure that a unique set of ϕ and $\hat{\boldsymbol{e}}$ are determined – i.e. only one possible value for each element in ϕ and $\hat{\boldsymbol{e}}$ can satisfy these assumptions simultaneously. First, the distribution of model entrants $\hat{\boldsymbol{e}}$ by risk group is assumed to equal the distribution of individuals in the model $\hat{\boldsymbol{x}}$. Second, it is assumed that the average duration spent in each risk group $\boldsymbol{\delta}$ is known. Finally, the absolute number of individuals moving between two groups in either direction is assumed to be balanced. The resulting system of equations which can be used to resolve ϕ and $\hat{\boldsymbol{e}}$ is given in Appendix A.2.

The above assumptions specify which parameters are defined, but not what specific value they take. Those parameter values are summarized for this base model in Table 1. After resolving the system of equations, $\hat{\boldsymbol{e}}$ is equal to $\hat{\boldsymbol{x}}$ (assumed), and ϕ is:

$$\phi = \begin{bmatrix} * & 0.0833 & 0.0867 \\ 0.0208 & * & 0.0158 \\ 0.0058 & 0.0042 & * \end{bmatrix} \tag{12}$$

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

Epidemics are then simulated using these parameters in this model (and variants, described below). Each epidemic is initialized at $t = 0$ with $N_0 = 1000$ individuals, and initial individuals distributed among risk groups according to $\hat{\boldsymbol{x}}$. All initial individuals are susceptible $\mathcal{S}$ except for one infected person $\mathcal{I}$ in each risk group. Each epidemic is solved numerically in Python[3] using Euler's method with a time step of $dt = 0.1$ years. The model is assumed to be at equilibrium after $t = 500$ years, which was verified qualitatively.

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

3.2. Model Variants

Consider the risk group dynamics outlined in Box 1. The assumption b) in each case is taken to be generally more plausible: populations *are* heterogeneous in risk; turnover among risk groups *does* occur; and population growth *does* occur. These assumptions are reflected in the base model, described above. However, many epidemic models do not make these assumptions. The following experiments aim the illustrate the potential influence of such assumptions on model outputs. In each case, a model variant is constructed which makes the alternate assumption a): 1.1) populations *are not* heterogeneous in risk; 1.2) turnover among risk groups *does not* occur; and 1.3) population growth *does not* occur. An epidemic is then simulated with fixed parameter values using both the base model and the variant. Overall projected incidence and prevalence are then compared between the models. The details of each model variant are summarized as follows.

Experiment 1.1: Risk Heterogeneity. The first variant (V1) is defined to no longer consider heterogeneity in risk. The three risk groups are combined, and the new contact rate C for all individuals is defined as the weighted average of previously risk-stratified C_i. It is also no longer possible to include turnover, as there is only one risk group. The number of individuals who are initially infected remains as three.

Experiment 1.2: Population Growth. Another variant (V2) is defined which does not model population growth. In this case, the exit rate μ remains fixed but the entry rate ν is reduced to equal μ. This ensures that the average duration of individuals in the model μ^{-1} remains unchanged.

Experiment 1.3: Turnover. Finally, variant (V3) is defined with all turnover rates $\phi = 0$. Following Eq. (8), this is equivalent to assuming that the duration in each group is equal to the average duration of individuals in the model μ^{-1}. Since the distribution of risk behaviour in the entering population $\hat{\boldsymbol{e}}$ was assumed to be equal to that in the model $\hat{\boldsymbol{x}}$, no other modifications are required.

The model variants constructed for each experiment are summarized in Figure 3, and the corresponding parameters are given in Table 2.

3.3. Influence of Turnover

As noted above, the influence of turnover on equilibrium prevalence and incidence depends on several factors, including turnover magnitude and duration of infectiousness. To help understand trends in model outputs with respect to these factors, the base model is used to explore a range of turnover magnitudes and durations of infectiousness.

Experiment 2.1: Turnover Magnitude. First, the magnitude of turnover is considered alone, for fixed duration of infectiousness. As in similar experiments (Zhang et al., 2012; Henry and Koopman, 2015), the rates of turnover are scaled by a single parameter. However, since the model used here has $G = 3$ risk groups, it is not possible to simply multiply a set of base rates ϕ by a scalar factor; this would result in changes to

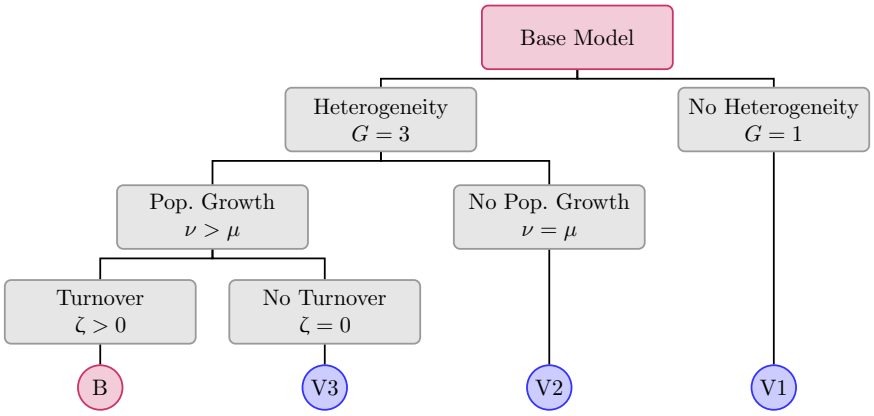


Figure 3: Summary of four model variants with respect to simulated risk group dynamics. Relative risk group size is the same in all variants. G: number of risk groups, ν: rate of population entry, μ: rate of population exit, ϕ: rates of population turnover.

Table 2: Parameters for model variants. All rates have units year^{-1} and durations are in years. Vectors correspond to parameters stratified by high, medium, and low risk groups.

Parameter	Base	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

[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

the equilibrium risk group sizes, which is avoided throughout this work. Rather, the rates of turnover are controlled by the duration of individuals in the high risk group δ_H in the following way. The distribution of model entrants $\hat{\boldsymbol{e}}$ is again assumed to equal the distribution of individuals in the model $\hat{\boldsymbol{x}}$. The absolute number of individuals moving between two groups in either direction is also assumed to be balanced, as before. The duration of individuals in the medium risk group δ_M is 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 scales with δ_H, but the value is not required to resolve ϕ; it can be determined from ϕ afterwards using Eq. (8). In this way, each value of δ_H can be used to compute a set of turnover rates ϕ whose elements all scale inversely with the duration in the high risk group δ_H. The value of δ_H is then varied from 3 to 33 years, and trends in the equilibrium

incidence and prevalence are plotted for each group, as well as overall.

Experiment 2.2: Turnover and Treatment Rate. Next, the above experiment is repeated for a range of treatment rates. The treatment rate controls the duration of infectiousness $\delta_{\mathcal{I}}$ as in $\delta_{\mathcal{I}} = \tau^{-1}$. Treatment rate τ is varied from 1 to 0.05, implying a duration of infectiousness of 1 to 20 years. The duration in the high risk group δ_H is varied from 3 to 33 years as before, and trends equilibrium incidence and prevalence are again shown, this time using 2D surface plots.

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, some potential implications of omitting turnover from a fitted model are explored. In particular, the influence of turnover on two model outputs are estimated, before and after model fitting: the inferred level of risk heterogeneity in the population; and the transmission population attributable fraction (TPAF) (Mishra et al., 2016) of the high risk group. TPAF estimates the proportion of cumulative new infections which are attributable to prevention gaps among a specific population.[4]

Experiment 3.1: Inferred Risk Heterogeneity. First, the Base model (turnover) and model variant V3 (no turnover) are both 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.[5] The ratio of fitted contact rates C_H / C_L represents the degree of risk heterogeneity in the population which must be present in order to observe the given prevalence ratio. Comparing the fitted contact rates with and without turnover, the influence of turnover on inferred risk heterogeneity is shown.

Experiment 3.2: TPAF of the High Risk Group. The level of risk heterogeneity has presumed implications for prioritization of risk groups for interventions. The TPAF of a risk group provides an estimate of the importance of prioritizing the group. Thus, differences in inferred risk heterogeneity due to turnover from Experiment 3.1 are hypothesized to result in differences in the estimated TPAF of the high risk group. To test this hypothesis, the TPAF of the high risk group is calculated and compared across the model variants with and without turnover, before and after fitting to group-specific prevalence, as described above.

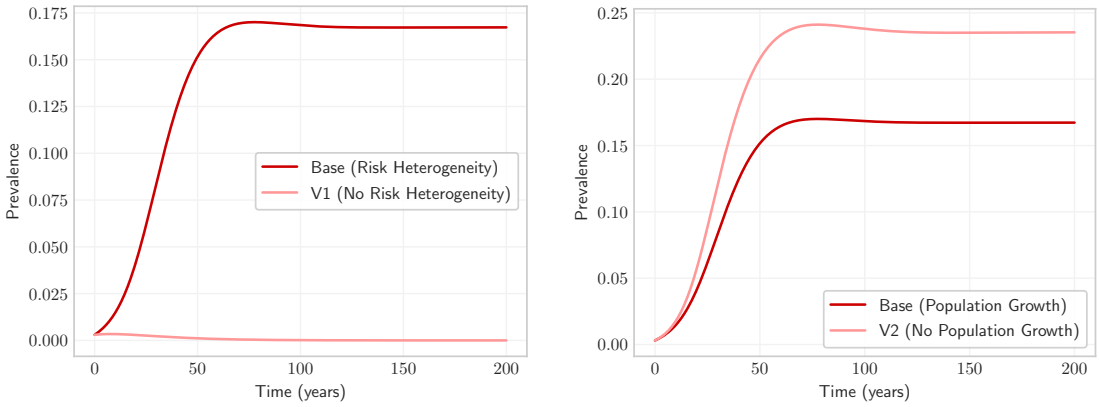
4. Results

In this section we summarize the results of the experiments described above.

---

[4] To estimate TPAF of the high risk group, transmission "from" the high risk group is turned off, not "to".

[5] 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

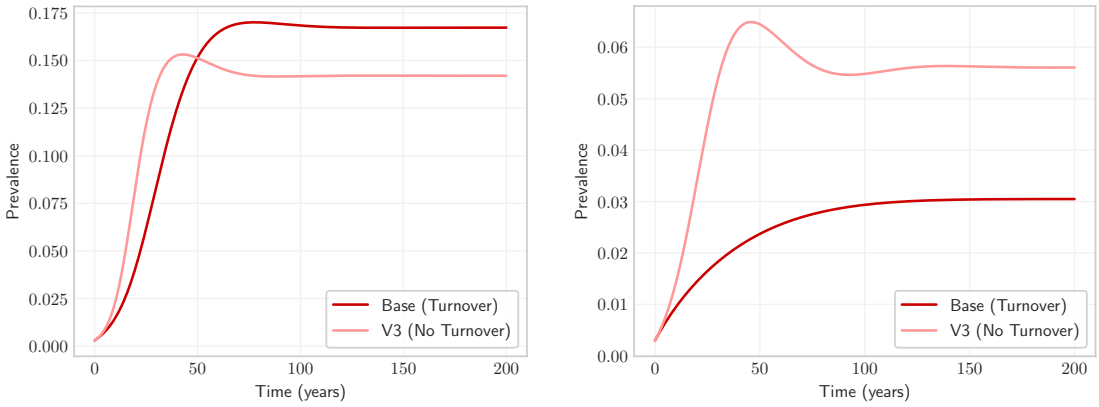
4.1. Model Variants

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 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.



(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.

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.

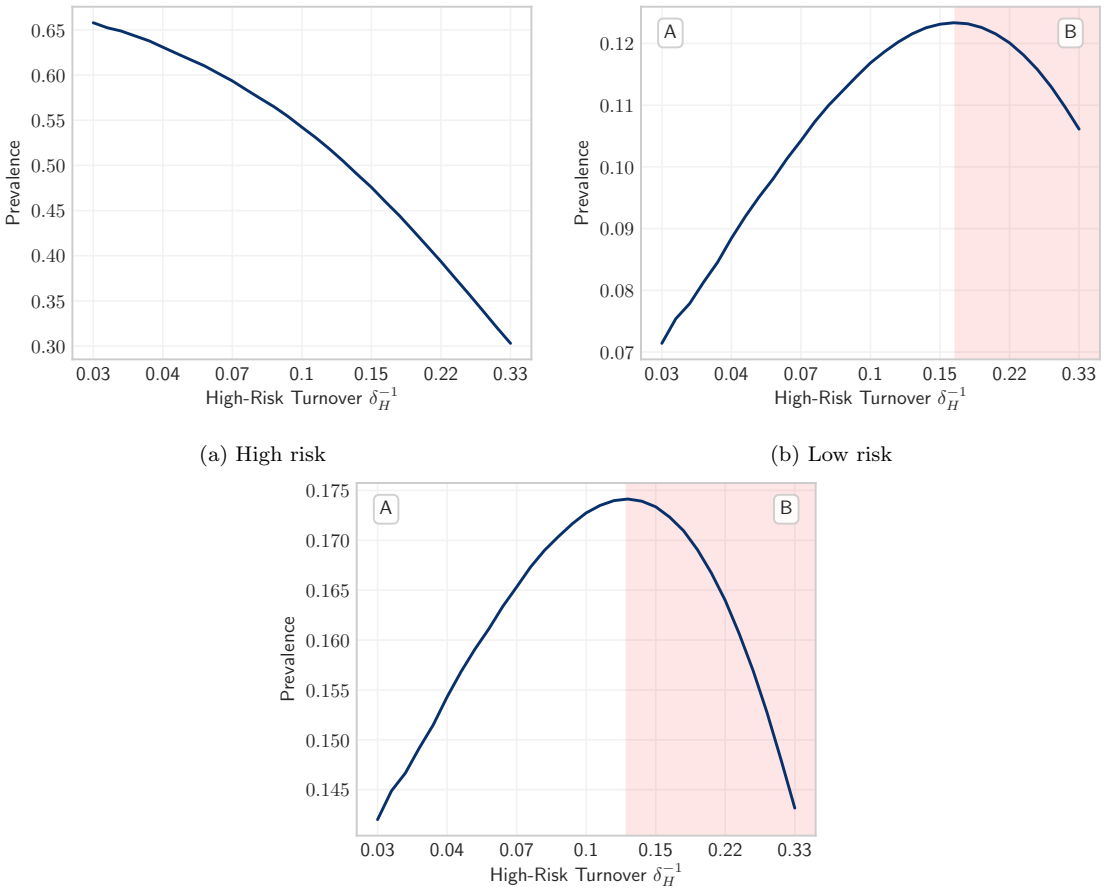
4.2. Influence of Turnover

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).[6] However, at high rates of turnover, turnover 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 (11). As shown in Appendix A.3, the driving component in this expression is the proportion

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

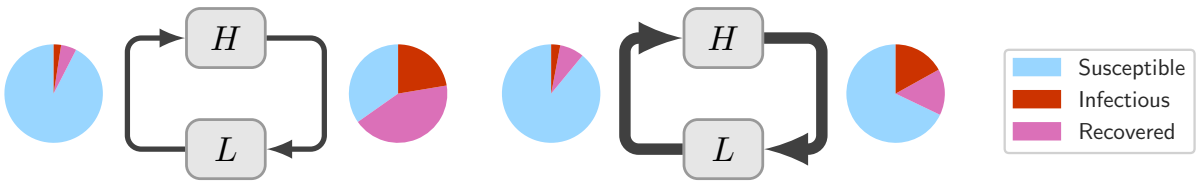


(a) High risk

(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.



(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.

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.

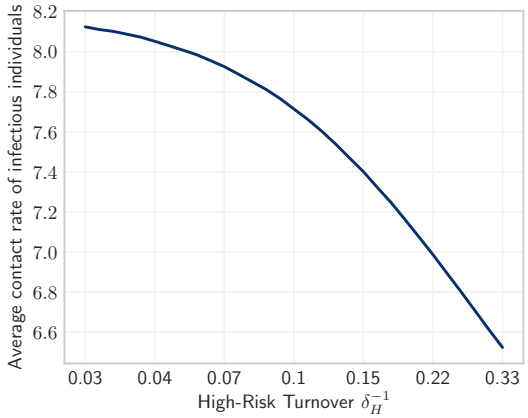
As shown in Figure 8a, turnover decreases $\hat{C}_{\mathcal{I}}$ the average contact rate among infectious individuals. This is because turnover results in a net movement of infected individuals from high to low risk (Figure 7) (Henry and Koopman, 2015). However, at low to moderate rates of turnover,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

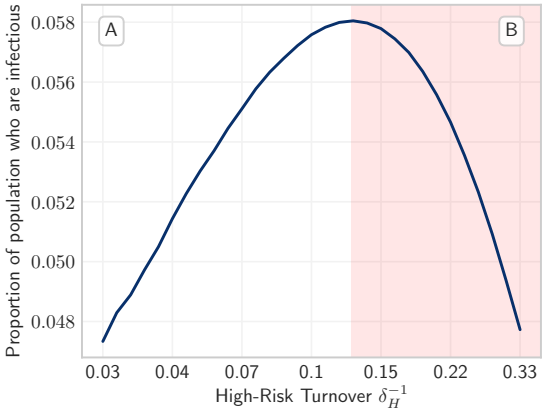
Finally, it is worth noting that the rates of turnover which maximize incidence (Figure 8d) are lower than the rates of turnover which maximize prevalence among the lowest risk group, as well as prevalence overall (Figure 6b and 6c) That is, incidence "peaks" first with respect to turnover. This is because, 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).

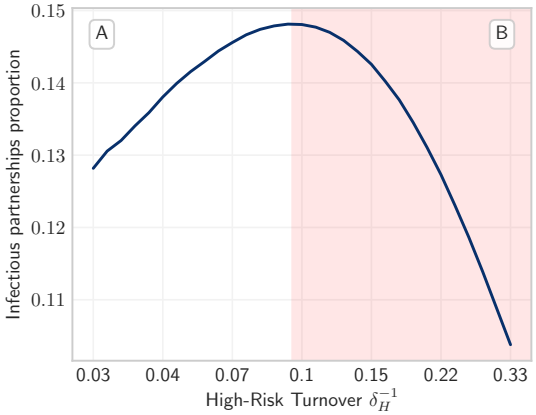
Figure 9d also shows that the rate of turnover which maximizes incidence decreases with increasing treatment. That is, as treatment rates increase, turnover is more likely to decrease incidence than it is to increase incidence (region B grows). This effect can be explained as follows. Recall that the mechanism by



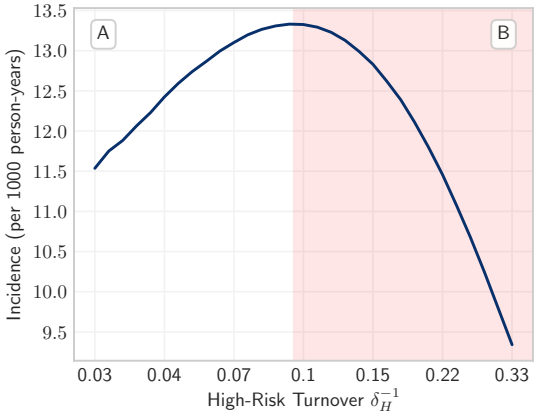
(a) Average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$



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



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

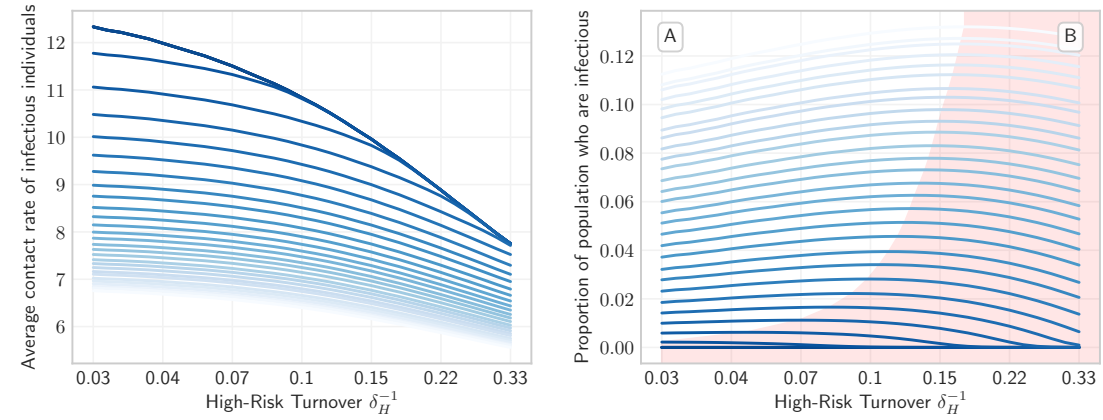


(d) Overall incidence λ

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.

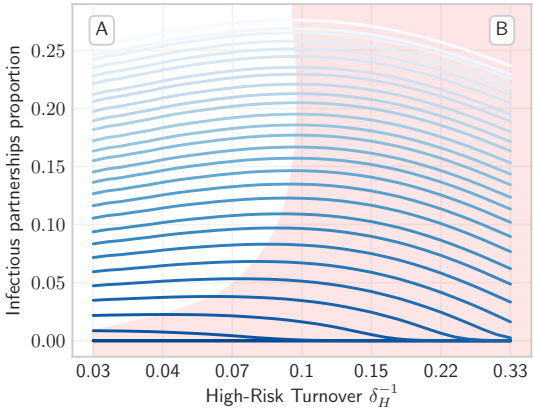
which turnover decreases incidence is through reduction of the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$, due to net movement of infectious individuals from high to low risk (Figure 7). If treatment increases the concentration of infections in the high risk group, then the 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

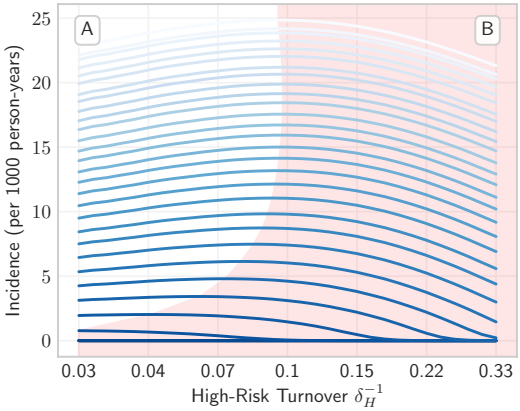


(a) Average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$

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



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

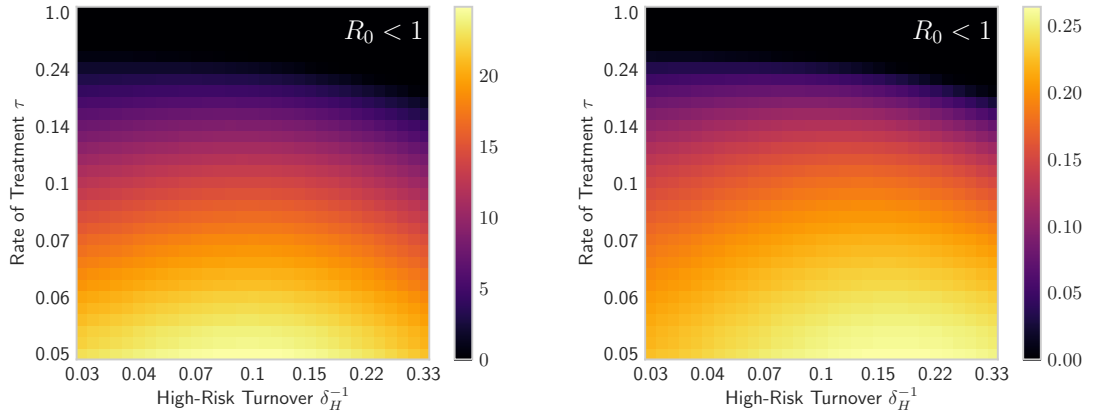


(d) Overall incidence λ

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.

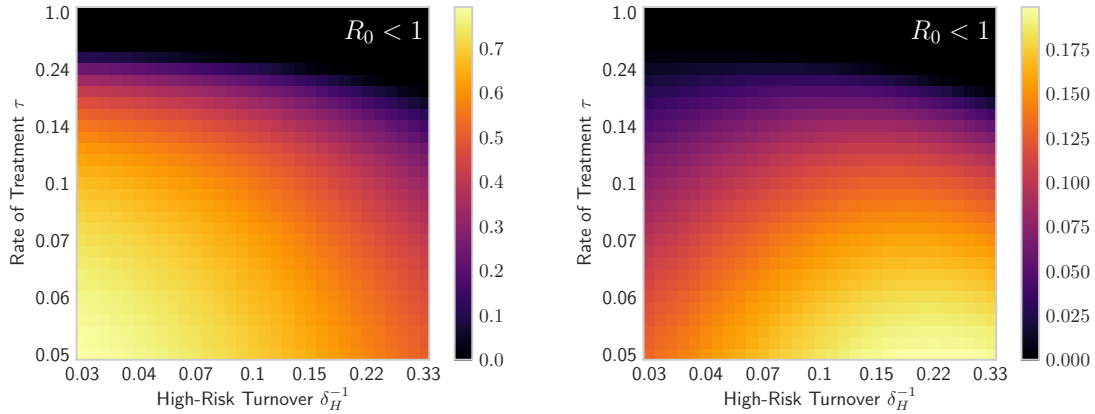
respect to both turnover ϕ and treatment rate τ.[7] Treatment consistently decreases equilibrium incidence (as noted above), as well as prevalence, at all rates of turnover. As suggested in Experiment 2.1, prevalence among the highest risk group (Figure 10c) also decreases with turnover for any rate 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

[7] 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.[8]

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.3. 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

[8] 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.3), 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

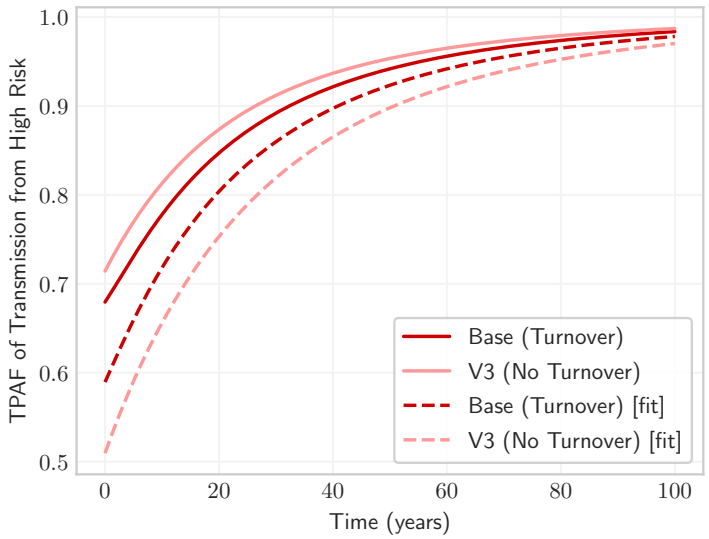


Figure 11: Transmission population attributable fraction (TPAF-from) of the high risk group with and without turnover, and with and without fitted C_i to group-specific prevalence.

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

5. Discussion

6. Conclusion

That's it! We've hopefully illustrated the importance of including population heterogeneity, growth, and turnover in deterministic compartmental epidemic models. Such features can have enormous effects on the model projections, and almost always constitute a more accurate representation of real-world systems. We've also provided tools for implementing and parameterizing these dynamics, considering data which are typically available, and assuming steady-state risk group sizes. So go forth, and build better models.

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

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

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

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

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.

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

Mishra, S., 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.

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

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

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

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

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

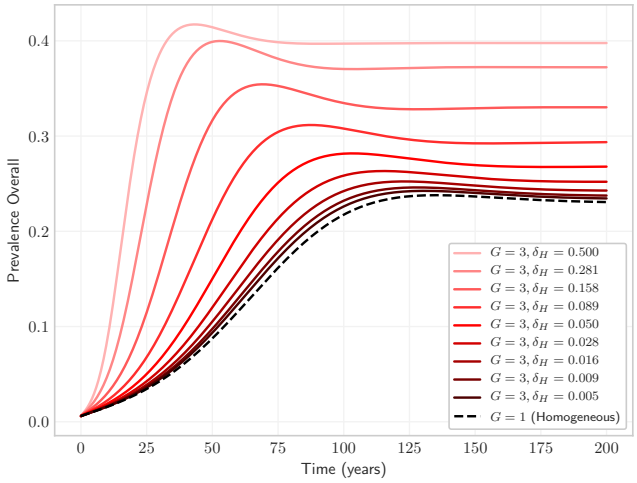
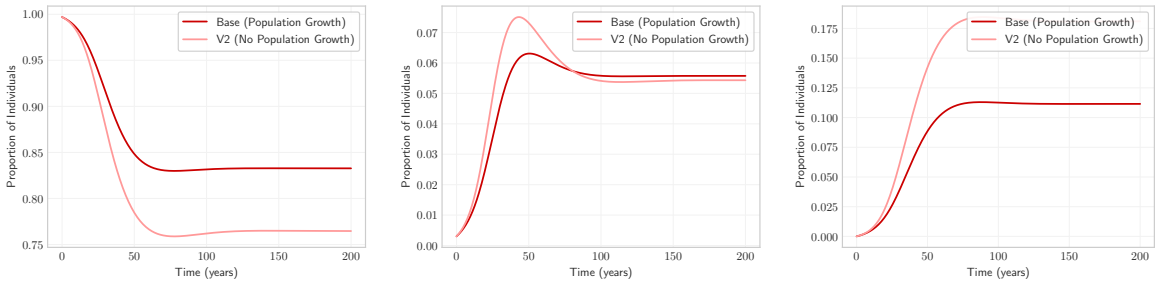


Figure B.1: Overall prevalence predicted by a heterogeneous system for a range of high turnover rates. Note how the heterogeneous model ($G = 3$) converges on a homogeneous model ($G = 1$) with very high turnover rates. Compared to the Base model, transmission probability is increased to $\beta = [TBD]$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

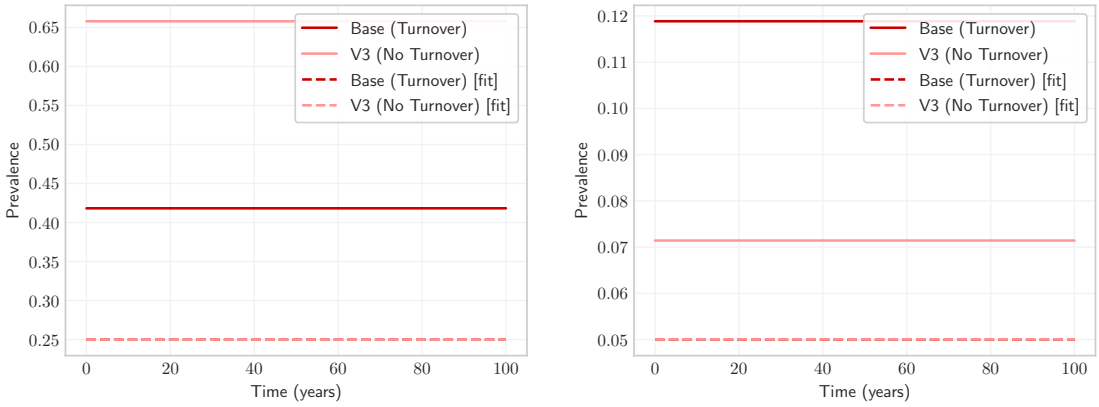
Appendix B.2. Health States With and Without Population Growth



(a) Susceptible (b) Infectious (c) Recovered

Figure B.2: Proportion of health states in the population with and without population growth

Appendix B.3. Equilibrium Prevalence Before and After Model Calibration



(a) High risk

(b) Low risk

Figure B.3: Equilibrium prevalence among high and low risk groups with and without turnover, and with and without fitted C_i to group-specific prevalence.