Risk Group Turnover in Epidemic Models

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

[a] *MAP Centre for Urban Health Solutions, Unity Health Toronto*
[b] *Deptartment of Epidemiology, Johns Hopkins Bloomberg School of Public Health*
[c] *TB HIV Care, South Africa*
[d] *Department of Medicine, Division of Infectious Disease, University of Toronto*
[e] *Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto*
[f] *Instituof Medical Sciences, University of Toronto*

Abstract

Keywords: mathematical modelling, risk heterogeneity, STI, HIV

* Corresponding author (mishras@smh.ca)

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

Contents

1. Introduction

Heterogeneity in transmission risk is a critical feature of epidemics of sexually transmitted infections (STI) (Anderson and May, 1991). This heterogeneity is often demarcated by identifying sub-populations whose risks of acquisition and onward transmission are the highest, such that their unmet STI prevention and treatment needs can sustain local epidemics (Yorke et al., 1978). An important indicator in the appraisal of STI epidemics is the contribution of high-risk groups to the overall epidemic (Case et al., 2012).

Indicators of "contribution" are used to help prioritize or target interventions to reach groups at highest risk (Case et al., 2012; Shubber et al., 2014). Transmission models are increasingly being used to quantify "contribution" by simulating counterfactuals where transmission from and/or to specific subgroups is stopped, and the relative difference in cumulative infections in the total population over various time-periods is measured (Mishra et al., 2016; Mukandavire et al., 2018). This is often referred to as the *transmission population attributable fraction* (TPAF) and counterfactuals have included setting susceptibility to zero and/or setting infectiousness to zero to calculate the TPAF (Mishra et al., 2012). The TPAF is then interpreted as the fraction of all new infections that stem, directly and indirectly, from a failure to prevent acquisition and/or a failure to provide effective treatment in a particular risk group (Mishra et al., 2016).

An epidemiologic phenomenon that is often overlooked in transmission models, but is well-described in the context of sexual behaviour, is the movement between risk groups (Watts et al., 2010). Such movement is often referred in the STI epidemiology literature as "turnover" (Watts et al., 2010). For example, turnover may reflect entry and retirement from formal sex work – a period of time in an individual's sexual life-course that represents a greater risk of STI susceptibility and onward transmission (Watts et al., 2010). Similarly, individuals may have a larger number of sexual partnerships or experience greater STI-associated vulnerabilities during specific times within their sexual life-course (Marston and King, 2006).

Risk group turnover has been shown to influence the predicted equilibrium prevalence of an STI (Stigum et al., 1994; Zhang et al., 2012); the fraction of transmissions occurring during acute HIV infection (Zhang et al., 2012); the basic reproductive number R_0 of HIV (Henry and Koopman, 2015); and the level of universal treatment required to achieve HIV epidemic control (Henry and Koopman, 2015). Yet how, and the extent to which, turnover influences the TPAF remains unknown.

Implementations of risk group turnover in compartmental transmission models also vary widely. In works by Koopman et al. and Stigum et al., rates of movement between two risk groups were balanced analytically based on the size of the groups; Boily et al. (2015) used a 100-year burn-in period to equilibrate a complex system of demographic transitions before introduction of HIV into the modelled system; others restricted the system to unidirectional turnover – e.g. from high to low risk (Eaton and Hallett, 2014).

Challenges in implementing turnover include incorporation of data-driven epidemiologic constraints. For example, data may suggest that the relative sizes of subgroups in the model (such as the population of sex

workers) have remained constant over time. Data may also suggest that heterogeneity in risk behaviour of individuals entering into the model (reflecting sexual debut) may be different from the risk behaviour of individuals already in the model (reflecting sexually active adults). Data may indicate the average duration of a high-risk period of one's sexual life-course (Watts et al., 2010), or how their behaviour changes following that period. Such data should be reflected in implementations of turnover, but it is not always clear how to do so. Moreover, without an exact analysis of the turnover implementation, a burn-in period may be required to equilibrate the system, resulting in potentially unwanted changes to the initial group sizes.

In this paper, we aim to examine the influence of risk group turnover on the TPAF of an illustrative STI without STI-attributable mortality. First, we propose a solution to the challenges outlined above of including epidemiologic data while avoiding the need for a burn-in period. That is, we introduce a unified framework for parameterizing risk group turnover, based on available data. We then examine the mechanisms by which turnover influences group-specific STI incidence and prevalence (Experiment 1). We then examine how inclusion/exclusion of turnover influences the values parameters related to heterogeneity during model fitting (Experiment 2). Finally, we compare the TPAF of the highest-risk groups estimated from two different settings: one with and one without turnover, and from two models that reproduce the same setting: one model with and one model without turnover (Experiment 3).

2. System

In this section we introduce a unified system of equations to describe risk group turnover in deterministic epidemic models with heterogeneity in risk. We then describe how the unified approach can be used in practical terms, based on different assumptions and data available for parameterizing turnover in risk. We then conclude by framing previous approaches to this task using the proposed system.

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 $N = \sum_i x_i$, and the relative population size of each group is denoted as $\hat{x}_i = x_i/N$. Individuals enter the population at a rate ν per year, and exit at a rate μ per year. We assume that individuals entering into the population originate from another exogenous population $\boldsymbol{e} = \{e_1, \ldots, e_G\}$. We make this assumption so that the exogenous population $\boldsymbol{e}$ may have different relative risk group sizes from the system population $\boldsymbol{x}$.[1] That is, $\hat{e}_i$ may not necessarily equal $\hat{x}_i$. However, since the entry rate ν is relative to the size N of population $\boldsymbol{x}$, it will be convenient and otherwise inconsequential to assume that population $\boldsymbol{e}$ also has size N. That

[1] We could equivalently stratify the rate of entry ν by risk group while keeping the exogenous population $\boldsymbol{e}$ equal to the system population $\boldsymbol{x}$. However, we find this formulation more difficult to work with in the subsequent sections.

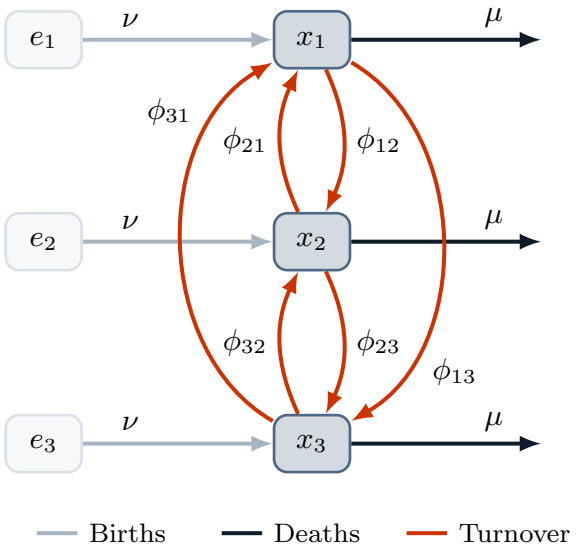


Figure 1: System of risk groups and flows between them for $G = 3$

way, the total number of individuals entering into population $\boldsymbol{x}$ per year is given by νN, and the number of individuals entering into group x_i specifically is given by $\hat{e}_i \nu N$.

Turnover transitions may then 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 the diagonal elements as $*$ since they represent transitions from a group to itself.

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

Risk groups, transitions, 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 a function of the rates of population entry $\nu(t)$ and exit $\mu(t)$, given an initial size N_0. We allow the proportion entering the system to vary by risk group via $\hat{\boldsymbol{e}}$, while the exit rate has the same value for each group. We assume that there is no disease-attributable death. Because the values of ν and μ are the same for each risk group, they can be estimated independent of $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{e}}$, and 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 (Malthus, 1798):

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

Census data, such as (DataBank, 2019), can be used to source the total population size in a given geographic setting over time $N(t)$, thus allowing Eqs. (3) and (4) to be used to estimate $\mathcal{G}(t)$.

If the population size is assumed to be constant, then $\mathcal{G}(t) = 0$ and $\nu(t) = \mu(t)$. If population growth occurs at a stable rate, then $\mathcal{G}$ is fixed at a constant value which can be estimated via Eq. (4) using any two values of $N(t)$, separated by a time interval τ:

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

If the rate of population growth $\mathcal{G}$ varies over time, then Eq. (5) can be reused for consecutive time intervals, and the complete function $\mathcal{G}(t)$ approximated piecewise by constant values. The piecewise approximation can be more feasible than exact solutions using Eq. (3), and can reproduce $N(t)$ accurately for small enough intervals τ, such as one year.

Now, given a value of $\mathcal{G}(t)$, either $\nu(t)$ must be chosen and $\mu(t)$ calculated using Eq. (2), or $\mu(t)$ must be chosen, and $\nu(t)$ calculated. Most modelled systems assume a constant duration of time that individuals spend in the model $\delta(t)$ (Anderson and May, 1991) which is related to the rate of exit μ by:

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

In the context of sexually transmitted infections, the duration of time usually reflects the average sexual life-course of individuals from age 15 to 50 years, such that $\delta = 35$ years. The duration δ may also vary with time to reflect changes in life expectancy. The exit rate $\mu(t)$ can then be defined as $\delta^{-t}(t)$ following Eq. (6), and the entry rate $\nu(t)$ 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 required 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.

The number of risk groups G specifies the number of 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:

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

We then 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. Specifically, each row in A and $\boldsymbol{b}$ defines a constraint: an assumed mathematical relationship involving one or more elements of $\hat{\boldsymbol{e}}$ and ϕ. For example, a simple constraint could be to assume the value $\hat{e}_2 = 0.20$. Each of the following sections introduces a type of constraint, including: assuming a constant group size, specifying elements of $\boldsymbol{\theta}$ directly, assuming an average duration in a group, and assuming relative rates of turnover. Constraints may be selected and combined together based on availability of data and plausibility of assumptions. However, a total of $M = G^2$ constraints must be defined in order to obtain a "unique solution": exactly one value of $\boldsymbol{\theta}$ which satisfies all constraints. The values of $\hat{\boldsymbol{e}}$ and ϕ can then be calculated algebraically by solving Eq. (8) with $\boldsymbol{\theta} = A^{-1}\boldsymbol{b}$, for which many algorithms exist (LAPACK, 1992).

Constraint Type 1: Constant group size. One epidemiologic feature that epidemic models consider is whether or not the relative sizes of risk groups are constant over time (Henry and Koopman, 2015; Boily et al., 2015). Assuming constant group size implies a stable level of heterogeneity over time. To enforce this assumption, 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:

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

Eq. (9) is written in terms of absolute population sizes $\boldsymbol{x}$ and $\boldsymbol{e}$, but can be written as proportions $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{e}}$ when dividing by N. Working with proportions is useful, as N need not be known. If we assume that the proportion 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 rate of population growth of the risk group, $\mathcal{G}x_i$. Substituting $\frac{d}{dt}x_i = \mathcal{G}x_i$ 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 ensure risk groups do not change size over time. However, a unique solution requires an additional $G(G-1)$ constraints. For $G = 3$, this corresponds to 6 additional constraints.

Constraint Type 2: Specified elements. The simplest type of additional constraint is to directly specify the values of individual elements in $\hat{\boldsymbol{e}}$ or ϕ. Such constraints may be appended to $\boldsymbol{b}$ and A as an additional row k using indicator notation.[2] That is, with b_k as the specified value v, and A_k as the indicator vector, with 1 in the same position as the desired element in $\boldsymbol{\theta}$:

$$b_k = v; \qquad A_k = [0, \ldots, 1, \ldots, 0] \tag{12}$$

For example, for $G = 3$, if it is known that 20% of individuals enter directly into risk group x_2 upon entry into the model ($\hat{e}_2 = 0.20$), then $\boldsymbol{b}$ and A can be augmented with:

$$b_k = \begin{bmatrix} 0.20 \end{bmatrix}; \qquad A_k = \begin{bmatrix} \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{13}$$

since $\hat{e}_2$ is the second element in $\boldsymbol{\theta}$. If the data suggest zero turnover from group i to group j, then Eq. (13) 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 $G - 1$ constraints, as the last element will be either redundant or violate the sum-to-one rule. As shown in Appendix A.3, the sum-to-one rule is actually implicit in Eq. (11), so it is not necessary to supply a constraint like $1 = \sum_i \hat{e}_i$.

Constraint Type 3: Group duration. Type 1 Constraints assume that the relative population size of each group remains constant. Another epidemiologic feature that epidemic models considered is whether or not the duration of time spent within a given risk group remains constant. For example, in STI transmission models that include formal sex work, it can be assumed that the duration in formal sex work work remains

[2] Indicator notation, also known as "one-hot notation" is used to select one element from another vector, based on its position. An indicator vector is 1 in the same location as the element of interest, and 0 everywhere else.

stable over time, such as in (Mishra et al., 2014; Boily et al., 2015). The duration δ_i is defined as the inverse of all rates of exit from the group:

$$\delta_i = \left(\mu + \sum_j \phi_{ij}\right)^{-1} \tag{14}$$

Estimates of the duration in a given group can be sourced from cross-sectional survey data where participants are asked about how long they have engaged in a particular practice – such as sex in exchange for money (Watts et al., 2010). Data on duration may also be sourced from longitudinal data, where repeated measures of self-reported sexual behaviour, or proxy measures of sexual risk data, are collected (USAID, 2019; ICAP, 2019). Data on duration in each risk group can then be used to define ϕ by rearranging Eq. (14) 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 another row k:

$$b_k = \begin{bmatrix} 5^{-1} - \mu \end{bmatrix}; \qquad A_k = \begin{bmatrix} \cdot & \cdot & \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{15}$$

Note that, similar to specifying all elements of $\hat{\boldsymbol{e}}$, specifying δ_i may result in conflicts or redundancies with other constraints. A conflict means it will not be possible to resolve values of ϕ which simultaneously satisfy all constraints, while a redundancy means that adding one constraint does not help resolve a unique set of values $\boldsymbol{\theta}$. For example, for $G = 3$, if Type 2 Constraints are used to specify $\phi_{12} = 0.1$ and $\phi_{13} = 0.1$, and $\mu = 0.05$, then by Eq. (14), we must have $\delta_1 = 4$. Specifying any other value for δ_1 will result in a conflict, while specifying $\delta_1 = 4$ is redundant, since it is already implied. There are innumerable situations in which this may occur, so we do not attempt to describe them all. Section 2.2.2 describes how to identify conflicts and redundancies when they are not obvious.

Constraint Type 4: Relative rates of turnover. In many cases, it may be difficult to obtain estimates of a given turnover rate ϕ_{ij} for use in Type 2 Constraints. However, it may be possible to estimate relative relationships between rates of turnover, such as:

$$r\,\phi_{ij} = \phi_{i'j'} \tag{16}$$

where r is a ratio relating the values of ϕ_{ij} and $\phi_{i'j'}$. For example, for $G = 3$, let T_1 be the total number of individuals entering group x_1 due to turnover. If we know that 70% of T_1 originates from group x_2, while 30% of T_1 originates from group x_3, then $0.7\,T_1 = \phi_{23}\,x_2$ and $0.3\,T_1 = \phi_{13}\,x_1$, and thus: $\phi_{23}\left(\frac{0.3\,x_2}{0.7\,x_1}\right) = \phi_{13}$. This constraint can then be appended as another row k in $\boldsymbol{b}$ and A like:

$$b_k = \begin{bmatrix} 0 \end{bmatrix}; \qquad A_k = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \left(\frac{0.3\,x_2}{0.7\,x_1}\right) & \cdot & 1 & \cdot & \cdot \end{bmatrix} \tag{17}$$

The example in Eq. (17) is based on what proportions of individuals entering a risk group j came from which former risk group i, but similar constraints may be defined based on what proportions of individuals exiting a risk group i enter into which new risk group j. It can also be assumed that the absolute number

Table 1: Summary of constraint types for defining risk group turnover

	Name	Eq.	E.g.	Data requirements
1.	Constant group size	(10)	(11)	all values of $\hat{x}_i$ and ν
2.	Specified elements	(12)	(13)	any value of $\hat{e}_i$ or ϕ_{ij}
3.	Group duration	(14)	(15)	any value of δ_i
4.	Relative rates of turnover	(16)	(17)	any relationship between two turnover rates ϕ_{ij} and $\phi_{i'j'}$

ν: rate of population entry; ϕ_{ij}: rate of turnover from group i to group j; $\hat{x}_i$: proportion of individuals in risk group i; $\hat{e}_i$: proportion of individuals entering into risk group i; δ_i: average duration spent in risk group i.

of individuals moving between two risk groups is equal, in which case the relationship is: $\phi_{ij}\left(\frac{x_i}{x_j}\right) = \phi_{ji}$. All constraints of type 4 will have $b_k = 0$.

Solving the System. Table 1 summarizes the four types of constraints described above. Given a set of sufficient constraints on $\boldsymbol{\theta}$ to ensure exactly one solution, the system of equations Eq. (8) can be solved using $\boldsymbol{\theta} = A^{-1}\boldsymbol{b}$. The resulting values of $\hat{\boldsymbol{e}}$ and ϕ can then be used in the epidemic model.

However, we may find that we have an insufficient number of constraints, implying that there are multiple values of the vector $\boldsymbol{\theta}$ which satisfy the constraints. An insufficient number of constraints may be identified by a "rank deficiency" warning in numerical solvers of Eq. (8) (LAPACK, 1992). Even if A has G^2 rows, the system may have an insufficient number of constraints because some constraints are redundant. In this situation, we can pose the problem as a minimization problem, namely:

$$\boldsymbol{\theta}^* = \arg\min f(\boldsymbol{\theta}), \quad \text{subject to: } \boldsymbol{b} = A\,\boldsymbol{\theta};\ \ \boldsymbol{\theta} \geq 0 \tag{18}$$

where f is a function which penalizes certain values of $\boldsymbol{\theta}$. For example, $f = ||\cdot||_2$ penalizes large values in $\boldsymbol{\theta}$, so that the smallest values of $\hat{\boldsymbol{e}}$ and ϕ which satisfy the constraints will be resolved.[3]

Similarly, we may find that no solution exists for the given constraints, since two or more constraints are in conflict. Conflicting constraints may be identified by a non-zero error in the solution to Eq. (8) (LAPACK, 1992). In this case, the conflict should be resolved by changing or removing one of the conflicting constraints.

2.3. Previous Approaches

Few epidemic models of sexually transmitted infections with heterogeneity in risk have simulated turnover among risk groups, and those models which have simulated turnover have done so in various ways. In this

[3] Numerical solutions to such problems are widely available, such as the Non-Negative Lease Squares solver (Lawson and Hanson, 1995), available in Python: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.nnls.html`.

section, we review four prior implementations of turnover and each study's objectives for the implementation – e.g. constant relative group sizes over time. We then highlight how the approach proposed in Section 2.2 could be used to achieve the same objectives.

Stigum et al. (1994) simulated turnover among $G = 2$ risk groups in a population with no exogenous entry or exit ($\nu = \mu = 0$ and hence $\hat{\boldsymbol{e}}$ is not applicable). Turnover between the groups was balanced in order to maintain constant risk group sizes (Type 1 Constraint),[4] while the rate of turnover from high to low was specified as κ (Type 2 Constraint). Thus, the turnover system used by Stigum et al. (1994) can be written in the proposed framework as:

$$\begin{bmatrix} 0 \\ \kappa \end{bmatrix} = \begin{bmatrix} \hat{x}_1 & -\hat{x}_2 \\ 1 & \cdot \end{bmatrix} \begin{bmatrix} \phi_{12} \\ \phi_{21} \end{bmatrix}, \qquad \hat{e}_1 = \hat{e}_2 = 0 \tag{19}$$

Henry and Koopman (2015) also simulated turnover among $G = 2$ risk groups, but considered exogenous entry and exit, both at a rate μ. The authors used the notation f_i for our $\hat{x}_i$, and assumed that the exogenous population had the same distribution of risk groups as the system population: $\hat{e}_i = f_i$ (Type 2 Constraint). The authors further maintained constant risk group sizes (Type 1 Constraint) by analytically balancing turnover between the two groups using: $\phi_{12} = \omega\hat{x}_2$; $\phi_{21} = \omega\hat{x}_1$. However, it can be shown that this analytical approach is also the solution to the following combination of Type 1 and 2 Constraints:

$$\begin{bmatrix} 0 \\ \omega f_2 \end{bmatrix} = \begin{bmatrix} f_1 & -f_2 \\ 1 & \cdot \end{bmatrix} \begin{bmatrix} \phi_{12} \\ \phi_{21} \end{bmatrix}, \qquad \hat{e}_i = f_i \tag{20}$$

Eaton and Hallett (2014) simulated turnover among $G = 3$ risk groups, considering exogenous entry from a population with a unique distribution of risk groups $\hat{\boldsymbol{e}}$. Turnover was considered from high-to-medium, high-to-low, and medium-to-low risk, all with an equal rate ψ; the reverse transition rates were set to zero (six total Type 2 Constraints). In the absence of turnover in the other direction, risk group sizes were maintained using the values of $\hat{e}_i$, computed using Type 1 Constraints as follows:

$$\begin{bmatrix} \nu x_1 + 2x_1\psi \\ \nu x_2 - x_1\psi + x_2\psi \\ \nu x_3 - x_1\psi - x_2\psi \end{bmatrix} = \begin{bmatrix} \nu & \cdot & \cdot \\ \cdot & \nu & \cdot \\ \cdot & \cdot & \nu \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}, \qquad \begin{aligned} \phi_{12} = \phi_{13} = \phi_{23} = \psi \\ \phi_{21} = \phi_{31} = \phi_{32} = 0 \end{aligned} \tag{21}$$

Boily et al. (2015) simulated turnover among $G = 7$ risk groups, – TODO.

In sum, the framework for modelling turnover presented in this section aims to generalize all previous implementations. In so doing, we hope to clarify the requisite assumptions, dependencies on epidemiologic data, and relationships between previous approaches. In the following Experiments, we leverage this framework to explore the potential influences of turnover on simulated epidemics.

[4] Due to its simplicity, this constraint is actually an example of both Type 1 and Type 4 Constraints.

Table 2: Summary of experiments

Experiment	Outputs	Stratifications	Turnover	Treatment	Fitting
1.1.	prevalence	overall	any vs none	fixed, 2 rates	none
1.2.	prevalence, incidence	overall, by group	varied	fixed, 1 rate	none
1.3.	prevalence, incidence	overall, by group	varied	varied	none
2.	inferred contact rate	by group	any vs none	fixed, 1 rate	prevalence by group
3.1.	TPAF	highest risk	any vs none	fixed, 1 rate	none
3.2.	TPAF	highest risk	any vs none	fixed, 1 rate	prevalence by group

TPAF: transmission population attributable fraction

3. Experiment

We aimed to determine and understand the influence of risk group turnover on the contribution of the highest risk group to the overall epidemic, as measured by the transmission population attributable fraction (TPAF). However, to understand the underlying mechanism, we first examined the influence of turnover on the equilibrium incidence and prevalence projected among risk groups, as well as overall. Since the influence may depend on the duration of infectiousness, we also explored the sensitivity of these results to different treatment rates. Next, we examined how turnover may influence inferred model parameters through model fitting. Finally, we compared the TPAF of the highest risk group with and without turnover, before and after model fitting. Our experiments can be summarized as in Table 2.

3.1. Model & Simulations

To run our experiments, we developed a deterministic single-sex SIT 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 treated $\mathcal{T}$ (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 year so that individuals in risk group i are assumed to form contacts at a rate C_i per year. The probability of contact formation ρ_{ik} between individuals in group i and individuals in risk group k is assumed to be proportionate to the total number of available contacts within each group:

$$\rho_{ik} = \frac{C_k x_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} x_{\mathrm{k}}} \tag{22}$$

Figure 2: Modelled health states. $\mathcal{S}$: susceptible; $\mathcal{I}$: infected; $\mathcal{T}$: treated; λ: force of infection; τ: treatment.

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:

$$\lambda_i = C_i \sum_k \rho_{ik} \, \beta \, \frac{\mathcal{I}_k}{x_k} \tag{23}$$

Individuals are assumed to transition from the infectious $\mathcal{I}$ to treated $\mathcal{T}$ health-state at a rate τ per year, reflecting diagnosis and treatment. The treatment rate is not stratified by risk group. Individuals in the treated $\mathcal{T}$ health-state are not infectious nor susceptible, and 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 $\hat{\boldsymbol{e}}$ were computed using the methods outlined in Section 2.2.2 and 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 $\hat{\boldsymbol{e}}$ is computed.

Using the above three assumptions, we need to specify the values of $\hat{\boldsymbol{x}}$, $\boldsymbol{\delta}$, ν, and μ. Such parameters could be derived from data as described in Section 2.2.2; however, in this experiment, we use the illustrative values summarized in Table 3. After resolving the system of equations, $\hat{\boldsymbol{e}}$ is equal to $\hat{\boldsymbol{x}}$ (assumed), and ϕ is:

$$\phi = \begin{bmatrix} * & 0.0833 & 0.0867 \\ 0.0208 & * & 0.0158 \\ 0.0058 & 0.0042 & * \end{bmatrix} \tag{24}$$

We then simulated epidemics 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 treated 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[5] using Euler's method with a time step of $dt = 0.1$ years. The

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

Table 3: Model parameters. All rates have units year^{-1}; durations are in years; parameters stratified by risk group are written [high, medium, low] risk.

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 by risk group	[0.05 0.20 0.75]
$\hat{\boldsymbol{e}}$	proportion of entering individuals risk by risk group	[0.05 0.20 0.75]
$\boldsymbol{\delta}$	average duration spent in each risk group	[5 15 25]
C	rate of contact formation among individuals in each risk grouop	[25 5 1]
ν	rate of population entry	0.05
μ	rate of population exit	0.03

full system of model equations is given in Appendix A.1. All comparative analyses are then conducted at equilibrium, defined as a steady state with <1% difference in incidence per year.

3.2. Experiment 1: Influence of turnover on equilibrium incidence and prevalence

Experiment 1 examined the influence of turnover on equilibrium incidence and prevalence, where incidence is defined as λ_i from Eq. (23), and prevalence is defined as $\hat{\mathcal{I}}_i = \dfrac{\mathcal{I}_i}{\mathcal{X}_i}$.

Experiment 1.1: Overall prevalence with vs without turnover. First, we compare the overall prevalence predicted by the model with and without turnover. The model with turnover is as described above. The model without turnover has all rates $\phi = 0$. Following Eq. (14), this means that in the model without turnover, the time spent within each risk group is equal to the average duration in the modelled population μ^{-1}. The comparison of prevalence is repeated for two different treatment rates, in order to illustrate variability in the relative difference.

Experiment 1.2: Influence of turnover rates on incidence and prevalence. Second, we determined the influence of different turnover rates on equilibrium incidence and prevalence 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 groups $\hat{\boldsymbol{x}}$. Thus, we controlled the rates of turnover using the duration of individuals in the high risk group δ_H, such that a shorter period of time in the high risk group corresponded to higher rates of turnover among all groups. 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 $\kappa = 0.3$. 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. (14). In this way, each value of δ_H was used to define a unique 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 was then varied from 33 to 3 years to examine the influence of different turnover rates.

Experiment 1.3: Influence of turnover rates at various treatment rates. Next, we conducted a 2-way sensitivity analysis 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 33 to 3 years as in Experiment 1.2. We examined 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.

3.3. Experiment 2: Inferred risk heterogeneity with vs without turnover

In Experiment 2, we examined the influence of turnover on the parameter values inferred via model fitting. Specifically, we fit the model to 20% infection prevalence among the high risk group, 3% among the low risk group, and 5% overall, with and without turnover . We fit the contact rates C of all risk groups by minimizing the negative log-likelihood of each predicted prevalence versus the target.[6] We then compared the inferred contact rates C in the model with versus without turnover. 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 produces the given infection prevalence.

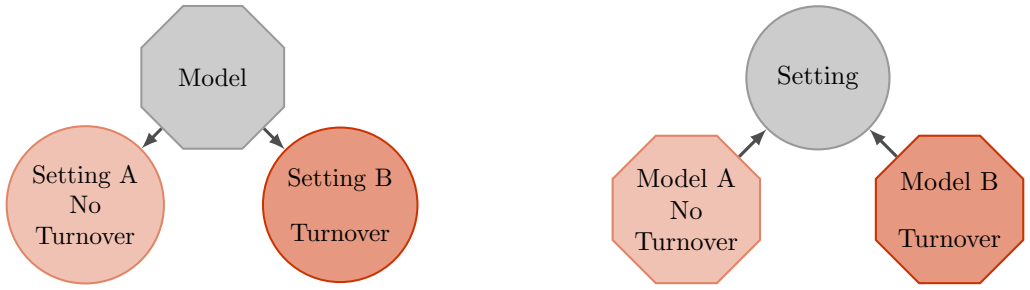
3.4. Experiment 3: Influence of turnover on the TPAF of the high risk group

Finally, Experiment 3 sought to examine how the estimated contribution of highest risk group to overall transmission, as measured by the transmission population attributable fraction (TPAF), varies with versus without turnover. The TPAF of a risk group i is defined as:

$$\text{TPAF}_i(t) = \frac{I_0(t) - I_i(t)}{I_0(t)} \tag{25}$$

where $I_0(t)$ is the cumulative number of new infections by time t under usual conditions, and $I_i(t)$ is the cumulative number of new infections assuming no transmission from risk group i. Both $I_0(t)$ and $I_i(t)$ are

[6] Sample sizes of 50, 750, and 1000 were assumed to generate binomial distributions for the high, low, and overall prevalence targets respectively, and the minimization was performed using the SLSQP method (Kraft, 1988) from the SciPy Python package (`scipy.optimize.minimize`).



(a) Experiment 3.1: same parameters, different prevalence (no model fitting)

(b) Experiment 3.2: same prevalence, different parameters (fitted model)

Figure 3: Two approaches to comparing TPAF with and without turnover

calculated starting from a system at equilibrium. There are two ways to consider the comparison of TPAF with versus without turnover; these are illustrated in Figure 3, and explained in the following sections.

Experiment 3.1: Two settings. First, we compared two hypothetical settings, which had identical populations (including behaviour), except that setting A had no turnover: $\phi = 0$, while setting B had turnover: ϕ from Eq. (24). As a result, model parameters were identical (except turnover), but the infection prevalence predicted for each setting was different. Following equilibration of the model in both settings, the TPAF of the high risk group was then estimated.

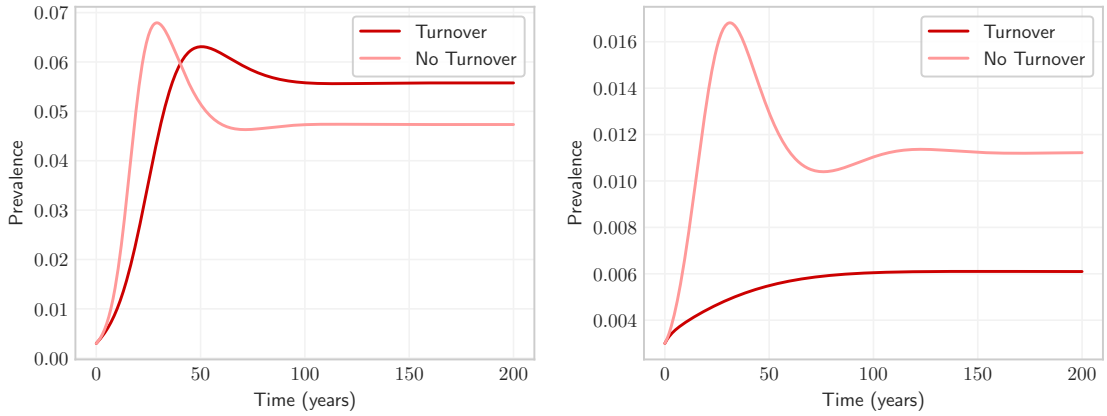
Experiment 3.2: Two models. Second, we compared two models, which were identical in structure except that Model A had no turnover and Model B had turnover. In this case, both models were fitted to the same "setting", as defined by overall and group-specific equilibrium infection prevalence (from Experiment 2). As a result, prevalence was the same in both models, but the group-specific contact rates inferred via fitting were different. As before, the TPAF of the high risk group was then estimated in each model after equilibration.

4. Results

4.1. Experiment 1: Influence of turnover on equilibrium incidence and prevalence

First, we present general trends in equilibrium prevalence and incidence with respect to turnover.

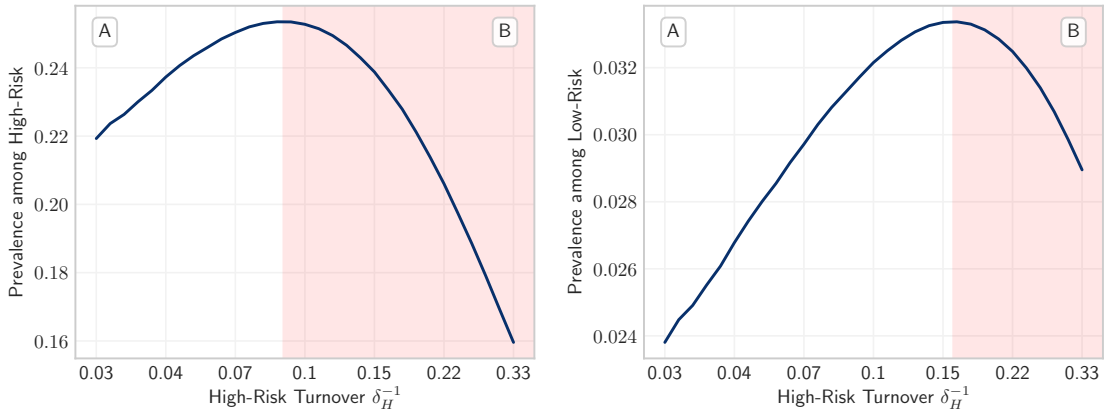
Experiment 1.1: Overall prevalence with vs without turnover. Figure 4 shows the influence of turnover on overall prevalence at two different treatment rates. In both scenarios, turnover appeared to slow the initial epidemic growth as indicated by overall prevalence. However, at equilibrium, the influence of turnover on overall prevalence depended on the treatment rate: prevalence was higher with turnover for $\tau = 0.1$ (Figure 4a), while it was higher without turnover for $\tau = 0.2$ (Figure 4b). Experiments 1.2 and 1.3 aimed 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.



(a) Treatment rate $\tau = 0.1$

(b) Treatment rate $\tau = 0.2$

Figure 4: Overall projected prevalence with and without risk group turnover under two different treatment rates τ.



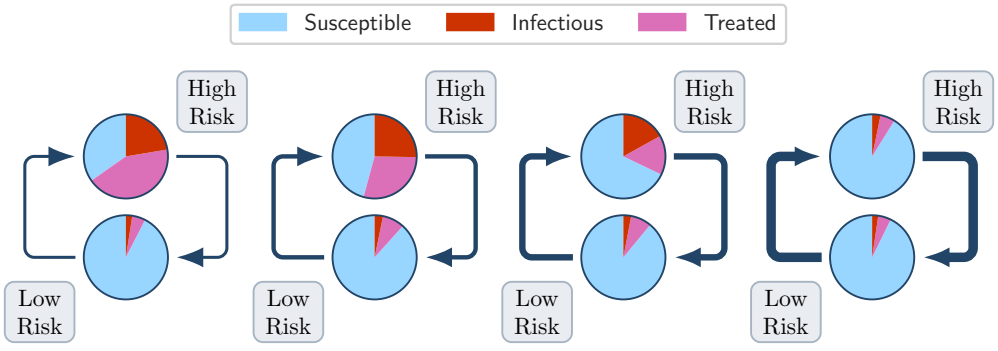
(a) High risk

(b) Low risk

Figure 5: Equilibrium prevalence among high and low risk groups versus turnover, as controlled by the duration in the high risk group δ_H. Turnover shown in log scale.

Experiment 1.2: Influence of turnover rates on incidence and prevalence. Figure 5 illustrates trends in equilibrium prevalence versus turnover among the high and low risk groups for fixed treatment rate ($\tau = 0.1$). For both groups, the same profile was observed: at low turnover, increasing turnover increased prevalence, up to a maximum value (region A), and increasing turnover beyond this point (region B) then decreased prevalence. In the high risk group (Figure 5a), this transition occurred at a lower rate of turnover, while in the low risk group (Figure 5b), the transition occurred at a higher rate of turnover. This peaked prevalence profile can be explained by the interaction between two factors: the movement of individuals between risk groups and incidence.

The movement of individuals between risk groups is depicted in Figure 6 for four rates of turnover and a



(a) Low turnover (b) Moderate turnover (c) High turnover (d) Very high turnover

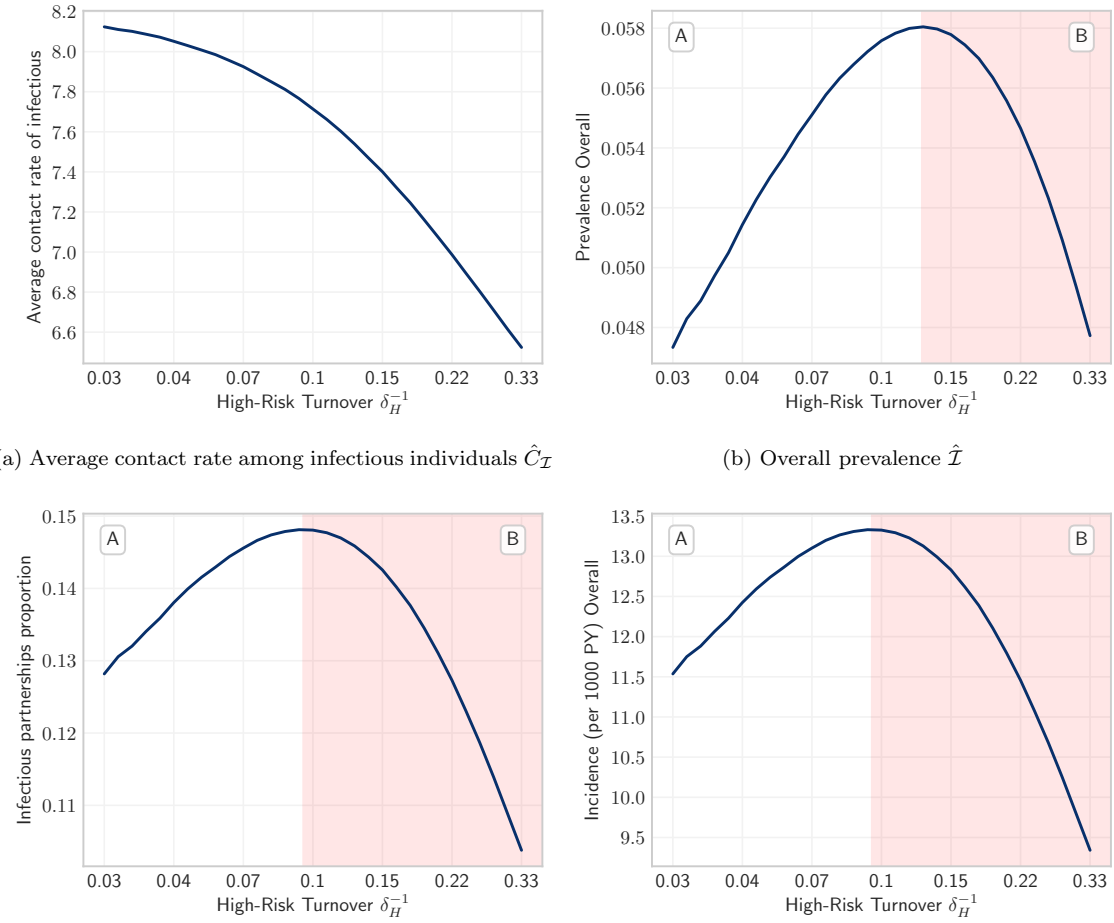
Figure 6: Average health states of individuals moving between high and low risk groups due to different rates of turnover.

simplified system (two risk groups). Recall that, by our assumptions, the distribution of health states among individuals leaving a risk group was equal to the distribution of health states within the group. Therefore, a higher proportion of individuals leaving the high risk group were infectious as compared to individuals entering the high risk group, who were mostly susceptible. Thus, in the high risk group, turnover yielded a net replacement of infectious individuals with susceptible individuals. However, turnover similarly replaced treated individuals with susceptible individuals in this group. If incidence is sufficiently high, infection of these susceptible individuals can outpace the loss of infectious individuals via turnover. As a result, prevalence among the high risk group can actually increase with increasing turnover. In fact, this is what we observed in this model for moderate rates of turnover (Figure B.2b), yielding the increase in prevalence among high risk individuals shown in region A of Figure 5a.[7] Among the low risk group, turnover yielded a net exchange susceptible individuals for infectious and treated individuals. As a result, moderate rates of turnover also increased prevalence among the low risk group, as shown in region A of Figure 5b.

Next, and in order to explain the reversal of these trends at higher rates of turnover (region B), we examined the second factor in the influence of turnover on infection prevalence: incidence. Consider the force of infection equation, Eq (23). As shown in Appendix A.4, the dynamic 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 sub-factors: $\hat{C}_{\mathcal{I}}$ the average contact rate among infectious individuals, and $\hat{\mathcal{I}}$ overall prevalence. Thus the influence of turnover on incidence can be understood through the influence of turnover on these two sub-factors.[8]

[7] If prevalence is defined as to include both infectious individuals and individuals on treatment, such as in the context of HIV, then infection prevalence among the high risk group monotonically decreased with increasing turnover (Knight et al., 2019).

[8] Since we assumed proportionate mixing among risk groups, and only consider heterogeneity in contact rate C_i, incidence in each risk group is the same as overall except for a scale factor proportional to C_i.



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

(b) Overall prevalence $\hat{\mathcal{I}}$

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

(d) Overall incidence λ

Figure 7: 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.

As shown in Figure 7a, turnover decreased the first sub-factor: $\hat{C}_{\mathcal{I}}$ the average contact rate among infectious individuals. This is because, as noted above, turnover causes in a net movement of infectious individuals from high to low risk. However, at low to moderate rates of turnover, turnover increased the second sub-factor: $\hat{\mathcal{I}}$ overall prevalence (Figure 7b, region A). Now, under the conditions in region A, overall prevalence $\hat{\mathcal{I}}$ increased faster with turnover than the average contact rate of infectious people $\hat{C}_{\mathcal{I}}$ decreased. Thus, as a product of these two sub-factors, $\boldsymbol{C}_{\mathcal{I}}$ increased with turnover in region A (Figure 7c) and incidence increased proportionally (Figure 7d).

It therefore follows that the peak in incidence, and the transition between regions A and B in Figure 7c occurred when the dominating sub-factor between $\hat{\mathcal{I}}$ and $\hat{C}_{\mathcal{I}}$ reversed. That is, the transition occurred when the average contact rate of infectious people $\hat{C}_{\mathcal{I}}$ decreased with turnover faster than prevalence $\hat{\mathcal{I}}$

increased with turnover. Then, as rates of turnover continued to increase, declining incidence also reversed the upward trend in prevalence, and incidence and prevalence decreased across all groups in a snowball effect. This mechanism then explains the observations shown in region B throughout.

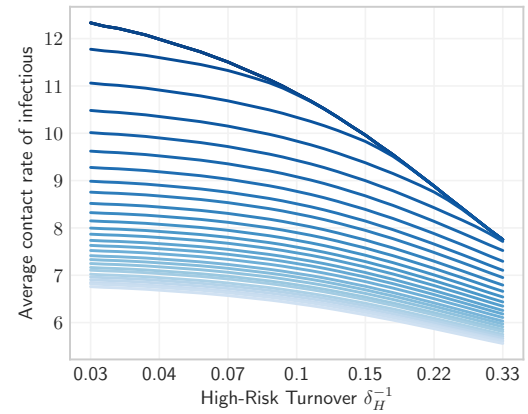
Finally, we note that the rates of turnover which maximized incidence (Figure 7d) were lower than those which maximized overall prevalence (Figure 7b). This is because incidence represents the number of new infections per susceptible, not per individual; so a system at slightly higher prevalence can be maintained by a slightly lower incidence, provided the proportion of susceptibles has decreased more than prevalence has increased. We also note that infection prevalence among the high risk group peaked at lower rates of turnover than prevalence among the low risk group. This is because turnover yields a net movement of infectious individuals from high to low risk; so prevalence among the high risk group can decrease with turnover even as incidence increases, whereas prevalence among the low risk group can increase with turnover even as incidence decreases.

Experiment 1.3: Influence of turnover rates at various treatment rates. So far, the influence of turnover on equilibrium incidence and prevalence was explored for a single treatment rate. Experiment 1.3 explored additional treatment rates τ.

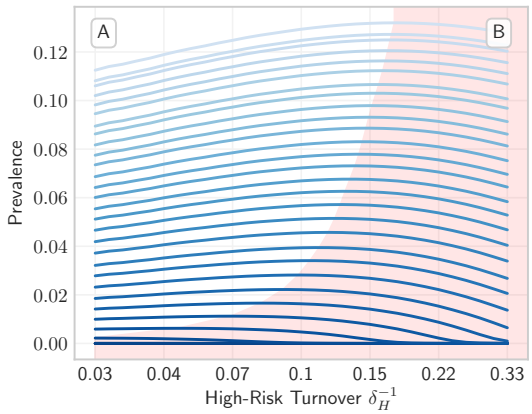
First, the sub-factors of incidence are again shown in Figure 8. Increasing the treatment rate τ increased the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$ (Figure 8a). This is because increasing treatment concentrated infections in the highest risk group (Figure B.7), so that $\hat{C}_{\mathcal{I}}$, on average, increased. However, increasing treatment reduced overall prevalence $\hat{\mathcal{I}}$ (Figure 8b), due to the herd immunity effect of treated individuals. The dominant effect was that of $\hat{\mathcal{I}}$, which tended towards zero faster than $\hat{C}_{\mathcal{I}}$ tended towards infinity, and so incidence declined with treatment (Figure 8d).

Figure 8d also shows that the rate of turnover which maximized incidence decreased with treatment. That is, as treatment rates increased, turnover was more likely to decrease incidence than it was to increase incidence (region B grew). This effect can be explained as follows. Recall that the mechanism by which turnover decreased incidence was 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 6). If treatment increased the concentration of infections in the high risk group, then the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$ would be more sensitive to redistribution of those infectious individuals to lower risk groups via turnover. In Figure 8a, 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 was more likely to decrease incidence because movement of infectious individuals from high to low risk had a larger impact on the average contact rate of infectious individuals.

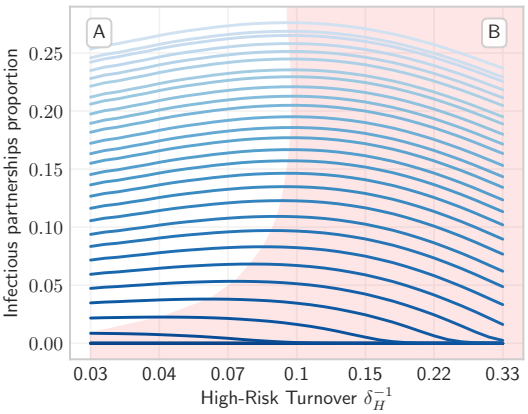
Finally, Figure 9 summarizes trends in overall equilibrium incidence and group-specific prevalence with



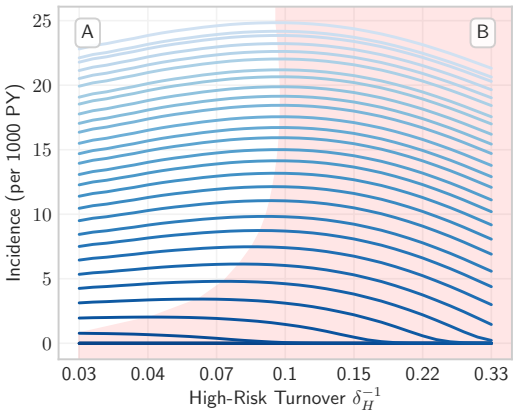
(a) Average contact rate among infectious individuals $\hat{C}_{\mathcal{I}}$



(b) Overall prevalence $\hat{\mathcal{I}}$



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

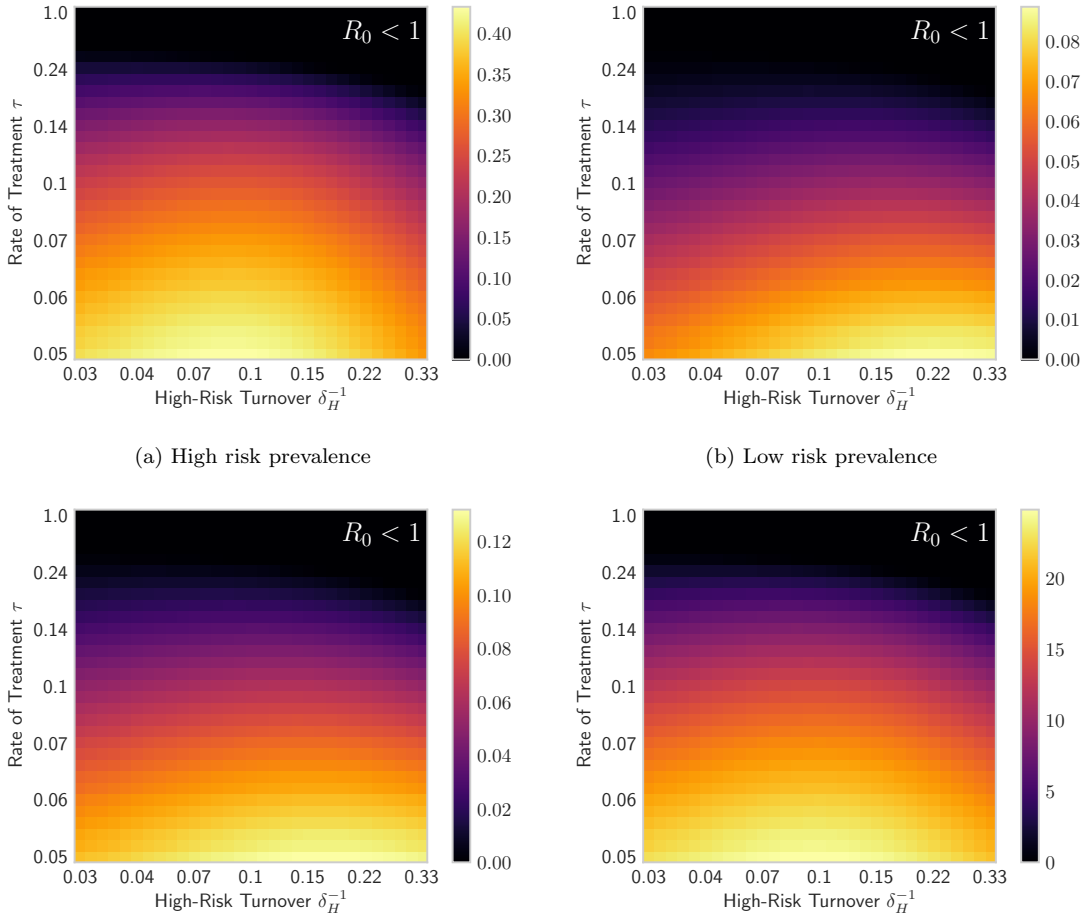


(d) Overall incidence λ

Figure 8: Incidence and the dynamic factors of incidence versus turnover, 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.[9] Treatment consistently decreased equilibrium incidence (as noted above), as well as prevalence, at all rates of turnover. As suggested in Experiment 1.2, infection prevalence increased with turnover for moderate rates of turnover and treatment, among all groups, and overall. However, for higher rates of turnover, prevalence among each risk group peaked, and then declined. As shown in Figure 8, the rate of turnover at which the peak occurs decreased with treatment rate. Finally, for high rates of treatment and/or turnover, the product of the average contact rate of infectious individuals $\hat{C}_{\mathcal{I}}$ and prevalence $\hat{\mathcal{I}}$ was too low to sustain the epidemic. That is, the basic reproductive number R_0

[9] Figures 9c and 9d are the surface projections of the profiles shown in Figures 8b and 8d, respectively.



(a) High risk prevalence

(b) Low risk prevalence

(c) Overall prevalence

(d) Overall incidence

Figure 9: Equilibrium prevalence and incidence for different rates of turnover ϕ (log scale) and treatment τ (log scale).

declined to less than one, and no epidemic was observed.[10]

4.2. Experiment 2: Inferred risk heterogeneity with vs without turnover

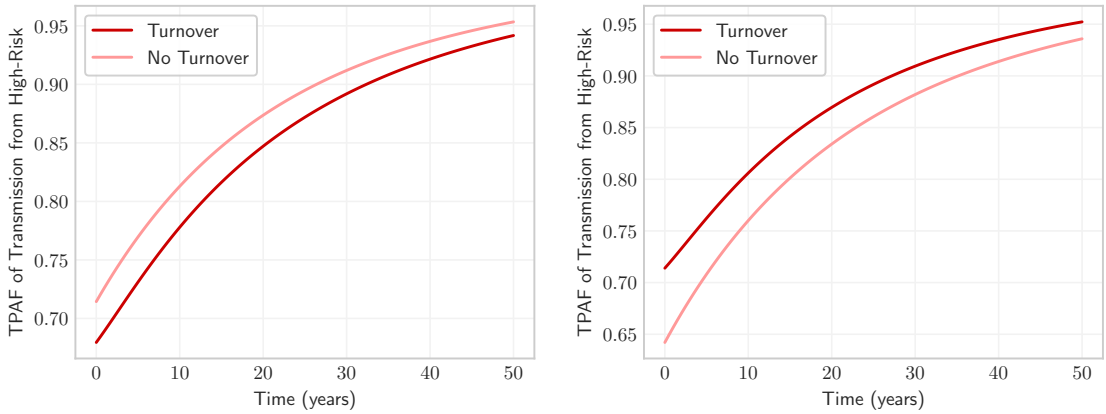
Before model fitting, the predicted prevalence ratio between high and low risk groups was lower with turnover than without: 6.7 vs 9.2. This reflects the "homogenizing" effect of turnover on the average risk experienced by individuals in the model. As shown in Figure B.7b, the high-to-low prevalence ratio consistently declined with turnover for all turnover and treatment rates explored. Thus, when fitting the model to target prevalence values, the fitted contact rates C would have to compensate for this difference.

After fitting the contact rates, both models predicted the desired equilibrium infection prevalence values of 20%, 3%, and 5% among the high risk group, low risk group, and overall. However, in order to do so,

[10] In fact, it can be shown that for extreme rates of turnover, a heterogeneous system (e.g. Full model) will converge on a homogeneous system (e.g. model V1). This result is shown in Figure B.9.

Table 4: Equilibrium contact rates C and prevalence P among the high H and low L risk groups predicted by the models with and without turnover, before and after model fitting.

Context	C_H	C_L	C_H / C_L	P_H	P_L	P_H / P_L
No Turnover (Setting A)	25.0	1.0	25.0	22%	2%	**9.2**
Turnover (Setting B)	25.0	1.0	25.0	22%	3%	**6.7**
No Turnover [fit] (Model A)	23.5	1.6	**15.1**	20%	3%	6.6
Turnover [fit] (Model B)	24.3	1.0	**23.8**	20%	3%	6.7



(a) No fitting

(b) Fitted contact rates

Figure 10: 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.

the ratio of fitted contact rates between high and low risk groups (C_H / C_L) was higher with turnover than without: 23.8 vs 15.1. That is, the inferred level of risk heterogeneity was higher in the model with turnover than in the model without turnover. This is because, 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. These results are also summarized in Table 4.

4.3. Experiment 3: Influence of turnover on the TPAF of the high risk group

Finally, we compared the predicted TPAF of the high risk group with and without turnover in: two settings (same parameters, different prevalence); and two models (same prevalence, different parameters). These results are shown in Figure 10. The TPAF approaches 1 for all models over the 50 year period, indicating that unmet treatment needs of the high risk group are central to epidemic persistence in all scenarios. Additionally, no TPAFs intersect during this period, so relative differences between TPAFs can be described irrespective of time horizon.

Experiment 3.1: Two settings. Figure 10a shows the estimated TPAF of the high risk group in two different settings – with and without turnover. In this case, the estimated TPAF is lower in Setting B with turnover versus in Setting A without turnover. This can be attributed to the larger equilibrium prevalence ratio without model fitting (Experiment 2, Table 4), which results in more onward transmission from the high prevalence high risk group. In other words, the importance of reaching the high risk group in a context without turnover is higher than in a context with turnover, all other factors being equal.

Experiment 3.2: Two models. Figure 10b shows the estimated TPAF of the high risk group by two different models – with and without turnover – both fitted to the same prevalence. Now the estimated TPAF is higher in Model B with turnover versus in Model A without turnover. This result, opposite to Experiment 3.1, 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, differences in prevalence in the high risk group no longer contribute to an increased TPAF estimate by Model A without turnover. Second, as shown in Experiment 2, the ratio of fitted contact rates C_H / C_L in model B with turnover are higher than in Model A without. This affords a higher risk of onward transmission to the high risk group in Model B with turnover, and thus an increased TPAF. This result then implies that models which fail to capture turnover dynamics which are present in reality may underestimate the TPAF of high risk groups. Consequently, the importance of prioritizing high risk groups to achieve epidemic control may be similarly underestimated by such models.

5. Discussion

Epidemic models have rarely considered turnover of individuals among risk groups. We developed a unified framework for modelling turnover in risk, and used the framework to explore the influence of turnover on the contribution of the highest risk group to onward transmission in an illustrative STI epidemic without STI-attributable mortality. We found that this contribution is lower in settings with turnover, but that failure to model turnover when simulating settings with turnover could result in underestimation of the contribution.

Turnover framework. The proposed framework provides a flexible way to parameterize risk group turnover based on available epidemiologic data and/or assumptions. The approach has four advantages. First, the framework defines how specific epidemiologic data and assumptions can be used as constraints to help define rates of turnover (Table 1). These data and assumptions are further discussed in the next paragraph. Second, the framework allows such constraints to chosen and combined in a flexible way (as rows in the linear system of equations $\boldsymbol{b} = A\boldsymbol{\theta}$), depending on which data are available, or which assumptions are most plausible. While it is necessary that constraints do not conflict one another, it is not necessary that

a complete set of G^2 constraints be defined (where G is the number of risk groups), since optimization techniques can be used to calculate, for example, the smallest possible values of the parameters which satisfy the given constraints. Third, this flexible approach also allows the framework to scale to any number of risk groups, G. Fourth, we have shown how several previous implementations of turnover (Stigum et al., 1994; Eaton and Hallett, 2014; Henry and Koopman, 2015) can be recreated exactly using the proposed framework. In so doing, we highlight which specific assumptions are the same and which are different across the different implementations.

The major data needs of this approach include the following four categories. First, the proportion of total individuals in each risk group $\hat{x}_i$ is required. In the context of STIs, risk group size estimates may be obtained from representative surveys of the modelled population such as demographic health surveys, which collect data on self-reported behaviours or engagement in networks associated with variable risk (USAID, 2019). For example, one risk group may defined by any engagement in casual sex within the past year. For marginalized populations, such as sex workers, estimates of population size are generated using various mapping and enumeration methods (Abdul-Quader et al., 2014). Second, the proportion of exogenous individuals who enter into each risk group can be used to specify elements $\hat{e}_i$ as Type 2 Constraints. Such proportions could be obtained the same way as $\hat{x}_i$ above, except using only data from individuals who recently became sexually active. For example, among women who became sexually active in the past year, what proportion also engaged in casual sex within the past year. If data on sexual debut is not available, then recent entry into sexual activity could be approximated using a suitable age range. Third, the average duration of time spent within each risk group δ_i can be used to define Type 3 Constraints. Cross-sectional survey questions asked of female sex workers such as "for how many years have you been a sex worker?" may be used to obtain estimates of duration in sex work, with the recognition that such data are censured (Watts et al., 2010). Longitudinal, or cohort studies that track the self-reported sexual behaviour over time can also provide estimates of duration within variable periods of risk (Fergus et al., 2007). Fourth, similar data, if available, can also be used to estimate the rates of transition between specific risk groups (ϕ).

Influence of turnover on incidence & prevalence. We found that turnover influences the overall equilibrium STI incidence and prevalence, as shown by previous works (Stigum et al., 1994; Zhang et al., 2012; Henry and Koopman, 2015). However, unlike previous works, we also illustrated the influence of turnover on group-specific incidence and prevalence, and demonstrated mechanistically how this occurs. We can summarize this influence of turnover, including reduced ratio of STI prevalence between the highest and lowest risk groups, as "reducing heterogeneity in risk" via movement of individuals between risk groups. Henry and Koopman (2015) demonstrated that such reduction in risk heterogeneity through turnover decreases the basic reproductive number and thus, means epidemic control could be easier to achieve. Our findings on the mechanisms by which increasing turnover reduces the STI treatment rate to achieve zero prevalence further

supports these insights from Henry and Koopman (2015).

Implications for interventions. We found that the TPAF of the highest risk group would be lower in settings without turnover. This is important to consider when comparing whom to prioritize in different regions with different epidemiologic features, such as STI prevalence. That is, a setting with turnover may require less of a focused approach on those at highest risk. Such an implication may be counterintuitive, as a shorter period of higher risk among a smaller group means fewer person-years of intervention required. However, as described in the paragraph above, turnover reduces heterogeneity in risk and reduces the STI prevalence ratio between the highest and lowest risk groups, decreasing the contribution of the highest risk group to overall transmission. Public health implications of the above finding are two-fold. First, if epidemiologists and modelers were to generate a geographic map of TPAF across different settings with different rates of turnover, then turnover could be an important source of variability in the projected TPAF between settings. Second, if decision-makers were allocating a fixed ceiling of STI prevention and treatment resources for a particular high-risk group (e.g. sex workers) across settings, based on the TPAF, then settings with zero or lower rates of turnover may benefit the most from those resources. The latter implication is a general insight, and does not account for: transmission between sites; how costs of STI interventions vary of the sexual life-course of individuals; and other sources of variability in TPAF not explored in our study.

In contrast to a comparison of settings with and without turnover, our comparison of models with and without turnover, calibrated to the same epidemic, showed that ignoring turnover (if it exists in a given setting) will cause the TPAF to be underestimated. This is because heterogeneity in risk must be higher in the presence of turnover than in a model without turnover in order to produce the same epidemic features. Although we examined a single parameter to capture risk (frequency of partner change), these insights would be generalizable to any other component of susceptibility or infectivity, because the risk per susceptible individual (force of infection) includes both biological transmission probabilities and frequency of partner change (Anderson and May, 1991). In the context of models with assortative mixing (individuals are more likely to form partnerships with other individuals in their risk group), the difference between the TPAFs estimated by the model with vs. without turnover was even larger (Appendix B.7). The public health implication of models ignoring turnover which is present is that the TPAF of the highest-risk group will be systematically underestimated, potentially misguiding resources away from the highest-risk group. Follow on research should quantify the size of this potential bias in TPAFs generated from models without turnover, and characterize the epidemiologic conditions under which the bias would be largest.

Limitations. There are five limitations of the study that are important when considering the implications of our proposed unified approach to turnover, and the implications of turnover on TPAF. First, our approach does not account for infection-attributable mortality, such as HIV-attributable mortality. It is well-established that HIV-attributable mortality will reduce the relative size of the higher risk groups, and that

alone can cause an epidemic to decline (Boily and Mâsse, 1997). As such, many models of HIV transmission, in particular those that include very small ($< 3\%$ of the population) high risk groups, such as female sex workers, often do not constrain the relative size of the sub-group populations to be stable over time (Pickles et al., 2013). Second, we assumed a single-sex population and did not stratify by age. In the context of real-world STI epidemics, the relative size of risk groups may differ by both sex and age, such as the number of females who sell sex, versus the number of males who sell sex, versus the number of males who pay for sex (clients). Additional work on the proposed framework could address these two important limitations to the generalizability of the approach across STI models.

Third, we did not include individuals becoming re-susceptible – an important feature of many STIs such as syphilis and gonorrhoea (Fenton et al., 2008). As shown by Fenton et al. (2008) and Pourbohloul et al. (2003), the re-supply of susceptible individuals following STI treatment will fuel an epidemic, and so the influence of turnover on STI incidence, prevalence, and TPAF may be different, and warrants future study. Fourth, our analyses were restricted to equilibrium STI prevalence and incidence. The influence of turnover at different phases of an epidemic – growth, mature, declining – are expected to vary, and thus represents an important topic for future investigation. Finally, our analyses reflected an illustrative STI epidemic in a population with illustrative risk strata. Future work should explore more realistic systems for specific STIs, such as in (Johnson and Geffen, 2016).

Conclusion. In conclusion, turnover in risk will influence epidemic model outputs, including projected incidence, prevalence, and measures of the contribution of high-risk groups to overall STI transmission. Turnover should therefore be considered in transmission models with heterogeneity in risk. Although the data-needs remain potentially challenging, a failure to meet this challenge could lead to misguided information on the importance of addressing the unmet needs of high risk populations – such as those engaged in sex work – to achieve population-level transmission reduction.

Acknowledgements

We would like to thank Kristy Yiu for her logistical support, and attendees of the STI & HIV 2019 World Congress, including Ian Spicknall, Janneke Heijne, Geoff Garnett, and Sevgi Aral, for their thoughtful discussions and feedback on this work.

Funding

The study was supported by: the Canadian Institutes of Health Research Foundation Grant FN 13455; the National Institutes of Health, Grant number: NR016650; the Center for AIDS Research, Johns Hopkins University through the National Institutes of Health, Grant number: P30AI094189.

Conflicts of Interest

Declarations of interest: none.

References

Abu S Abdul-Quader, Andrew L Baughman, and Wolfgang Hladik. Estimating the size of key populations: Current status and future possibilities. 9(2):107–114, mar 2014. DOI 10.1097/COH.0000000000000041.

Roy M Anderson and Robert M May. Infectious diseases of humans: dynamics and control. *Infectious diseases of humans: dynamics and control.*, 1991.

Marie Claude Boily and Benoît Mâsse. Mathematical models of disease transmission: A precious tool for the study of sexually transmitted diseases. *Canadian Journal of Public Health*, 88(4):255–265, 1997. DOI 10.1007/bf03404793.

Marie Claude Boily, Michael Pickles, Michel Alary, Stefan Baral, James Blanchard, Stephen Moses, Peter Vickerman, and Sharmistha Mishra. 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, mar 2015. DOI 10.1097/QAI.0000000000000437.

Kelsey Case, Peter Ghys, Eleanor Gouws, Jeffery Eaton, Annick Borquez, John Stover, Paloma Cuchi, Laith Abu-Raddad, Geoffrey Garnett, and Timothy Hallett. Understanding the modes of tranmission model of new HIV infection and its use in prevention planning. *Bulletin of the World Health Organization*, 90(11):831–838, nov 2012. DOI 10.2471/blt.12.102574.

DataBank. Population estimates and projections, 2019. URL https://databank.worldbank.org/source/population-estimates-and-projections.

Jeffrey W. Eaton and Timothy B. Hallett. Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence. *Proceedings of the National Academy of Sciences*, 111(45): 16202–16207, nov 2014. DOI 10.1073/pnas.1323007111.

Kevin A. Fenton, Romulus Breban, Raffaele Vardavas, Justin T. Okano, Tara Martin, Sevgi Aral, and Sally Blower. Infectious syphilis in high-income settings in the 21st century. *The Lancet Infectious Diseases*, 8(4):244–253, apr 2008. DOI 10.1016/S1473-3099(08)70065-3.

Stevenson Fergus, Marc A Zimmerman, and Cleopatra H Caldwell. Growth trajectories of sexual risk behavior in adolescence and young adulthood. *American Journal of Public Health*, 97(6):1096–1101, jun 2007. DOI 10.2105/AJPH.2005.074609.

Geoffrey P. Garnett and Roy M. Anderson. Balancing sexual partnership in an age and activity stratified model of HIV transmission in heterosexual populations. *Mathematical Medicine and Biology*, 11(3):161–192, jan 1994. DOI 10.1093/imammb/11.3.161.

Christopher J. Henry and James S. Koopman. Strong influence of behavioral dynamics on the ability of testing and treating HIV to stop transmission. *Scientific Reports*, 5(1):9467, aug 2015. DOI 10.1038/srep09467.

ICAP. PHIA Project, 2019. URL https://phia.icap.columbia.edu.

Leigh F. Johnson and Nathan Geffen. A Comparison of two mathematical modeling frameworks for evaluating sexually transmitted infection epidemiology. *Sexually Transmitted Diseases*, 43(3):139–146, mar 2016. DOI 10.1097/OLQ.0000000000000412.

Jesse Knight, Linwei Wang, Huiting Ma, Sheree Schwartz, Stefan Baral, and Sharmistha Mishra. The influence of risk group turnover in STI/HIV epidemics: mechanistic insights from transmission modeling. In *STI & HIV 2019 World Congress*, Vancouver, BC, Canada, 2019. URL https://sti.bmj.com/content/95/Suppl_1/A83.3.

J S Koopman, J A Jacquez, G W Welch, C P Simon, B Foxman, S M Pollock, D Barth-Jones, A L Adams, and K Lange. The role of early HIV infection in the spread of HIV through populations. *Journal of Acquired Immune Deficiency Syndromes*, 14(3):249–58, mar 1997. URL http://www.ncbi.nlm.nih.gov/pubmed/9117458.

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

LAPACK. LAPACK: Linear Algebra PACKage, 1992. URL http://www.netlib.org/lapack.

Charles L Lawson and Richard J Hanson. *Solving least squares problems*, volume 15. SIAM, 1995.

Thomas Robert Malthus. *An Essay on the Principle of Population*. 1798.

Cicely Marston and Eleanor King. Factors that shape young people's sexual behaviour: a systematic review. *Lancet*, 368 (9547):1581–1586, nov 2006. DOI 10.1016/S0140-6736(06)69662-1.

Sharmistha Mishra, Richard Steen, Antonio Gerbase, Ying Ru Lo, and Marie Claude Boily. Impact of High-Risk Sex and Focused Interventions in Heterosexual HIV Epidemics: A Systematic Review of Mathematical Models. *PLoS ONE*, 7(11): e50691, nov 2012. DOI 10.1371/journal.pone.0050691.

Sharmistha Mishra, Michael Pickles, James F Blanchard, Stephen Moses, and Marie Claude Boily. Distinguishing sources of HIV transmission from the distribution of newly acquired HIV infections: Why is it important for HIV prevention planning? *Sexually Transmitted Infections*, 90(1):19–25, feb 2014. DOI 10.1136/sextrans-2013-051250.

Sharmistha Mishra, Marie-Claude Boily, Sheree Schwartz, Chris Beyrer, James F. Blanchard, Stephen Moses, Delivette Castor, Nancy Phaswana-Mafuya, Peter Vickerman, Fatou Drame, Michel Alary, and Stefan D. Baral. 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, aug 2016. DOI 10.1016/j.annepidem.2016.06.004.

Christinah Mukandavire, Josephine Walker, Sheree Schwartz, Marie-Claude Boily, Leon Danon, Carrie Lyons, Daouda Diouf, Ben Liestman, Nafissatou Leye Diouf, Fatou Drame, Karleen Coly, Remy Serge Manzi Muhire, Safiatou Thiam, Papa Amadou Niang Diallo, Coumba Toure Kane, Cheikh Ndour, Erik Volz, Sharmistha Mishra, Stefan Baral, and Peter Vickerman. Estimating the contribution of key populations towards the spread of HIV in Dakar, Senegal. *Journal of the International AIDS Society*, 21:e25126, jul 2018. DOI 10.1002/jia2.25126.

Annett Nold. Heterogeneity in disease-transmission modeling. *Mathematical Biosciences*, 52(3-4):227–240, dec 1980. DOI 10.1016/0025-5564(80)90069-3.

Michael Pickles, Marie Claude Boily, Peter Vickerman, Catherine M Lowndes, Stephen Moses, James F Blanchard, Kathleen N Deering, Janet Bradley, Banadakoppa M Ramesh, Reynold Washington, Rajatashuvra Adhikary, Mandar Mainkar, Ramesh S Paranjape, and Michel Alary. Assessment of the population-level effectiveness of the Avahan HIV-prevention programme in South India: A preplanned, causal-pathway-based modelling analysis. *The Lancet Global Health*, 1(5):e289–e299, nov 2013. DOI 10.1016/S2214-109X(13)70083-4.

Babak Pourbohloul, Michael L. Rekart, and Robert C. Brunham. Impact of mass treatment on syphilis transmission: A mathematical modeling approach. *Sexually Transmitted Diseases*, 30(4):297–305, apr 2003. DOI 10.1097/00007435-200304000-00005.

Zara Shubber, Sharmistha Mishra, Juan F. Vesga, and Marie Claude Boily. The HIV modes of transmission model: A systematic review of its findings and adherence to guidelines. *Journal of the International AIDS Society*, 17(1):18928, jan 2014. DOI 10.7448/IAS.17.1.18928.

Hein Stigum, W. Falck, and P. Magnus. The core group revisited: The effect of partner mixing and migration on the spread of gonorrhea, chlamydia, and HIV. *Mathematical Biosciences*, 120(1):1–23, mar 1994. DOI 10.1016/0025-5564(94)90036-1.

USAID. The DHS Program, 2019. URL https://www.dhsprogram.com.

C. Watts, C. Zimmerman, A. M. Foss, M. Hossain, A. Cox, and P. Vickerman. Remodelling core group theory: the role of sustaining populations in HIV transmission. *Sexually Transmitted Infections*, 86(Suppl 3):iii85–iii92, dec 2010. DOI 10.1136/sti.2010.044602.

James A Yorke, Herbert W Hethcote, and Annett Nold. Dynamics and control of the transmission of gonorrhea. *Sexually Transmitted Diseases*, 5(2):51–56, 1978. DOI 10.1097/00007435-197804000-00003.

Xinyu Zhang, Lin Zhong, Ethan Romero-Severson, Shah Jamal Alam, Christopher J Henry, Erik M Volz, and James S Koopman. 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), nov 2012. DOI 10.1515/1948-4690.1041.

A. Supplemental Equations

Table A.1: Notation

Symbol	Definition
i	risk group index
j	risk group index for "other" group in turnover
k	risk group index for "other" group in incidence
$\mathcal{S}_i$	Number of susceptible individuals in risk group i
$\mathcal{I}_i$	Number of infectious individuals in risk group i
$\mathcal{T}_i$	Number of treated individuals in risk group i
N	total population size
ν	rate of population entry
μ	rate of population exit
ϕ_{ij}	rate of turnover from group i to group j
λ_i	force of infection among susceptibles in risk group i
τ	rate of treatment initiation among infected
$\hat{x}_i$	proportion of individuals in risk group i
$\hat{e}_i$	proportion of individuals entering into risk group i
δ_i	average duration spent in risk group i
C_i	contact rate among individuals in risk group i
β	probability of transmission per contact
ρ_{ik}	probability of contact formation between risk groups i and k

A.1. Model Equations

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

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

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

A.2. Complete Example Turnover System

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

A.3. Redundancy in specifying all elements of $\hat{e}$

Whenever it is assumed that risk groups do not change size (Type 1 Constraint), G rows of the form shown in Eq. (11) are added to $\boldsymbol{b}$ and A:

$$
\boldsymbol{b} = \left[\begin{array}{c} \nu x_1 \\ \nu x_2 \\ \nu x_3 \end{array}\right]; \qquad A = \left[\begin{array}{ccccccccc}
\nu & \cdot & \cdot & -x_1 & -x_1 & x_2 & \cdot & x_3 & \cdot \\
\cdot & \nu & \cdot & x_1 & \cdot & -x_2 & -x_2 & \cdot & x_3 \\
\cdot & \cdot & \nu & \cdot & x_1 & \cdot & x_2 & -x_3 & -x_3
\end{array}\right] \tag{11}
$$

These G rows can be row-reduced by summing to obtain:

$$
\begin{aligned}
[\nu x_1 + \nu x_2 + \nu x_3] &= [\nu e_1 + \nu e_2 + \nu e_3 + 0\,\phi_{12} + 0\,\phi_{13} + 0\,\phi_{21} + 0\,\phi_{23} + 0\,\phi_{31} + 0\,\phi_{32}] \\
\nu\,[x_1 + x_2 + x_3] &= \nu\,[e_1 + e_2 + e_3]
\end{aligned} \tag{A.5}
$$

which therefore implies that $\sum_i x_i = \sum_i e_i$, or equivalently $\sum_i \hat{x}_i = \sum_i \hat{e}_i = 1$. Thus, it is redundant to specify all G elements of $\hat{e}$, as the final element will be dictated by Type 1 Constraints.

A.4. Factors of Incidence

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

$$
\begin{aligned}
\lambda_i &= C_i \sum_k \rho_{ik} \beta \frac{\mathcal{I}_k(t)}{\mathcal{X}_k} \\
&= C_i \beta \sum_k \frac{C_k \mathcal{X}_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} \mathcal{X}_{\mathrm{k}}} \frac{\mathcal{I}_k(t)}{\mathcal{X}_k} \\
&= C_i \beta \underbrace{\frac{\sum_k C_k \mathcal{I}_k(t)}{\sum_{\mathrm{k}} C_{\mathrm{k}} \mathcal{X}_{\mathrm{k}}}}_{*}
\end{aligned} \tag{A.6}
$$

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:

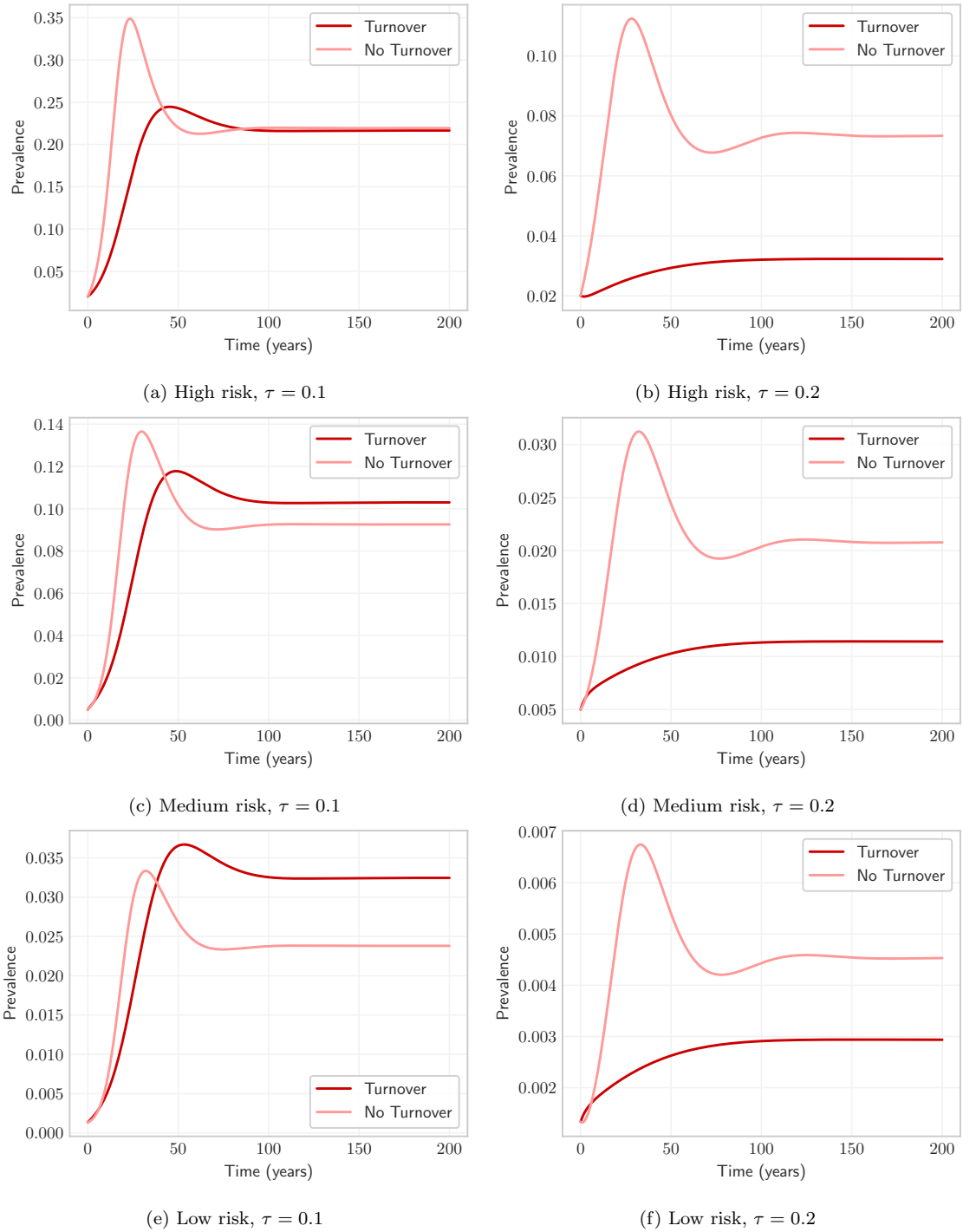
- The average contact rate among infectious individuals $\hat{C}_{\mathcal{I}} = \frac{\sum_k C_k \mathcal{I}_k}{\sum_k \mathcal{I}_k}$
- The proportion of the population who are infectious (prevalence) $\hat{\mathcal{I}} = \frac{\sum_k \mathcal{I}_k}{\sum_k \mathcal{X}_k}$
- The average contact rate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

$$\begin{aligned} \boldsymbol{C}_{\mathcal{I}} &= \hat{C}_{\mathcal{I}} \times \hat{\mathcal{I}} \times \hat{C}^{-1} \\ &= \frac{\sum_k C_k \mathcal{I}_k}{\sum_k \mathcal{I}_k} \times \frac{\sum_k \mathcal{I}_k}{\sum_k \mathcal{X}_k} \times \frac{\sum_k \mathcal{X}_k}{\sum_k C_k \mathcal{X}_k} \\ &= \frac{\sum_k C_k \mathcal{I}_k}{\sum_{\mathrm{k}} C_{\mathrm{k}} \mathcal{X}_{\mathrm{k}}} \end{aligned} \tag{A.7}$$

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.6). In fact, the incidence in each group individually is proportional to $\boldsymbol{C}_{\mathcal{I}}$, as C_i is only factor depending on i.

B. Supplemental Results

B.1. Prevalence with and without turnover



(a) High risk, $\tau = 0.1$

(b) High risk, $\tau = 0.2$

(c) Medium risk, $\tau = 0.1$

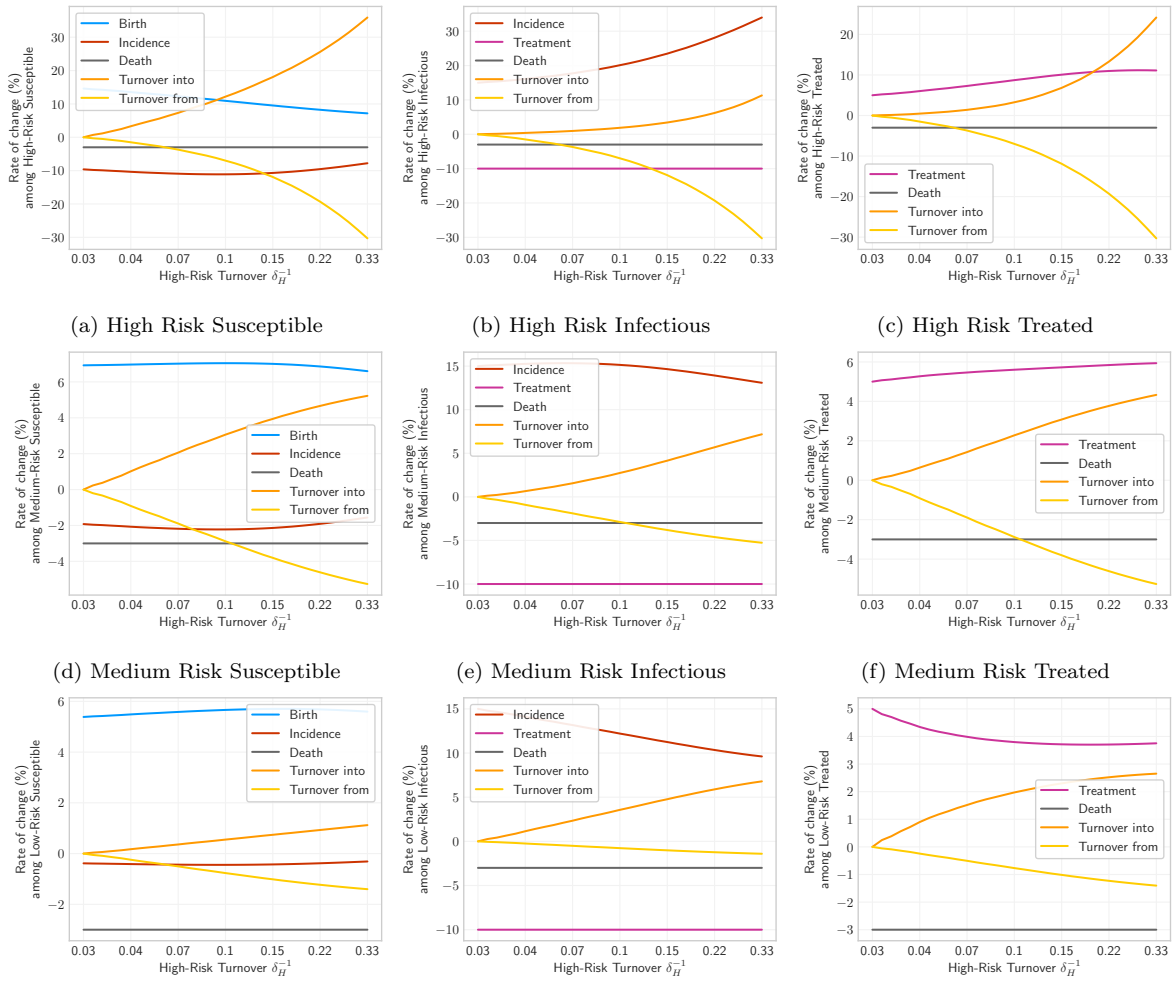
(d) Medium risk, $\tau = 0.2$

(e) Low risk, $\tau = 0.1$

(f) Low risk, $\tau = 0.2$

Figure B.1: Comparison of projected prevalence for each group with and without risk group turnover under two different treatment rates τ. Note that differences in initial trajectories are attributable arbitrary initialization of health states.

B.2. Rates of transition at equilibrium



(a) High Risk Susceptible (b) High Risk Infectious (c) High Risk Treated

(d) Medium Risk Susceptible (e) Medium Risk Infectious (f) Medium Risk Treated

(g) Low Risk Susceptible (h) Low Risk Infectious (i) Low Risk Treated

Figure B.2: Rates of transition among risk groups and health states at equilibrium. All transition rates are shown relative to the named group, for both afferent and efferent transitions. Note that, although each system is at equilibrium, profiles should not sum to zero due population growth.

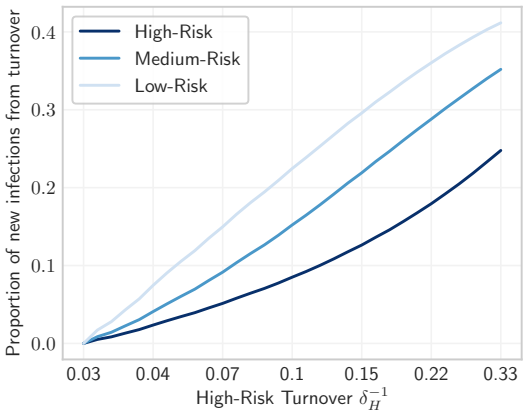
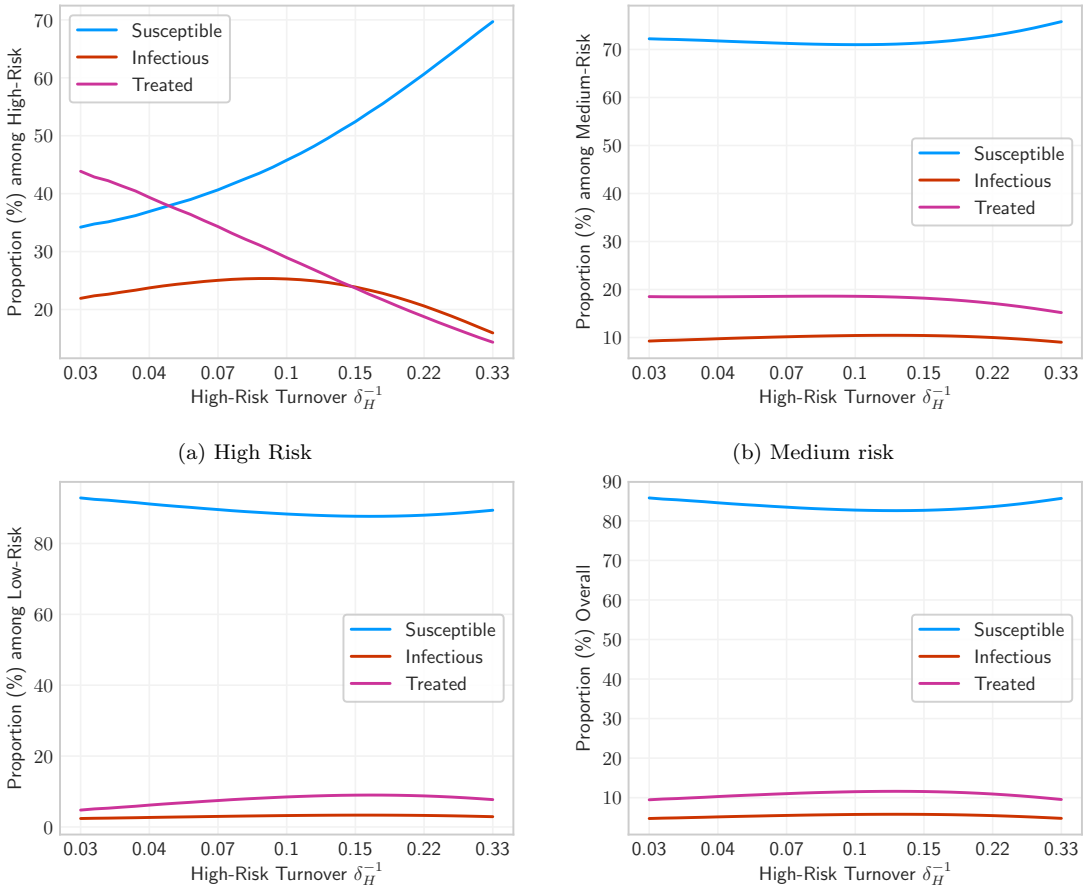


Figure B.3: Proportion of new infectious individuals in each risk group which are from turnover of infectious individuals, as opposed to infection of susceptible individuals in the risk group ($\tau = 0.1$).

B.3. Distribution of health states at equilibrium



(a) High Risk

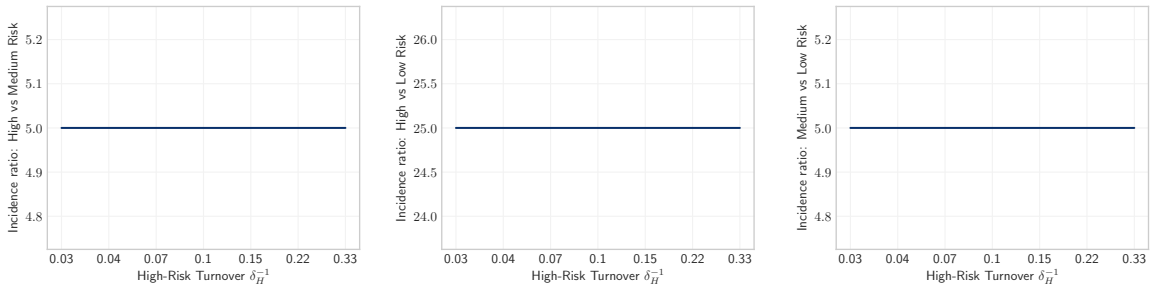
(b) Medium risk

(c) Low Risk

(d) Overall

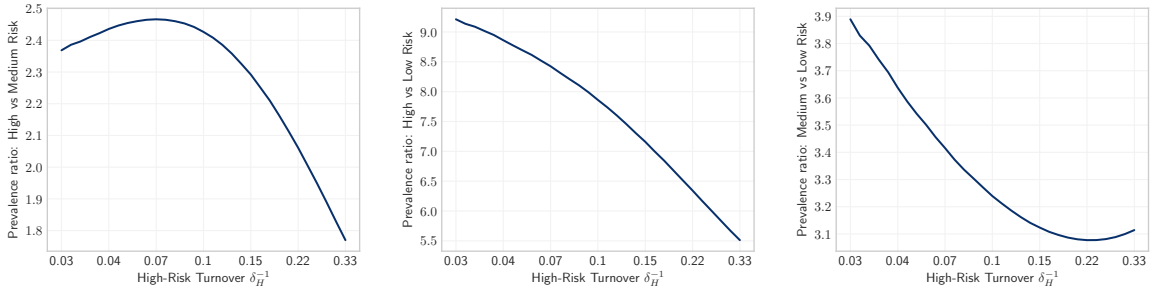
Figure B.4: Equilibrium incidence ratios between risk groups under different rates of turnover.

B.4. Equilibrium Incidence and Prevalence Ratios



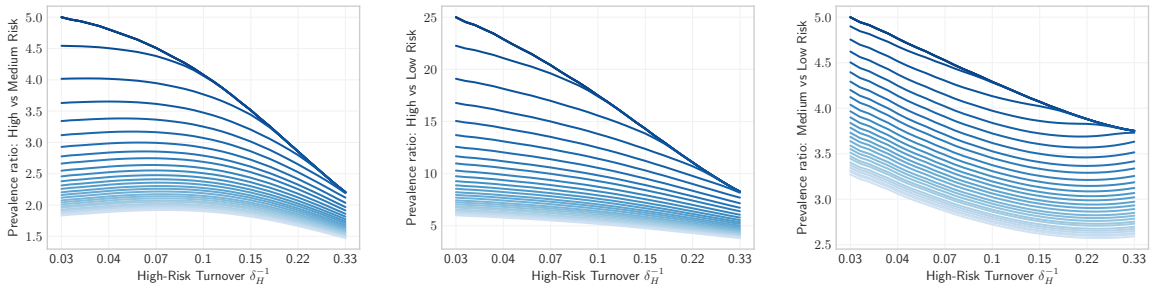
(a) High vs Medium Risk (b) High vs Low risk (c) Medium vs Low Risk

Figure B.5: Equilibrium incidence ratios between risk groups under different rates of turnover ϕ for a treatment rate $\tau = 0.1$. Incidence ratios do not depend on turnover nor treatment – see Eq. (23).



(a) High vs Medium Risk (b) High vs Low risk (c) Medium vs Low Risk

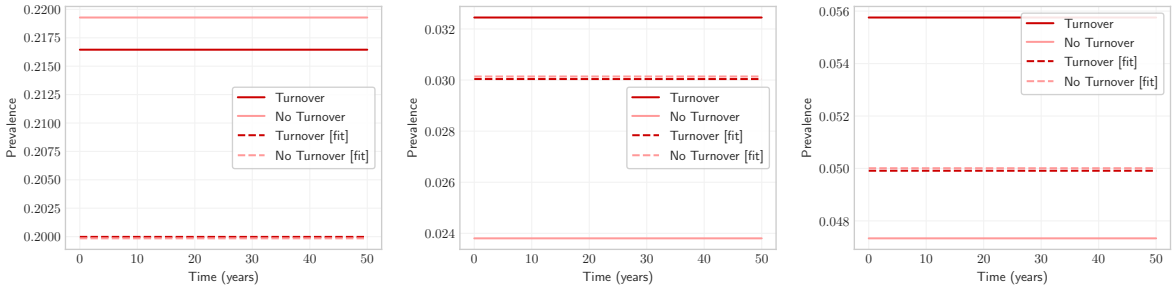
Figure B.6: Equilibrium prevalence ratios between risk groups under different rates of turnover ϕ for a treatment rate $\tau = 0.1$.



(a) High vs Medium Risk (b) High vs Low risk (c) Medium vs Low Risk

Figure B.7: Equilibrium prevalence ratios between risk groups under different rates of turnover ϕ and treatment τ. Turnover increases left to right; treatment increases light to dark.

B.5. *Equilibrium prevalence before and after model fitting*



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

Figure B.8: Equilibrium prevalence among high and low risk groups as well as overall, with and without turnover, and with and without fitted C_i to group-specific prevalence.

B.6. *Extreme turnover converges on a homogeneous system*

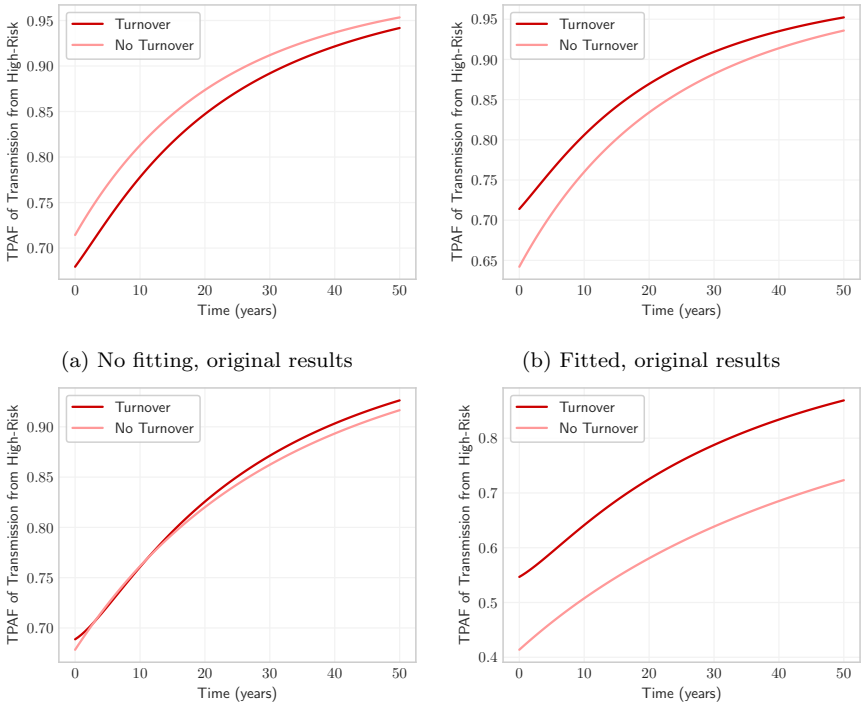
TODO: update

Figure B.9: 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.

B.7. *Experiment 3 under Assortative Mixing*

This section presents preliminary analysis regarding the potential impacts of assortative mixing (individuals are more likely to form partnerships with other individuals in their risk group), on the results of Experiment 3. For reference, Figures B.10a and B.10b show the original results of Experiment 3 (copied from the main text).

Assortative mixing increases the concentration of transmission within the high risk group, decreasing the TPAF, regardless of turnover. However, turnover counteracts this effect by redistributing infectious individuals from high to low risk, permitting greater onward transmission from infectious formerly high risk individuals. Therefore, while turnover may influence the TPAF of the high risk group through the STI prevalence before model fitting, and the inferred risk heterogeneity after model fitting, we hypothesize that the secondary effect of turnover to circumvent assortative mixing would increase the TPAF in both cases.



(a) No fitting, original results

(b) Fitted, original results

(c) No fitting, assortative mixing

(d) Fitted, assortative mixing

Figure B.10: TPAF of the high risk group with and without turnover, and with and without fitted contact rates to group-specific prevalence. Comparison of original results, under proportionate mixing, with no disease-attributable mortality, and without reinfection (a, b), with a model considering assortative mixing (c, d).

To explore this hypothesis, we re-ran Experiment 3 under an assortative mixing scenario, with $\epsilon = 0.5$ (Nold, 1980). Before model fitting and under assortative mixing (Figure B.10c), the TPAF of the high risk group with turnover is similar to the TPAF without turnover, whereas under proportional mixing it was lower with turnover than without turnover (Figure B.10a). After model fitting and under assortative mixing (Figure B.10d), the TPAF of the high risk group is even higher with turnover than without turnover, as compared to proportional mixing (Figure B.10b). Both of these results support our hypothesis above, though more comprehensive experiments are needed.