The Initialization and Sensitivity of Multigroup Models for the Transmission of HIV *

James M. Hyman
Theoretical Division, MS-B284
Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545

Jia Li
Department of Mathematical Sciences
University of Alabama in Huntsville
Huntsville, AL 35899

E. Ann Stanley
Theoretical Division, MS-B284
Los Alamos National Laboratory
Los Alamos, NM 87545

October 17, 2000

Abstract

We study two multigroup mathematical models of the spread of HIV. In the differential infectivity model, the infected population is divided into groups according to their infectiousness, and HIV is primarily spread by a small, highly infectious, group of superspreaders. In the staged-progression model, every infected individual goes through a series of infection stages and the virus is primarily spread by individuals in an initial highly infectious stage or in the late stages of the disease. We demonstrate the importance of choosing appropriate initial conditions, and define a new approach to distributing the initial population among the subgroups so as to minimize the artificial transients in the solutions due to unbalanced initial conditions. We demonstrate that the rate of removal in and out of a population is an important, yet often neglected, effect. We also illustrate the importance of distinguishing between the number of partners a person has and the number of contacts per partner. By assuming that people with many partners have fewer contacts per partner than people

*This research was supported by the Department of Energy under contracts W-7405-ENG-36 and the Applied Mathematical Sciences Program KC-07-01-01.

with few partners, we found that the epidemic is less sensitive to the partner acquisition rate than one might expect. However, because the probability of transmission of HIV per contact is low, the epidemic is very sensitive to the number of contacts per partner. Modeling this distinction is particularly important when estimating the impact of programs which encourage people to have fewer sexual partners.

1 Introduction

Numerous studies show that the viral burden during HIV infection varies in two ways: temporally within a single individual, and, during the chronic stages, between individuals. (See [9] for references.) Recent studies of heterosexual couples in Africa found that the probability of transmission per sexual contact increases along with increased viral levels in the bloodstream, at least within the studied populations [6,20], and that the viral load is the chief predictor of the risk of heterosexual transmission of HIV. The variations in viral levels and their connection with variations in the infectiousness of an individual introduce questions about the effectiveness of methods for controlling the epidemic. In order to study some of these questions, in [9] we proposed a differential infectivity (DI) model that accounts for differences in infectiousness between individuals during the chronic stages, and the correlation between viral loads and rates of developing AIDS. This study compared the DI model with a simple version of a well-established staged-progression (SP) model, in which every infected individual goes through the same series of stages. The SP model accounts for time variations in infectiousness in the same individual. The SP simulations provided insight into the numerical computations of Jacquez et al. [12, 13], showing that, when partner acquisition rates are high, many of the infections early in the epidemic are caused by those in the initial acute infectious stage. The DI model simulations demonstrated that a small number of highly infectious individuals can have a disproportionate impact on the epidemic, even when they have a short life expectancy.

The importance of sensitively analysis to understanding the impact of uncertainties model parameters has also been investigated by Blower, Sanchez and Dowlatabadi [4, 22] using Latin Hypercubes. Our investigations emphasize the sensitivity of the internal dynamics of the epidemic to uncertainties in the parameters and the initial conditions and complements their approach. Usually there is little data on how infections are distributed among subgroups, so researchers have to select an initial distribution of the infected population based on some intuitive, possibly arbitrary, justification. However, It is well known that initial conditions are important in multigroup models [11]. If the choice of initial distribution effects the results, then the researcher has to be extremely careful when studying the sensitivity of the models to parameters and when cross-comparing models. results. Here we use numerical simulations to illustrate the extreme sensitivity of the timing of an epidemic in multiple-group models to the initial distribution of infected individuals among the different groups. We propose two initialization methods which are "natural" and robust, and allow results from different models to be compared.

Previously we showed that the DI and SP models are very sensitive to the probability of transmission per contact. In this paper we investigate the sensitivity of these two models to the sexually active removal rate, the number of partners per year, and number of contacts per partner. The sexually active removal rate encompasses both physical movement and behavior changes that bring people in and out of the population at risk. It can vary enormously between populations, and is much larger than the natural death rate often used in modeling studies. We demonstrate that both epidemic models are very sensitive to this seemly simple and often neglected parameter.

Most modeling studies of sexually transmitted diseases recognize the partner acquisition rate as one of the most sensitive parameters. The common assumption that the probability of transmission is directly proportional to the number of partners is usually inappropriate when analyzing the impact of reducing the number of partners on the course of an epidemic. The probability of transmission per partner is a function of the probability of transmission per contact *and* the number of contacts per partner. Because the number of contacts per partner tends to be smaller when the partner acquisition rate is larger, the probability of transmission per partner decreases as the number of partners per unit time increases. This decrease will be most dramatic when the probability of transmission per contact is very small, as it is for HIV, and reduces the sensitivity of the epidemic to the partner acquisition rate and has obvious implications for control programs designed to slow the epidemic by encouraging people to have fewer partners. We conclude that, for HIV spread in high-risk populations, the epidemic is as sensitive to the average number of contacts per partner as it is to the number of partners per unit time.

2 The DI and SP Model Formulations

The DI model is shown on the left in Figure 1. Individuals enter a specific group when they become infected and stay in that group until they are no longer transmitting the disease. Infectivity and progression rates depend upon the group. In contrast, in the SP model, shown on the right in Figure 1, all infected individuals enter the same group, and then pass serially through a series of groups before leaving the sexually active population due to illness or other factors. Their infectivity also depends on their group. In [9] we motivated both models and derived explicit formulas for the reproductive numbers and the endemic steady states, including the fraction of infections being caused by each group at equilibrium.

Equations for the DI model are

$$\begin{aligned} \frac{dS}{dt} &= \mu(S^0 - S) - \lambda S, \\ \frac{dI_i}{dt} &= p_i \lambda S - (\mu + \nu_i) I_i, \qquad i = 1, \cdots, n, \\ \frac{dA}{dt} &= \sum_{j=1}^{n} \nu_j I_j - \delta A, \end{aligned} \tag{2.1}$$

where $\sum_{i=1}^{n} p_i = 1$.

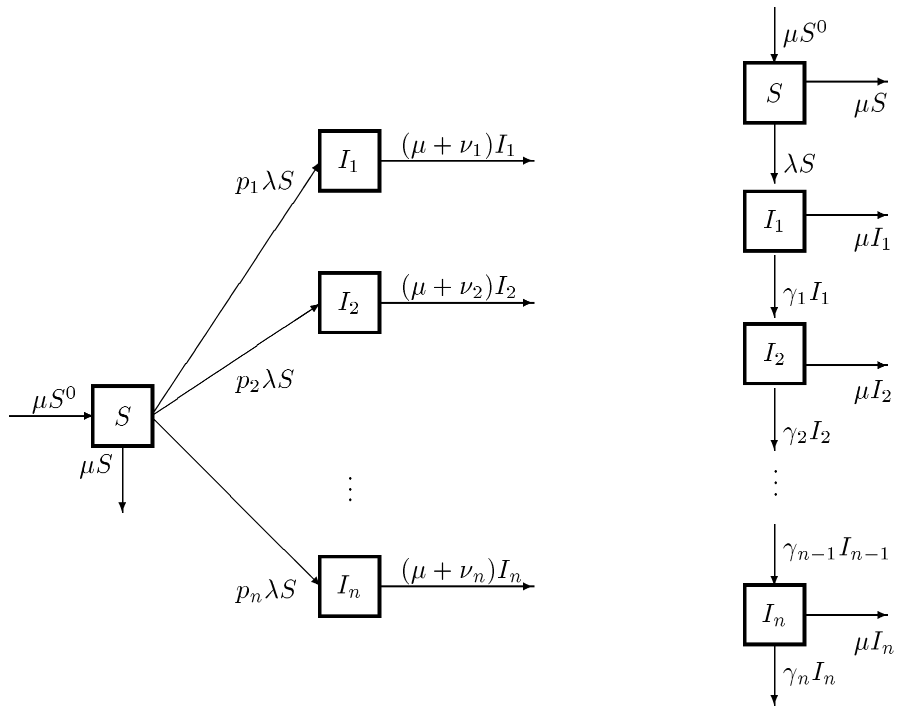

Figure 1: The DI model, shown on the left, divides the infected population into groups according to their infectiousness and differences in rates of developing AIDS. Infected people stay in the same group the whole time they are in the population. In the SP model, shown on the right, every infected individual goes through the same series of stages. This allows us to account for a short early highly infectious stage equivalent to the acute phase of infection, a middle period of low infectiousness, and a late chronic stage with higher infectiousness.

The SP model equations are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S^0 - S) - \lambda S, \\
\frac{dI_1}{dt} &= \lambda S - (\gamma_1 + \mu) I_1, \\
\frac{dI_i}{dt} &= \gamma_{i-1} I_{i-1} - (\gamma_i + \mu) I_i, \qquad 2 \le i \le n, \\
\frac{dA}{dt} &= \gamma_n I_n - \delta A.
\end{aligned} \tag{2.2}$$

In both of these models the susceptible population is assumed to be homogeneous, and variations in susceptibility, risk behavior, and many other factors associated with the dynamics of HIV spread are neglected. This simplicity allows us to isolate and examine the differences between the mechanisms captured by the two models. We also assume that the population we are studying is an isolated, high-risk, subset of a larger population. The larger embedding population is relatively free of HIV and provides a constant source of uninfected individuals entering the high-risk population at a rate μS^0. These models are appropriate for a homosexual population of a major city, or for a group of highly active heterosexual individuals, but might not be appropriate for populations where there is a substantial level of contacts between high-risk groups and lower-risk groups, or where the virus is spreading primarily between people with fairly long-term relationships.

When no virus is present in the population the population of susceptible individuals, S, has a constant steady state S^0. This equilibrium is maintained by the constant inflow of individuals plus a constant per person rate of outflow, in which each individual remains in the population an average of μ^{-1} years, where μ is the total removal rate. In the presence of infection, individuals are infected by HIV at a per capita rate $\lambda(t)$.

For the DI model, the infected population is subdivided into n subgroups, I_1, I_2, $\cdots$, I_n. Upon infection, an individual enters subgroup i with probability p_i and stays in that group until becoming inactive in transmission. The rate, ν_i, of leaving a subgroup depends on behavior changes induced by HIV-related illnesses, a positive HIV test, or other factors. This model accounts for the time-independent differences in viral load between individuals and the differences in rates of developing AIDS that individuals with different viral loads have. (See [9] for references.)

For the SP model we subdivide the infected population into subgroups I_1, I_2, $\cdots$, I_n with different infection stages. Infected individuals enter the first subgroup I_1 and then gradually progress from subgroup I_1 to subgroup I_n. We define γ_i to be the average rate of progression from subgroup i to subgroup $i+1$, for $i = 1, \cdots, n-1$, and γ_n as the rate at which infected individuals in subgroup I_n become sexually inactive or uninfectious due to end-stage disease or behavior changes. This model captures the time-dependence of viral loads within each individual and the distribution of times to AIDS seen in populations (see [9] for references).

The rate of infection, λ, depends upon the transmission probability per partner of individuals in subgroup i, β_i, the proportion of individuals in the subgroup, I_i/N,

Table 1: Reproductive number R_0, mean duration of infection in group i, $\bar{\tau}_i$, mean duration of infection for the whole population $\bar{\tau}$, mean transmission probability $\bar{\beta}$, equilibrium infection rate λ^*, susceptible population S^*, equilibrium infected group population I_i^*, equilibrium total infected population I_T^*, and equilibrium relative impact ρ_i^* for both models.

Name	DI Model	SP Model	Name	DI Model	SP Model
R_0	$r\bar{\tau}\bar{\beta}$	$r\bar{\tau}\bar{\beta}$	S^*	$\dfrac{\mu S^0}{\mu+\lambda^*}$	$\dfrac{\mu S^0}{\mu+\lambda^*}$
$\bar{\tau}_i$	$\dfrac{1}{\mu+\nu_i}$	$\dfrac{1}{\mu+\gamma_i}$	I_i^*	$p_i\bar{\tau}_i S^*\lambda^*$	$q_i\bar{\tau}_i S^*\lambda^*$
$\bar{\tau}$	$\sum_{i=1}^{n} p_i\bar{\tau}_i$	$\sum_{i=1}^{n} q_i\bar{\tau}_i$	I_T^*	$S^*(R_0-1)$	$S^*(R_0-1)$
$\bar{\beta}$	$\sum_{i=1}^{n} p_i\beta_i\bar{\tau}_i/\bar{\tau}$	$\sum_{i=1}^{n} q_i\beta_i\bar{\tau}_i/\bar{\tau}$	λ^*	$\dfrac{R_0-1}{\bar{\tau}}$	$\dfrac{R_0-1}{\bar{\tau}}$
q_i	undefined	$\prod_{j=1}^{i-1}\gamma_j\bar{\tau}_j$	ρ_i^*	$\dfrac{p_i\beta_i\bar{\tau}_i}{\bar{\beta}\bar{\tau}}$	$\dfrac{q_i\beta_i\bar{\tau}_i}{\bar{\beta}\bar{\tau}}$

where $N = S + I$, and $I = \sum\limits_{j=1}^{n} I_j$, and the number of partners of an individual per unit time, r. For both models, a simple random mixing assumption leads to

$$\lambda(t) = \sum_{i=1}^{n}\lambda_i(t), \quad \lambda_i(t) = r\beta_i\frac{I_i(t)}{N(t)}. \tag{2.3}$$

For both models we denote the subgroup of removed people by A. People in A are assumed to die at a rate $\delta \geq \mu$.

The *relative impact* of the group is defined by the relative fraction of individuals being infected by each group:

$$\rho_i(t) = \frac{\lambda_i(t)}{\lambda(t)} = \frac{\beta_i I_i(t)}{\sum\limits_{j=1}^{n}\beta_j I_j(t)}. \tag{2.4}$$

Analyzing the stability of the infection-free equilibrium ($S = S^0, I_i = 0$) gives the epidemic threshold condition, which defines the reproductive number as

$$R_0 = r\bar{\tau}\bar{\beta} \tag{2.5}$$

for both models. The mean duration of infection, $\bar{\tau}$, and the mean probability of transmission per partner, $\bar{\beta}$, are defined in Table 1 for each model. In each case, if $R_0 < 1$, the infection-free equilibrium is the only equilibrium and is locally asymptotically stable. If $R_0 > 1$, the infection-free equilibrium is unstable, a small initial infection will spread until the population converges to a unique endemic equilibrium ($S = S^* > 0, I_i = I_i^* > 0$), given in Table 1. (See [8, 9] for details.)

Table 2: Variable parameters: In the sensitivity studies these quantities are varied to determine the impact that changes in these parameters have on the solution of the models.

Description	**Formula**	**Baseline Value**	**Range**
Sexually active removal rate	α	0.05 yrs^{-1}	(0.02, 0.2) yrs^{-1}
Mean duration of infection (when $\alpha = 0$ in the DI model)	$\bar{\tau}$	12 years	(8.6, 19) yrs
Partner acquisition rate	r	5 partners/yr	(2, 10) partners/yr
Contacts per partner parameter	η	1.0	(0.8, 1.4)

3 Parameter Ranges

Tables 2 – 4 show the baseline parameters that we use in the studies in this paper and the ranges for those that we vary. Many of these parameters are the same as those in [9]. Here we give a brief review of that information and more detailed discussions on some parameters, in particular the number of contacts per partner.

3.1 Parameters Common to Both Models

3.1.1 Total Removal Rate

The total removal rate, μ, is the sum of the natural death rate, d, and the sexually active removal rate, α, which is the rate individuals leave the high-risk population due to both physical migration and changes in sexual behavior. For the natural death

Table 3: Fixed Parameters: These parameters were chosen based on the studies and calculations cited in the text, and do not change as the parameters in Table 2 are varied.

Description	**Formula**	**Baseline Value**
Initial Conditions		
Initial population size	$N(0)$	S^0
Initial infected population	$I_T(0)$	$0.01S^0$
Normalized infection-free equilibrium	S^0	1
Basic parameters		
Natural death rate	d	0.02 yrs^{-1}
Mean duration of infection when $\alpha = 0$ in the DI model	$\bar{\tau}$	12 years
DI parameters		
Distribution of the newly infected	$\boldsymbol{p}$	(0.05, 0.33, 0.5, 0.12)
Progression rates by group	$\boldsymbol{\nu}$	(0.19, 0.096, 0.058, 0.028)
Relative per contact transmission	$\boldsymbol{\zeta}$	$(10^3, 10^2, 10, 1)z^D$
Infectivity adjustment factor	z^D	5.1×10^{-5}
SP parameters:		
Progression rates by group	$\boldsymbol{\gamma}$	$(13.0, 0.23553, 0.23553, 0.47)$
Relative per contact transmission	$\boldsymbol{\zeta}$	$(100, 1, 1, 10)z^S$
Infectivity adjustment factor	z^S	9.08×10^{-4}

rate we assume that individuals in the population are young adults and can expect to live an average of 50 more years, so $d = 0.02$ yrs^{-1}.

The average number of years that people engage in high-risk behavior is unknown, and probably varies greatly between populations. We assume a baseline of 20 years ($\alpha = 0.05$ yrs^{-1}), and study how changing α^{-1} from 5 to 50 years affects our model results. This parameter is often neglected in models for the spread of sexually transmitted diseases, but, as we shall see, can play a key role in determining how fast the HIV virus spreads. Note that increasing α decreases the mean time that an individual spends in the sexually active population in different ways for each model (see the formulas for $\bar{\tau}$ in Table 1). This means that, even when the two models have the same reproductive numbers at one value of α, if everything else is held fixed, and α is changed the same amount in both models, they will no longer have the same R_0.

Table 4: Derived Parameters at Baseline: These parameters are derived from the parameters given in Tables 2 and 3.

Description	**Formula**	**Baseline Value**
Duration of Infection	$\bar{\tau}$	7.3 yrs
Mean probability of transmission per contact	$\bar{\zeta}$	0.003
Number of contacts per partner	$c(r=5)$	21.8 contacts per partner
DI parameters		
Probability of transmission per partner	$\boldsymbol{\beta}$	(0.68, 0.105, 0.011, 0.0011)
Mean probability of transmission per partner	$\bar{\beta}$	0.053
Reproductive number	R_0	1.93
SP parameters:		
Probability of transmission per contact	$\boldsymbol{\beta}$	(0.87, 0.0196, 0.0196, 0.1802)
Mean probability of transmission per contact	$\bar{\beta}$	0.051
Reproductive number	R_0	1.88

3.1.2 Mean Duration of Infection

From [14–16], the mean time from infection to AIDS lies between 8.6 and 19 years for people in developed countries and people undergoing treatment therapies. All of these studies keep people in the study regardless of their sexual behavior, so their data is only applicable when $\alpha = 0$. We take 12 years as the baseline value for $\bar{\tau}$ when $\alpha = 0$, in the DI model.

3.1.3 Partner Acquisition Rate and Number of Contacts Per Partner

While HIV can spread even in populations with fairly low partner acquisition rates, r, (for example, in African populations in which HIV has spread, reported partner acquisition rates may be as low as 1 or 2 per year [3], and a study in Bangkok found many of the infected women had had only one sex partner in their lives [23]), our model does not take long-term partnerships into account, and thus is designed to study high risk populations where most partnerships are fairly short-term, and the impact of spread to long-term partners can and is being neglected. Such a population might be highly-active populations of homosexual men or heterosexuals, or communities of prostitutes and the clientele which visit them regularly. In the latter case, we would be neglecting spread from the clientele to their wives and girlfriends, and spread from these women to other men. This spread outside the active core group would have to be calculated separately as a consequence of the core spread. Early studies of high-risk homosexual men found averages of between 5 and 25 sex partners per year (see [7] for references and analysis), but today a more typical high-risk population might average far fewer new partners per year. We take the baseline value of r to be 5 partners per year, and investigate the sensitivity of our model to values of r ranging from 2 to 10 partners per year.

The parameter r enters the model both as a multiplicative factor and through the dependence of the transmission probabilities per partner, β_i, on the average number of contacts per partner, c, which in turn depends on the number of contacts per partner ($c = c(r)$).

If ζ_i is the transmission probability per contact in group i, the probability that a susceptible individual will not be infected by a single contact with an infected individual is $1 - \zeta_i$. Hence the probability that a susceptible individual will avoid infection when they have $c(r)$ contacts with an infected partner is $(1 - \zeta_i)^{c(r)}$ [1], and the probability that a susceptible individual will fail to avoid infection when he/she has an infected partner is

$$\beta_i = 1 - (1 - \zeta_i)^{c(r)}. \tag{3.1}$$

This is the probability of transmission per partner from an infected person in group i. Note that Eqn. (3.1) is an improvement over the approach used to estimate β_i in our previous paper [9], and will change our results somewhat by decreasing the importance of the most infectious groups relative to the other groups.

Sexual behavior surveys indicate that people in long-term relationships have an average of 1-3 contacts per week with their partner and that people with hundreds of partners per year rarely have multiple contacts with a typical partner. Thus the average number of contacts with each partner would be a decreasing function of the number of partners per year.

Our functional choice for $c(r)$ is somewhat arbitrary because of the lack of sociological data on this issue. A simple decreasing function of the partner acquisition rate that gives approximately two contacts per week for people with one partner per year, and decreases to about one contact per partner as r gets large is

$$c(r) = 104r^{-\eta} + 1, \tag{3.2}$$

where η is a parameter which controls how fast this function decreases.

We take $\eta = 1$ at baseline, which means that for small values of r people have approximately two contacts per week (e.g., if people have one new partner every two years, then $r = 0.5$, and with $\eta = 1$ we have that $c(0.5) = 209$ total contacts) and for $r = 20$ partners/yr they are averaging slightly more than 6 contacts with each partner, and averaging about 120 contacts per year. On the other hand, if $\eta < 1$ then people in relationships which last longer than a year have fewer than 2 contacts per week with their partner, and people with more than one partner per year have more contacts with them than in the $\eta = 1$ case. If $\eta > 1$, people with less than one new partner per year average more than 2 per week, and when $r > 1$ people have fewer contacts with each partner than when $\eta = 1$. This dependence of $c(r)$ on r and η is shown in Section 5.2, Figure 6, where at baseline $c(5) = 21.8$ contacts per partner.

3.2 Transmission Probability

In [9] we used a single parameter $\bar{\zeta}$ for the mean transmission probability per sexual contact, and determined the population's average probability of transmission, $\bar{\beta}$, using the formula $1-(1-\bar{\zeta})^{c(r)}$. The β_i for each infected population group was proportional to the relative viral load of its group, and was varied in such a way to keep $\bar{\beta}$ fixed.

Recognizing that it is the transmission probability per sexual contact for people in group i, ζ_i, which is proportional to a person's viral load, rather than the actual probability of transmission per partner, here we determine the β_i differently. We fix

$\bar{\zeta}$ and determine the ζ_i such that they give this value for $\bar{\zeta}$. Then the probability of transmission per infected partner from group i is given by Eq. (3.1), and the β_i are given from the ζ_i by formula (3.1).

The mean transmission probability per contact $\bar{\zeta}$ for the DI model is

$$\bar{\zeta} = \sum_{i=1}^{n} p_i \frac{\bar{\tau}_i}{\bar{\tau}} \zeta_i, \tag{3.3}$$

and for the SP model it is

$$\bar{\zeta} = \sum_{i=1}^{n} q_i \frac{\bar{\tau}_i}{\bar{\tau}} \zeta_i. \tag{3.4}$$

The data which we present below gives the average viral levels for each group. The actual infectiousness per contact which goes along with that viral level is unknown, but we assume that the relative infectiousness goes up the same amount as the viral load does. This allows us to determine the ζ_i except for a multiplicative constant which is the same for every ζ_i in the model. This multiplicative constant is then determined from formulas (3.3) and (3.4) by specifying $\bar{\zeta}$.

Estimates of $\bar{\zeta}$ range from 0.0003 (lowest value estimated for female-to-male transmission) to 0.08 (highest value estimated for male-to-male transmission) [21]. In [9] we showed that the reproductive number and transient dynamics of the epidemic are very sensitive to ζ. We found that both models are so sensitive to ζ that it only made sense to study a small fraction of its range. At the bottom of the range, $R_0 \ll 1$ and the epidemic is below threshold. At the top of ζ's range, the epidemic is rapid and devastating ($R_0 \gg 1$). Here $\bar{\zeta} = 0.003$ at baseline.

3.3 Other Parameters

The remaining parameters for the DI model were obtained from the HIV progression study reported in [16], following which we divide the infected population into four groups, ranging from the highest viral load to the lowest. We assume a connection between infectivity and viral load, so that, like the average level of viral particles in the serum, each group's infectivity decreases by a factor of 10 from the one before it. We use values from this paper to obtain

$$\mathbf{p} = (0.05, 0.33, 0.5, 0.12)^T, \quad \boldsymbol{\zeta}^D = (1000, 100, 10, 1)^T z^D,$$

where the scalar parameter z^D is to be chosen so that the mean probability of transmission per contact is 0.003. Here bold letters are used to denote the vector parameters and the vectors are defined using matrix transpose notation.

In [9] we showed that $\bar{\tau}$ is equal to 12 years when $\alpha = 0$ when $\boldsymbol{\nu} = (0.19, 0.096, 0.058, 0.028)^T$, where the relative values of the progression rates were obtained from [16]. When we account for changes in sexual activity, the mean value for $\bar{\tau}$ at baseline becomes 7.3 years.

From equation (3.3), the above parameter choices give $\bar{\zeta} = 59.0 z^D$. By specifying $\bar{\zeta} = 0.003$, $z^D = 5.1 \times 10^{-5}$. We then hold this value for z^D fixed as we vary other parameters, as we assume that the probability of transmission per contact is unaffected by those other parameters. $\bar{\zeta}$ thus becomes a derived parameter, shown in Table 4.

Here we choose the duration of each SP stage to ensure that both models have the same value for $\bar{\tau}$ at baseline and to fit what is known about them. The initial highly infectious stage lasts 4 weeks, the final moderately infectious stage lasts 2.13 years, and the middle two stages each last 4.25 years: $\gamma_1 = 13$ yrs^{-1}, $\gamma_2 = \gamma_3 = 0.2353$ yrs^{-1}, $\gamma_4 = 0.47$ yrs^{-1}. Setting α to be its baseline value of 0.05 yrs^{-1} gives $\bar{\tau} = 7.3$ years.

Given the data on viral loads [12, 17, 19], we assume for the SP model that

$$\boldsymbol{\zeta}^S = (100, 1, 1, 10)^T z^S.$$

From equation (3.4), this gives $\bar{\zeta} = 3.31 z^S$. Specifying $\bar{\zeta} = 0.003$ then allows us to calculate that $z^S = 9.08 \times 10^{-4}$. This gives the ζ_i for the SP model, which we hold fixed for all calculations in this paper. $\bar{\zeta}$ then becomes a derived parameter for some of our analysis. Table 3 shows the parameters which do not vary in any simulation, and Table 4 shows the parameters values at baseline for all derived parameters.

4 Initialization Procedures for the Models

An added complication in any model with infected subgroups is the dependence of the behavior and the timing of the transient solutions of the epidemic model on the initial distribution of the infected populations among the subgroups. We will demonstrate that, even when the total number of infected individuals is fixed, the timing of the epidemic can be shifted by up to 25 years by varying the distribution of the initial infected population. This important observation is often overlooked in simulation studies comparing multigroup models. We develop a robust, systematic procedure for determining the initial distribution of the infected population when $R_0 > 1$ based on the progression of a natural epidemic that sets the timing of different multigroup models and allows them to be quantitatively compared.

Let $f_i = I_i(0)/I(0)$. The total initial infected population, $I(0) = \sum_{i=1}^n I_i(0)$, is fixed (in our case, we take $I(0) = 0.01$) and the goal is to prescribe a robust procedure to define the fractions, f_i, where $\sum_{i=1}^n f_i = 1$.

4.1 The Initialization Methods

One of the simplest ways to define these fractions is to use the *equilibrium fraction initialization procedure* (EFIP). In this approach the relative fractions are defined based on the distribution of the infected groups at the endemic equilibrium. That is, $f_i = I_i^*/I^*$, where I_i^* and I^* are given in Table 1. Using the formulas in Table 1, we obtain $f_i = p_i\bar{\tau}_i/\bar{\tau}$ for the DI model and $f_i = q_i\bar{\tau}_i/\bar{\tau}$ for the SP model. This method is simple to implement and effective when R_0 is near one. However, in other cases this initialization can generate a rapid initial transient in the solution.

For models of the spread of infectious agents such as HIV, which have been spreading into populations that were originally free of this disease and thus have an initial background level of zero infections, we propose two procedures for defining the f_i when $R_0 > 1$ based on the idea that in nature the epidemic would have spread into the population from only a few infected individuals. The epidemic then grows from this small seed until by the time a noticeable number of individuals are infected, a natural balance between the infected subgroups is reached (thus defining f_i). When $R_0 > 1$, then the *numerical preinitialization procedure* (NPP) is a simple approach to approximating the natural balance in the population that exists when an epidemic started in the past. In this procedure we distribute the initial infected population based on what it would be if a very small initial infected population was introduced in the distant past and grew to infect a given percentage of the population.

For example, to use the NPP simulation to distribute a 1% initial infection rate, the preinitialization simulation is started with a much smaller infected population, say 0.01% of the population infected, and the equations are integrated until 1% of its population has become infected. This occurs at some time t^P (which depends on how we initialized this preinitialization simulation). At time t^P we stop the simulation and use the current distribution of the infected population in the preinitialization simulation to define the relative fraction of the infected population in each group for the actual simulation that we study. Thus $f_i = I_i^P(t^P)/I^P(t^P)$, where the superscript P denotes the solution and ending time of the preinitialization simulation. In numerical experiments we observed that the resulting initial conditions are almost independent of the distribution used in the preinitialization simulation.

Even though the numerical preinitialization procedure is robust for our two models in that it is insensitive to how the initial infected population is distributed, it still contains heuristic parameters and can be complicated to implement numerically. We now define an approach, that we call the *natural initialization procedure* (NIP), to define the distribution of the initial infected population, based on the instantaneous balance between the infected groups at $t = 0$.

Suppose that we move our solution forward in time only a short period of time t_ϵ and then use this new distribution which the equations have naturally found to redefine f_i. Since t_ϵ is small, we can linearize this solution about $t = 0$. To do this, we define functions $g_i(t) = I_i(t)/I(t)$, where the constraint $\sum_{i=1}^{n} g_i(t) = 1$ for all t holds by definition. If our first guess for f_i is f_i^1, our first set of solutions to the model we are studying gives $g_i^1(t)$, with the initial condition $g_i^1(0) = f_i^1$. Linearizing gives

$$g_i^1(t) \approx f_i^1 + t\dot{g}_i^1(0) := h_i^1(t), \tag{4.1}$$

where $\dot{g}_i^1(0) = \frac{d}{dt} g_i^1(t)\Big|_{t=0}$, and this quantity is found by substituting the initial conditions $I(0) = I_0$, $S(0) = S_0$, $I_i(0) = f_i^1 I(0)$.

We then use this approximation at a small and arbitrarily chosen value $t = t_\epsilon$ to define a new initial condition, $f_i^2 = h_i^1(t_\epsilon)$, and repeat this procedure, finding a solution $g_i^2(t)$ with initial conditions $I(0) = I_0$, $S(0) = S_0$, $I_i(0) = f_i^2 I_0$, approximating the solution for small times by $h_i^2(t) = f_i^2 + t\dot{g}_i^2(0)$, and obtaining $f_i^3 = h_i^3(t_\epsilon)$. Continuing in this manner with the I_0 and S_0 remaining the same every time, we have an iteration scheme for f_i

$$f_i^{j+1} = f_i^j + t_\epsilon \dot{g}_i^j(0). \tag{4.2}$$

Now suppose that the sequence $\{f_i^j\}$ converges to f_i^* and the solution for the $I_i(t)$ which has f_i^* as its initial condition is I_i^*. Then $\lim_{j\to\infty} \dot{g}_i^j(0) = 0$. Since

$$\dot{g}_i^j(t) = \frac{d}{dt}\left(\frac{I_i^j(t)}{I^j(t)}\right) = \frac{\dot{I}_i^j(t)\cdot I^j(t) - \dot{I}^j(t)\cdot I_i^j(t)}{I^j(t)^2},$$

$$\dot{I}_i^*(0) I^*(0) = \dot{I}^*(0) I_i^*(0). \tag{4.3}$$

This implies that f_i^* satisfies the nonlinear equation

$$f_i^* = \frac{I_i^*(0)}{I^*(0)} = \frac{\dot{I}_i^*(0)}{\dot{I}^*(0)}. \tag{4.4}$$

The right hand side of this equation, $\dot{I}_i^*(0)/\dot{I}^*(0)$, is a function which depends on the model equations, the choice of initial conditions for I_0 and S_0, as well as the f_i^*.

We solve the nonlinear equations (4.4) for f_i^* by a simple function iteration. Let f_i^{*k} denote the kth approximation to f_i^*, and $I_i^k(0) = f_i^{*k} I(0)$. Then

$$f_i^{*k+1} = \theta f_i^{*k} + (1-\theta)\frac{\dot{I}_i^k(0)}{\dot{I}^k(0)}$$

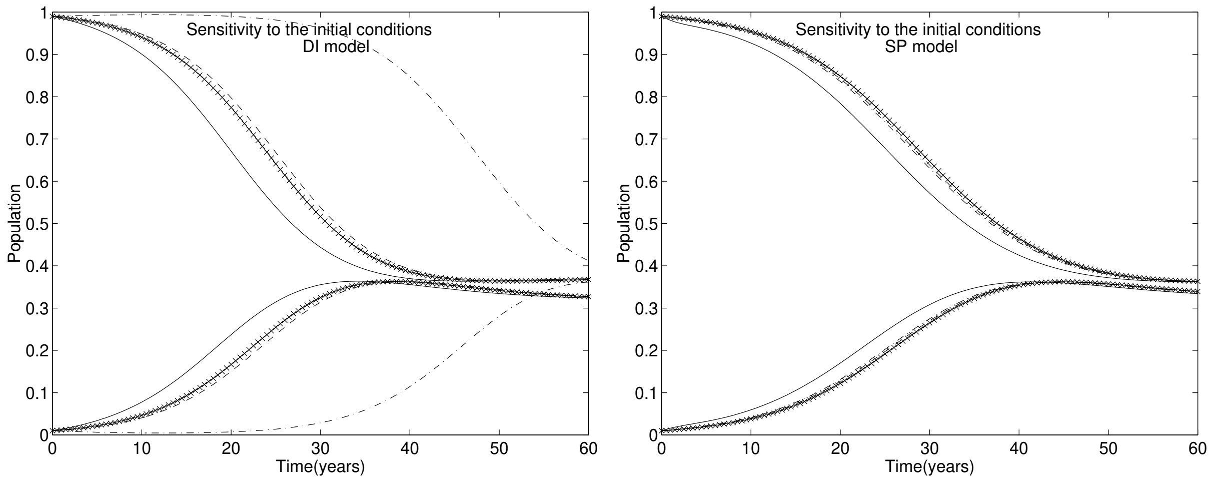

Figure 2: Plot of the susceptible and infected populations with different initial distributions. Solid line: all initially in the group with the largest ρ_i^*. Dash-dot: all initially in the group with the smallest ρ_i^*. Dashed: EFIP method. Crossed lines: NPP and NIP methods. Even a small infected population can hasten or delay the onset of the epidemic in the DI model by over 25 years. Surprisingly the onset of the epidemic in the SP model for the same conditions varied by only 4 years.

where $\theta \in (0, 1)$ is a relaxation parameter. In all cases we have tested this scheme converges to a root of (4.4) when θ is chosen sufficiently close to 1.

To monitor the convergence of the iteration, we noted that equation (4.3) implies that the relative growth rates of all the subgroups

$$\frac{\dot{I}_i^*(0)}{I_i^*(0)} = \frac{\dot{I}^*(0)}{I^*(0)}$$

are equal at convergence. We found this to be an excellent convergence test when the integration converged slowly.

4.2 Comparisons of the Different Initializations

In Figures 2 and 3 we compare the results from four different ways of distributing the initial infected populations. These figures demonstrate that different initial distributions of even a small infected population can hasten or delay the onset of the epidemic in the DI model by over 25 years. Surprisingly, the onset of the epidemic in the SP model for the same conditions varied by only 4 years.

In all of these examples we use the baseline parameters given in Tables 2 – 3, and begin the simulation at $t = 0$ with 1% of the population infected, 99% susceptible

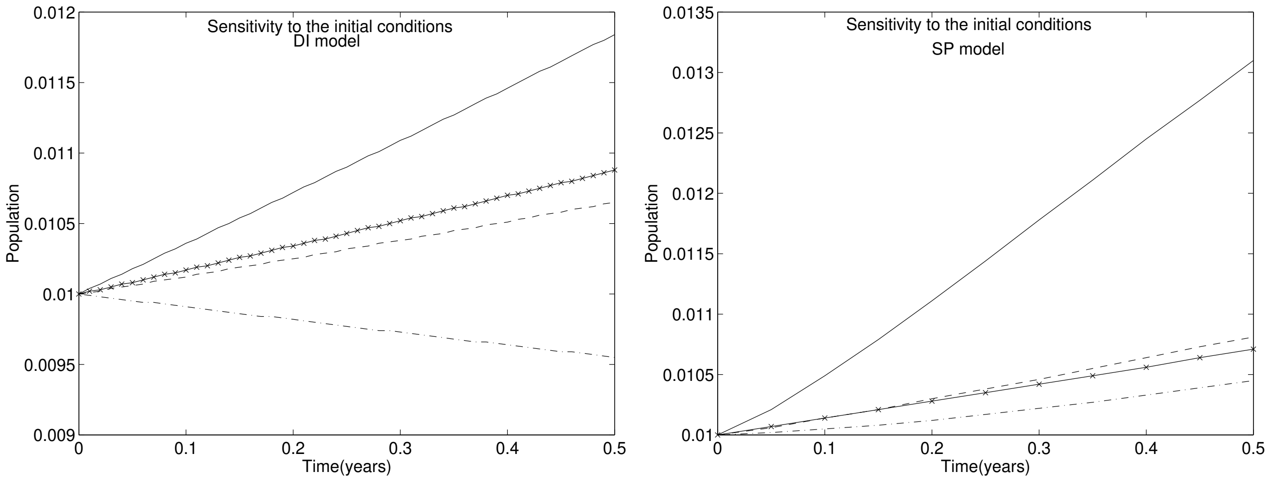

Figure 3: Plot of the infected populations for the first 6 months with different initial distributions. Solid line: all initially in the group with the largest ρ_i^*. Dash-dot: all initially in the group with the smallest ρ_i^*. Dashed: EFIP method. Crossed lines: NPP method.

and a total population of $S^0 = N(0) = 1$ ($I(0) = 0.01$, $S(0) = 0.99$). We investigate the impact of distributing the initial population with one of the following methods:

1. All of the infections are initially in the group with the largest relative impact, ρ_{max}, at equilibrium. For this fast starting situation the parameters are $\mathbf{f} = (0, 1, 0, 0)$ for the DI model and $\mathbf{f} = (0, 0, 0, 1)$ for the SP model.

2. All of the infections are initially in the group with the smallest relative impact, ρ_{min}, at equilibrium. For this slowly starting situation the parameters are $\mathbf{f} = (0, 0, 0, 1)$ for the DI model and $\mathbf{f} = (0, 0, 1, 0)$ for the SP model.

3. The EFIP method.

4. The NPP method, with EFIP initialization used for the preinitialization. (For this example the NIP and NPP solutions agreed within a fraction of 1%.)

The variability in the initial progression of the epidemic shown in Figures 2 and 3 demonstrates how important it is to make a well-reasoned choice for the initial distribution of the infected population when studying models with multiple infection groupings. These simulations were all run with identical parameters and the same total initial infected population of 1%. Figures 2 and 3 indicate that using the EFIP the epidemic preceded the NPP and NIP solutions by less that a year, but we caution against assuming that this result will hold for other models, or even for these two models when $R_0 \gg 1$.

To analyze the impact of the initialization procedure on the timing of the epidemic, we first focus on the initial growth rates. These initial growth rates are the slopes of the solution curves at $t = 0$ in Figures 2 and 3. Initial growth for epidemic models which are systems of ODEs was discussed in [11], and is well-known to depend on the initialization of the model. Note that even though all parameters and the total number of initial infected people are the same in each simulation, because the f_i are not the same the initial slopes can be different. For the DI model,

$$\left.\frac{dI(t)}{dt}\right|_{t=0} = \left(r \sum_i^n \beta_i f_i \frac{S(0)}{N(0)} - \mu - \sum_i^n \nu_i f_i \right) I(0).$$

and for the SP model

$$\left.\frac{dI(t)}{dt}\right|_{t=0} = \left(r \frac{S(0)}{N(0)} \sum_{i=1}^n \beta_i f_i(0) - \mu - \gamma_n f_n(0) \right) I(0).$$

The initial slopes for the three numerical simulations where f_i are defined explicitly are given in Table 5. From this table we see that there are huge differences in initial growth rate, depending upon how the infected population is initially distributed between the different groups. If the early infections were in the most infectious subgroup, the epidemic can take off quickly and get a head start over the situation where all the infections are in the least infectious group. In the DI model, the simulations indicate that this head start is maintained, but in the SP model, much of it is quickly lost.

The initialization f_i's for the NPP depend on the preinitialization method used. In practice we found that the f_i's obtained by using different preinitialization schemes differed from each other and the NIP method by less than 0.0004. Thus for this example, the NPP method is very insensitive to the distribution of the infected population in the preinitialization simulation. Even though the EFIP is significantly different from the other methods for this example, as R_0 approaches one the f_i in all of these methods converge to the same values.

Table 5: The initial slopes can differ depending upon the initial distribution of the infected population. In fact, the slope for the DI model when the initial 1% infected population is all in the least infectious group (group 4, ρ_{min}) is initially negative before the epidemic finally takes off. Note that the f_i for the NPP and NIP methods are extremely close, but significantly different than the EFIP method.

Model	Method	**f**	$\dot{I}(0)$
DI	ρ_{max}	(0.0000, 1.0000, 0.0000, 0.0000)	0.00337
DI	ρ_{min}	(0.0000, 0.0000, 0.0000, 1.0000)	-0.00093
DI	EFIP	(0.0263, 0.2719, 0.5343, 0.1675)	0.00116
DI	NPP	(0.0357, 0.3037, 0.5210, 0.1396)	0.00158
DI	NIP	(0.0358, 0.3039, 0.5209, 0.1394)	0.00158
SP	ρ_{max}	(0.0000, 0.0000, 0.0000, 1.0000)	0.00355
SP	ρ_{min}	(0.0000, 0.0000, 1.0000, 0.0000)	0.00026
SP	EFIP	(0.0104, 0.4847, 0.3635, 0.1414)	0.00115
SP	NPP	(0.0188, 0.5904, 0.2982, 0.0926)	0.00135
SP	NIP	(0.0189, 0.5900, 0.2983, 0.0928)	0.00135

Because of this extreme sensitivity to the initial distributions, we have demonstrated that one has to be careful drawing conclusions between the timing of the epidemic of two different models when there are multiple subgroups. In particular, when comparing the SP and DI models, the initial distributions of the infected population may be as important in the course of the epidemics as the differences between the models. The NIP eliminates this problem, and we use it for all of the remaining simulations in this paper.

5 Sensitivity Studies

In [9] we presented some baseline simulations for these models, and showed that the size and speed of the epidemic is sensitive to the choice of $\bar{\zeta}$. In the SP model the relative impacts, ρ_i, early in the epidemic are also sensitive, with group 1 causing more and more of the early infections as $\bar{\zeta}$ is increased. In this section we further explore the sensitivity of the DI and SP models to variations in α, r and η. We use the baseline parameters from Tables 2–4 unless otherwise stated. First we examine the dependence of the solutions on the sexually active removal rate of people in and out of the susceptible population, and then we show how the relationship between the average number of partners per unit time and the number of contacts per partner in a homogeneous randomly mixing population can affect the epidemic.

5.1 Sensitivity to the Sexually Active Removal Rate

Often sexually-transmitted disease epidemic models neglect the impact of people moving in and out of the sexually active population. In our model this would be equivalent to assuming that the sexually active removal rate, α, is 0. However, it is likely that in many populations the average person does not maintain high risk behaviors for long periods of time. If people spend less time in the sexually active population, they have less time to either become infected or to infect, and thus the size and extent of the epidemic should decrease as α increases. Behavioral intervention programs attempt to affect α by convincing people to leave the high risk population. These intervention programs do this by attempting to convince people to either abstain from sex altogether, or else to have few partners, and to be careful to choose low risk partners, through the strategy of “know your partner”.

In this section, we hold all parameters in Tables 2 and 3 at their baseline values, except α. Thus p_i, ν_i, and γ_i are held fixed, but $\bar{\tau}$ varies, decreasing as α increases. We also assume that the probability of transmission per contact in each group is not affected by behavior changes, so that the ζ_i and β_i of both models do not vary. However, the equations for the mean transmission probabilities imply that $\bar{\zeta}$ and $\bar{\beta}$ change as α varies, and because there are so many ways in which α comes into play in their expressions, it is not possible to predict how they depend on α. Since the reproductive number depends on both $\bar{\tau}$ and $\bar{\beta}$, it is not possible to predict if it will increase or decrease with α. Figure 4 shows how $\bar{\tau}$, $\bar{\beta}$ and R_0 vary with α for the two models. Notice that $\bar{\beta}$ changes as a function of α differently for the two different models, being monotonically increasing in the DI model, and nonmonotonic in the SP model. Both R_0 and $\bar{\tau}$ are decreasing and concave upward.

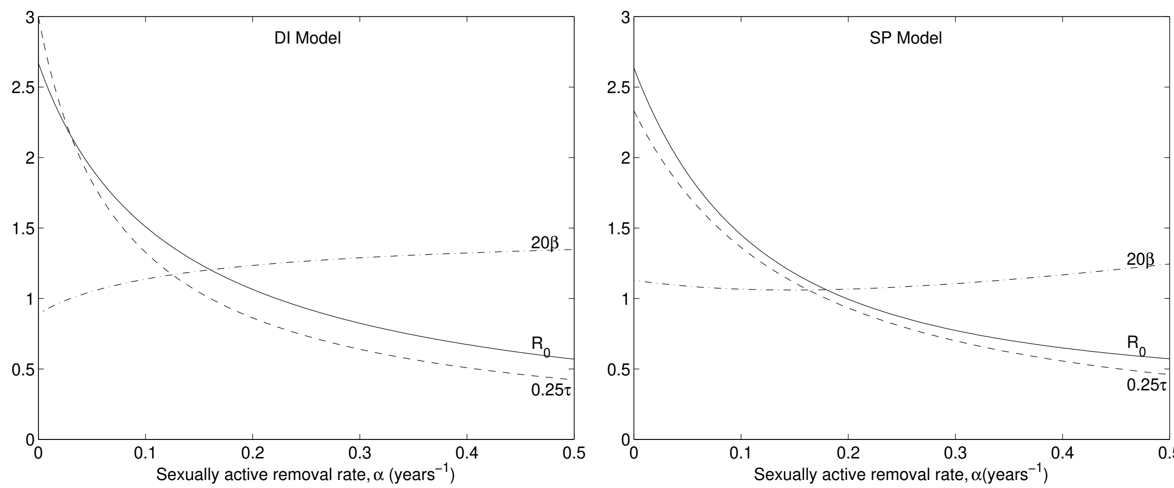

Figure 4: The reproductive number R_0 as a function of α is shown by solid lines for the DI model on the left and the SP model on the right. The dashed lines show the mean duration of infection, $\bar{\tau}$, divided by four and dash-dot lines show $20\bar{\beta}$. Note that $\bar{\tau}$ decreases more than R_0 because $\bar{\beta}$ is not held fixed.

In Figure 5 we show the difference between having a mean time in the sexually active population of 10 years versus 20 and 50 years. As expected, the epidemic becomes more severe and dramatic as α decreases, and people stay in the population for longer and longer periods of time. There are fewer new, uninfected people coming into the population to replenish the susceptibles, and those who are in the population have a greater chance of becoming infected and infecting others before they leave it.

For the DI model, group 2 always causes the most spread, with group 1 causing the second largest number of infections. Thus the relative importance of each group in spreading the epidemic changes little with the removal rate varying, although the group with the shortest life expectancy and highest infectiousness, group 1, becomes slightly more important as the removal rate increases.

In the SP model the initial values of ρ_i are insensitive to changes in α, but α does influence which groups cause the most infections once the epidemic is established. Groups 1 and 4 cause most of the infections at large α during the entire epidemic, with group 4 causing slightly more than group 1, especially as time goes on, and group 2 causing nearly one quarter of the infections throughout the whole epidemic. At small α group 1 is only important in the early epidemic, eventually becoming the

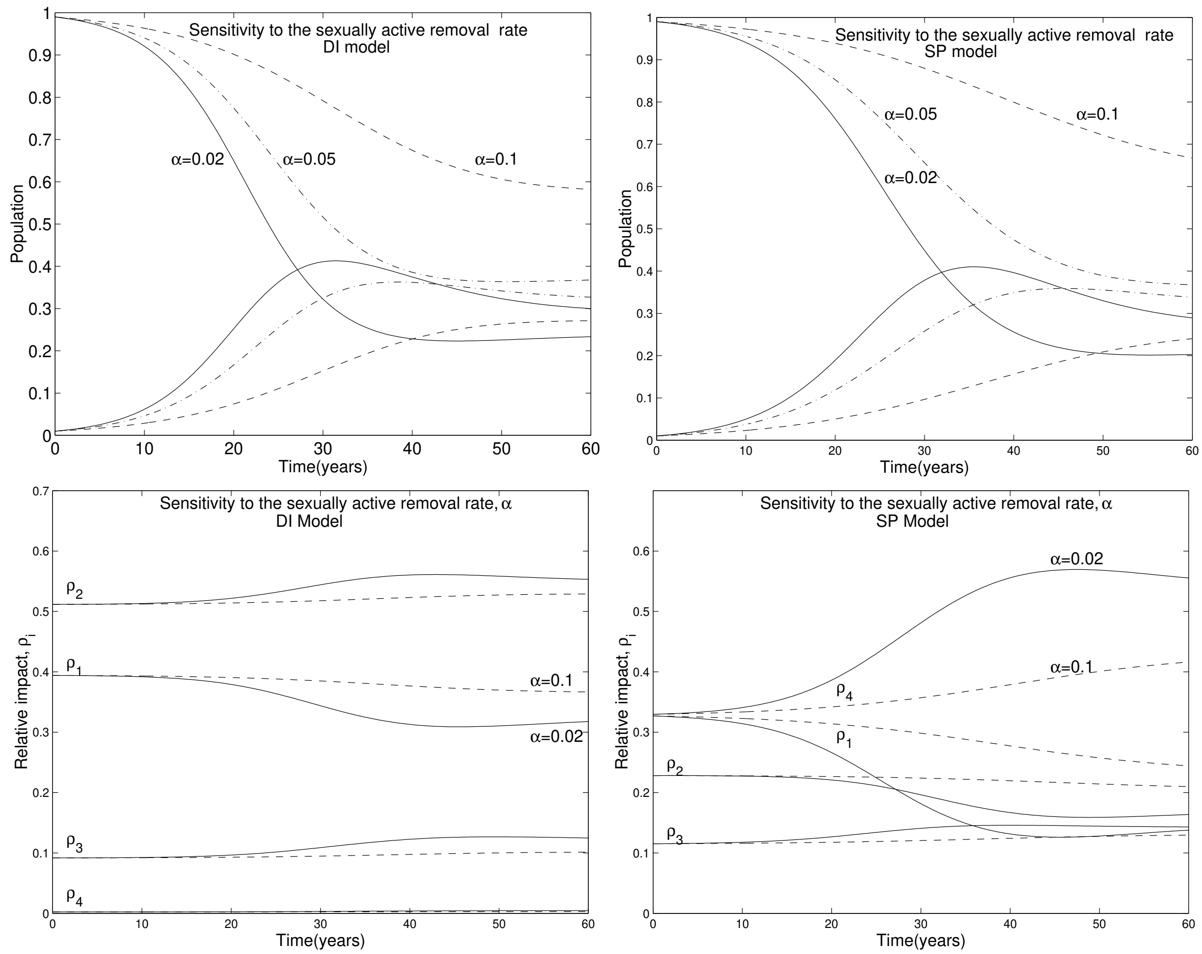

Figure 5: The epidemic spreads faster as α increases and all other parameters are held fixed. Solid lines are $\alpha = 0.02$ yrs^{-1}, dash-dot lines are $\alpha = 0.05$ yrs^{-1}, and dash-dash lines are $\alpha = 0.1$ yrs^{-1}. For the infection-rate plots we leave off the middle, baseline, case of $\alpha = 0.05$ yrs^{-1} for readability. We see that the speed and intensity of the epidemic are dramatically changed by changing α for both models. Note particularly the large difference between assuming a mean stay of 10 years and a mean stay of 20 years. In the DI model the relative impact of the groups is not affected much by α, but the SP model shows a dramatic change in who is infecting people once the epidemic is mature. Although it doesn't change the ρ_i in the early epidemic, the longer people stay in this high-risk population, the less important those who are recently infected (groups 1 and 2) become to the long-term spread, and the more important those in stage 4 become. In fact, with $\alpha = 0.02$ yrs^{-1}, over 50% of the spread after 40 years is caused by group 4.

least important group, and group 4 comes to dominate the spread. Since group 4 is the group most likely to be identified due to the duration of their infection, this would imply that this more severe epidemic might be controlled more easily by screening programs than the less severe epidemic in a less stable population.

5.2 Sensitivity to the Partner Acquisition Rate and the Number of Contacts per Partner

Campaigns against the spread of HIV generally target two other aspects of the spread besides the time spent in the sexually active population: the infectiousness of each contact (for example, through encouraging condom use or needle sterilization) and the partner acquisition rate. In [9], we showed that the epidemic was extremely sensitive to changes in the infectiousness of individual contacts. This is true for both models and implies that condom use, strain variations, and even differences in mucosal health can have a big impact on how fast HIV spreads, even when the differences in these factors between populations are small. Here we examine the sensitivity of HIV spread to the other target of intervention programs, the partner acquisition rate.

As we discussed in Section 3, when people have fewer partners they will most likely have more contacts with each partner. This assumption decreases the sensitivity of our models to the partner acquisition rate, compared to models which assume a fixed probability of transmission per partner independent of r, because our assumptions about $c(r)$ cause the probability of transmission per partner to decrease as the number of partners increases.

The impact of this decrease in the number of partners for three different choices of η in a population with a single infected group ($n = 1$) that has $\zeta = 0.003$ is shown in Figure 6. This figure also shows how our choice of $c(r)$ affects the reproductive number in this homogeneous population as r varies. The reproductive number for this single group model is $R_0(r) = r\bar{\tau}\left(1 - (1-\zeta)^{104r^{-\eta}+1}\right)$.

These plots show that it is not only possible for the reproductive number to increase very slowly once $r > 1$, as in the $\eta = 1.0$ case, but R_0 could even be nonmonotonic in r. In the $\eta = 1.2$ case the reproductive number increases with r for

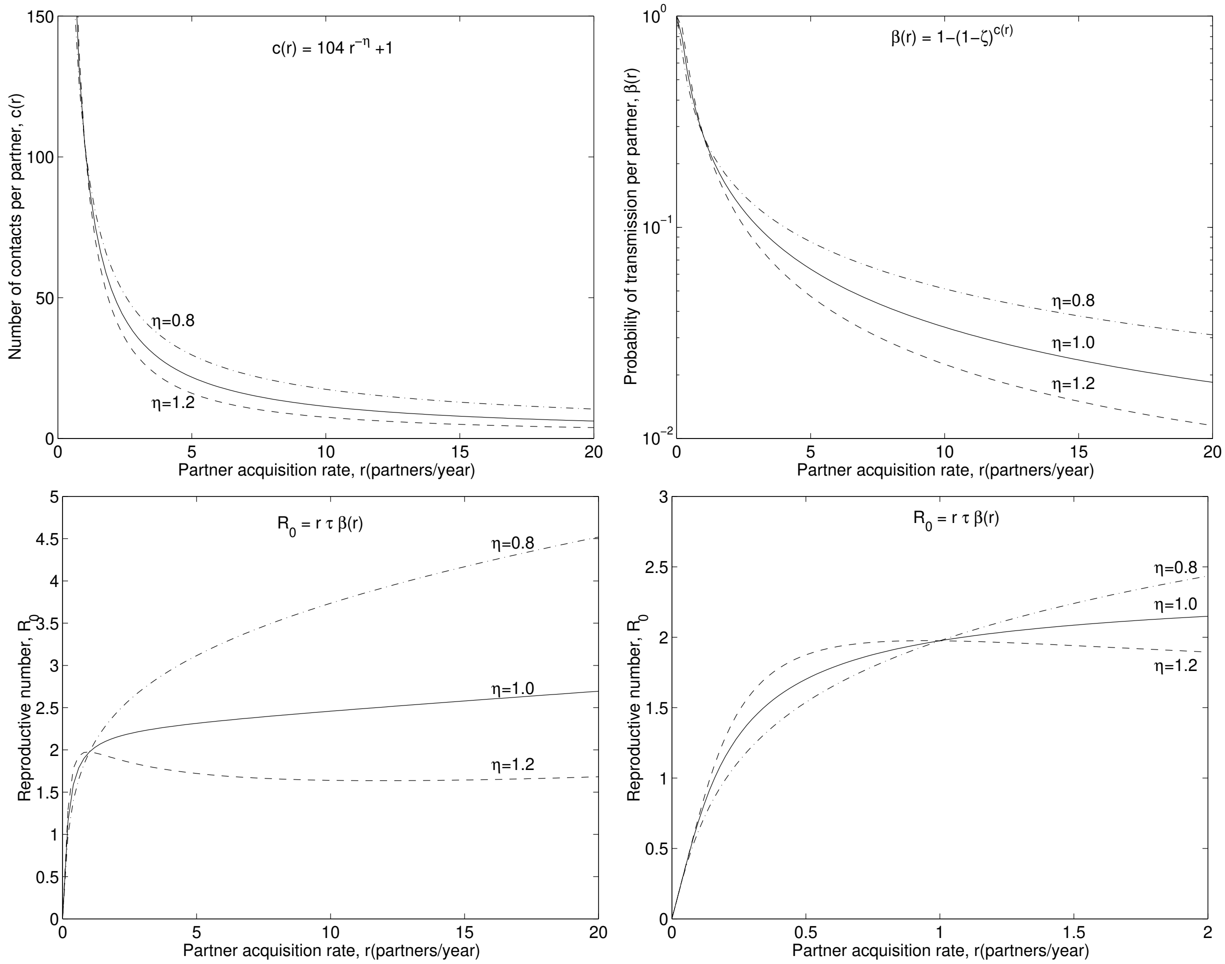

Figure 6: The functionality of $c(r)$ affects the behavior of the epidemic. The first plot shows $c(r) = 104r^{-\eta} + 1$ for three different choices of η, and the other three plots show the impact of these three choices of η on the probability of transmission per partner and the reproductive number. In each plot dash-dot is $\eta = 0.8$, the solid line is the baseline choice of $\eta = 1.0$ and dash-dash is $\eta = 1.2$. In all cases we have taken $\zeta = 0.003$. Although $c(r)$ and $\beta(r)$ are similar for the three cases, the reproductive number, for which we have used the DI baseline value of $\bar{\tau} = 7.3$ yrs^{-1}, varies enormously between the 3 choices. In order to see the region where R_0 is nonmonotonic with $\eta = 1.2$, we have magnified that region in the 4th plot.

$r < 0.95$ partners/year, but then it decreases for r between 0.95 and 11.9 partners/yr, after which it increases again. In the Appendix we show that R_0 can decrease as r increases only when $\eta > 1$ and $\zeta < 1 - (1/\eta)\exp((\eta - 1)/\eta)$. This implies that, if the number of contacts per partner drops rapidly enough as the number of partners per unit time increases, the epidemic may spread more easily in a population where people have few partners (small r) than in one with moderate values of r.

In Figure 7 we explore this possibility in more detail, and demonstrate that there is a small drop in R_0 between this local maximum and minimum. This small drop in R_0 can be significant if there are some values of τ, ζ and η for which the epidemic is above threshold when r is in the range of 1-2 partners per year and below threshold as r increases. This situation is illustrated when $\eta = 1.2$ and $\zeta = 0.0017$ in Figure 8. A more common situation is when R_0 is insensitive to changes in r and reducing the number of partners has only minimal effect on R_0. In these situations control programs which rely solely upon decreasing the number of partners might not be effective if as people decrease their partner acquisition rate they also increase the number of contacts they have with each partner.

Given the importance of the implication that convincing people to reduce their partner acquisition rates might not be an effective way to control the epidemic, it is important to determine if $\eta > 1$ is a realistic assumption. Looking at the graphs of of $c(r)$ for various values of r and η, it is difficult to pick the values out, so here are some special cases:

- if $\eta = 1$, $c(2) = 53$, $c(5) = 22$, $c(10) = 11$ and $c(20) = 6$;
- if $\eta = 0.8$, $c(2) = 61$, $c(5) = 30$, $c(10) = 17$ and $c(20) = 10$;
- if $\eta = 1.2$, $c(2) = 46$, $c(5) = 16$, $c(10) = 8$ and $c(20) = 4$.

Although we do not have data to directly support our parameter choices, it seems more reasonable to us that a person with 5 new partners a year averages 16 or fewer contacts with each partner ($\eta \geq 1.2$) than that they average more than 16 contacts with each partner ($\eta < 1.2$). This implies that many high-risk populations will be in

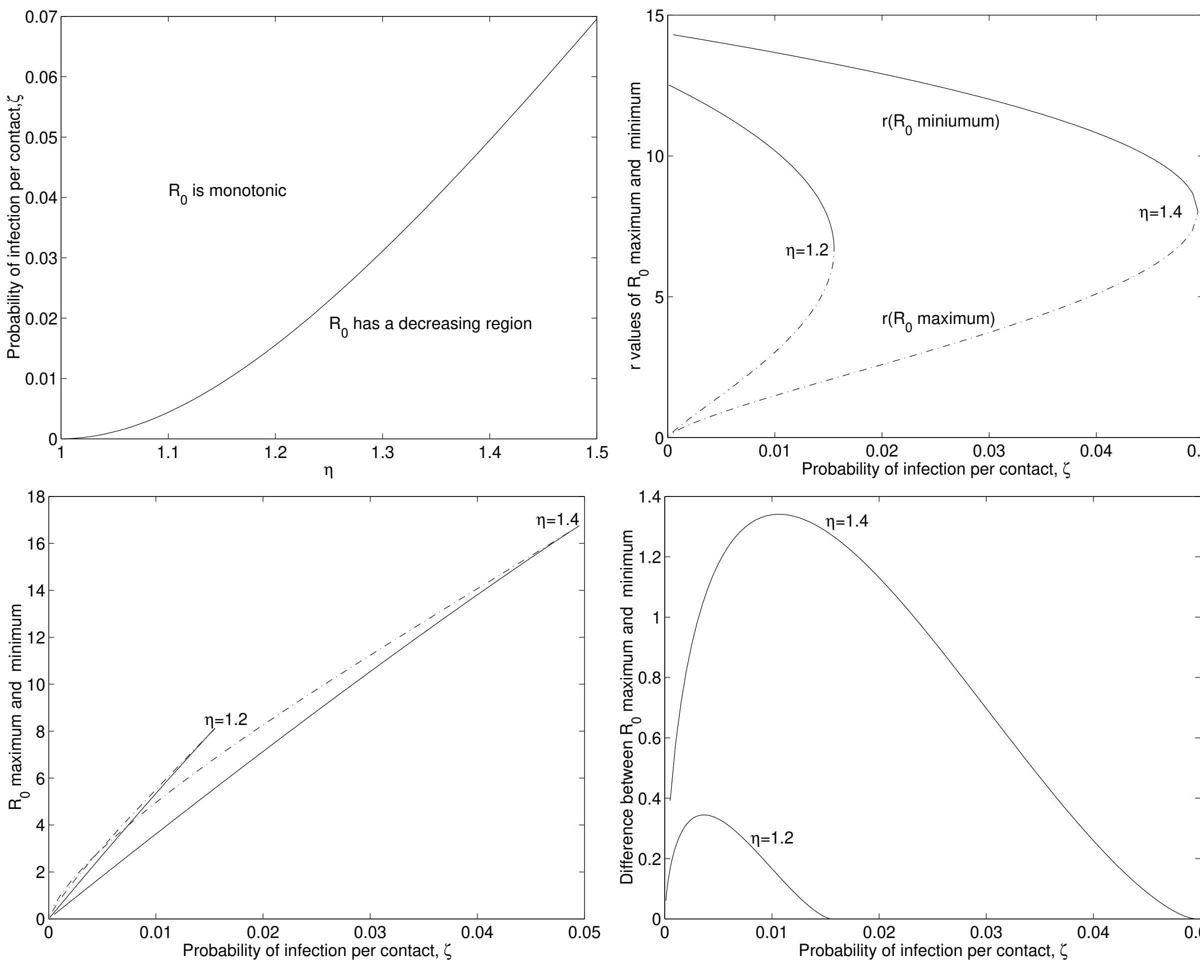

Figure 7: This figure explores the parameter ranges for which R_0 has a region in which it decreases as r increases. In the first plot, we see the parameter values of η and ζ for which this occurs: all values below the curve. The second plot shows the r values for which the maximum and minimum in R_0 happen as a function of ζ for $\eta = 1.2$ and 1.4. Note how much of the range of r these curves encompass. We see that the curve in the first plot is defined by the places where $r_{\pm}$ converge to each other. The bottom left plot then gives the maximum and minimum values of R_0 as ζ varies for the two values of η and $\bar{\tau} = 7.3$yrs. We see the differences between these values in the last plot. These bottom two plots indicate that R_0 varies only over a small range for these cases.

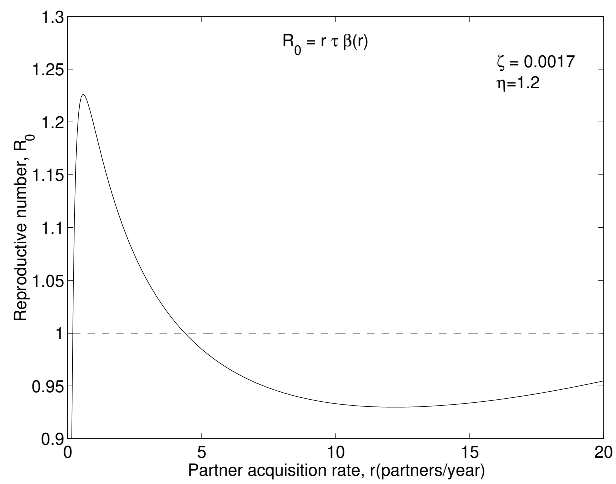

Figure 8: The reproductive number can be a nonmonotonic function of the partner acquisition rate. Here $\eta = 1.2$, $\zeta = 0.0017$, and everything else is the same as in Figure 6. For this case the epidemic spread at small r but dies out when r is greater than 5 partners per year.

parameter ranges where reducing partner acquisition rates is not the most effective way to control the epidemic, unless they can be reduced to very low levels.

These analytical results only hold when there is a single infected group, and change in the 4 group models, each of which has a different ζ_i and a β_i which varies with r and η. Numerically we observe that the $\bar{\beta}$ for each model varies somewhat less with η than in the homogeneous case, and the reproductive number is monotonic in r for $\eta = 0.8$, 1.0 and 1.2 when all other parameters are held at baseline. However, if η is increased to 1.4, R_0 is no longer monotonic. Also, at $\eta = 1.2$, the reproductive number changes very slowly with r once r is greater than 4, while at $\eta = 1.0$ it is increasing more rapidly than for the homogeneous case. Thus it takes a greater functional decrease in contacts per partner with r to get the effects seen in a homogeneous population. These results are shown in Figure 9 for the DI model, and the results for the SP model are nearly identical.

Note, however, that even when the reproductive number is insensitive to changes in r, it is possible that other aspects of the dynamics are sensitive. We examine this by taking the two cases $\eta = 1.0$ and $\eta = 1.2$. For the $\eta = 1.2$ case the reproductive number is monotonic in r, but changing fairly slowly for $r > 1$. We hold all other

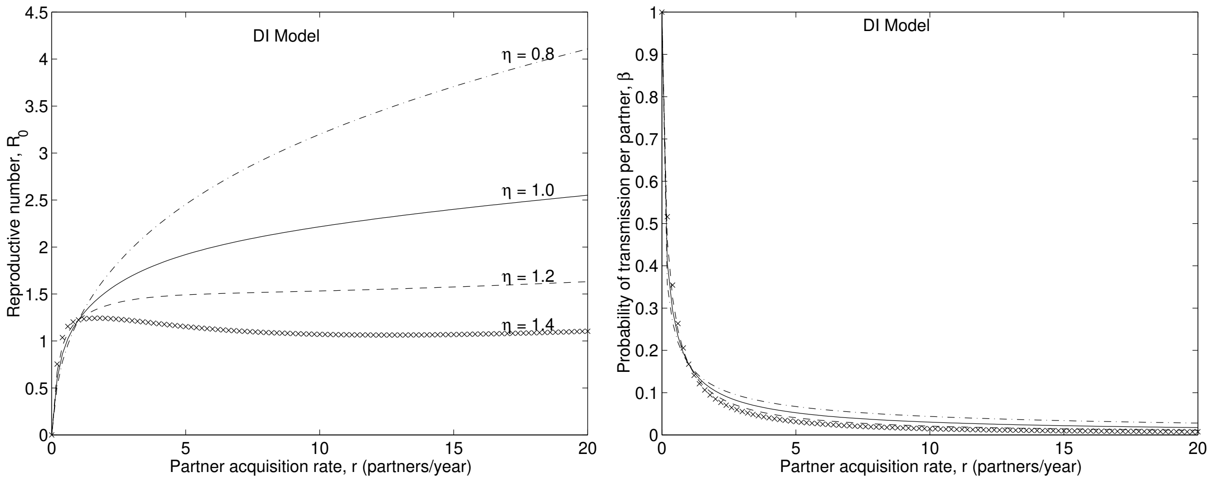

Figure 9: The sensitivity of the DI model's reproductive number and probability of transmission per contact to the partner acquisition rate and the functional form of the number of contacts per partner for the case we study in the paper of 4 groups. Note that for the cases where $\eta > 1$, R_0 is almost independent of r for $r > 1$.

parameters at baseline and vary r. Our numerical simulation results are shown in Figures 10 and 11 for both models. We studied the three cases of $r = 2$, 5, and 10 partners per year. In both models the epidemic is much faster and more severe as r increases, and the change is greatest for r going from 2 to 5. For our nonhomogeneous population, the reproductive number in the DI model goes from 1.48 at $r = 2$ partners per year to 1.86 at $r = 5$ partners per year. It reaches 2.15 at $r = 10$ partners/yr. The SP model has a similar increase in R_0. It is well-known that when the reproductive number is close to 1, even small changes in parameters can lead to a large change in epidemic behavior, and that appears to be the case here.

This dramatic change in behavior is evident in the plots of the relative impact of the groups. In the DI model, when r is small, group 2 causes over 60% of the infections throughout the entire simulation, and group 1 causes most of the rest. At the larger partner acquisition rate of 10 partners per year, group 1 becomes much more important, despite its shorter life expectancy.

For the SP model, the change is so dramatic that it is difficult to even make a clear comparison. At the larger partner acquisition rate, nearly all of the early infections are caused by the just-infecteds, while in the less active case group 4 is the main

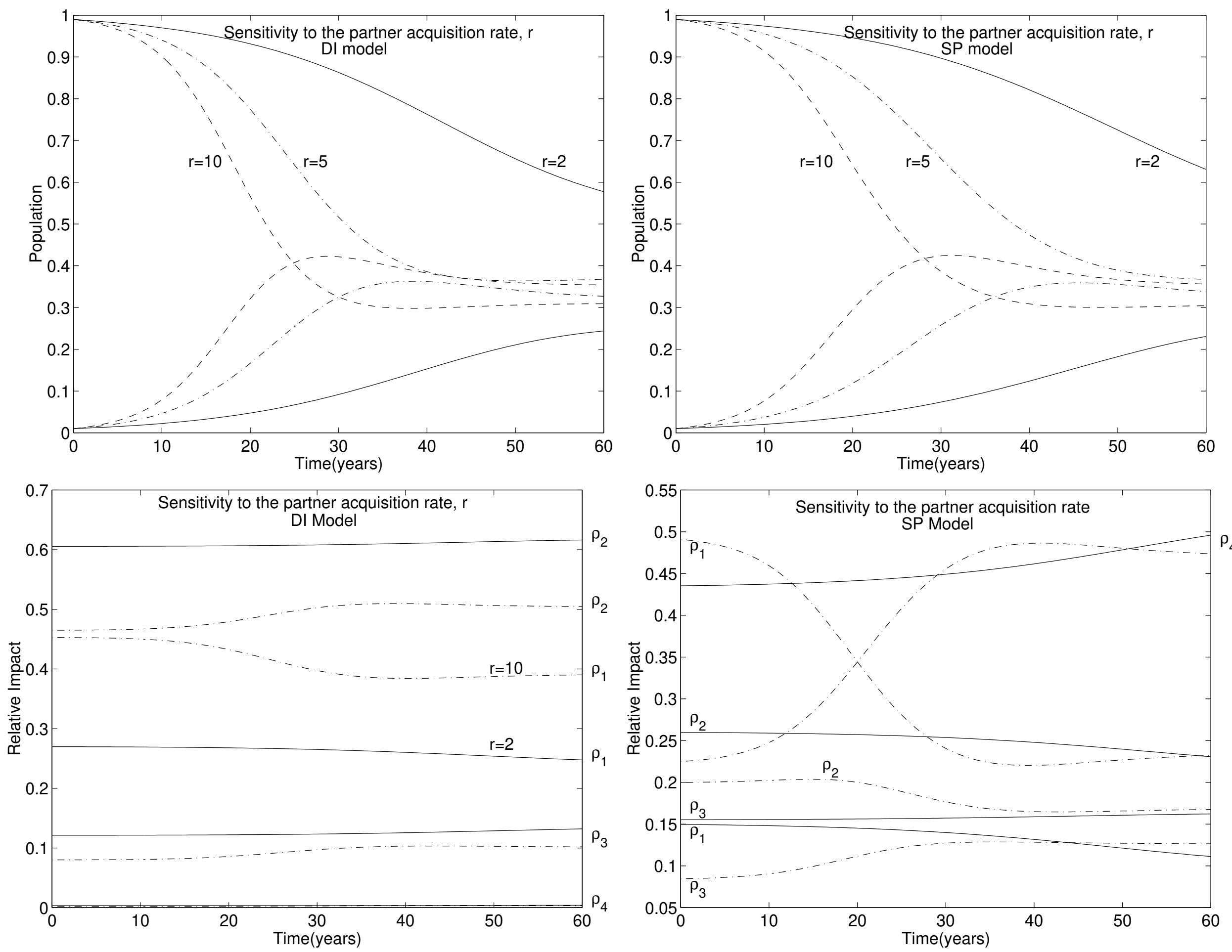

Figure 10: The sensitivity of the two models to the partner acquisition rate with all other parameters held at baseline, and $c(r) = 104r^{-1} + 1$. Solid lines: r=2; dash-dot lines: $r = 10$; dash lines: $r = 50$. ($r = 10$ case not shown in the ρ plots for readability) Note that while both models do show an increase in severity of the epidemic with increased r, these are extremely different population behaviors, and we might expect much more change than this. Note also that while the order of the relative impact of each group in the DI model remains the same (although group 1 becomes more important and group 2 less important as r increases), in the SP model the importance of group 1 to the early epidemic increases dramatically with increased partner acquisition rates. We would expect this, since people can only transmit the infection if they acquire a new partner during the very short newly-infected period.

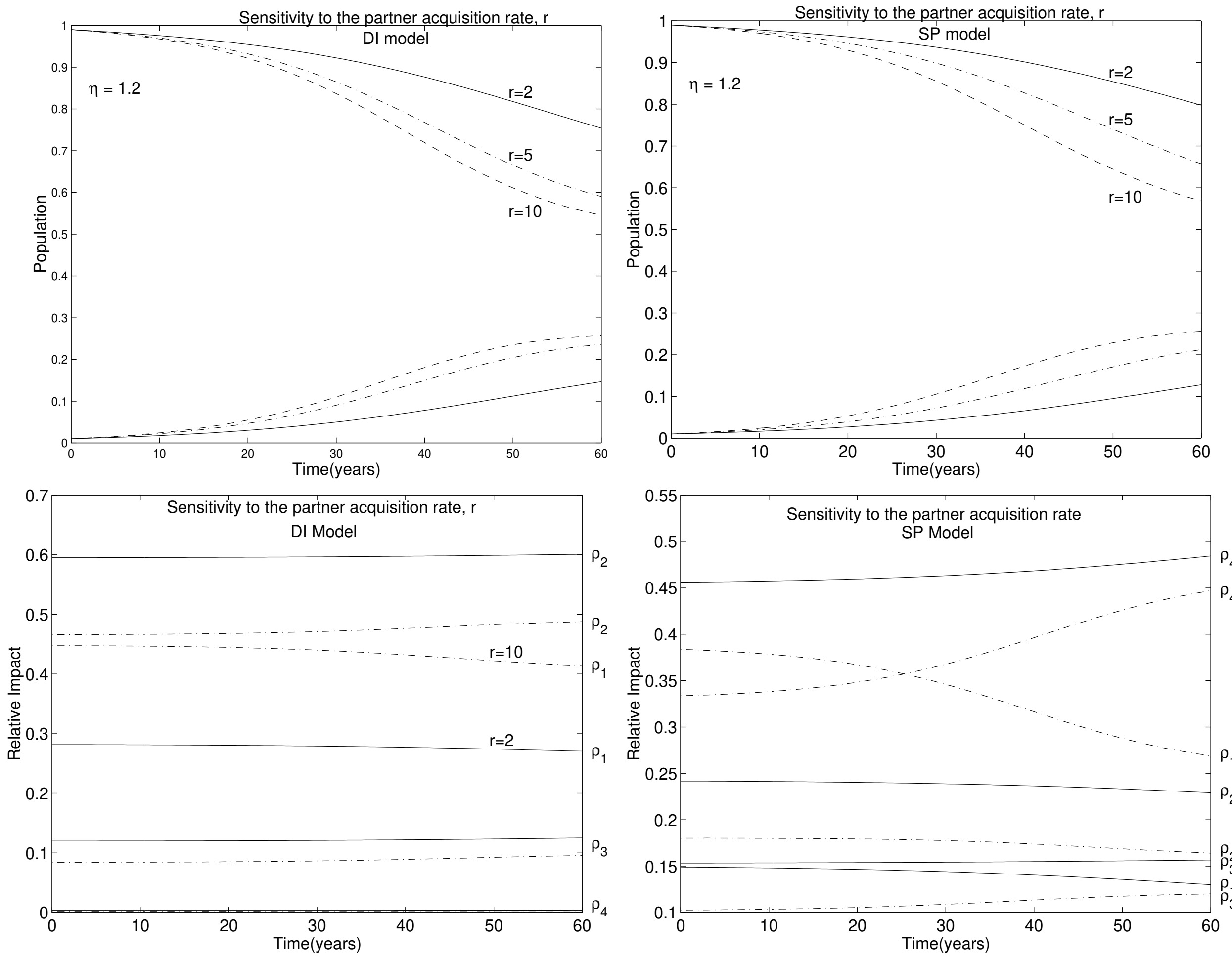

Figure 11: The sensitivity of the two models to the partner acquisition rate with all other parameters held at baseline, and $c(r) = 104r^{-1.2} + 1$. Solid lines: r=2; dash-dot lines: $r = 10$; dash lines: $r = 50$. (The $r = 10$ case is not shown in the ρ plots for readability.) Note that while both models do show an increase in severity of the epidemic with increased r, these are extremely different population behaviors, and we might expect much more change than this. Note that the changes in the relative impacts are similar to those seen in Figure 10.

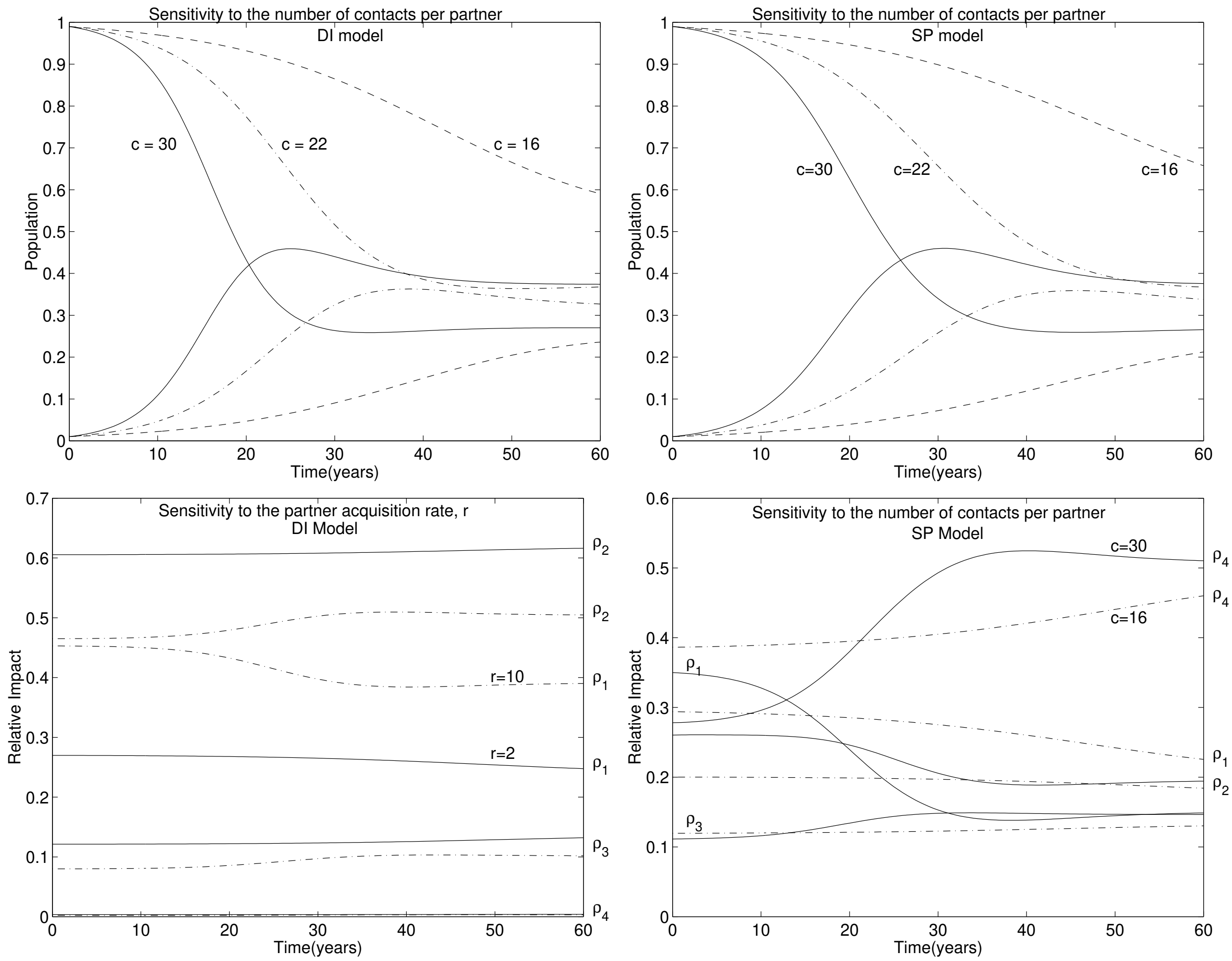

Figure 12: The sensitivity of the two models to the number of contacts per partner. Here all parameters are at baseline, except c. Solid lines are for $c = 30$ ($\eta = 0.8$), dash-dot lines for $c = 22$ ($\eta = 1.0$) and dash lines for $c = 16$ ($\eta = 1.2$) (in the ρ plots, $c = 22$ is not shown for readability). Note that the epidemic gets much more severe and rapid as c increases. The groups causing most of the infections in the DI model remain groups 1 and 2, but the interior dynamics of the SP model change dramatically. Group 4 is causing the largest fraction of the infections at the smaller value of c, and group 1 causes the largest number of early infections at the largest value. Group 2 is important, causing up to 26% of infections but 3 never causes more than 15% of infections.

transmitter throughout the entire epidemic.

The reproductive number shown in Figure 6 is sensitive to our functional choice of $c(r)$, a function about which very little is known in any population. To investigate the sensitivity of the transient dynamics of the epidemic to $c(r)$, we fixed $r = 5$ and varied the number of contacts per partner. We choose the values of c to correspond to the three different functions of $c(r, y)$ shown in Figure 6, which give $c = 16$, 22 and 30 average contacts per partner.

In Figure 12, we show that the fractions infected by different groups in the DI model change very little as c varies, but there is a big change in the fractions infected by the different groups in the SP model early in the epidemic. As the number of contacts per partner increases, the probability of transmission per partner increases, and the newly-infected individuals (group 1) play more and more of a role in the first 10 years of the epidemic, with the long term infected individuals (group 4) playing less and less of a role. With few average contacts per partner, the largest group of infections are caused by group 4, while with more average contacts per partner, the early epidemic is almost entirely driven by group 1. The fractions infected by the two middle groups are also affected, but not nearly as dramatically.

6 Summary and Conclusions

In this paper we have studied a number of important issues with the DI and SP models, and shown the sensitivity of the transmission dynamics to the following:

- Initial conditions: The early epidemic in multigroup models is extremely sensitive to the initial distribution of the infected populations among the subgroups, and this early variation may be maintained throughout the epidemic. This sensitivity can be dealt with by using the NIP or the NNP to approximate the *natural initial conditions*.
- Sexually active removal rate: As the movement of people in and out of the susceptible population increases, the epidemic slows because an infected person

is not around as long to spread the disease. This often ignored rate must be accurately accounted for before a model is used for quantitative predictions or used to help guide control efforts.

- Partner acquisition rate: By building in the assumption that people with many partners have fewer contacts per partner than people with few partners, we found that the epidemic is less sensitive to the partner acquisition rate than one might expect. Prevention campaigns which focus solely on encouraging people to have fewer partners may have minimal impact on the epidemic.

- Number of contacts per partner: Since the probability of transmission of HIV per contact is low, the epidemic is sensitive to our assumptions about the number of contacts per partner. One can argue that the epidemic is as sensitive to this quantity as it is to the number of partners per year.

Our simulations demonstrate that the early behavior of both our multigroup models can be different, depending upon how we distribute the initial population between infected groups. It is obvious that conclusions cannot be drawn from numerical simulations of models with multiple infected groups unless the initial conditions have been chosen carefully. In a real world epidemic, one or a few people would bring the infection in, and it would spread outward from there, but public health officials would not notice and begin to study the epidemic until it has reached a certain size. This concept led us to the idea of approximating the natural initial conditions.

The sexually active removal rate has been neglected in most models of the spread of sexually transmitted diseases. This is not important when studying a disease which has a short incubation period, like gonhorrhea, but its effects are large given the long duration of HIV infection. Many factors bring people in and out of a given high risk population, including the physical migration of people in and out of a population, such as the gay population in a large city, and behavior changes that occur as individuals age or find long term partners and settle down. We have examined only the possibility that infected individuals leave our population, but there is certainly also a possibility that infected individuals enter it from other populations.

We argue that when people have fewer partners, they are likely to have more contacts with each partner. Given this, lowering the partner acquisition rate does not have as large an effect on the spread of HIV as one might expect. Partner acquisition rates may have to drop to very low levels before the disease spread is stopped. The epidemic is more sensitive to the infectiousness per contact [9]. This implies that controlling the epidemic by convincing people to use condoms would have more impact than controlling it by convincing people to have fewer partners. We also showed that the epidemic is fairly sensitive to our assumptions about how many contacts people have per partner. It makes a significant difference to the spread of the epidemic if people with 5 or so partners per year are having an average of 16 contacts with each partner or an average of 22 contacts. Obviously it would make even more difference if they would only have 5 contacts with each partner. Thus attitudes and behaviors towards partners are factors which should be taken into consideration in prevention programs.

These studies raise questions about how effective contact-tracing programs are in controlling treatable sexually transmitted diseases. These programs ask infected individuals to name people with whom they have had contact and where the virus may have been transmitted during the contact. These potentially-infected people are then contacted by a trained person and encouraged to be tested. The effectiveness of the program in identifying the infected population depends upon the etiology of the underlying disease. If the underlying epidemic is close to the SP model, then contact tracing will identify most individuals only after they are past the initial most infectious stage and are no longer as likely to transmit the infection. For those situations where most infections are being transmitted by group 4, this may not be a problem, but in many situations substantial numbers of cases are being caused by group 1. If, however, the underlying epidemic is closer to the DI model, where some individuals are highly infectious during the entire course of the infection, then these superspreaders will be quickly identified by contact tracing. We will explore these issues in [10].

References

[1] Ackerman, E., L. R. Elveback, and J. P. Fox, Simulation of Infectious Disease Epidemics, Charles C Thomas Publisher, Illinois (1984).

[2] Baltimore, D., *Lessons from people with nonprogressive HIV infection*, N. Engl. J. Med., 332 (1995), 259–260.

[3] Berkeley, S. F., R. Widy-Wirski, S. I. Okware, R. Downing, M. J. Linnan, K. E. White, and S. Sempala, *Risk factors associated with HIV infection in Uganda*, J. Infect. Dis., 160 (1989) 22–29.

[4] Blower, S.M. and H. Dowlatabadi, *Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: an HIV Model, as an Example*, Int. Stat. Rev, 62 (1994), 229–243.

[5] Cao, Y, L. Qin, L. Zhang, J. Safrit, D. D. Ho, *Virologic and immunologic characterization of long-term survivors of HIV type 1 Infection*, N. Engl. J. Med., 332 (1995), 201–208.

[6] Fideli, U., S. Allen, R. Musonda, J. Meinzen-Derr, D. Decker, L. Li and G. M. Aldrovandi, *Virologic Determinants of Heterosexual Transmission in Africa*, Abstract, 7th Conference on Retroviruses and Opportunistic Infections, 2000.

[7] Hyman, J. M. and E. Ann Stanley, *Using Mathematical Models to Understand the AIDS Epidemic* Math. Biosci., 90 (1988), 415–473.

[8] Hyman, J. M. and Jia Li, *An intuitive formulation for the reproductive number for the spread of diseases in heterogeneous populations*, Math. Biosci., 167 (2000), 65–86.

[9] Hyman, J. M., Jia Li, and E. A. Stanley, *The differential infectivity and staged progression models for the transmission of HIV*, Math. Biosci., 155 (1999), 77–109.

[10] Hyman, J. M., Jia Li, and E. A. Stanley, *The impact of random screening and contact tracing in reducing the spread of HIV*, in preparation

[11] Jacquez, J. A. and C. P. Simon, *AIDS: the epidemiological significance of two different mean rates of partner change*, IMA J. Math. Med. Biol., 7 (1990), 27–32.

[12] Jacquez, J. A., J. S. Koopman, C. P. Simon, and I. M. Longini, *Role of the primary infection in epidemics of HIV infection in gay cohorts*, JAIDS 7 (1994), 1169–84.

[13] Jacquez, J. A., C. P. Simon, and J. Koopman, *Core groups and the R_0s for subgroups in heterogeneous SIS and SI models*, In: Epidemic Models: Their Structure and Relation to Data (Mollison ed.), Cambridge Univ. Press, Cambridge (1995), 279–301.

[14] Longini, I. M., W. S. Clark, M. Haber, and R. Horsburgh, *The stages of HIV infection: Waiting times and infection transmission probabilities*, In: Mathematical Approaches to AIDS Epidemiology, (Castillo-Chavez, Levin, and Shoemaker, eds.), Lecture Notes in Biomathematics Vol. 83, Springer-Verlag, New York (1989), 111–137.

[15] Mellors, J. W., A. Munoz, J. V. Giorgi, J. B. Margolick, C. J. Tassoni, P. Gupta, L. A. Kingsley, J. A. Todd, A. J. Saah, R. Detels, J. P. Phair, C. R. Rinaldo, *Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection*, Ann. Intern. Med. (1997), 126, 946–954.

[16] O'Brien, T. R., W. A. Blattner, D. Waters, M. E. Eyster, M. W. Hilgartner, A.R. Coher, N. Luban, A. Hatzakis, L. M. Aledort, P. S. Rosenberg, W. J. Milet, B. L. Kroner, J. J. Goedert, *Serum HIV-1 RNA levels and time to development of AIDS in the multicenter hemophilia cohort study*, JAMA, 276(1996), 105–110.

[17] Piatak, M., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K. C. Luk, B. H. Hahn, G. M. Shaw, and J. D. Lifson, *High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR*, Science, 259(1993), 1749–1754.

[18] Pratt, R. D., J. F. Shapiro, N. McKinney, S. Kwok, S. A. Spector, *Virologic characterization of primary HIV-1 infection in a health care worker following needlesick injury*, J. Infect. Dis., 172 (1995), 851–854.

[19] Quinn, T. C., *Acute primary HIV infection*, JAMA, 278 (1997), 58–62.

[20] Quinn, T.C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. J. Li, F. Wabwire-Mangen, M. Meehan, T. Lutalo, R. H. Gray, *Viral Load and Heterosexual Transmission of HIV Type 1*, New England Journal of Medicine, 342 (2000), 921–929.

[21] Royce, R. A., A. Sena, W. Cates, M. S. Cohen, *Sexual transmission of HIV*, N. Engl. J. Med., 336 (1997), 1072–1078.

[22] Sanchez, M. A. and S. M. Blower, *Uncertainty and Sensitivity Analysis of the Basic Reproductive Rate: Tuberculosis as an Example*, Am. J. Epidem., 145 (1997), 1127–1137.

[23] Siriwasin, W., N. Shaffer, A. Roongpisuthipong, P. Bhiraleus, P. Chinayon, C. Wasi, S. Singhanati, T. Chotpitayasunondh, S. Chearskul, W. Pokapanichwong, P. Mock, B. G. Weninger, T. D. Mastro, *HIV prevalence, risk, and partner serodiscordance among pregnant women in Bangkok*, JAMA, 280 (1998) 49–54.

[24] Wong, M. T., M. J. Dolan, E. Kozlow, R. Doe, G. P. Melcher, D. S. Burke, R. N. Boswell, M. Vahey, *Patterns of virus burder and T cell phenotype are established early and are correlated with the rate of disease progression in HIV type 1 infected persons*, J. Infect. Dis., 173 (1996), 877–887.

Appendix

A Relationship between R_0 and r

In this Appendix we examine the behavior of the reproductive number for a single-infected-group population, with $c(r)$ given by Eqn (3.2), and derive the results cited in Section 5.2. In this case, the reproductive number given by Eqn (2.5) is

$$R_0 = r\beta\tau,$$

where we drop the overbars since we have a single group. We can rewrite formula (3.1), dropping the subscripts because we have a single group and defining a new parameter $\omega = -\ln(1-\zeta)$ so that

$$\beta = 1 - (1-\zeta)^c = 1 - e^{-\omega c}.$$

Here ζ is a parameter between zero and one, and $\omega > 0$. From Eqn (3.2)

$$c(r) = 104r^{-\eta} + 1, \tag{A.1}$$

and therefore

$$R_0(r) = r(1 - e^{-\omega c(r)})\tau.$$

Note that

$$\begin{aligned} \lim_{r\to 0} c(r) &= \infty \quad \text{implies} \quad \lim_{r\to 0} R_0(r) = 0, \\ \lim_{r\to\infty} c(r) &= 1 \quad \text{implies} \quad \lim_{r\to\infty} R_0(r) = \infty, \end{aligned}$$

and that $R_0(r) \geq 0$ is finite for all finite values of $r \geq 0$. $R_0(r)$ is a continuous function, and its derivatives with respect to r are

$$\begin{aligned} R_0'(r) &= (1 - e^{-\omega c(r)} + r\omega c'(r)e^{-\omega c(r)})\tau \\ &= \left(1 - e^{-\omega c(r)}(1 + r\omega 104\eta r^{-\eta-1})\right)\tau \qquad \text{(A.2)} \\ &= \left(1 - e^{-\omega c(r)}\left(1 + \omega\eta(c(r)-1)\right)\right)\tau, \end{aligned}$$

and

$$\begin{aligned} R_0''(r) &= \left((\omega c'(r)\,(1 + \omega\eta(c(r) - 1)) - \omega\eta c'(r))\, e^{-\omega c(r)} \right) \tau \\ &= \left(\omega c'(r)\,(1 - \eta + \omega\eta(c(r) - 1))\, e^{-\omega c(r)} \right) \tau. \end{aligned} \tag{A.3}$$

As r increases from zero, then there are three possible ways that R_0 can vary:

1. R_0 is monotonically increasing in r with $R_0'(r) > 0$.

2. R_0 has one or more maxima and minima for $r > 0$. The maxima and minima have to appear in pairs, and thus there will be even numbers of extreme values where $R_0'(r) = 0$.

3. R_0 is monotonically nondecreasing in r, but has one or more points where the slope of the curve of $R(r)$ is zero. In this case, there will be r-values where $R_0'(r) = 0 = R_0''(r)$.

Since $c'(r) = -104\eta r^{-\eta-1} < 0$ for r and η positive, $R_0''(r) = 0$ if

$$1 - \eta + \omega\eta(c(r) - 1) = 0. \tag{A.4}$$

Substituting the formula (A.1) for $c(r)$ into (A.4) and rearranging the terms shows that $R_0''(r) = 0$ only when

$$\frac{\eta - 1}{\omega\eta} = 104 r^{-\eta}. \tag{A.5}$$

Eqn (A.5) has a unique positive solution

$$r^{**} = \left(\frac{104\omega\eta}{\eta - 1} \right)^{1/\eta} \tag{A.6}$$

if and only if $\eta > 1$. That is, there is at most one pair of maximum and minimun for $R_0(r)$, and they can only occur when $\eta > 1$. When $\eta \leq 1$, $R_0''(r) > 0$ and $R_0(r)$ is monotonically increasing in r.

The plots of $R_0'(r)$ for various values of η and ω indicate that there are two real roots $r = r^{\pm}(\eta, \omega)$ of $R_0'(r) = 0$ only within a particular range of these two parameters. Thus $R_0(r)$ has both a maximum and minimum only within this range. The roots $r^{\pm}$ merge along a curve in $\eta - \omega$ space, and along this curve $R_0'(r) = 0 = R_0''(r)$. Solving both of these equations simultaneously, and eliminating r, gives this curve.

Substituting (A.4) into the expression for $R_0'(r)$ and setting it to zero, we get

$$R_0'(r^{**}) = 0 = (1 - e^{-\omega-1+1/\eta}\eta)\tau = (1 - (1-\zeta)e^{-1+1/\eta}\eta)\tau. \tag{A.7}$$

Solving (A.7) for ζ then gives

$$\zeta^{**} = 1 - (1/\eta)e^{1-1/\eta}. \tag{A.8}$$

Then two roots of $R_0'(r) = 0$ exist for $\zeta < \zeta^{**}$, one root for $\zeta = \zeta^{**}$ and no roots for $\zeta > \zeta^{**}$.

The r at which this double root occurs is then given by substituting (A.8) for ζ into the expression (A.6) for r^{**} to get

$$r_{double} = \frac{104(\eta\ln(\eta) + 1 - \eta)}{\eta - 1}^{1/\eta}.$$

It is this curve which is plotted in the upper left-hand corner of Figure 7. Along this curve R_0 has a single inflection, above it R_0 is monotonic, and below it R_0 has a local maximum at a value $r = r^+$ and a local minimum at $r = r^-$, where $r^- > r^+$.

These values for $r^\pm$ are the roots of Eqn (A.2), and they depend on both η and ζ. They are solved for numerically and shown in the upper right-hand corner of Figure 7 as a function of ζ, for two values of η, where r^- is the solid line and r^+ the dashed line. Thus, when $\zeta < \zeta^{**}$, $R_0(r)$ starts out from zero at $r = 0$, increases until $r = r^+$, decreases until $r = r^-$, and then increases monotonically for $r > r^-$, going to infinity as $r \to \infty$.