Published in final edited form as:

J Theor Biol. 2017 September 07; 428: 1–17. doi:10.1016/j.jtbi.2017.06.004.

Risk Stratification in Compartmental Epidemic Models: Where to Draw the Line?

Sze-chuan Suen, PhD¹, Jeremy D. Goldhaber-Fiebert, PhD², and Margaret L. Brandeau, PhD³

¹Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA

²Stanford Health Policy, Centers for Health Policy and Primary Care and Outcomes Research, Stanford University, Stanford, CA, USA

³Department of Management Science and Engineering, Stanford University, Stanford, CA, USA

Abstract

Economic evaluations of infectious disease control interventions frequently use dynamic compartmental epidemic models. Such models capture heterogeneity in risk of infection by stratifying the population into discrete risk groups, thus approximating what is typically continuous variation in risk. An important open question is whether and how different risk stratification choices influence model predictions of intervention effects. We develop equivalent Susceptible-Infected-Susceptible (SIS) dynamic transmission models: an unstratified model, a model stratified into a high-risk and low-risk group, and a model with an arbitrary number of risk groups. Absent intervention, the models produce the same overall prevalence of infected individuals in steady state. We consider an intervention that either reduces the contact rate or increases the disease clearance rate. We develop analytical and numerical results characterizing the models and the effects of the intervention. We find that there exist multiple feasible choices of risk stratification, contact distribution, and within- and between-group contact rates for models that stratify risk. We show analytically and empirically that these choices can generate different estimates of intervention effectiveness, and that these differences can be significant enough to alter conclusions from cost-effectiveness analyses and change policy recommendations. We conclude that the choice of how to discretize risk in compartmental epidemic models can influence predicted effectiveness of interventions. Therefore, analysts should examine multiple alternatives and report the range of results.

Corresponding Author and Address for Reprints: Sze-chuan Suen, 3715 McClintock Ave, GER 240, Los Angeles, CA 90089-0193, Phone: 314-853-1784, ssuen@usc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Risk stratification; Compartmental epidemic models; Infectious disease; Cost-effectiveness analysis

1 INTRODUCTION

Cost-effectiveness analyses (CEAs) of health interventions frequently employ models of disease in a population. Because of variability among individuals, disease models used to support such CEAs are often stratified to reflect population heterogeneity: for example, by age, gender, current health status, risk behaviors or other risk factors. However, the way in which the population is stratified in such an analysis can affect estimates of the likely effectiveness and cost-effectiveness of the interventions being evaluated. For example, an analysis of breast cancer screening policies that does not take into account a woman's age will not yield accurate estimates of cost-effectiveness [1] nor will an analysis that ignores BRCA-1 status (BRCA-1 is a genetic mutation that is highly predictive of breast cancer).

While many studies examine behavior and equilibrium solutions of heterogeneous models of epidemics [2, 3, 4, 5, 6], far fewer studies focus on the effect of modeling heterogeneous mixing on control policy, although these exist as well [7, 8]. Shepard and Zeckhauser show that ignoring heterogeneity in CEA can lead to skewed estimates of intervention cost-effectiveness [9]. Several other studies find that unstratified models can overestimate intervention effectiveness [10, 11, 12]. One study shows that the gain from targeting high-risk groups for treatment increases as the basic reproduction number increases [13]. Henry and Koopman quantify the variation in intervention effects from population heterogeneity in conjunction with heterogeneity over time for HIV [14]. Feng provides an overview of the effect of population heterogeneity in infectious disease modeling [15]. Bosch et al. demonstrate using dengue that multiple model specifications can reproduce observed disease patterns, even when incorporating seasonal-forcing and protective-immunity into the model [16]. However, none of these studies use a distribution of risk to characterize heterogeneous behavior in subpopulations, nor do they use such a distribution to couple risk-group dynamics with overall population dynamics. We suggest that using such a framework allows additional insight into how stratification choices influence subpopulation disease prevalence and transmission rates. It also allows us to identify when stratification choices can generate different epidemic dynamics in the presence of an intervention. In the next few paragraphs, we discuss the motivation for using such a framework.

To perfectly model the spread of a communicable disease in a population, complete information on the transmission network is needed. However, such data is unlikely to be available unless the population of interest is very small: to fully identify the transmission network, we require detailed information on who contacts whom and how often. Documenting contacts for infectious diseases can be difficult. For diseases such as influenza, one would need information on potential airborne contacts which may not be observed by the individual, and for sexually transmitted diseases, individuals are often unwilling or unable to report how many risky sexual contacts they have had, and with whom. Thus,

compartmental epidemic models are often used to model epidemic spread. To reflect population heterogeneity, such models often stratify the population into different risk groups (the limiting case, with a stratification for every individual, corresponds to the network model that would be developed if full network information were available). Population stratification by risk level is a simplification of a complex mixing process where individuals can have widely disparate numbers of risky contacts.

This paper examines the impact of population stratification in compartmental epidemic models used to support CEAs of health interventions for communicable diseases. We assume that the disease is in an endemic steady state and that, although the contact network is not known, data is available on the distribution of number of contacts that individuals have; this is a measure of the distribution of transmission risk in the population. We develop a novel model framework that uses a continuous contact distribution to describe the frequency of risky contacts and we incorporate this information into a susceptible-infected-susceptible (SIS) disease process. The distribution provides the relationship between subpopulation risk-group characteristics and population characteristics. For instance, the weighted average transmission rates by each group, determined by the contact distribution, must be consistent with observed epidemic transmission in the total population, since these groups together form the total population. We show that multiple model stratifications may perfectly fit the observed data. For example, the same endemic steady state may be obtained by a stratification of the population that assumes 5% of individuals are in a high-risk group and the remaining 95% of individuals are low risk, and a different stratification with 10% and 90% of individuals in these groups, respectively.

Using this framework, we compare differences in heterogeneous mixing within models calibrated to the same observable data. We develop analytical results that characterize how the choice of who is in each risk group interacts with the shape of the contact distribution, influences the within-group equilibrium prevalence, and limits the possible within- and between-group transmission rates (but cannot uniquely determine them, and can change estimates of intervention effect).

We further examine the effect of stratification choice on the estimation of intervention effects using a numerical example. We parametrize our model framework with data on gonorrhea dynamics in the population of men who have sex with men (MSM) in the United States. We consider a range of interventions and demonstrate that although multiple models may generate consistent epidemic characteristics absent intervention, they may predict very different estimates of intervention effectiveness. Our numerical example diverges from the existing literature by exploring differences caused by varying the magnitude, type, and targeted nature of an intervention, and we provide evidence that the differences in the estimated intervention effect could be significant enough to alter conclusions of a CEA and potentially change policy recommendations.

Our findings suggest that it is important for modelers to ask themselves “where to draw the line?” when stratifying models of infectious disease to predict intervention outcomes, as the choice of how the model should be stratified, and how finely, has implications for the estimated effect of control policies.

At the heart of the matter is a model selection problem: which model should be used when multiple models would be reasonable? A variety of techniques have been developed to answer this question by picking the model that has the best fit to the data while still being parsimonious. For instance, one can use statistical measurements such as the Akaike or Bayesian information criterion (AIC or BIC) [17, 18, 19, 20, 21]. This evaluation allows the modeler to eliminate models with poor fit to the data and may be a particularly fruitful approach in epidemic modeling when multiple sources of information can be used to inform transmission patterns. Bayesian model selection can also be a useful tool in this context if the modeler can solicit prior distributions on model inputs; alternatively, uninformative priors can be used [22, 23, 24, 25, 26, 27]. These methods have been used in epidemic modeling; for instance, studies have used these approaches for estimating disease parameters from social contact and serological data or age-dependent transmission parameters [18, 20, 28]. Multimodel inference and model averaging techniques can also be considered for reporting results when it is unclear that a single best model should be chosen [29].

Model selection issues continue to persist when a contact distribution for risky behavior is introduced. Several models may equally likely fit the observable data even after eliminating models with poor fit. Some of these well-fitting models will have the same number of parameters, as we demonstrate below, and therefore have similar AIC values. Despite their similarities in terms of fitting the data, we show that these models can generate different predictions of intervention outcome. We suggest that it is important for epidemic modelers to recognize this model selection issue given uncertainty around the underlying contact structure. Modelers can consider using the aforementioned model selection techniques in this context, and should report results generated from multiple possible models that fit the observed data well. These techniques have been well documented elsewhere, and demonstrating them is beyond the scope of this study. Instead, we focus on showing theoretically and numerically that stratification choices can lead to variation in the predicted intervention outcome. Our findings suggest that modelers should incorporate the level of model stratification into their sensitivity analyses as a test of structural sensitivity to more fully understand the consequences of their modeling decisions. While the importance of model uncertainty and the effect of modeling decisions is becoming more recognized [30], previous work has not examined the effect of risk group choice in the context of a population risk distribution.

In the following sections, we consider an unstratified epidemic model and two types of risk-stratified models using a novel model framework that uses a contact distribution in addition to an SIS disease model. The first stratified model divides a population into a low-risk and a high-risk group (allowing for disease transmission within and between groups), where low-risk individuals have less potentially disease transmitting contact with other individuals and have lower disease prevalence. In the second stratified model, the population is divided into n equally sized groups, where each group has a different average probability of making potentially infectious contacts. In the absence of intervention, all models are consistent with observable disease characteristics such as incidence, prevalence, average number of contacts in one period in that community, the probability of disease transmission for a contact between a susceptible individual and an infected individual, and average clearance rate. We examine the estimated impact of an intervention that either reduces the transmission rate or

increases the disease clearance rate. We develop analytical results characterizing the stratified and unstratified models and the effects of the intervention. We then numerically examine model predictions of intervention effect arising from different stratifications of the population using an example reflective of gonorrhea in the MSM population in the United States.

2 STRATIFIED AND UNSTRATIFIED EPIDEMIC MODELS

We consider several dynamic compartmental SIS (susceptible-infected-susceptible) models representing the spread of a communicable disease in a closed population of N individuals. We assume that prior to any intervention, the disease has reached a steady-state prevalence in the population. We consider an SIS model in which the population is captured by a single risk group (Figure 1a), one in which the population is stratified into two risk groups (Figure 1b), and one with n equally sized risk groups. Each model contains two disease states, susceptible and infected. Infected individuals can recover from the disease and become susceptible again. The models define the epidemic's evolution over time as individuals transition between compartments due to disease transmission and clearance.

2.1 Model Specification

We assume that information on the full contact network is not known but that the distribution of the number of contacts in the population is known. In particular, the number of contacts (k) that individuals have per unit of time is described by a probability density function $f(k)$ with non-negative support.

Unstratified Model—We define the unstratified SIS system as follows:

$$\frac{di}{dt} = \beta i s - \nu i \quad (1)$$

$$\frac{ds}{dt} = -\beta i s + \nu i \quad (2)$$

where disease prevalence i equals the number of infected individuals (I) divided by the total number of individuals in the population (N), s is the fraction of individuals who are susceptible, and therefore equal to $1 - i$, and β and ν are the transmission and clearance rates, respectively (see Figure 1). Since knowing the dynamics of the infected population completely characterizes the dynamics of the susceptible population (because $s = 1 - i$), we will limit the remainder of our exposition to the infected population.

Because we assume an endemic steady state prior to any intervention, the right-hand sides of equations (1) and (2) both equal zero and i is non-zero. Solving these equations, the disease prevalence i can be expressed as:

$$i = \frac{\beta - \nu}{\beta} \quad (3)$$

Consistent with classical SIS formulations [31, p. 18], we define $\beta = -\bar{k} \ln(1 - c)$, where $\bar{k}$ is the average number of contacts in one period, $\bar{k} = \int_0^\infty k f(k) dk$, and c is the probability of disease transmission for a contact between a susceptible individual and an infected individual.

Stratified Models—The two-group stratified model is similar. We define the high-risk group as the proportion α of the population with the most per-period contacts; the remainder of the population is low risk (see Figure 2). Suppose distribution function f accurately describes the real-world contact distribution. Often, the complete density function cannot be observed, and only limited information is available from survey data. We assume that, in addition to the mean number of contacts $\bar{k}$, the modeler has information about the conditional mean number of contacts per period in the high- and low-risk groups, which depends on α .

The mean number of contacts per period in the high- and low-risk groups is given by

$$\bar{k}_{high} = \int_{k^*}^{\infty} k f(k) dk \quad (4)$$

and

$$\bar{k}_{low} = \int_0^{k^*} k f(k) dk \quad (5)$$

where

$$k^* = F^{-1}(1 - \alpha) \quad (6)$$

and F^{-1} is the inverse cumulative distribution function for the pdf f . Note that

$$\bar{k} = \alpha \bar{k}_{high} + (1 - \alpha) \bar{k}_{low} \quad (7)$$

We define β_{rs} as the transmission rate to susceptibles in group r ($r = 1, 2$) from infecteds in group s ($s = 1, 2$) and k_{rs} as the average number of contacts per unit time that members of group r have with members of group s . For example, k_{12} represents the average number of contacts per unit time that high-risk individuals have with low-risk individuals. Using the

same relationship between contacts and transmission rate as in the unstratified model, where $\beta = -k \ln(1 - c)$, we can write the transmission rates in the high- and low-risk groups as:

$$b_1 = -\bar{k}_{high} \ln(1 - c) = -(\bar{k}_{11} + \bar{k}_{21}) \ln(1 - c) = \beta_{11} + \beta_{21} \quad (8)$$

$$b_2 = -\bar{k}_{low} \ln(1 - c) = -(\bar{k}_{22} + \bar{k}_{12}) \ln(1 - c) = \beta_{22} + \beta_{12} \quad (9)$$

where we have assumed for simplicity that c , the probability of disease transmission given a contact between a susceptible individual and an infected individual, is the same as in the unstratified model for both groups. This means that the riskiness of the groups is determined wholly by the number of contacts and not, for example, by heterogeneity in other risk behaviors (such as condom usage, in the case of sexually transmitted diseases) or genetics.

Analogous to the unstratified model, we let I_1 , S_1 and N_1 be the number of infecteds and susceptibles and total number of people, respectively, in the high-risk group. Then we have:

$$\begin{aligned} N_1 &= \alpha N \\ i_1 &= \frac{I_1}{N_1} \\ s_1 &= \frac{S_1}{N_1} = 1 - i_1 \end{aligned} \quad (10)$$

Similarly, for the low-risk group, we have:

$$\begin{aligned} N_2 &= (1 - \alpha) N \\ i_2 &= \frac{I_2}{N_2} \\ s_2 &= \frac{S_2}{N_2} = 1 - i_2 \end{aligned} \quad (11)$$

The terms i_1 and i_2 represent the disease prevalence in the high- and low-risk groups, respectively. Thus,

$$i = \alpha i_1 + (1 - \alpha) i_2 \quad (12)$$

where i is the overall prevalence defined above in the unstratified SIS model.

The system dynamics can be written as follows (see Appendix):

$$\frac{di_1}{dt} = \beta_{11} i_1 (1 - i_1) + \frac{1 - \alpha}{\alpha} \beta_{12} i_2 (1 - i_1) - \nu i_1 \quad (13)$$

$$\frac{di_2}{dt} = \beta_{22}i_2(1-i_2) + \frac{\alpha}{1-\alpha}\beta_{21}i_1(1-i_2) - \nu i_2 \quad (14)$$

These equations fully capture the population dynamics since $N_1 = S_1 + I_1$ and $N_2 = S_2 + I_2$, so $\frac{ds_1}{dt}$ and $\frac{ds_2}{dt}$ are omitted for brevity. Equations (13) and (14) both equal zero, as the system is assumed to be in steady state prior to the intervention. For simplicity we assume that the clearance rate for both groups is the same as the overall clearance rate, ν . As before, since $s_1 = 1 - i_1$ and $s_2 = 1 - i_2$, we limit our discussion to i_1 and i_2 .

We assume that the high-risk group has higher disease prevalence than the low-risk group ($i_1 > i_2$). We define the ratio of cross-group transmission rates β_{21} and β_{12} as:

$$w = \frac{\beta_{12}}{\beta_{21}} \quad (15)$$

The term w can be interpreted as a measure of discordance of between-group transmission rates; $w = 1$ means that the transmission rate from the high- to low-risk group is equal to that from the low- to high-risk group. A low value of w indicates that transmission from low- to high-risk individuals is much smaller than transmission from high- to low-risk individuals.

We assume that the population size N , the disease clearance rate ν , the probability of transmission given a discordant contact c , the overall disease prevalence i , and the average per-period number of contacts in the overall population $\bar{k}$ are known from empirical data. We assume that the pdf of the number of contacts $f(k)$ is unknown, with a form assumed by the analyst, as is w . Any unobserved model parameters therefore need to be consistent with these observed data given the disease dynamics. Therefore we hereafter refer to N , ν , c , i , and $\bar{k}$ as constants; b_1 , b_2 , i_1 , i_2 , β_{11} , β_{12} , β_{22} and β_{21} as unknowns; the equations relating them (equations (4), (5), (8), (9), (12)–(15)) as the constraints; and the relationship between the unstratified SIS model and stratified model (equations (3), (6), (7)) as the unstratified-stratified system constraints.

We also consider a stratified model with n groups of equal size (unlike the two-group model just described, where relative risk group size is given by the parameter α). As for the two-group model, the column sums of the β 's correspond to a transmission rate that is governed by the contact distribution, where $\bar{k}_r$ is the mean number of contacts for group r as determined by distribution f :

$$b_r = -\bar{k}_r \ln(1-c) = \sum_{j=1}^n \beta_{jr} \quad (16)$$

The system dynamics are similar to that for the two-group model:

$$\frac{di_r}{dt} = \frac{1}{n} \sum_{j=1}^n \beta_{rj} i_j (1 - i_r) - \nu i_r \quad \text{for } r=1, 2, \dots, n \quad (17)$$

2.2 Stratified and Unstratified Epidemic Models: Analytical Results

In this section we present analytical results characterizing the relationship between the stratified and unstratified models. We provide equations characterizing the steady-state solution to the stratified system. We show that multiple parameter sets can satisfy the constraints for equivalence of the stratified and unstratified models. All proofs are provided in the Appendix.

We first consider the case of two independent populations in the two-group stratified model; that is, $\beta_{12} = \beta_{21} = 0$. This may be a reasonable approximation of urban and rural populations with little or no contact and where overall, but not regional, incidence rates are known. This may also be reasonable for epidemics with risk groups that have little interaction with one another: for example, for the HIV epidemic, MSM have higher risk of acquiring HIV than the general population and low rates of sexual contact with non-MSM. Examining the case of two independent populations in the two-group stratified model offers insight into how the choice of α can influence the system. For the following result, we consider uniform and exponential contact distributions, as these distributions have analytical inverse cumulative density functions.

Proposition 1—*Assuming independent populations in the two-group stratified model, that is, $\beta_{12} = \beta_{21} = 0$, and either a uniform or exponential distribution for f , and given an observed overall prevalence i , then the choice of α uniquely identifies the parameters of the contact distribution f .*

Proposition 1 illustrates the highly constrained relationship between overall prevalence, α , and the contact distribution. The contact distribution f that results from a given value of α may not always be consistent with reasonable assumptions about contact patterns: for instance, f could have unreasonably high upper bounds. Thus, the choice of α may be constrained if the model aims to match observed overall prevalence and real-world information about contact patterns.

We next turn to reconciling the two-group stratified and unstratified SIS systems with no cross-transmission, and with no assumptions about the form of the contact distribution.

Proposition 2—*Assuming independent populations in the two-group stratified model, that is, $\beta_{12} = \beta_{21} = 0$, it is not possible to find a stratified SIS model that generates the same non-zero, non-saturated steady state ($i_1 > 0, i_2 > 0$) as the non-stratified SIS model.*

We now consider the more general case of a two-group stratified model in which the populations are not independent, i.e., β_{12} and β_{21} do not necessarily equal 0. We treat α , b_1 , b_2 , i_1 , and i_2 as known, and solve for the β 's (β_{11} , β_{12} , β_{21} , and β_{22}).

Proposition 3—*The β 's that satisfy the constraints for equivalence of the two-group stratified and unstratified models (e.g., equations (3)–(9), (12)–(15)) are either an infinite or null set.*

Proposition 3 tells us that if any sets of β satisfy the constraints for a given (i_1, i_2) pair, there will be multiple sets of β that satisfy the constraints.

Given the modeler's choice of α and contact distribution f , we can find b_1 and b_2 using equations (8) and (9). Using the constraints (equations (3)–(9), (12)–(15) where equations (13) and (14) equal zero), we can calculate an analytical solution for i_1 and i_2 in the risk-stratified steady-state system:

Proposition 4—*Given α and a contact distribution f , the pre-intervention steady-state solution to the two-group risk-stratified system such that it is equivalent in steady state to the unstratified system (e.g., satisfies equations (3)–(9), (12)–(15)) is given by the following equations for i_1 and i_2 :*

$$\begin{aligned} i_1 &= \frac{\alpha^2(b_1 - b_2)^2 - b_1\nu + b_2(-b_2 + \nu)}{2\alpha(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)} \\ &+ \frac{\sqrt{(\alpha b_1 + b_2 - \alpha b_2)^4 - 2(-1 + 2\alpha)(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)^2\nu + (b_1 - b_2)^2\nu^2}}{2\alpha(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)} \\ i_2 &= \frac{-2\alpha(b_1 - b_2)^2 + \alpha^2(b_1 - b_2)^2 + b_2^2 + b_1\nu - b_2(2b_1 + \nu)}{2(-1 + \alpha)(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)} \\ &+ \frac{\sqrt{(\alpha b_1 + b_2 - \alpha b_2)^4 - 2(-1 + 2\alpha)(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)^2\nu + (b_1 - b_2)^2\nu^2}}{2(-1 + \alpha)(b_1 - b_2)(\alpha(b_1 - b_2) + b_2)} \end{aligned} \quad (18)$$

This is a novel result; to the authors' knowledge no previous literature provides the steady-state two-group risk-stratified SIS solution analytically incorporating contact distributions.

Proposition 4 shows how the choice of α and f (which controls b_1 and b_2) determine i_1 and i_2 . The next proposition establishes that the converse is not true: multiple choices of α and f can generate the same steady state.

Proposition 5—*Multiple sets of α and f (and therefore b_1 and b_2) can generate an equivalent endemic steady state (i_1, i_2) in the two-group stratified model.*

Proposition 5 holds even if w , the ratio of $\frac{\beta_{12}}{\beta_{22}}$, is known from empirical data and equation (15) is used as a constraint. In our numerical analyses we use equation (15) as a constraint.

We develop an analogous result for the n -group stratified model.

Proposition 6—*For the n -group model, there are either zero or infinitely many between-group transmission rate β matrices that will generate the same non-zero, non-saturated equilibrium steady state.*

Proposition 6 establishes that there are multiple calibrations of contact networks that will achieve the same steady state, if any exist.

We have shown that multiple modeling choices can generate risk-stratified systems that have the same steady state as the unstratified system prior to any intervention. This means that given the available data that describes the steady state, it is possible to parameterize multiple stratified models that perfectly fit all the available data. This implies that within a stratification level (i.e., within all two-group stratification models or n -group stratification models), the AIC and BIC values will be the same, as these models have the same number of parameters. However, these different modeling choices for the stratified system can lead to different estimates of intervention outcomes. This is the subject of the next section.

3 DISEASE CONTROL INTERVENTION

In this section we evaluate the effect on disease incidence and prevalence of an untargeted intervention that decreases the transmission rate across the population. This may occur if the contact rate is decreased uniformly across the population. For example, for a sexually transmitted disease epidemic in a high school, the intervention might be a sexual health education program delivered to a fraction of all students that has the effect of reducing the overall number of unprotected sexual contacts. This would result in a decrease of transmission rates. A reduction in the infectious period for fractions of each of the subgroups would also generate reductions in transmission rates would be an equivalent interpretation of this intervention.

3.1 Modeling the Intervention

We model the reduction in transmission rate intervention as a decrease y in the average number of per-period contacts. For the unstratified model we have

$$\bar{k}^{trt} = \bar{k} - y \quad (19)$$

where the superscript trt denotes numbers of contacts after the intervention.

The transmission rate in the unstratified model becomes:

$$\beta^{trt} = -(\bar{k} - y) \ln(1 - c) \quad (20)$$

For the two-group stratified model we have

$$\begin{aligned} \bar{k}_{high}^{trt} &= \bar{k}_{high} - y \\ \bar{k}_{low}^{trt} &= \bar{k}_{low} - y \end{aligned} \quad (21)$$

where the subscripts *high* and *low* denote the high- and low-risk groups, respectively. We show in the Appendix that this is the same magnitude of intervention in both the unstratified and stratified models.

In the stratified model the reduction in contacts in the high- and low-risk groups is proportionally split across the within- and between-group contacts so as to maintain pre-intervention relative levels of contact, reflecting the untargeted nature of the intervention:

$$\begin{aligned}\beta_{11}^{trt} &= -(\bar{k}_{11} - \frac{\bar{k}_{11}}{k_{11} + k_{21}} y) \ln(1-c) \\ \beta_{21}^{trt} &= -(\bar{k}_{21} - \frac{k_{21}}{k_{11} + k_{21}} y) \ln(1-c) \\ \beta_{12}^{trt} &= -(\bar{k}_{12} - \frac{k_{12}}{k_{22} + k_{12}} y) \ln(1-c) \\ \beta_{22}^{trt} &= -(\bar{k}_{22} - \frac{k_{22}}{k_{22} + k_{12}} y) \ln(1-c)\end{aligned}\quad (22)$$

The terms b_1 and b_2 can be written as (proof in Appendix):

$$\begin{aligned}b_1^{trt} &= \beta_{11}^{trt} + \beta_{21}^{trt} \\ b_2^{trt} &= \beta_{22}^{trt} + \beta_{12}^{trt}\end{aligned}\quad (23)$$

Note that $\frac{\bar{k}_{11}}{k_{11} + k_{21}} = \frac{\beta_{11}}{\beta_{11} + \beta_{21}}$, and similarly for the rest of the k 's.

The intervention for the n -group model is the natural extension of this to n groups:

$$\beta_{pr}^{trt} = -(\bar{k}_{pr} - \frac{\bar{k}_{pr}}{\sum_{j=1}^n k_{jr}} y) \ln(1-c) \quad (24)$$

The change in overall prevalence with the intervention for the unstratified model is (see Appendix):

$$\frac{di^{trt}}{dt} = -(\bar{k} - y) \ln(1-c) i(1-i) \quad (25)$$

For the two-group stratified model the change in overall prevalence is:

$$\begin{aligned}\frac{di^{trt}}{dt} &= \alpha \frac{di_1^{trt}}{dt} + (1-\alpha) \frac{di_2^{trt}}{dt} \\ &= \alpha \left[\left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) + \frac{1-\alpha}{\alpha} \left(\frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_1) \right] \\ &\quad + (1-\alpha) \left[\left(\frac{y\beta_{22}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_2) + \frac{\alpha}{1-\alpha} \left(\frac{y\beta_{21}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_2) \right]\end{aligned}\quad (26)$$

and for the n -group model is:

$$\frac{di^{trt}}{dt} = \frac{1}{n} \sum_{j=1}^n \frac{di_j^{trt}}{dt} \quad (27)$$

We define the total number of cases averted as the per-period difference between the number of disease cases with and without the intervention, summed over the analysis period. We can evaluate both the immediate effect of the intervention on overall prevalence ($\frac{di^{trt}}{dt}$) in the stratified and unstratified systems as well as the total cases averted over the duration of the analysis period to determine whether and how these quantities change with different choices of α and contact distribution.

3.2 Health Intervention: Analytical Result

For the case of an exponential contact distribution and an untargeted intervention that decreases the number of contacts, we can prove analytically that different modeling choices – including the choice of α as well as the choice of other parameter values from the set of values for which the stratified and unstratified models are equivalent – lead to different values for the reduction in new disease cases caused by the intervention.

Proposition 7—*Assuming an exponential form for the contact distribution f , and an untargeted intervention that decreases the number of contacts uniformly across the population, the rate at which the intervention reduces prevalence in the two-group stratified model for any given value of α changes with choice of unknowns from among the set of $b_1, b_2, i_1, i_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$ that satisfy the equivalence constraints (equations (3)–(9), (12)–(15)) for the stratified and unstratified models.*

Proposition 7 demonstrates that modeling choices can influence estimated intervention effects. We explore this further with an empirical example using different contact distributions in the next section. We also consider the case of an untargeted intervention that increases the disease clearance rate and a targeted intervention that decreases the number of contacts only in the high-risk group.

4 NUMERICAL ANALYSES

4.1 Numerical Example: Gonorrhea in Men who have Sex with Men

We consider a highly stylized model of gonorrhea in the population of men who have sex with men (MSM) in the United States. SIS dynamics are well suited to modeling gonorrhea, as patients can recover and return to a susceptible state. Previous models of gonorrhea have used SIS dynamic structures [32, 33]. We parameterize the unstratified model with inputs taken from published data (Table 1). In the two-group stratified model, we consider values of α between 0 and 50%, along with two choices of contact distribution form (beta and uniform). For each α and contact distribution, we use the constants in Table 1 along with the calculated b_1 and b_2 values and calibrate the pre-intervention β s and within-group prevalence to the observed steady-state prevalence. We stop the calibration process when the observed steady state prevalence is within $2.2204e^{-16}$ of the desired prevalence (0.164). In the two-group model, we can validate this process by comparing the calibrated i_1 and i_2 values against those given in equation (18). In all cases we find the calibrated values to be within 0.03 of the values calculated in equation (18). We then calculate the number of disease cases averted over the analysis period (365 days) for these different α 's and contact distributions. We also generate two instances of the n -group stratified model using $n = 10$,

where the pre-intervention steady state is the same as in the unstratified and two-group stratified model. The disease and population behavior are assumed to be in steady state prior to intervention.

We simulate a closed population of $N = 10,000$ individuals over one year. We solve for the number of average per-period contacts ($\bar{k}$) in the population using the observed population prevalence (i) and assuming that the unstratified SIS model is in steady state prior to the intervention:

$$\begin{aligned} i &= \frac{\beta - \nu}{\beta} \\ &= \frac{-k \ln(1-c) - \nu}{-k \ln(1-\nu)} \\ \Rightarrow k &= \frac{v}{-(1-i) \ln(1-c)} \quad (28) \end{aligned}$$

Using the parameter values in Table 1, this results in an average contact rate of about one unprotected sexual encounter per month.

For our simulation results, we use beta and uniform distributions. The beta distribution can have a shape consistent with expectations about contact distributions in real life (infrequent very low or very high values, potentially skewed with a long right tail), and the uniform distribution may be of interest since it represents the “uninformed” contact distribution where all frequencies of contact are given equal weight.

For the two-group stratified model, we employ a calibration procedure to determine model parameters. For each choice of α and contact distribution and assuming w is close to 1 ($\pm 5\%$ tolerance), we find b_1 , b_2 . We use an interior-point search algorithm [37] to find the stratified transmission rates and within-group prevalence. Specifically, we find values of i_1 and i_2 between 0 and 1 that satisfy the contact constraints (equations (8) and (9)), along with the system dynamics (equations (13) and (14)) and overall prevalence (equation (12)). This generates the same high-risk and low-risk prevalence as given by equation (18), and can also be used for n -group models. We identify multiple sets of β 's that satisfy the constraints given particular values for b_1 , b_2 , i_1 , and i_2 . We evaluate the number of cases averted over the course of one year (365 periods) for each set.

We perform a similar procedure for the 10-group stratified model, and identify two sets of β 's that achieve the observed steady state using the same contact distribution: we refer to these as Network Model 1 and Network Model 2. The β 's can be thought of as a measure of how “connected” the subgroups are: a larger transmission rate β_{pr} between group p and group r indicates higher “connectedness”. The two sets of β 's used for the empirical example are depicted in Figure 3. As the figure shows, the transmission rates between groups are quite different in the two networks even though both achieve the same steady-state prevalence before the intervention begins.

4.2 Numerical Analysis Findings

As one would expect, not all choices of α and f generate feasible sets of β 's. This can be understood intuitively, using the two-group stratified model as an example: choosing a

contact distribution and a value for α fixes the number of total high- and low-risk contacts (b_1 and b_2), but this level of mixing may not achieve the observed overall steady-state prevalence (i_1 , i_2) given the probability of transmission c and clearance rate ν . For instance, if the clearance rate is low and the number of contacts is very high, it may be impossible to maintain a low steady-state prevalence.

For cases where it is possible to find a feasible stratified SIS system that matches observed prevalence prior to the intervention, we find that within-group and overall prevalence after the intervention may not be the same for models with different α values or for stratified and unstratified models, despite the models having the same pre-intervention steady-state prevalence. We illustrate this in Figure 4. Here we have assumed a beta contact distribution (panel a) with $\bar{k} = .0315$ per day and 16.4% baseline steady-state prevalence. We create two different two-group stratified models that we calibrate to the unstratified model, one with $\alpha = 14\%$ (thus $i_1 = 63\%$, $i_2 = 9\%$) and one with $\alpha = 19\%$ (thus $i_1 = 59\%$, $i_2 = 3\%$) (panel b). This calibration process generates the same i_1 and i_2 values as if we had used equation (18) (see Appendix). We also calibrate the 10-group network models to reflect $\bar{k} = .0315$ per day and 16.4% baseline steady-state prevalence. We consider an untargeted intervention that reduces the average number of contacts per period by $y = 0.01$.

We observe from Figure 4 (panel c) that, after the intervention, overall prevalence in the stratified models is significantly higher than in the unstratified model (approximately 12% – 16% in the stratified models vs. 3% in the unstratified model by the end of the analysis period). Although the untargeted intervention reduces contacts in the stratified models by the same amount as in the unstratified model, the intervention appears less effective in the stratified models. We also observe that prevalence after the intervention (panel c) is slightly different in the two-group stratified models (13.42% in the model with $\alpha = 14\%$ vs. 14.24% in the model with $\alpha = 19\%$) as well as in the two 10-group models (12.6% and 15.9% for Network Models 1 and 2, respectively). Figure 4 also shows that finer stratifications do not necessarily mean lower estimated intervention effectiveness, even for an untargeted intervention: Network Model 1, with 10 risk groups, generates a larger estimated intervention effectiveness than the two-group stratified models. This may be unsurprising given that Network Model 1 has more homogeneous transmission rates between groups than does Network Model 2, and therefore behaves more like an unstratified model than Network Model 2.

This result is not isolated to a single distribution or type of distribution. In Figure 5 we consider a beta distribution and a uniform distribution with $\bar{k} = .0315$ per day (panels a and d) and (as for our other examples) 16.4% baseline steady-state prevalence. For each of the distributions we create two-group stratified models for four feasible choices of α ranging from 14% to 33%. We consider an untargeted intervention that reduces the average number of contacts by $y = 0.01$. We choose a single β set for each α and calculate cases averted over time for each model (panels b and e). For each distribution, we similarly find two 10-group networks (Networks 1 and 2, depicted in Figure 3, for the beta distribution and Networks 3 and 4, depicted in Appendix Figure 2, for the uniform distribution). For both distributions, fewer cases are averted in the stratified models than in the unstratified model. Moreover, as the value of α increases for the two-group models, the difference between cases averted in

the unstratified and stratified models increases and the intervention appears increasingly less effective in the stratified models (panels c and f). Variation in the number of cases averted suggests that there can be significant variation in the estimated intervention effect from different risk stratification choices.

From our theoretical results, we know that when a feasible stratified system exists, multiple sets of within- and between-group contact rates (the β s) can satisfy the constraints. Each set can generate a different intervention effect. Inclusion of this type of uncertainty results in additional variation in estimated intervention effectiveness. For the example in Figure 5 we sample four sets of feasible β s for each α value. The set of β s generates a range of cumulative cases averted over one year for each value of α (shown in the four bars for each color α in panels c and f). For instance, for the beta contact distribution and $\alpha = 14\%$, the number of cumulative cases averted ranges from approximately 100,000 to 170,000 depending on the β s chosen (panel c). This shows that calibration uncertainty can exacerbate the effect of uncertainty due to modeling choices around risk stratification. This variation is also seen in the two 10-group models, which were calibrated with different choices of β . The cumulative cases averted differ by 74,000 (91,000 in Network Model 1 vs. 17,000 in Network Model 2), even though both used the same number of risk groups and were generated using the same contact distribution (Figure 5 panels c and f). The cumulative number of cases averted in the unstratified model is approximately 330,000, more than twice as many as estimated using the stratified models.

We next consider an intervention that reduces contacts by twice as much ($y = 0.02$ instead of $y = 0.01$). Pre-intervention, the average contact rate is about one per month ($\bar{k} = 0.0315$). With the original intervention ($y = 0.01$) this is reduced to about two contacts every three months (average contact rate of 0.0215). Doubling the reduction in contacts reduces the rate to about one contact every three months (0.0115). Appendix Figure 3 shows cases averted over time and cumulative cases averted, analogous to Figure 5 for the unstratified and stratified models. As before, the stratified models predict a lesser intervention effect than the unstratified model, with the difference between the two-group stratified and unstratified models increasing as α increases. We also observe that while the intervention doubles the reduction in contacts, the estimated reduction in new gonorrhea cases does not double. For example, for the intervention with $y = 0.01$, a beta contact distribution with $\alpha = 14\%$ yields 607,000 cases averted after one year (Figure 5 panel b), whereas when $y = 0.02$, 856,000 cases are averted after one year (Appendix Figure 3 panel b). This is to be expected, as the disease dynamics are nonlinear and the same infection cannot be prevented twice.

We obtain similar results when we consider an intervention that decreases the average number of contacts proportionally for all risk groups. Figure 6 shows cumulative cases averted and overall prevalence after one year for an intervention that decreases all transmission rates β by the percentages shown on the x-axis. We choose one set of β s for each α level in the two-group stratified model. Figure 6 shows that a larger percentage decrease in transmission rates in the stratified models compared to the unstratified model is needed to achieve the same intervention outcome. For instance, to reduce prevalence to 12% after one year of intervention, the unstratified model predicts that a 9% reduction in transmission rates is needed, whereas a two-group stratified model using the exponential

distribution with $\alpha = 14\%$ predicts that a 22% reduction in transmission rates is needed and one with $\alpha = 29\%$ predicts that a 30% reduction is needed. However, finer stratifications do not necessarily mean a smaller estimated intervention effect. For small decreases in the transmission rate, the intervention is more effective at reducing overall prevalence after one year in the 10-group network models than in the two-group stratified models (Figure 6, right panel).

It is unsurprising that a stratified model would require a larger magnitude of intervention to achieve the same results as an unstratified model, as individuals are preferentially mixing instead of having equal probability of mixing with anyone, and thus much of the intervention effort is wasted. To illustrate this, suppose we have a two-group model with a particular value of α such that the prevalence within the high-risk group is very high and prevalence in the low-risk group is effectively zero. Decreasing β_{11} will not avert many cases (since most of the high-risk population is already infected), nor will decreasing β_{22} and β_{12} (since low-risk individuals do not account for many new infections). Only by decreasing transmission from high-risk to low-risk individuals, β_{21} , are meaningful numbers of cases averted. This intuition also applies with a larger number of risk groups. With small intervention magnitudes, the intervention is able to act on multiple risk groups that have moderate levels of prevalence and thus behaves similarly to the two-group model; however, as the intervention magnitude rises, the prevalence within a greater number of these risk groups goes to zero, and it becomes harder for additional increases in the intervention to have effect. Of course, at some point the intervention magnitude rises enough that all transmission from and to all risk groups goes to zero, at which point all model stratifications avert the same number of cases.

We additionally evaluate the effect of an intervention that increases the clearance rate by 20% ($\nu = 0.0216$ instead of $\nu = 0.018$), assuming the same contact distributions as in the previous examples. As for the interventions that reduce contacts, the estimated intervention effect for the two-group stratified model varies with α , with the intervention effect decreasing as α increases (Appendix Figure 4), and fewer cases are averted in the stratified models than in the unstratified model.

In all of the examples described above, the unstratified model predicts a larger number of cases averted by the (untargeted) intervention than do the stratified models. Thus an untargeted intervention would appear more cost-effective when an unstratified model is used than when a stratified model is used. These examples show that it is important to examine the estimates from both unstratified and stratified models when evaluating intervention effects. Moreover, because multiple sets of β 's can be feasible for a given risk stratification level α and contact distribution f , it is important to examine the effects of these choices as well.

We now examine a targeted intervention that reduces contacts only in the high-risk group for the two-group stratification and applies the intervention to the two groups with the highest transmission rate in the 10-group networks (to be similar to the two-group models with interventions that reach the top 14% or 19% of risky individuals). Figure 7 shows cumulative cases averted and overall prevalence after one year as a function of the reduction γ in

average number of per-period contacts for the following scenarios: an unstratified model and an untargeted intervention; a two-group stratified model using an exponential contact distribution with $\alpha = 14\%$ or $\alpha = 29\%$ and an untargeted or targeted intervention; and Network Model 1 or Network Model 2 with an untargeted or targeted intervention. The figure shows reduction in contacts ranging up to $y = 0.011$; at that point, all high-risk contacts are eliminated by the targeted intervention when $\alpha = 14\%$. One set of β 's is chosen for each level of α for the two-group stratifications and for each network.

Not surprisingly, for the two-group stratified models the targeted intervention averts more cases than the untargeted intervention for any considered level of contact reduction y , with the intervention effect greater for smaller levels of α . Moreover, as seen in our previous examples, estimated cases averted by the untargeted intervention are smaller when a stratified model is used than when an unstratified model is used.

Interestingly, we find that a targeted intervention can sometimes be less effective than an untargeted intervention, as seen in our results for Network Model 1. Since so much transmission occurs in the lower-risk groups in Network Model 1, focusing only on the higher-risk groups is not as effective as focusing on all risk groups in this case. A similar result is seen comparing targeted interventions in two-group models to untargeted interventions in the unstratified model: for some values of intervention effectiveness y and stratification level α , fewer cases are estimated to be averted by the targeted intervention (modeled by necessity with a stratified model) than for an untargeted intervention modeled using the unstratified model; and for other values of y and α , the opposite is true. The results in Figures 6 and 7 show that the magnitude of difference between estimates from stratified models and the unstratified model is affected by the magnitude of the intervention (i.e., the amount by which the intervention reduces risky contacts).

Again, these effects have much to do with the initial disease prevalence within each risk group. If the disease is highly concentrated within one risk group, a targeted intervention will be effective, but if there are multiple risk groups with moderate prevalence, as in the network models, targeting only the highest-risk groups will do very little to eradicate disease within the other risk groups. This illustrates the importance of the modeler's original choice of risk stratification, which heavily influences (or determines completely, in the two-group case) the within-group prevalence of disease.

5 DISCUSSION

Incorporating risk stratification in models of real-world epidemics is often important in order to accurately model epidemic dynamics, particularly when small subpopulations have a disproportionately large share of the disease burden (e.g., hepatitis C and injection drug users [38]; HIV and men who have sex with men [39, 40]; tuberculosis (TB) and populations without access to TB treatment [41, 42]). However, as we have shown, assumptions about risk stratification can affect estimates of intervention effectiveness. Moreover, multiple feasible choices of risk stratification, contact distribution, and within- and between-group contact rates will yield a stratified model with the same steady state as an unstratified model. We have shown analytically and empirically that these choices can generate different

estimates of intervention effectiveness. Even when models calibrate well to observed prevalence prior to the intervention, predictions of the intervention effect for different levels of population risk stratification may vary widely. In some cases the intervention effect measured by a stratified model is smaller than when an unstratified model is used, and in other cases the opposite is true.

In summary, multiple models may fit the limited observable data perfectly and still generate different predictions surrounding intervention effectiveness. We argue that in these cases modelers should consider multiple model specifications in reporting model outcomes in order to fully convey the effect of structural uncertainty on the results. For instance, when one model cannot be selected as definitive, one can report the range of results, as we do here, or assign likelihoods to multiple models or weigh the perfectly-fitting models by various weights and average the outcomes to generate an average intervention estimate. We direct readers to Burnham and Anderson for discussion on Akaike weights, model selection uncertainty, and multimodel inference [29], and Miao et al. for an example for how these methods can be applied in reporting epidemic model values [19]. In general, considering multiple models in the context of data-driven modeling highlights the importance of structural uncertainty and its relationship to accurately predicting and quantifying the robustness of intervention effects.

Our study has several limitations. We considered interventions that reduce contact rates or increase the disease clearance rate. Health interventions can act through a variety of other mechanisms: for example, treatment can lower infectivity and reduce the death rate among individuals who receive treatment. We would expect results similar to those we have found to hold for other types of health interventions.

We have only examined risk stratification in the context of an SIS model, assuming that the disease is in steady state prior to any intervention. If the pre-intervention disease pattern showed transient effects, a much greater number of model parameters and model stratifications could potentially fit the data. Additionally, SIS models may only be applicable to a limited set of diseases such as gonorrhea, influenza or hepatitis C, but may not be appropriate for diseases with more complex natural histories or treatment patterns such as TB or HIV; these diseases might be more appropriately modeled with SIT, SEIR or SIR structures. Despite their additional complexities, we would expect to see results similar to those we have obtained for the simple SIS model, with non-risk-stratified models predicting greater intervention effect for an untargeted intervention than risk-stratified models, and mixed results for targeted interventions. In non-risk-stratified models, interventions have the same marginal effect on all individuals. Consider the case of an untargeted intervention that reduces contacts: in a risk-stratified model, high-risk individuals, who have the most contacts and are more likely to be infected and therefore more likely to drive the epidemic, have a relatively smaller decrease in their average contacts than their low-risk counterparts. The importance of the relative marginal effect of the intervention within each risk group is underscored by our finding that, when a stratified model is used, a targeted intervention reduces disease incidence more than an untargeted intervention.

The marginal differences, compounded by the nonlinear epidemic dynamics, lead to divergent estimates of the intervention effect. This differential mixing generates variation in intervention effects even when transmission rates are proportionally reduced for all risk groups or when the intervention is targeted. Similarly, for an intervention that increases the clearance rate, the change in total cases cleared per unit time is the same for any level of risk stratification, but the discrepancy in initial prevalence (the high-risk group has higher prevalence) means that a relatively larger number of cases are cleared in the high-risk group than in the low-risk group. The remaining infected individuals then mix with within- and between-group susceptibles at different rates, generating a different estimated treatment effect than in the unstratified model. It is therefore unlikely that adding disease compartments or using other parameters to capture intervention effects would eliminate the effect of risk stratification on outcomes when the effect is present for a basic structure like the SIS model.

Our analyses mainly aim to examine the effects of alternative stratification choices, but our results also have implications for estimation of mixing patterns. We find that significant variation in estimated intervention effectiveness occurs even when there is a good approximation of the number of contacts individuals have: the two 10-group network models with risk stratifications generated using the same contact distribution information generated very different intervention effectiveness predictions; and even for the models with only two risk groups, different between- and within-group transmission rates generated variation in intervention effectiveness (Figure 5, panels c and f). It is therefore important to empirically measure not only the average number of contacts in the different subgroups of a population, but to identify which other subgroups they are contacting. This would allow for more accurate estimation of the β matrix, which still has many degrees of freedom even when the average number of contacts, disease dynamics, and pre-intervention prevalence are all known. Even if full contact network information is not available, estimates of within- and cross-group transmission rates would reduce modeling uncertainty. Studies of subpopulation contact patterns (e.g., [43, 44]) should be encouraged and used by modelers.

In many cases, detailed information on mixing patterns is not available. Modelers should be aware that assumptions made for risk-stratified models in the absence of such data can generate varying predictions of intervention cost-effectiveness. Our results suggest the need for increased scrutiny of population stratification modeling choices. Use of an unstratified model may lead to overestimation of intervention effectiveness for the case of an untargeted intervention, and either underestimation or overestimation for the case of a targeted intervention – and in both cases, underestimation of the uncertainty in the results. A stratified model may lead to more accurate estimates, but stratification choices can generate variation in predicted effects. Modelers should therefore explore several stratified and unstratified model choices, with multiple definitions of high- and low-risk groups and multiple values for group-specific transmission rates. Uncertainty due to these effects should then be reported and incorporated into standard sensitivity analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources: Financial support for this study was provided in part by a National Science Foundation Graduate Research Fellowship under grant DGE-114747 (Sze-chuan Suen), the National Institute on Aging (K01-AG037593; PI: Goldhaber-Fiebert), and by Stanford's Freeman Spogli Institute Underdevelopment Action Fund (PI: Goldhaber-Fiebert). Margaret Brandeau was supported by Grant Number R01-DA15612 from the National Institute on Drug Abuse. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

References

1. Eddy, DM. Screening for Cancer: Theory, Analysis and Design. Prentice-Hall; Englewood Cliffs, New Jersey: 1980.
2. Lajmanovich A, Yorke JA. A deterministic model for gonorrhea in a nonhomogeneous population. *Math Biosci.* 1976; 28(3–4):221–236.
3. Busenberg S, Castillo-Chavez C. A general solution of the problem of mixing subpopulations and its application to risk- and age- structured epidemic models for the spread of AIDS. *IMA J Math Appl Med Biol.* 1991; 8(1):1–29. [PubMed: 1875096]
4. Jacquez JA, Simon CP, Koopman J, Sattenspiel L, Perry T. Modeling and analyzing HIV transmission: the effect of contact patterns. *Math Biosci.* 1988; 92(2):119–199.
5. Del Valle SY, Hyman JM, Chitnis N. Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. *Math Biosci Eng.* 2013; 10(5–6):1475–1497. [PubMed: 24245626]
6. Kaplan EH, Cramton PC, Paltiel AD. Nonrandom mixing models of HIV transmission. *Mathematical and Statistical Approaches to AIDS Epidemiology, Lecture Notes in Biomathematics.* 1989; 83:218–239.
7. Oles K, Gudowska-Nowak E, Kleczkowsky A. Efficient control of epidemics spreading on networks: balance between treatment and recovery. *PLoS One.* 2013; 8(6):e63813. [PubMed: 23750205]
8. Kleczkowsky A, Oles K, Gudowska-Nowak E, Gilligan CA. Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. *J R Soc Interface.* 2011; 9(66):158–169. [PubMed: 21653570]
9. Shepard, DS., Zeckhauser, RJ. The choice of health policies with heterogeneous populations. In: Fuchs, VR., editor. *Economic Aspects of Health.* University of Chicago Press; Chicago, Illinois: 1982. p. 253–298.
10. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Med Decis Making.* 2002; 22(3):218–227. [PubMed: 12058779]
11. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. *Med Decis Making.* 2003; 23(5):379–396. [PubMed: 14570296]
12. Bentley TG, Weinstein MC, Kuntz KM. Effects of categorizing continuous variables in decision-analytic models. *Med Decis Making.* 2009; 29(5):549–556. [PubMed: 19596867]
13. Choe S, Lee S. Modeling optimal treatment strategies in a heterogeneous mixing model. *Theor Biol Med Model.* 2015; 12:28. [PubMed: 26608713]
14. Henry CJ, Koopman JS. Strong influence of behavioral dynamics on the ability of testing and treating HIV to stop transmission. *Sci Rep.* 2015; 5:9467. [PubMed: 25902018]
15. Feng, Z. *Applications of Epidemiological Models to Public Health Policymaking: The Role of Heterogeneity in Model Predictions.* World Scientific Publishers; Singapore: 2014.
16. Ten Bosch QA, Singh BK, Hassan MRA, Chadee DD, Michael E. The role of serotype interactions and seasonality in dengue model selection and control: insights from a pattern matching approach. *PLoS Negl Trop Dis.* 2016; 10(5):e0004680. [PubMed: 27159023]
17. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974; 19(6):716–723.
18. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978; 6(2):461–464.

19. Miao H, Dykes C, Demeter LM, Wu H. Differential equation modeling of HIV viral fitness experiments: model identification, model selection, and multimodel inference. *Biometrics*. 2009; 65(1):292–300. [PubMed: 18510656]
20. Goeyvaerts N, Hens N, Ogunjimi B, Marc Aerts M, Shkedy Z, Van Damme P, Beutels P. Estimating infectious disease parameters from data on social contacts and serological status. *J R Stat Soc Ser C Appl Stat*. 2010; 59(2):255–277.
21. Ponciano JM, Capistran MA. First principles modeling of nonlinear incidence rates in seasonal epidemics. *PLoS Comput Biol*. 2011; 7(2):e1001079. [PubMed: 21379320]
22. Gelfand A, Dey DK. Bayesian model choice: asymptotics and exact calculations. *J Royal Stat Soc Series B*. 1994; 56(3):501–514.
23. Gelman, A., Carlin, JC., Stern, HS., Rubin, DB. *Bayesian Data Analysis*. 2. Chapman and Hall/CRC Press; New York: 2003.
24. Raftery AE. Bayesian model selection in social research (with discussion). *Soc Meth*. 1995; 25:111–195.
25. Kass RE, Raftery AE. Bayes factors. *J Amer Stat Assoc*. 1995; 90(430):773–795.
26. Key, JT., Pericchi, LR., Smith, AFM. Bayesian model choice: what and why?. In: Bernardo, JM.Berger, JO.Dawid, AP.Smith, AFM., et al., editors. *Bayesian Statistics*. Vol. 6. Oxford University Press; Oxford, United Kingdom: 1999. p. 343–370.
27. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial (with discussion). *Stat Sci*. 1999; 14(4):382–417.
28. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol*. 2006; 164(10): 936–944. [PubMed: 16968863]
29. Anderson DR, Burnham KP. Multimodel inference: understanding AIC and BIC in model selection. *Soc Meth and Res*. 2004; 33(2):261–304.
30. Bilcke J, Chapman R, Atchison C, Cromer D, Johnson H, Willem L, Cox M, Edmunds WJ, Jit M. Quantifying parameter and structural uncertainty of dynamic disease transmission models using MCMC: an application to rotavirus vaccination in England and Wales. *Med Decis Making*. 2015; 35(5):633–647. [PubMed: 25623063]
31. Keeling, MJ., Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press; Princeton, New Jersey: 2008.
32. Yorke JA, Hethcote HW, Nold A. Dynamics and control of the transmission of gonorrhea. *Sex Transm Dis*. 1978; 5(2):51–56. [PubMed: 10328031]
33. Capasso, V. *Lecture Notes in Biomathematics*. Springer-Verlag; Berlin: 2008. *Mathematical Structures of Epidemic Systems*, corrected 2nd printing.
34. Morin BR, Medina-Rios L, Camacho ET, Castillo-Chavez C. Static behavioral effects on gonorrhea transmission dynamics in a MSM population. *J Theor Biol*. 2010; 267(1):35–40. [PubMed: 20670632]
35. Jolly AM, Wylie JL. Gonorrhoea and chlamydia core groups and sexual networks in Manitoba. *Sex Transm Infect*. 2002; 78(Suppl 1):i145–i151. [PubMed: 12083435]
36. Centers for Disease Control and Prevention. [accessed 05.05.17] STDs in men who have sex with men. 2012. <http://www.cdc.gov/std/stats12/msm.htm>
37. MathWorks. [accessed 05.05.17] R2015b: fmincon. 2015. <http://www.mathworks.com/help/optim/ug/fmincon.html>
38. Cipriano LE, Zaric GS, Holodniy M, Bendavid E, Owens DK, Brandeau ML. Cost effectiveness of screening strategies for early identification of HIV and HCV infection in injection drug users. *PLoS One*. 2012; 7(9):e45176. [PubMed: 23028828]
39. Juusola JL, Brandeau ML, Owens DK, Bendavid E. The cost-effectiveness of preexposure prophylaxis for HIV prevention in the United States in men who have sex with men. *Ann Intern Med*. 2012; 156(8):541–550. [PubMed: 22508731]
40. Punyacharoensin N, Edmunds WJ, De Angelis D, Delpech V, Hart G, Elford J, Brown A, Gill N, White RG. Modelling the HIV epidemic among MSM in the United Kingdom: quantifying the contributions to HIV transmission to better inform prevention initiatives. *AIDS*. 2015; 29(3):339–349. [PubMed: 25686682]

41. Suen SC, Bendavid E, Goldhaber-Fiebert JD. Cost-effectiveness of improvements in diagnosis and treatment accessibility for tuberculosis control in India. *Int J Tuberc Lung Dis*. 2015; 19(9):1115–1124. [PubMed: 26260835]
42. Xu B, Diwan VK, Bogg L. Access to tuberculosis care: what did chronic cough patients experience in the way of healthcare-seeking? *Scand J Public Health*. 2007; 35(4):396–402. [PubMed: 17786803]
43. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 2008; 5(3):e74. [PubMed: 18366252]
44. Edmunds WJ, O’Callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc Biol Sci*. 1997; 264(1384):949–957. [PubMed: 9263464]

APPENDIX

1 Derivations

Derivation of equation (13), stratified system dynamics

In the unstratified system, we have

$$\frac{dI}{dt} = \beta I \frac{N-I}{N} - \nu I \iff \frac{di}{dt} = \beta i(1-i) - \nu i \quad (1)$$

We derive the analogous equations for the stratified system: we express the dynamics in terms of the population fractions i_1, i_2 using the stratified system dynamics expressed in terms of the number of infected individuals I_1 and I_2 in the high- and low-risk populations, respectively. All formulations are consistent with classical models of stratified and unstratified SIS models [31, p.18, 58].

$$\begin{aligned} \frac{dI_1}{dt} &= \beta_{11} I_1 \frac{N_1 - I_1}{N_1} + \beta_{12} I_2 \frac{N_1 - I_1}{N_1} - \nu I_1 \\ \Rightarrow \frac{1}{N_1} \frac{dI_1}{dt} &= \frac{1}{N_1} \beta_{11} I_1 (1 - i_1) + \frac{1}{N_1} \beta_{12} I_2 (1 - i_1) - \nu \frac{I_1}{N_1} \\ \Rightarrow \frac{di_1}{dt} &= \beta_{11} i_1 (1 - i_1) + \frac{1}{N_1} \frac{N_2}{N_2} \beta_{12} I_2 (1 - i_1) - \nu i_1 \\ &= \beta_{11} i_1 (1 - i_1) + \frac{N_2}{N_1} \beta_{12} i_2 (1 - i_1) - \nu i_1 \\ &= \beta_{11} i_1 (1 - i_1) + \frac{(1-\alpha)N}{\alpha N} \beta_{12} i_2 (1 - i_1) - \nu i_1 \\ \Rightarrow \frac{di_1}{dt} &= \beta_{11} i_1 (1 - i_1) + \frac{1-\alpha}{\alpha} \beta_{12} i_2 (1 - i_1) - \nu i_1 \end{aligned} \quad (13)$$

The derivation of $\frac{di_2}{dt}$ from $\frac{dI_2}{dt}$ follows similarly.

Intervention magnitude: stratified vs unstratified system

We show that the magnitude of intervention is the same in the stratified model as in the unstratified model:

$$\begin{aligned}
\bar{k}^{trt} &= \bar{k} - y \\
\Rightarrow \bar{k}^{trt} &= \alpha \bar{k}_{high} + (1-\alpha) \bar{k}_{low} - (\alpha + 1 - \alpha)y \\
&= \alpha \bar{k}_{high} - \alpha y + (1-\alpha) \bar{k}_{low} - (1-\alpha)y \\
&= \alpha (\bar{k}_{high} + y) + (1-\alpha) (\bar{k}_{low} - y) \\
&= \alpha \bar{k}_{high}^{trt} + (1-\alpha) \bar{k}_{low}^{trt}
\end{aligned} \tag{19}$$

Derivation of equation (23), relationship between contacts and transmission rates with the intervention

$$\begin{aligned}
b_1^{trt} &= -\bar{k}_{high}^{trt} \ln(1-c) \\
&= -(\bar{k}_{high} - y) \ln(1-c) \\
&= -\left(\bar{k}_{11} + \bar{k}_{21} - \left(\frac{\bar{k}_{11}}{\bar{k}_{11} + \bar{k}_{21}} + \frac{\bar{k}_{21}}{\bar{k}_{11} + \bar{k}_{21}}\right) y\right) \ln(1-c) \\
&= -\left(\bar{k}_{11} - \left(\frac{\bar{k}_{11}}{\bar{k}_{11} + \bar{k}_{21}}\right) y\right) \ln(1-c) - \left(\left(\bar{k}_{21} + \frac{\bar{k}_{21}}{\bar{k}_{11} + \bar{k}_{21}}\right) y\right) \ln(1-c)
\end{aligned}$$

Using the definitions given by equation (22), this implies:

$$b_1^{trt} = \beta_{11}^{trt} + \beta_{21}^{trt} \tag{23}$$

and similarly for b_2^{trt} . Note that $\frac{k_{11}}{k_{11} + k_{21}} = \frac{B_{11}}{B_{11} + B_{21}}$, and similarly for the rest of the k 's.

Derivation of equation (26), overall intervention effect $\frac{di}{dt}$

The dynamics for the overall system with the intervention can be written as:

$$\frac{di}{dt} = \frac{d}{dt} (\alpha i_1^{trt} + (1-\alpha) i_2^{trt}) = \alpha \frac{di_1^{trt}}{dt} + (1-\alpha) \frac{di_2^{trt}}{dt}$$

where

$$\begin{aligned}
\frac{di_1^{trt}}{dt} &= \beta_{11}^{trt} i_1 (1-i_1) + \frac{1-\alpha}{\alpha} \beta_{12}^{trt} i_2 (1-i_1) - \nu i_1 \\
\frac{di_2^{trt}}{dt} &= \beta_{22}^{trt} i_2 (1-i_2) + \frac{\alpha}{1-\alpha} \beta_{21}^{trt} i_1 (1-i_2) - \nu i_2
\end{aligned}$$

We plug in the definition of $\beta_{11} = -k_{11} \ln(1-c)$ and k ratio into the definition of β_{11} :

$$\begin{aligned}
\beta_{11}^{trt} &= -\left(\frac{-\beta_{11}}{\ln(1-c)} - \frac{\beta_{11}}{\beta_{11} + \beta_{21}} y\right) \ln(1-c) \\
&= \beta_{11} + \frac{y \beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c)
\end{aligned}$$

Similarly,

$$\beta_{12}^{trt} = \beta_{12} + \frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c)$$

Plugging these into the high-risk group intervention dynamics (equation (13)), we obtain:

$$\begin{aligned} \frac{di_1^{trt}}{dt} &= \left(\beta_{11} + \frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) + \frac{1-\alpha}{\alpha} \beta_{12}^{trt} i_2(1-i_1) - \nu i_1 \\ &= \beta_{11} i_1(1-i_1) + \frac{1-\alpha}{\alpha} \beta_{12}^{trt} i_2(1-i_1) - \nu i_1 + \left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) \\ &= \beta_{11} i_1(1-i_1) + \frac{1-\alpha}{\alpha} \left(\beta_{12} + \frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_1) - \nu i_1 + \left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) \\ &= \beta_{11} i_1(1-i_1) + \frac{1-\alpha}{\alpha} \beta_{12} i_2(1-i_1) - \nu i_1 + \left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) + \frac{1-\alpha}{\alpha} \left(\frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_1) \\ &= \frac{di_1}{dt} + \left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) + \frac{1-\alpha}{\alpha} \left(\frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_1) \end{aligned}$$

Similarly,

$$\frac{di_2^{trt}}{dt} = \frac{di_2}{dt} + \left(\frac{y\beta_{22}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_2) + \frac{\alpha}{1-\alpha} \left(\frac{y\beta_{21}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_2)$$

Since we assume the system is in steady state prior to the intervention, we have $\frac{di_1}{dt} = 0$ and

$\frac{di_2}{dt} = 0$. This means that the change in prevalence with the intervention is:

$$\begin{aligned} \frac{di}{dt} &= \alpha \frac{di_1^{trt}}{dt} + (1-\alpha) \frac{di_2^{trt}}{dt} \\ &= \alpha \left[\left(\frac{y\beta_{11}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_1) + \frac{1-\alpha}{\alpha} \left(\frac{y\beta_{12}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_1) \right] + (1-\alpha) \left[\left(\frac{y\beta_{22}}{\beta_{22} + \beta_{12}} \ln(1-c) \right) i_2(1-i_2) + \frac{\alpha}{1-\alpha} \left(\frac{y\beta_{21}}{\beta_{11} + \beta_{21}} \ln(1-c) \right) i_1(1-i_2) \right] \end{aligned} \quad (26)$$

2 Proofs of Propositions

Proof of Proposition 1

Since we assume there is no cross-group transmission, we have $\beta_{11} = b_1$ and $\beta_{22} = b_2$ from equations (8) and (9) and all parameters are known: the overall prevalence i and clearance rate ν are assumed to be available from empirical data, and α and f (which determine b_1 and b_2) are chosen by the modeler. We find the steady-state prevalence for high- and low-risk groups using equation (3), and the overall prevalence using equation (12):

$$i = \alpha \frac{\beta_{11} - \nu}{\beta_{11}} + (1-\alpha) \frac{\beta_{22} - \nu}{\beta_{22}} \quad (29)$$

We solve equation (29) for α :

$$\alpha = \frac{b_1(b_2 i - b_2 + \nu)}{\nu(b_1 - b_2)} \quad (30)$$

We can additionally characterize f , assuming a distributional form. Suppose we assume a uniform contact distribution ranging from lower bound u to upper bound z (pdf of $f(k) = 1/(z - u)$, cdf $F(k) = \frac{k-u}{z-u}$, and inverse cdf $F^{-1}(k) = k(z-u) + u$). We use equation (6) to find $k^* = \alpha u + z - \alpha z$. Plugging k^* and f into equations (4) and (5), and noting that $b_1 = \beta_{11} = -k_{high} \ln(1 - c)$ and $b_2 = \beta_{22} = -k_{low} \ln(1 - c)$, we have:

$$\begin{aligned} b_1 &= -\frac{1}{2} \alpha (2z - \alpha(z - u)) \ln(1 - c) \\ b_2 &= \frac{1}{2} (\alpha - 1) (y + z - \alpha(z - u)) \ln(1 - c) \end{aligned}$$

After substituting this into equation (30), we can uniquely solve for the upper bound z when we reasonably assume that the minimum number of contacts per period is zero ($u = 0$):

$$z = \frac{(6 - 4\alpha)\nu}{(\alpha - 2)(\alpha - 1)(i - 1) \ln(1 - c)}$$

Appendix Figure 1 (panel a) shows a plot of z versus α for example values of i , c , and ν . For every value of α , there is a unique uniform contact distribution that satisfies the constraints.

We can repeat this process for an exponential distribution with pdf $f(k; \lambda) = \lambda \exp(-\lambda k)$ and inverse cumulative distribution function $F^{-1}(1 - a) = \frac{-\ln[1 - (1 - a)]}{\lambda}$. We obtain:

$$\begin{aligned} b_1 &= -\frac{\alpha(1 - \ln(\alpha)) \ln(1 - c)}{\lambda} \\ b_2 &= -\frac{(1 - \alpha + \alpha \ln(\alpha)) \ln(1 - c)}{\lambda} \\ \lambda &= -\frac{(-1 + i)(-1 + \ln(\alpha))(1 - \alpha + \alpha \ln(\alpha)) \ln(1 - c)}{\nu(2 - 2\alpha + (-1 + 2\alpha) \ln(\alpha))} \end{aligned}$$

Again, λ is unique given a choice of α and constants, so the contact distribution is completely constrained by choice of risk stratification. Appendix Figure 1 (panel b) shows a plot of $\frac{1}{\lambda}$ versus α for example values of i , c , and ν .

Appendix Figure 1.

α determines distribution shape in a system without cross-transmission. For uniform (panel a) and exponential distributions (panel b), choice of α determines the shape of the distribution. We show the change in the contact distribution parameter (z or $1/\lambda$) for $i = 0.164$, $\nu = 1/55$, $c = 0.5$ (the values used for the gonorrhea in MSM example). Resulting probability density functions for selected parameter values are illustrated in the subplots.

Proof of Proposition 2

If we assume that the epidemic is well characterized by a non-stratified SIS model, we can incorporate the unstratified-stratified system constraints. Multiplying both sides of equation (7) by $-\ln(1 - c)$ and substituting equations (8) and (9), we obtain:

$$\beta = (\alpha)b_1 + (1 - \alpha)b_2$$

Substituting this into equation (3), we can solve for α :

$$\alpha = \frac{b_2(1 - i) - \nu}{(b_1 - b_2)(i - 1)}$$

This must hold at the same time as equation (30). With these two equations, we can solve for b_1 and b_2 :

$$b_1 = b_2 = \frac{\nu}{1 - i}$$

This is a contradiction, as b_1 and b_2 cannot be equal under any contact distribution with non-zero variance and $0 < \alpha < 1$. We therefore conclude that it is impossible to simultaneously represent the epidemic with an unstratified SIS model and a stratified SIS model with no cross-group transmission.

Proof of Proposition 3

We first set aside the unstratified-stratified system constraints and characterize the unknowns using the stratified system constraints. Given α and f , the terms b_1 and b_2 can be calculated from equations (8) and (9). We then use the relationship between the transmission rates and outgoing-contact rates (equations (13) and (14)) along with the system dynamics (equations (8) and (9)) to form the following linear equation:

$$u = \begin{bmatrix} \nu i_1 \\ \nu i_2 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} i_1(1-i_1) & \frac{1-\alpha}{\alpha} i_2(1-i_1) & 0 & 0 \\ 0 & 0 & \frac{\alpha}{1-\alpha} i_1(1-i_2) & i_2(1-i_2) \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} \quad (31)$$

The matrix (A) has rank 3, so we identify a single vector (n) that spans its nullspace.

$$n = \begin{bmatrix} (1-\alpha)i_2 \\ -\alpha i_1 \\ -(1-\alpha)i_2 \\ \alpha i_1 \end{bmatrix}$$

Let A^{-1} be the pseudoinverse of A and h be any scalar value. For any choice of i_1 and i_2 that satisfy equation (12), all values of $A^{-1}\vec{u} + hn$ that exist in the positive orthant are possible $\vec{\beta}$. It is also possible that no solutions are in the positive orthant, in which case there are no $\vec{\beta}$ for that set of i_1, i_2 and constants. Additional information on the relationship between $\vec{\beta}$'s would be required to limit the $\vec{\beta}$'s to one of the feasible sets of $\vec{\beta}$'s, given that feasible sets exist.

Proof of Proposition 4

To solve for i_1 and i_2 we develop two equations in these two unknowns. First, we can encapsulate the constraints in equation (31) by noting that the row space must be orthogonal to the right null space in any linear transformation, so $u \cdot \text{null}(A^T) = 0$:

$$\begin{bmatrix} \nu i_1 \\ \nu i_2 \\ b_1 \\ b_2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\alpha}{(i_1-1)} \\ \frac{(1-\alpha)}{(i_2-1)} \\ \alpha i_1 \\ (1-\alpha)i_2 \end{bmatrix} = \vec{0} \quad (32)$$

Second, we can multiply both sides of equation (7) by $-\ln(1-c)$ to find $\alpha b_1 + (1-\alpha)b_2 = \beta$. Plugging this into equation (3), which is equal to equation (12), we see that:

$$\alpha i_1 + (1-\alpha)i_2 = i = \frac{\beta - \nu}{\beta} = \frac{\alpha b_1 + (1-\alpha)b_2 - \nu}{\alpha b_1 + (1-\alpha)b_2} \quad (33)$$

We find the solution to the system of equations (32) and (33) in Mathematica using the following code:

```
(* First we define the equations *)
Amat = {{i1*(1 - i1), ((1 - a)/a)*i2*(1 - i1) , 0, 0},
{0, 0, (a/(1 - a))*i1*(1 - i2), i2*(1 - i2)},
{1, 0, 1, 0},
{0, 1, 0, 1}}
firstConstraint = NullSpace[Transpose[Amat]]
u = {{i1 v}, {i2 v}, {b1}, {b2}}
*Next we solve for i1 and i2*)
sols = Solve[{firstConstraint.u == 0,
a*i1 + (1 - a)* i2 == (a*b1 + (1 - a)* b2 - v)/(a*b1 + (1 - a)* b2)},
{i1, i2}] // FullSimplify
```

This generates the output:

$$\begin{aligned} & \text{sols=} \\ & \text{Solve[\{firstConstraint.outVec==0, a * i1+(1-a) * i2==(a * b1+(1-a) * b2-v)/(a * b1 * (1-a) * b2)\},} \\ & \quad \{i1, i2\}]/\text{FullSimplify} \\ & \quad \left\{ \left\{ i1 \rightarrow 1 - \frac{v}{ab1+b2-ab2}, i2 \rightarrow 1 - \frac{v}{ab1+b2-ab2} \right\}, \left\{ i1 \rightarrow \left(a^t (b1-b2)^t - b1v + \right. \right. \right. \\ & \quad \left. \left. b2(-b2+v) - \sqrt{(ab1+b2-ab2)^4 - 2(-1+2a)(b1-b2)(a(b1-b2)+b2)^t v + (b1-b2)^t v^t} \right) / \right. \\ & \quad \left. (2a(b1-b2)(a(b1-b2)+b2)), i2 \rightarrow \left(-2a(b1-b2)^t + a^t (b1-b2)^t + b2^t + b1v - b2(2b1+v) - \right. \right. \\ & \quad \left. \left. \sqrt{(ab1+b2-ab2)^4 - 2(-1+2a)(b1-b2)(a(b1-b2)+b2)^t v + (b1-b2)^t v^t} \right) / \right. \\ & \quad \left. (2(-1+a)(b1-b2)(a(b1-b2)+b2)) \right\}, \left\{ i1 \rightarrow \left(a^t (b1-b2)^t - b1v + b2(-b2+v) + \right. \right. \\ & \quad \left. \left. \sqrt{(ab1+b2-ab2)^4 - 2(-1+2a)(b1-b2)(a(b1-b2)+b2)^t v + (b1-b2)^t v^t} \right) / \right. \\ & \quad \left. (2a(b1-b2)(a(b1-b2)+b2)), i2 \rightarrow \left(-2a(b1-b2)^t + a^t (b1-b2)^t + b2^t + b1v - b2(2b1+v) + \right. \right. \\ & \quad \left. \left. \sqrt{(ab1+b2-ab2)^4 - 2(-1+2a)(b1-b2)(a(b1-b2)+b2)^t v + (b1-b2)^t v^t} \right) / \right. \\ & \quad \left. (2(-1+a)(b1-b2)(a(b1-b2)+b2)) \right\} \} \end{aligned}$$

The output demonstrates that solving these two equations for the two unknowns i_1 and i_2 generates three solution pairs. As with the β 's, solutions must be non-negative and we must have $i_1 > i_2$ for feasibility, since prevalence cannot be negative and we assume that the high-risk group has higher prevalence than the low-risk group. Two of the solutions are infeasible: the first is the same as the unstratified system ($i_1 = i_2 = i$) and the two remaining are complementary quadratic solutions, one of which is negative or imaginary. The last therefore is the unique feasible solution, as provided in equation (18).

Proof of Proposition 5

Proof by example. Using a clearance rate of 0.1 ($\nu = 0.1$), the values $(a, b_1, b_2) = (0.5, 0.7333, 0.0667)$, $(0.4, 0.769231, 0.128205)$ and $(0.3, 0.802469, 0.185185)$ all yield $(i_1, i_2) = (0.8, 0.7)$ when plugged into equation (18).

Proof of Proposition 6

As in Proposition 3, we can express our dynamics and constraints as a linear equation, except this time there are n groups, each with q individuals. This generates a $2n \times n^2$ matrix:

$$\begin{bmatrix} n\nu i_1 \\ n\nu i_2 \\ n\nu i_3 \\ \vdots \\ b_1 \\ b_2 \\ b_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} i_1(1-i_1) & i_2(1-i_1) & i_3(1-i_1) & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & i_1(1-i_2) & i_2(1-i_2) & i_2(1-i_3) & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & i_1(1-i_2) & i_2(1-i_2) & i_2(1-i_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \vdots \\ \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{nn} \end{bmatrix} \quad (34)$$

The matrix has rank $n^2 - 2n + 1$ and is not full rank for n larger than 1, so there exist a nontrivial null-space. Similar to Proposition 3, this implies that there will be infinite feasible sets of β if one feasible set is found.

Proof of Proposition 7

We assume the modeler has chosen a feasible set of values $(i_1, i_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$ consistent with choice of a and f and observed data on overall prevalence i . We find a second feasible set of $(i_1, i_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$ that satisfies the constraints. Let $\text{null}(A)$ be a unit vector in the null space of A , J be the Jacobian function, and M be the matrix formed from the left hand side of equations (8), (9), (12), (13), (14), (15), after substitution with equations (4) and (5) and exponential contact distribution $f(k; \lambda)$. Then if the modeler's original set of unknowns was $x = [i_1, i_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, a, \lambda]$, a second feasible set of unknowns is $x + \text{null}(J(M))$. To compare this to the intervention effect generated by the original set of unknowns, we can take the dot product $\text{null}(J(M))$ with the gradient of $\frac{di}{dt}$ with respect to the unknowns. This product gives the change in the rate of prevalence decline with

the intervention, due to switching from using unknowns x to using unknowns $x + null(J(M))$. If the product is zero, the two sets of unknowns will have the same intervention effect.

We again use Mathematica to evaluate the product, using the following code:

```
(*Define the intervention metric, pdf, inverse cdf, and find b1 and b2*)
TrtMetric [i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] :=
a*((B11/(B11 + B21)*trtdec)*Log[1 - c]*i1*(1 - i1)
+ ((1 - a)/a)*(B12/(B22 + B12)*trtdec)*Log[1 - c]*i2*(1 - i1) )
+ (1 - a)* ((B22/(B22 + B12)*trtdec)*Log[1 - c]*i2*(1 - i2)
+ (1/(1 - a))*(B21/(B11 + B21)*trtdec)*Log[1 - c]* i1*(1 - i2))
pdfFunc[y ] := lambda *Exp[-lambda*y]
cdfInv[p ] := -Log[1 - (p)]/lambda
b1Val = Integrate[y*pdfFunc[y], y, cdfInv[1 - a], Infinity]
bland2 = {b1 -> b1Val[[1]], b2 -> Integrate[y*pdfFunc[y], {y, 0, cdfInv[1
- a]}}}
(*Form constraint matrix*)
f1[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] :=
B11*i1*(1 - i1) + ((1 - a)/a)*B12*i2*(1 - i1) - v*i1
f2[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] :=
B22*i2*(1 - i2) + (a/(1 - a))*
B21*i1*(1 - i2) - v*i2
f3[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] := a*i1 + (1 - a)* i2
f4[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] :=
B11 + B21 - Evaluate[b1 /. bland2]
f5[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] :=
B22 + B12 - Evaluate[b2 /. bland2]
f6[i1 , i2 , B11 , B12 , B21 , B22 , a , lambda ] := B12/B21
(*Find the gradient of the intervention metric with respect to the
unknowns*) gradTrtMetric =
D[TrtMetric[i1, i2, B11, B12, B21, B22, a, lambda], {{i1, i2, B11, B12, B21,
B22, a, lambda}}}
(*Find the Jacobian of the constraints with respect to the unknowns*)
jacobianMat =
D[{f1[i1, i2, B11, B12, B21, B22, a, lambda],
f2[i1, i2, B11, B12, B21, B22, a, lambda],
f3[i1, i2, B11, B12, B21, B22, a, lambda],
f4[i1, i2, B11, B12, B21, B22, a, lambda],
f5[i1, i2, B11, B12, B21, B22, a, lambda],
f6[i1, i2, B11, B12, B21, B22, a, lambda]},
{{i1, i2, B11, B12, B21, B22, a, lambda}}}
(*Find null space of Jacobian*)
nullOfJac = NullSpace[jacobianMat]; // Simplify
```

The null space of the Jacobian is spanned by two vectors, and we can show that both of these vectors are not orthogonal to the intervention by taking the dot product, which does not simplify to zero. Because the analytical form of the dot product is long and complicated, we instead provide numeric examples of the products at $i_1 = 0.6$, $i_2 = 0.4$, $\beta_{11} = 0.9$, $\beta_{12} = 0.3$, $\beta_{21} = 0.5$, $\beta_{22} = 0.8$, $\alpha = 0.2$, $c = 0.5$, and $\nu = 0.1$:

```
dotProd1 = nullOfJac[[1]].gradTrtMetric //Simplify
dotProd2 = nullOfJac[[2]].gradTrtMetric //Simplify
dotProd1 /. {i1 -> 0.6, i2 -> 0.4, B11 -> 0.9, B12 -> 0.3,
B21 -> 0.5, B22 -> 0.8, a -> 0.2, lambda -> 0.2, c -> 0.5,
v -> 0.1, trtdec -> 0.01}
dotProd2 /. {i1 -> 0.6, i2 -> 0.4, B11 -> 0.9, B12 -> 0.3,
B21 -> 0.5, B22 -> 0.8, a -> 0.2, lambda -> 0.2, c -> 0.5,
v -> 0.1, trtdec -> 0.01}
```

The products dotProd1 and dotProd2 equal 0.105 and 0.060, respectively. This means that a unit change along the vector represented by the system parameters (i_1 , i_2 , β_{11} , β_{12} , β_{21} , β_{22} , α , λ) generates a 6–10% change in the intervention metric, $\frac{di}{dt}$ due to the intervention. This shows that the change in the rate of infection due to the intervention varies with choice of risk-stratification parameters.

3 Calibration of Two-group Model

From Table 1, we have $c = 0.5$, $\nu = 0.018$, and overall prevalence $i = 0.164$. Using an exponential distribution, we can find the b_1 and b_2 values (using equations given in the proof for Proposition 1) given an α value. For instance, if $\alpha = 14\%$, then $b_1 = 0.0538$ and $b_2 = 0.0167$. Evaluating equation (18) from Proposition 4 at these values, we find $(i_1, i_2) = (0.63, 0.09)$. If $\alpha = 19\%$, then $b_1 = 0.0492$ and $b_2 = 0.0154$. Evaluating equation (4) with these values gives $(i_1, i_2) = (0.59, 0.03)$. These are the steady state values depicted in Figure 4 (panels b and c) prior to the intervention.

Highlights

- We examine effects of risk stratification choices in compartmental epidemic models.
- We compare unstratified and risk-stratified SIS models.
- We show that risk stratification choices change intervention outcome estimates.
- These changes can be significant in magnitude.
- Analysts should examine multiple alternatives and report the range of results.

a. Unstratified SIS Model

b. Stratified SIS Model

Figure 1.

Schematics of the unstratified model (a) and two-group stratified model (b). Individuals become infected due to transmission at rate β and clear infection at rate ν (solid arrows). In the stratified model, individuals always remain in their original high- or low-risk groups but can become infected due to cross-group transmission (dashed arrows).

Figure 2. Stratification of population into high- and low-risk groups in the two-group model. A fraction α of the population is defined to be high risk. This group has a higher number of contacts ($\bar{k}_{high}$) than the low-risk group ($\bar{k}_{low}$) and higher disease prevalence. The overall average number of contacts in the population is $\bar{k}$.

Figure 3. Depiction of the transmission rates in the 10-group models as directed graphs, where arrows represent transmission from one group to another. The thickness of the arrow represents the magnitude of the transmission rate.

Figure 4.

Effect of an untargeted intervention that reduces the average number of contacts by 0.01. Within-group prevalence in the two-group stratified model for two levels of risk stratification, $\alpha = 14\%$ and $\alpha = 19\%$, (panel b) and overall prevalence over time (up to 365 days) for an unstratified model, both two-group stratified models, and both 10-group network models (panel c). We use a beta contact distribution (panel a). The stratified models are calibrated to an unstratified model with 16.4% baseline steady-state prevalence.

Figure 5.

Disease cases averted for different contact distributions f , different risk stratification choices α , and different sets of β 's. We consider two different contact distributions, a beta distribution and a uniform distribution (panels a and d, respectively). We consider an untargeted intervention that reduces the average number of contacts by 0.01. We show cases averted over time for multiple α values in the two-group stratified model, where a single choice of β 's was used for each α , and for the two 10-group network models (panels b and e). We show cumulative cases averted over one year (panels c and f), where we sample over four sets of feasible β 's for each level of α for the two-group models.

Figure 6.

Cumulative cases averted over one year (left panel) and overall prevalence after one year (right panel) for an untargeted intervention that reduces each β by a percentage. One set of β 's is chosen for each level of α in the two-group stratified models and one set of β 's is chosen for each 10-node network model.

Figure 7.

Cumulative cases averted over one year (left panel) and overall prevalence after one year (right panel) as a function of the reduction γ in average number of per-period contacts for the following scenarios: an unstratified model and an untargeted intervention; a two-group stratified model with $\alpha = 14\%$ or $\alpha = 29\%$ and an untargeted or targeted intervention; and Network Model 1 or Network Model 2 with an untargeted or targeted intervention. The figure shows intervention magnitude ranging up to $\gamma = 0.011$; at that point, all high-risk contacts are eliminated by the targeted intervention when $\alpha = 14\%$. One set of β 's is chosen for each level of α .

Table 1

Model inputs for stylized example of gonorrhea in men who have sex with men (MSM)

Variable	Description	Value	Source
c	Probability of transmission from discordant contact	0.500	[34]
ν	Clearance rate (1/average duration with disease in days)	0.018	[32], [35]
w	Ratio of between-group transmission rates	1.000	Assumed
y	Transmission rate decrease due to intervention	0.010	Assumed
i	Overall observed population prevalence	0.164	[36]