DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Basic Methods for Modeling the Invasion and Spread of Contagious Diseases

Wayne M. Getz and James O. Lloyd-Smith

Abstract. The evolution of disease requires a firm understanding of heterogeneity among pathogen strains and hosts with regard to the processes of transmission, movement, recovery, and pathobiology. In this and a companion chapter (Getz et al. this volume), we focus on the question of how to model the invasion and spread of diseases in heterogeneous environments, without making an explicit link to natural selection–the topic of other chapters in this volume. We begin in this chapter by providing an overview of current methods used to model epidemics in homogeneous populations, covering continuous and discrete time formulations in both deterministic and stochastic frameworks. In particular, we introduce Kermack and McKendricks SIR (susceptible, infected, removed) formulation for the case where the removed (R) disease class is partitioned into immune (V class) and dead (D class) individuals. We also focus on transmission, contrasting mass-action and frequency-dependent formulations and results. This is followed by a presentation of various extensions including the consideration of the latent period of infection, the staging of disease classes, and the addition of vital and demographic processes. We then discuss the relative merits of continuous versus discrete time formulations to model real systems, particularly in the context of stochastic analyses. The overview is completed with a presentation of basic branching process theory as a stochastic generation-based model for the invasion of disease into populations of infinite size, with numerical extensions generalizing results to populations of finite size. In framework of branching process theory, we explore the question of minor versus major stochastic epidemics and illuminate the relationship between minor epidemics and a deterministic theory of disease invasion, as well as major epidemics and the deterministic theory of disease establishment. We conclude this chapter with a demonstration of how the basic ideas can be used to model containment policies associated with the outbreak of SARS in Asia in the early part of 2003.

1. Introduction

The pillars upon which modern epidemiological theory for directly-transmitted infectious microparasitic disease (primarily viral and bacterial) is built are the deterministic model of Kermack and McKendrick [**33**] (also see Hethcote [**28**]), the stochastic chain binomial models of Reed and Frost (unpublished lecture notes), and Soper [**51**] later enriched through the application of Galton-Watson branching process theory (see [**11**]). The field has only come of age over the past three

decades, marked first by Baileys seminal text [**5**], *The Mathematical Theory of Infectious Diseases and its Application*, later by Anderson and Mays synthetic tome [**1**], *Infectious Diseases of Humans: Dynamics and Control*, and more recently with three research survey volumes by the broader community, arising from the Isaac Newton Institutes 1993 focus on infectious disease modeling [**24, 30, 45**]. Also, texts by Daley and Gani [**11**], Diekmann and Heesterbeek [**15**], and Thieme [**53**] provide well-crafted introductory and advanced mathematical presentations, while several edited volumes provide an overview of current interests and directions (e.g. [**14**] and [**29**], this volume).

An important current area of research is the development of theory and methods to model epidemics in heterogeneous systems. Heterogeneity due to spatial and other population structures, such as age, social groups, or genetic variation, can arise in many different ways, and modeling studies dealing with various kinds of heterogeneity are being published at an increasing rate. Heterogeneity is grist for the evolutionary mill. In this and the next chapter, however, we focus only on the development of more cohesive theoretical and methodological approaches to heterogeneity, because further maturation of these fields are still needed to approach some of the most challenging problems in the coevolution of host-pathogen systems.

The primary goal of this chapter is to provide a didactic overview of the basic underpinnings of our own studies of HIV, TB and other disease, presented in the next chapter [**23**].

2. SIR Models

Underlying all dynamical systems models of epidemiological processes is the S-I framework of Kermack and McKendrick [**33**] that was foreshadowed by the work of Enko [**18**]. Within this framework, a population infected by a microparasite (e.g. a bacterial or viral pathogen, or protist) is divided at its most basic level into susceptible and infected (assumed to also be infective) groups, with numbers or densities represented at time t by the continuous variables $S(t)$ and $I(t)$ respectively. At a trivial level, numbers are easily converted to densities once the area A confining the population is known; but, in general, aggregations in the density of populations across landscapes—or metapopulations when well-structured—has a critical influence on the epidemiology (see discussion in [**8, 12, 13, 43, 44**]). For deterministic models, a density representation is preferable because, as discussed in the next section on transmission, it focuses our attention on assumptions regarding the rate at which each susceptible individual acquires the disease as a function of the state of the system. In stochastic models, however, a numbers representation is often preferable. Descriptions of epidemics using continuous deterministic variables are only useful when the numbers in each disease class are relatively large. This, of course, is not the case in the initial stages of an epidemic when $I(t)$ is best described as a jump process on the integers.

At the heart of the original Kermack-McKendrick model is the *transmission function* βSI, with *transmission parameter* $\beta > 0$, which arises from the assumption that the rate at which susceptible individuals become infected is proportional to the densities or numbers of the susceptible and infected populations—i.e. transmission is a *mass action* process. (Note that some terminological confusion has arisen on this point through de Jong, Diekmann, and Heesterbeek [**13**], referring to βSI transmission as mass action when S and I are densities and pseudo mass action

when S and I are numbers.) Most presentations of the Kermack-McKendrick model explicitly or implicitly assume that all individuals are infectious immediately on becoming infected (i.e. no latent period exists), and that infectious individuals are removed from the population at a rate α to enter a class of size $R(t)$ of *recovered* (assumed to also be immune) or *removed* (dead) individuals. Some ambiguity has arisen in the past in models where it has not been specified whether the R class individuals are immune or dead. To avoid this ambiguity in our presentation, we split the R class into a V (recovered and immune—i.e. naturally vaccinated) and a D (dead) class, with flows from the I class at rates α_V and α_D respectively. Immunity is life-long only for a subset of diseases, including many so-called childhood diseases such as measles or chickenpox. Thus, for generality, we assume that individuals in the V class lose their immunity at rate ρ (also know as the relapse rate), to return to the S class. Under these assumptions, the model has the form

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \rho V & S(0) &= S_0 \\ \frac{dI}{dt} &= \beta SI - (\alpha_V + \alpha_D)I & I(0) &= I_0 \\ \frac{dV}{dt} &= \alpha_V I - \rho V & V(0) &= V_0 \\ \frac{dD}{dt} &= \alpha_D I & D(0) &= 0 \end{aligned} \tag{2.1}$$

In model (2.1), the total density of individuals alive at time t is $N(t) = S(t) + I(t) + V(t)$. The sum $N(t) + D(t)$ has the constant value $S_0 + I_0 + V_0$ throughout the epidemic because model (2.1) does not include a description of any demographic processes, which are assumed to be operating at much longer time scales than the epidemic itself. This model predicts that the invasion of a completely susceptible population by a single infected individual will give rise to an epidemic provided $S_0 > \frac{\alpha}{\beta} := S_T$ (:= means "by definition"), where $\alpha = \alpha_V + \alpha_D$ and S_T is the conceptually appealing, although hard to demonstrate [**38**], threshold density of susceptibles that is required for a disease to be able to invade a population when the transmission term is assumed to have a mass-action form.

This threshold density is related to the more general threshold criterion $R_0 > 1$, where R_0 is the basic reproductive number defined as the expected number of individuals that a typical infectious case will infect in a wholly susceptible population [**1, 27**]. For the mass action model shown above, $R_0 = \frac{\beta S_0}{\alpha}$.

If an epidemic does occur then, in the absence of any new susceptibles coming into the population, the number of infectious cases will peak but ultimately die out leaving a proportion $s_\infty := \lim_{t\to\infty} \frac{S(t)}{N_0}$ of the population uninfected. This proportion is the solution to the equation

$$s_\infty = e^{R_0(1-s_\infty)}. \tag{2.2}$$

See [**41, 42**] and [**15**] for an illuminating discussion of conditions under which this expression holds true.

3. Transmission

At the core of any epidemic is the pathogen transmission process, modeled in the above SIR formulation by a mass-action process with resulting form βSI. A more refined analysis views transmission as the concatenation of two processes:

a contact process focusing on the rate at which susceptible individuals encounter infected individuals, and a process relating the probability with which susceptible individuals become infected per susceptible-infective contact [**15, 26, 43**]. Let $C(N)$ be the rate at which each individual in a population contacts other individuals (assumed to depend on total population size N), and let p_T be the probability that a susceptible becomes infected during one such contact with an infective. In a randomly mixing population, the proportion of an individuals contacts that are with infectives is $\frac{I}{N}$. The expected transmission rate per-capita susceptible is then given by

$$\tau(I,N) = p_T C(N)\frac{I}{N}, \tag{3.1}$$

which is known in the epidemiology literature as the *hazard (or force) of infection*. Mass-action transmission (i.e. a transmission rate that is proportional to the product SI) arises when the rate at which susceptibles encounter infectives is proportional to the population density, that is $C(N) = cN$, with c the constant of proportionality, in which case the transmission parameter β is given by $\beta = p_T c$.

Although the mass-action formulation of transmission dominated epidemiological modeling into the early 1990's, by the end of the millennium most formulations were based on the assumption that contact rates are limited by behavioral or social factors and do not depend significantly on population density, with the simplest case being $C(N) = c'$. This assumption leads to so-called *frequency-dependent* transmission:

$$\tau(I,N) = p_T c'\frac{I}{N}, \tag{3.2}$$

with the interpretation that individuals contact each other at a fixed rate c', and if the density of all individuals is N, then a proportion $\frac{I}{N}$ of these contacts will be infective.

Use of the frequency-dependent rather than the mass-action formulation has profound implications for our understanding of epidemics [**22**]. For example, replacing mass action with frequency dependence in the SID model (2.1), we obtain

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{SI}{N} + \rho V & S(0) &= S_0\\
\frac{dI}{dt} &= \beta\frac{SI}{N} - (\alpha_V + \alpha_D)I & I(0) &= I_0\\
\frac{dV}{dt} &= \alpha_V I - \rho V & V(0) &= V_0\\
\frac{dD}{dt} &= \alpha_D I & D(0) &= 0
\end{aligned} \tag{3.3}$$

where $N(t) = S(t) + I(t) + V(t)$. The concept of a threshold population density no longer applies in this model: on introduction of an infected individual into the population an epidemic can occur if and only if $\beta > \alpha = \alpha_V + \alpha_D$. Also, for the case where recovery from disease ($\alpha_V = 0$) does not occur (e.g. HIV infections), unlike the mass action SIR model all susceptibles become infected during the course of an epidemic—that is, $s_\infty = 0$.

Mass-action and frequency-dependent transmission can be regarded as special cases of a more general transmission function [**43**]

$$\tau(I,N) = \left(\frac{p_T c'}{1+(K_T/N)}\right)\left(\frac{I}{N}\right), \tag{3.4}$$

where K_T is the population size or density at which transmission is half its maximum rate $p_T c'$ before accounting for the proportion of infective individuals in the population. Note that expression (3.4) is approximately mass action when $N \ll K_T$ (in this case $c = c'/K_T$ in the mass action expression given in equation (3.1) with $C(N) = cN$), is approximately frequency dependence when $N \gg K_T$ (cf. equation (3.2)), and is some interpolation of the two when N is the neighborhood of K_T. The influence of the interpolation can be controlled by a parameter $\gamma_T \geq 1$ using the expression

$$\tau(I,N) = \frac{p_T c'}{\left(1+(K_T/N)^{\gamma_T}\right)^{1/\gamma_T}}\left(\frac{I}{N}\right), \tag{3.5}$$

where, for γ_T close to 1 (say, $1 \leq \gamma_T \leq 3$), the region of interpolation is quite extensive; and, for increasing values of γ_T, we get an increasingly abrupt switch from mass action when $N < K_T$ to frequency dependence when $N > K_T$. An alternative formulation of a saturating transmission function, based on mechanistic considerations, is given by Heesterbeek and Metz [**26**].

In a well-mixed population with no bounds on contact rates, mass action is likely to be superior to frequency dependence as a model of transmission of airborne or casual contact infectious diseases, such as tuberculosis or influenza. (Note that random mixing and unbounded contact rates are very strong assumptions that probably only apply within homogeneous subunits of a population.) In socially or spatially structured populations, or when the disease is only transmitted by close contact, the overall transmission may be better modeled using the more general force of infection function represented by expression (3.5). An open question remains regarding which values of K_T and γ_T best reflect the aggregated properties of different types of spatial structure. Many models implicitly assume that $N \gg K_T$ and apply frequency-dependent transmission.

For a sexually-transmitted disease (STD), K_T is likely to be quite low so the frequency of intimate contact is unlikely to depend on population density for many populations. In this case, transmission should be modeled using the pure frequency-dependent form expressed in (3.2). However, monogamy is a distinguishing feature of many (though not all) sexual relationships, and the finite duration of monogamous partnerships can have significant impacts on STD spread [**16**]. In a recent study, we examined the contact process for STDs at a finer resolution, formulating a model of pair formation and dissolution [**40**]. We showed that frequency-dependent transmission can be derived rigorously from the pair-formation mechanism, but only by applying a strong timescale assumption that pairing processes are much faster than disease processes. The derivation is exact only for instantaneous partnerships; for partnerships of finite duration epidemics progress more slowly and may saturate at lower levels. Using simulations we found that frequency- dependent transmission is a reasonable depiction of STD transmission via monogamous partnerships only for relatively promiscuous populations: for faster bacterial STDs, such as gonorrhea and Chlamydia, partnerships had to last only days on average for the approximation to be reasonable, while for slower viral STDs, such as HIV or

herpes, partnerships needed to last a few months on average for the approximation to be reasonable.

Another important assumption made by most standard models of transmission is that the infection does not influence individuals' contact behavior. This is often not the case, because infected individuals may reduce their contacts due to physical effects of illness, social factors, or ethical concerns (e.g. [**34, 50**]), or the pathogen may manipulate host behavior to increase contact rates [**7**]. In the study described above, we extended the frequency-dependent transmission function to account for instances where individuals pairing behavior was a function of their disease class status (healthy (S) or sick (I)) [**40**]. In particular, if k_S and k_I are the relative rates at which healthy and sick individuals enter pairs and l_{SS}, l_{SI} and l_{II} are the relative rates at which SS, SI and II pairs break up (all rates on the same time scale, but the scale itself is arbitrary) then, assuming that infection is transmitted at the same average rate $p_T c'$ within all pairs, the frequency-dependent transmission expression (3.2) has the form

$$\tau(I, N) = p_T c' \frac{I}{N} \phi, \tag{3.6}$$

where the modifying factor ϕ depends only the proportions S/N and I/N, and the parameters k_S, k_I, l_{SS}, l_{SI} and l_{II}.

By way of example, if healthy and sick individuals pair up at different rates k_S and k_I, but all pairs break up at the same rate l irrespective of the disease status of the partners, then ϕ in expression (3.6) has the form

$$\phi = \frac{k_S k_I N}{k_S(k_I + l)S + k_I(k_S + l)I}.$$

The general case with different break up rates leads to more complex expressions [**40**].

4. Disease Class Extensions

The first obvious extension to the basic SIR model is to incorporate disease latency defined as the period from when an individual is infected to when they become infectious. Note that while symptoms often aid transmission, as in the case of coughing (lung infections) or the formation of pustules (poxes), the latent period is distinct from the incubation period between infection and the appearance of symptoms [**19**]. Indeed the proportion of transmission occurring before symptoms has a critical influence on options for disease control [**20, 48**]. In practice the onset of infectiousness is often difficult to measure, however, and the appearance of symptoms is sometimes used as a surrogate, particularly for novel diseases such as SARS when it emerged.

In a real population of individuals infected by a disease, the latent period will assume a distribution of values that is usually unimodal with a mode greater than zero [**19**]. The simplest approach conceptually is to model this with a fixed time delay. More generally, we can assume that the time from exposure to infectivity is characterized by a probability distribution $P_\Delta(t)$. In this more general case, just focusing on the transmission process terms, the equations for the susceptible and

infective classes have the form

$$\frac{dS}{dt} = -p_T C(N(t))S(t)\frac{I(t)}{N(t)} + \text{ non-transmission terms}$$
$$\frac{dI}{dt} = p_T \int_0^t C(N(u))S(u)\frac{I(u)}{N(u)}P_\Delta(t-u)du + \text{ non-transmission terms.}$$

This equation is an infinite-dimensional dynamical system with all the attendant numerical and analytical intricacies of such systems. A mathematically simpler approach is to add a disease class E of individuals that have been exposed to the disease and become infective at rate δ. In this case, ignoring equations and terms relating to removed or recovered individuals the model takes the form

$$\begin{aligned} \frac{dS}{dt} &= -p_T C(N)S\frac{I}{N} & S(0) &= S_0 \\ \frac{dE}{dt} &= p_T C(N)S\frac{I}{N} - \delta E & E(0) &= E_0 \\ \frac{dI}{dt} &= \delta E - \alpha I & I(0) &= I_0. \end{aligned} \tag{4.1}$$

In this approach, just focusing on a fixed cohort of individuals E_0 that are exposed at time $t = 0$ (i.e. ignoring new infections from transmissions), then the rate at which these individuals advance to the infectious class is given by the following equation and its associated solution:

$$\frac{dE}{dt} = -\delta E, \quad E(0) = E_0 \quad \Rightarrow \quad E(t) = E_0 e^{-\delta t}.$$

In this case, the so-called *residence time* in the exposed class E is exponentially distributed with mean $1/\delta$ [**52**]. The exponential distribution is a poor match to biological distributions of latent periods, however, because its mode is at $t = 0$ rather than in the vicinity of its mean at $t = 1/\delta$.

This latter problem can be resolved using a *distributed delay* (i.e. staging or box car) approach [**47**] in which we have n classes of exposed individuals E_i, $i = 1, \ldots, n$, and assume that individuals transfer through each class at a rate $n\delta$. In this case, equations (4.1) are extended to:

$$\begin{aligned} \frac{dS}{dt} &= -p_T C(N)S\frac{I}{N} & S(0) &= S_0 \\ \frac{dE_1}{dt} &= p_T C(N)S\frac{I}{N} - n\delta E_1 & E_1(0) &= E_{1\,0} \\ \frac{dE_i}{dt} &= n\delta(E_{i-1} - E_i) & E_i(0) &= E_{i\,0} \quad i = 2, \ldots, n \\ \frac{dI}{dt} &= n\delta E_n - \alpha I & I(0) &= I_0. \end{aligned} \tag{4.2}$$

The total residence time in the exposed class (i.e. the time from leaving S to entering I) is now gamma-distributed [**52**]. The mean residence time is still $1/\delta$, but the distribution is now modal near $1/\delta$ and has variance $1/(n\delta^2)$, which implies that as $n \to \infty$ the solution to model (4.2) approaches a fixed time lag (i.e. variance is 0) of duration $1/\delta$.

Of course, similar techniques can be employed to make the infectious period distribution more realistic as well. The dynamical consequences of using gamma-distributed latency and infectious periods were well-known in a queuing theory

context in the mid-1900s but were first laid out in detail for an epidemiological audience by Anderson and Watson [**2**]. Subsequent to this, the implications of exponential versus fixed infectious periods have been explored in greater depth (e.g. [**31, 32, 36, 37**]).

5. Including Demography

Demographic considerations fall into two categories: 1.) flows in and out of the population due to births, deaths, and migration; and 2.) disease-independent internal population structure due to age, developmental stage, or group structure (all of these may also have a spatial component). The former is easy to incorporate in the absence of any internal structures and, thus, we deal with it first. Deterministic models (2.1) and (3.3) pertain purely to epidemic processes in a population that is initially of size $N(0) = S_0 + I_0 + V_0$ and declines due to the effects of the disease induced death rate α_D on individuals in class I. Demographic flows are easily included into such models, as follows. In the absence of disease, assume that the population of interest is subject to a total birth or recruitment rate $f_b(N)$ and a per-capita mortality rate μ. Then in the absence of migration and age structure the demographic equation for the population is simply

$$\frac{dN}{dt} = f_b(N) - \mu N.$$

In the presence of disease, we need to decide whether all individuals participate equally in the birth process or whether infected individuals have an altered reproductive capacity. The most appropriate assumption will depend on the disease in question. For simplicity of exposition, assume that susceptible and infected individuals contribute equally to births and that newborn individuals are susceptible (i.e. the disease is not vertically transmitted and there is no immunity due to maternal antibodies), and that individuals are removed from the infective population at a rate $(\mu + \alpha)$. In this case, using the general transmission function defined by expression (3.1), equations (2.1) and (3.3) can be written as (cf. [**1, 22**])

$$\begin{aligned}
\frac{dN}{dt} &= f_b(N) - \mu N - \alpha_D I & N(0) &= N_0 \\
\frac{dI}{dt} &= p_T C(N)(N - I - V)\frac{I}{N} - (\mu + \alpha_D + \alpha_V)I & I(0) &= I_0 \\
\frac{dV}{dt} &= \alpha_V I - \mu V & V(0) &= V_0.
\end{aligned} \tag{5.1}$$

In this model, written in a form that accentuates the dynamics of the population as a whole, the susceptible class is determined by $S(t) = N(t) - I(t) - V(t)$.

The two simplest forms for the "input" function $f_b(N)$ births, are the following. In the case where $f_b(N)$ is interpreted as recruitment (i.e. as is useful for short-term predictions when N is "distanced" from the birth process by a long time delay such as in HIV models where N represents individuals age 15 to 50 and predictions are made 10 to 20 years ahead) then $f_b(N)$ is assumed either to be constant or an externally determined time-varying input of the form $f_b(N) = \lambda(t)$. In the case where $f_b(N)$ arises from a constant per-capita birth rate, this rate must balance the death rate—that is $f_b(N) = \mu N$— otherwise the population will grow or decline exponentially (cf. [**28**]). This of course can be corrected by ensuring births are

density dependent—as in the function

$$f_b(N) = \frac{bN}{1 + (N/K_b)^{\gamma_b}},$$

where we require $b > \mu$ (to ensure growth at low population densities) and $\gamma_b > 1$ (to ensure a sensible dependence on density—see [**21**]).

A treatment of age structure in the context of continuous time models is beyond the scope of this exposition because, as with the inclusion of time delays, continuous age- structure formulations are cast either in the context of McKendrick-von Foerster partial differential or integro-differential equations [**1, 10, 53**]. The mathematical properties and numerical solutions of such systems are much more difficult to obtain. Age structure, however, is easily incorporated in the context of discrete models. Further, as argued in subsequent sections, discrete time models better reflect the fact that empirical data are typically values obtained from averaging rates over predetermined discrete time intervals (e.g. daily, monthly, or annual birth, death and infection rates, and so on).

6. Discrete Time Formulations

We can find solutions to continuous time formulations of dynamic processes, such as the epidemiological models (2.1), (3.3) or (4.2), or we can, at least, analyze their behavior using the tools of calculus. Data used to estimate the parameters in these equations, however, are generally derived from events—such as births, deaths, new cases, cures, numbers vaccinated—recorded over appropriate discrete intervals of time (typically days for fast diseases such as SARS or influenza, weeks or months for slower diseases such as tuberculosis or HIV, and years for vital rates in seasonal breeders and long-lived species). Data reporting the proportion p_μ of individuals that die in a unit of time can be converted to a mortality rate parameter μ appearing in a differential equation model of the form $\frac{dN}{dt} = -\mu N$ by noting that the solution to this equation over any time interval $[k, k+1]$ is $N(k+1) = N(k)e^{-\mu}$. This implies that the proportion of individuals dying is

$$p_\mu = \frac{N(k) - N(k+1)}{N(k)} = \frac{N(k)(1 - e^{-\mu})}{N(k)} = 1 - e^{-\mu} \tag{6.1}$$

or, equivalently, $\mu = \ln\left(\frac{1}{1-p_\mu}\right)$.

It is often advantageous to formulate epidemiological and demographic models in discrete time. The primary advantage of differential equation models disappears once we resort to numerical simulation of systems rather than trying to obtain analytical results, which are difficult if not impossible to obtain for most detailed nonlinear models. Indeed, discrete time models can be implemented very naturally and easily in computer simulations, while numerical solutions of differential equations requires algorithms that use discretizing approximations. Also parameters in discrete time models can be more easily related to data that have been collated over discrete intervals (e.g. vital and transmission rates, etc.).

Discrete time equations, however, cannot properly account for the interactions of simultaneously nonlinear processes, such as individuals simultaneously subject to the processes of infection and death: in each time interval we either first account for infection and then natural mortality or vice versa. It does make a difference how we schedule things [**54**]. Alternatively we can treat the two processes simultaneously,

but then cannot accurately depict both processes occurring in one time step if transition rates are state-dependent. A further challenge arises in discrete time disease models because the force of infection depends on the size of the infectious population at each moment, which cannot be updated over the course of a time step. Fortunately, a good approximation can be obtained by using a piecewise linear modeling approach, as follows.

First we write down the continuous time model of interest. Consider, for example, equations (4.1) with a constant total recruitment rate λ to the susceptible class and a constant per-capita natural mortality rate μ. Then, using the notation $\tau(I,N) := p_T C(N) I/N$ (see equation (3.1)), these equations become

$$\begin{aligned} \frac{dS}{dt} &= \lambda - \mu S - \tau(I,N)S \qquad S(0) = S_0 \\ \frac{dE}{dt} &= \tau(I,N)S - (\delta+\mu)E \quad E(0) = E_0 \\ \frac{dI}{dt} &= \delta E - (\alpha+\mu)I \qquad\quad I(0) = I_0. \end{aligned} \tag{6.2}$$

Now assume over a small interval $t \in [k, k+1]$ that the proportional change in τ over this interval due to a change in $I(t)$ is sufficiently small that $\tau(I(t),N(t))$ is well approximated by $\tau_k = \tau(I(k),N(k))$ (e.g. over the time interval the change in $\tau(I(t),N(t))$ may be a few percent, but our estimates of the parameters c' and p_T in $\tau(I(t),N(t))$, as defined by expression (3.2), may have uncertainties that are several times as large). Obviously this assumption will influence the choice of time step duration, with shorter time steps required for faster-growing epidemics. Then replacing $\tau(I(t),N(t))$ with the constant τ_k for $t \in [k, k+1]$, equation (6.2) is an inhomogeneous linear system of ordinary differential equations, with solution given by

$$\begin{pmatrix} S(k+1) \\ E(k+1) \\ I(k+1) \end{pmatrix} = \exp\{A_k\} \begin{pmatrix} S(k) \\ E(k) \\ I(k) \end{pmatrix} + \left(\int_0^1 \exp\{A_k t\} dt \right) \begin{pmatrix} \lambda \\ 0 \\ 0 \end{pmatrix}, \tag{6.3}$$

where

$$A_k = \begin{pmatrix} -(\mu+\tau_k) & 0 & 0 \\ \tau_k & -(\mu+\delta) & 0 \\ 0 & \delta & -(\mu+\alpha) \end{pmatrix}. \tag{6.4}$$

Calculation of the exponential matrix function $\exp\{A_k\}$ and its integral requires that we first find the eigenvalues and eigenvectors of the matrix A_k itself, which is cumbersome to calculate and will generally not have a closed form solution for systems with more than four disease classes (unless A_k is triangular, in which case the eigenvalues are the diagonal entries themselves). The matrix $\exp\{A_k\}$ and its integral can be calculated numerically, but this calculation will have to be performed at each time step k, because of the dependence of τ_k on the current values of the state vector $(S(k), E(k), I(k))'$ (here $'$ denotes vector transpose).

A discrete version of equation (6.2) can be argued directly from first principles under the assumptions that individuals are recruited at the beginning of time interval $[k, k+1]$ and that individuals die at the same constant rate μ throughout

the time interval $[k, k+1]$. In this case we obtain the model

(6.5)
$$\begin{pmatrix} S(k+1) \\ E(k+1) \\ I(k+1) \end{pmatrix} = \begin{pmatrix} (1-p_\mu)(1-p_{\tau_k}) & 0 & 0 \\ (1-p_\mu)p_{\tau_k} & (1-p_\mu)(1-p_\delta) & 0 \\ 0 & (1-p_\mu)p_\delta & (1-p_\mu)(1-p_\alpha) \end{pmatrix} \times \begin{pmatrix} S(k) \\ E(k) \\ I(k) \end{pmatrix} + \begin{pmatrix} (1-p_\mu)\lambda \\ 0 \\ 0 \end{pmatrix},$$

which is iterated from the initial condition $(S_0, E_0, I_0)'$. The probabilities p_π are related to the corresponding rates π using the relationship expressed in (6.1), viz.

$$p_\pi = 1 - e^{-\pi} \quad \text{or} \quad 1 - p_\pi = e^{-\pi}, \quad \pi = \mu,\ \alpha,\ \delta,\ \tau_k. \tag{6.6}$$

While equations (6.5) are an approximation to exact solution (6.3), which itself is only exact if $\tau(I(t), N(t))$ is replaced by the constant $\tau_k = \tau(I(k), N(k))$, it is actually irrelevant how well equations (6.5) approximate equations (6.3) when solved precisely. The reason for this is that equations (6.3) are derived from a differential equation model that is not the "gold standard" for modeling epidemics; but, instead, equations (6.3) represent a highly simplified model that does not account for lags, latencies, or heterogeneity in the population being modeled—not to mention higher order processes taking place on faster time scales (one of which is the contact process discussed at the end of Section 3). No theoretical reason exists to prefer differential equation models over difference equation models: both have their strengths and weaknesses.

It is also worth noting at this point that all the deterministic models presented above can be interpreted as representing expected numbers of individuals in what are essentially stochastic epidemiological and demographic processes. Deterministic models provide reasonable realizations of stochastic models either when the size of the population is sufficiently large for the 'Law of Large Numbers' (proportions approach probabilities in the limit as population size approaches infinity) to prevail or they represent equations for the first order moments of the stochastic process in question when second and higher order moments are neglected (e.g. see [**11**, page 68]).

Discrete time models are more easily embedded in a Monte Carlo (i.e. stochastic) simulation framework than continuous time models. First the effects of demographic stochasticity can be simulated by treating the state variables as integers and then calculating the proportion that die as one realization of a set of appropriate Bernoulli trials that will produce a binomial distribution of outcomes. The underlying probabilities p_π themselves can be subject to stochastic variation specified by some appropriate probability distribution, and Monte Carlo simulation methods can be used to generate the statistics of associated distributions of possible solutions to an equation (6.5) when the parameters are interpreted as probabilities which themselves are drawn from statistical distributions.

Second, discrete time models are more flexible than ordinary differential equation models when it comes to fitting distributions reflecting the time spent in a particular disease stage. As we have seen, staging in continuous model can lead to gamma distributions on $[0, \infty)$ for the residence times of individuals in each disease class. Thus staging in continuous models allows us to construct a process that has a desired mean and variance, but also implies that the minimum time spent in a

class is zero, even though residence times close to zero may be extremely unlikely for processes with variance much smaller than the mean.

Staging in discrete models, however, allows us to easily set a minimum and maximum time in a class, as well as a desired mean and variance, as illustrated by the following example of a staged exposed class (cf. equation (4.2)). In these equations, as previously defined, p_{τ_k} and p_α are respectively the probabilities over one time step of a susceptible becoming infectious and an infectious individual being removed. In addition, with regard to staging, we define p_{δ_i} as the probability that an individual in exposed class i, makes the transition to exposed class $i+1$ (except for the case $i = n$). For added flexibility we also introduce the probability p_{θ_i} of moving directly from exposed class i into the infectious class. Note that the probabilities are formulated so that we first account for the proportion that become infectious (via p_{θ_i}) and only then consider what proportion of the remainder make the transition to the next exposed class (via p_{δ_i}). Accordingly, for $i = n$ the proportion moving into the infectious class over one time step is $p_{\theta_i} + (1 - p_{\theta_i})p_{\delta_i}$. Under these assumptions, the staged model has the form:

(6.7)

$$\begin{aligned}
S(k+1) &= S(k) - p_\tau(I(k), N(k))S(k) && S(0) = S_0\\
E_1(k+1) &= p_\tau(I(k), N(k))S(k) + (1 - p_{\theta_1})(1 - p_{\delta_1})E_1(k) && E_1(0) = E_{1\,0}\\
E_{i+1}(k+1) &= (1 - p_{\theta_{i+1}})(1 - p_{\delta_{i+1}})E_{i+1}(k) + (1 - p_{\theta_i})p_{\delta_i}E_i(k), && E_i(0) = E_{i\,0}\\
&\qquad\qquad i = 1, \ldots, n-1 &&\\
I(k+1) &= \sum_{i=1}^{n} p_{\theta_i}E_i(k) + (1 - p_{\theta_n})p_{\delta_n}E_n(k) + (1 - p_\alpha)I(k) && I(0) = I_0.
\end{aligned}$$

Consider the fate of a cohort of $E_{1\,0}$ individuals entering the exposed class at time $t = 0$. The rate at which these individuals enter class I is the element $I(k)$ in the solution to equations (6.7) from initial conditions, $S_0 = 0$, $E_{1\,0} > 0$, $E_{i\,0} = 0$, $i = 2, \ldots, n$, and $I_0 = 0$. In this case, it is easily shown that these $E_{1\,0}$ individuals will remain in the exposed class for a minimum of $k < n$ units of time whenever $p_{\theta_i} = 0$, $i = 1, \ldots, k$, and will all have become infectious by time $n + 1$, provided $p_{\theta_n} = 1$ and $p_{\delta_i} = 1$ for $i = 1, \ldots, n-1$. If $p_{\theta_n} < 1$, however, the final group of individuals will trickle into the infectious class at a geometrically decreasing rate over time. The value for n and the parameters p_{θ_i} and p_{δ_i}, $i = 1, \ldots, n$, can be selected to fit any empirically observed distribution.

7. Stochastic Branching Processes

During the invasion phase of an epidemic, the numbers of infected individuals are small and stochastic effects can play an important role. Branching processes are a family of stochastic models well-suited to modeling disease invasions in large populations. The theory of the branching process (also known as the Galton-Watson process: e.g. see [**9**]) was developed to explore the role of chance in demographic dynamics, originally to understand the extinction of notable family lines in England [**25**]. Consequently it is formulated in terms of generations, with its basic concept being the offspring distribution: the probability that a given "parent" (here equivalent to an infectious index case) will give rise to a given number of "offspring" (here equivalent to new infections). The core assumption is that the numbers of

offspring produced by different index cases are independent and identically distributed. In the context of epidemics, this amounts to assuming that the population is sufficiently large and well-mixed for depletion of the susceptible pool of individuals to be negligible, and that transmission conditions do not change with time. Thus a specific model may only be valid until control measures are introduced. The theory of branching processes is developed in detail elsewhere [**3, 4, 25**]; the treatment here is extracted essentially from Diekmann and Heesterbeek [**15**].

We define the offspring distribution $\{q_i\}_{i=0}^{\infty}$, where q_i is the probability that an infectious individual infects i other individuals. Thus we require $\sum_{i=0}^{\infty} q_i = 1$ and note that R_0, the mean number of cases contracting disease from each infective, is simply given by

$$R_0 = \sum_{i=0}^{\infty} i q_i. \tag{7.1}$$

A powerful tool for studying a branching process is the probability generating function $g(z)$ defined in terms of a dummy variable z by:

$$g(z) = \sum_{i=0}^{\infty} q_i z^i, \quad 0 \le z \le 1. \tag{7.2}$$

This power arises from the easily demonstrated properties

$$g(0) = q_0,\ g(1) = 1,\ \left.\frac{dg}{dz}\right|_{z=1} = R_0,\ \frac{d}{dz}g(z) > 0,\ \frac{d^2}{dz^2}g(z) > 0$$

and from the fact that the solutions z_k to the difference equation

$$z_k = g(z_{k-1}),\ z_0 = q_0,\ k = 1, 2, 3, \ldots,$$

are the probabilities that the disease traceable back to a single original infective dies out by generation k. Thus $z_\infty = \lim_{k\to\infty} z_k$ is the probability that an epidemic started by one individual will die out. It can also be shown that z_∞ is the smallest nonnegative root of the equation $z = g(z)$ and that $z_\infty = 1$ if and only if $R_0 \le 1$, in which case the epidemic will die out with certainty in a finite number of generations (essentially infecting a proportion of measure zero in an infinite population). When $0 < z_\infty < 1$ and there is initially one infective, then the disease causes a significant outbreak with probability $1 - z_\infty$.

Under the idealization of an infinite population, an epidemic that dies out in a finite number of generations is called a *minor epidemic*, while one that goes on to infect a a positive proportion of individuals in our infinite population (i.e. measurable in mathematical terms) is called a *major epidemic*. Thus, in contrast to the deterministic case which guarantees invasion of a pathogen to epidemic proportions, with possible long term establishment whenever $R_0 > 1$, in stochastic theory $R_0 > 1$ only implies that a *major* epidemic will occur with probability

$$P_{\text{epi}} = 1 - z_\infty, \tag{7.3}$$

whenever the initial number of infective individuals is one or, more generally, with probability $P_{\text{epi}}(a) = (1 - z_\infty)^a$ whenever the initial number of infective individuals is a.

In finite populations of size N, the distinction between minor and major epidemics can become less clear [**46**]. In very small populations such as households, the exact distribution of outbreak sizes can sometimes be calculated [**5**]. In larger

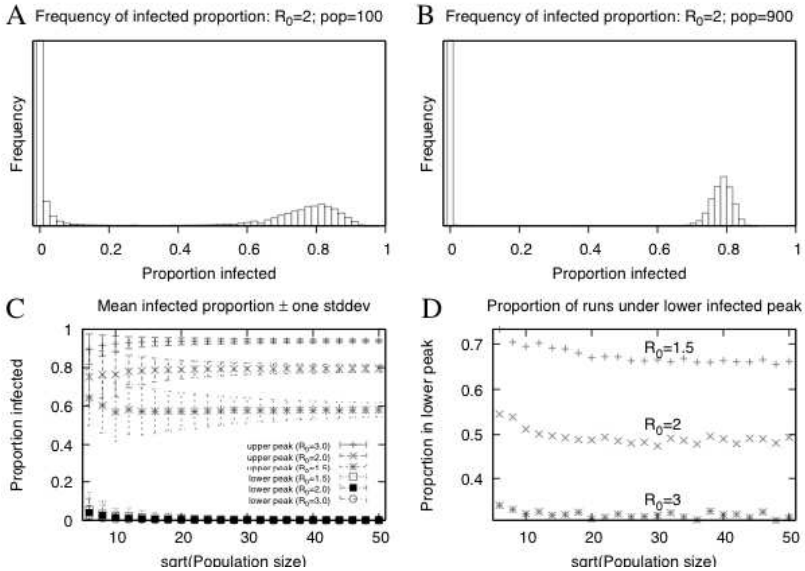

FIGURE 1. Results of individual-based stochastic simulations regarding the ultimate number of infections after the introduction of a single infected individual into a population of N-1 susceptible individuals (recall that $R_0 \approx \beta/\alpha$). **A.** $N = 100, \beta = 0.1, \alpha = 0.01$. **B.** $N = 900, \beta = 0.1, \alpha = 0.01$. **C.** The mean and standard deviation of the proportion of individuals in unsuccessful and successful invasions (*i.e.* relative areas in lower and upper peaks of histograms in **A** and **B**) are plotted as a function of $\sqrt{N}$ for the three cases $\beta = 0.075, 0.1$, and 0.15, with $\alpha = 0.05$ in all cases. **D.** Same three cases, but plotting the proportion of runs in the lower peak (10,000 simulations were used to generate each point). (We are indebted to Philip Johnson for producing this figure)

populations, a minor epidemic still goes extinct in a finite number of generations, while a major epidemic goes on to infect a proportion of the population approximated by the corresponding deterministic model, which for a basic SIR disease process without demographics is $i_\infty = 1 - s_\infty$, where s_∞ is the solution to equation (2.2). In this case, analytical results are difficult to generate, but the existence of a bimodal distribution of epidemic sizes can be demonstrated for sufficiently large N ([**55**], Figure 1A and B). Monte Carlo methods (discussed in the context of heterogeneous populations in the next chapter, [**23**]) can be used to generate simulated epidemics when N is finite. We can split the bimodal histograms obtained for sufficiently large N at the minimum frequency bin between the two modes and define p_{epi} and P_{epi} respectively to be the proportion of outbreaks in our population of size N falling to the left and to the right of this bin. Thus, as $N \to \infty$, $p_{\text{epi}} \to z_\infty$ and $P_{\text{epi}} \to 1 - z_\infty$ with modes of the proportion of individuals infected approaching 0 and i_∞ respectively (Figure 1).

Finally, we note that in a completely homogeneous infinite population where each individual has the same transmission rate and same fixed infectious period,

all individuals are expected to transmit the same number R_0 of new infections in each generation. The actual number of infections will vary due to stochasticity in transmission events occurring over the fixed infectious period, and will follow the Poisson distribution arising from the purely random transmission process with intensity parameter R_0 [**15**]. The probability generating function for this Poisson distribution is:

$$\underline{\text{Poisson}}: \quad g(z) = e^{R_0(z-1)}. \tag{7.4}$$

If we now apply our theory to a homogeneous population with no demography, subject to our standard SI process, we can show that the epidemic will die out if $R_0 \leq 1$, but will go on with probability $P_{\text{epi}} = 1 - z_\infty$ to infect a proportion $i_\infty = 1 - s_\infty$, where both z_∞ and s_∞ satisfy the same equation $x = e^{R_0(x-1)}$. (The root of this commonality can be explained via the theory of random graphs—see [**15**, Section 10.5.2].)

8. Group Structure and Containment of SARS

The assumption that a susceptible population is homogeneous is often too simplistic to address a number of important issues regarding the spread and containment of the spread of infectious disease. The theory we have outlined above is easily extended to a population in which the susceptible individuals are categorized into n (> 1) homogeneous subgroups. In this section, we develop this extension and then demonstrate its application to modeling the recent SARS outbreak in several cities in Asia, as well as Toronto, Canada.

We begin by considering a simple SIR process in a population composed of n subpopulations or groups. Let S_i and I_i respectively denote the density of susceptibles and infectives in population i, $i = 1, \ldots, n$. Let the constants c_{ij} denote the relative rates at which individuals in group i contact individuals in group j. Further, assume that the probability of transmission associated with each such contact is given by transmission parameter β_{ij}. Under the assumption of frequency-dependent transmission, the first two equations in system (3.3) generalize to

$$\begin{aligned}
\frac{dS_i}{dt} &= \left(\frac{\sum_{j=1}^{n} c_{ij}\beta_{ij}I_j}{\sum_{j=1}^{n} c_{ij}N_j} \right) S_i + \text{ non-transmission terms} \\
S_i(0) &= S_{i\,0}, \\
\frac{dI_i}{dt} &= \left(\frac{\sum_{j=1}^{n} c_{ij}\beta_{ij}I_j}{\sum_{j=1}^{n} c_{ij}N_j} \right) S_i + \text{ non-transmission terms} \\
I_i(0) &= I_{i\,0}, \quad i = 1, \ldots, n.
\end{aligned} \tag{8.1}$$

If we now discretize these equations, as discussed in Section 6 and use the vector notation $\mathbf{I} = (I_1, \ldots, I_n)'$ etc., equations (6.5) become

$$\begin{aligned} S_i(k+1) &= S_i(k) - e^{-\tau_i(\mathbf{I}(k),\mathbf{N}(k))} S_i(k) + \text{ non-transmission terms} \\ S_i(0) &= S_{i\,0}, \\ I_i(k+1) &= e^{-\tau_i(\mathbf{I}(k),\mathbf{N}(k))} S_i(k) + \text{ non-transmission terms} \\ I_i(0) &= I_{i\,0}, \quad i = 1, \ldots, n, \end{aligned} \tag{8.2}$$

where (cf. expression (3.2))

$$\tau_i(\mathbf{I}, \mathbf{N}) = \frac{\sum_{j=1}^{n} c_{ij}\beta_{ij} I_j}{\sum_{j=1}^{n} c_{ij} N_j} \tag{8.3}$$

To avoid the notation becoming too complex, we have dropped the convention of using subscripted p's (cf. equation (6.6)) to emphasize that the transition parameters in the model are probabilities.

This approach was used in a recent model of the SARS epidemic that threatened to sweep through parts of Asia and Canada during the first half of 2003 [**39**]. Other models also examined this important example of a modern emerging infectious disease, as reviewed elsewhere [**6, 17**]; the primary focus of ours, though, was on structuring the population into health care workers (HCWs) in a hospital setting and the general community served by that hospital. Each individual in the model was designated as a HCW or community member, and their status did not change regardless of other transitions in the model (e.g. infection, hospitalization, etc.). The reason for this grouping was that it became evident soon after the start of the epidemic that HCWs were at much greater risk for SARS than other individuals in all cities where epidemics were threatening to flare out of control: HCWs comprised 63% of SARS cases in Hanoi, 51% in Toronto, 42% in Singapore, 22% in Hong Kong and 18% in mainland China (see [**39**] for specific references). It was therefore important to study the processes driving HCW infection explicitly, and to assess specific interventions to combat SARS spread within the hospital and back to the general community.

Moving to a specialized notation for this particular SARS model, the indices $i = h$, c, and m were respectively used to denote HCW, community, and case-managed (quarantined and isolated) individuals (Figure 2a). The model was iterated on a daily basis (i.e. the basic unit of time was one day) and included ten one-day exposed classes (Figure 2b) with probability p_j of progressing to the symptomatic (here assumed equivalent to infectious) class on the j^{th} day since infection to simulate the empirically observed distribution of latent periods. The infectious state was broken into five subclasses with progression probabilities of $\{1, 1, r, r, r\}$ (Figure 2c) selected to fit empirical data on infectious periods. The parameters q_{ij} and h_{ij} denote the probabilities that exposed and infectious individuals are quarantined and hospitalized, respectively, where $i = c$ or h represents hosts from the community and HCW groups, and j represents the substage within the exposed or infectious class (because case management parameters may vary as a function of time since infection or appearance of symptoms).

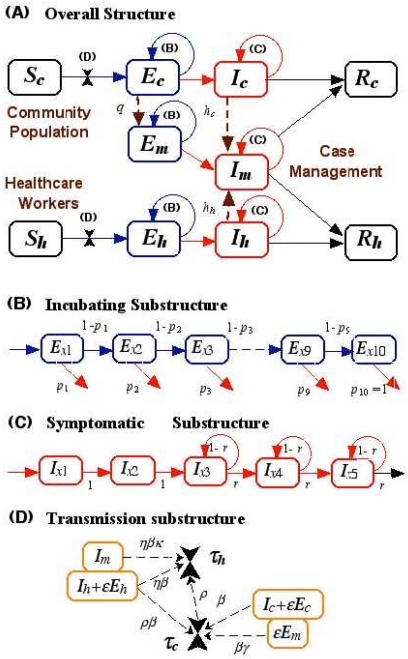

FIGURE 2. Flow diagram of the transmission dynamics of a SARS epidemic within a hospital coupled to that in a community (after Lloyd-Smith, Galvani and Getz, 2003). **A.** S: susceptible, E: incubating, I: symptomatic, R: removed. Subscripts h, c and m respectively represent individuals from healthcare worker, general community, and case-managed groups. E_m and I_m respectively are quarantined and isolated individuals. **B.** Incubating individuals (E_i, where $i = c, h, m$) were further structured into ten disease-age classes. Daily probabilities p_i of progressing to the symptomatic phase were linearly interpolated between $p_1 = 0$ and $p_{10} = 1$. **C.** For $0 < r < 1$, symptomatic individuals (I_i, where $i = c, h, m$) move through 5 stages to the final symptomatic class R_c or R_h, according to their group of origin. **D.** The transmission hazard rates for susceptible individuals S_i are denoted by $i(i = c, h)$, and depend on weighted contributions from community and HCW sources as described in the text.

The equations for both community (subscript c) and HCW (subscript h) pool are listed first, followed by the equations for the managed cases (subscript m). The managed case variables need a superscript $i = c, h$ in addition to the subscripts because recovered individuals are sent back to the pool of their origination. The variables and parameters used are depicted graphically in Figure 2. Transmission of the SARS coronavirus is represented by hazard rate functions τ_i for susceptible individuals in the i pool ($i = c, h$), which have the general form of expression (8.3). In particular, Lloyd-Smith, Galvani and Getz [**39**] formulated the parameters β_{ij}

and c_{ij} in expression (8.3) in terms of the basic transmission rate β (not to be confused with the above transmission probabilities β_{ij}) and a collection of parameters modifying transmission for different settings: ε, η, γ, κ, and ρ (all on the interval $[0,1]$). In particular, the reduced transmission rate of exposed (E) individuals (included because the extent of pre-symptomatic transmission of SARS was unknown when the model was created) is $\varepsilon\beta$. All transmission within the hospital setting occurs at a reduced rate $\eta\beta$ to reflect contact precautions adopted by all hospital personnel and patients, such as the use of masks, gloves and gowns. Additionally, quarantine of exposed individuals reduces their contact rates by a factor γ, yielding a total transmission rate of $\gamma\varepsilon\beta$, while specific isolation measures for identified SARS patients (I_m) in the hospital reduces their transmission by a further factor κ. Finally, we considered the impact of measures to reduce transmission rates between HCWs and community members by a factor ρ. Under these assumptions, the transmission hazards are:

$$\tau_c = \frac{\beta(I_c + \varepsilon E_c) + \rho\beta(I_h + \varepsilon E_h) + \gamma\beta\varepsilon E_m}{N_c}$$

and

$$\tau_h = \rho\tau_c + \frac{\eta\beta(I_h + \varepsilon E_h + \kappa I_m)}{N_h},$$

where E_i and I_i, $i = c, h$, represent sums over all sub-compartments in the incubating and symptomatic classes for pool j, and

$$N_h = S_h + E_h + I_h + V_h + I_m$$

and

$$N_c = S_c + E_c + I_c + V_c + \rho(S_h + E_h + I_h + V_h).$$

The detailed form of the SID equations that were formulated are:
Community and HCW equations:

$$\left.\begin{array}{l} S_i(t+1) = \exp\left(-\tau_i(t)\right) S_i(t) \\ E_{i1}(t+1) = \left[1 - \exp\left(-\tau_i(t)\right)\right] S_i(t) \\ E_{ij}(t+1) = (1-p_{j-1})(1-q_{ij-1})E_{ij-1}(t) \quad j = 2,\ldots,10 \\ I_{i1}(t+1) = \sum_{j=1}^{10} p_j(1-q_{ij})E_{ij}(t) \\ I_{i2}(t+1) = (1-h_{i1})I_{i1}(t) \\ I_{i3}(t+1) = (1-h_{i2})I_{i2}(t) + (1-r)(1-h_{i3})I_{i3}(t) \\ I_{ij}(t+1) = r(1-h_{i\,j-1})I_{i\,j-1}(t) + (1-r)(1-h_{ij})I_{ij}(t) \quad j = 4,5 \\ V_i(t+1) = V_i(t) + rI_{i5}(t) + rI^i_{m5}(t) \end{array}\right\} i = c, h,$$

$$\left.\begin{array}{l} E^i_{m,j}(t+1) = (1-p_{c\,j-1})\left(E^i_{m,j-1}(t) + q_{j-1}E_{c\,j-1}(t)\right) \quad j = 2,\ldots,10 \\ I^i_{m1}(t+1) = \sum_{j=1}^{10} p_j\left(E^i_{mj}(t) + q_{ij}E_{ij}(t)\right) \\ I^i_{m2}(t+1) = h_{i1}I_{i1}(t) + I^i_{m1}(t) \\ I^i_{m3}(t+1) = h_{i2}I_{i2}(t) + I^i_{m2}(t) + (1-r)\left[h_{i3}I_{i3}(t) + I^i_{m1}(t)\right] \\ I^i_{mj}(t+1) = r\left[h_{i\,j-1}I_{i\,j-1}(t) + I^i_{m\,j-1}(t)\right] \\ \qquad\qquad +(1-r)\left[h_{ij}I_{ij}(t) + I^i_{mj}(t)\right] \quad j = 4,5 \end{array}\right\} i = c, h.$$

In the analysis presented here, the probabilities q_{ij} and h_{ij} vary between 0 and a fixed value less than 1 and account for delays in contact tracing or case identification. In addition, we did not analyze scenarios where health care workers are quarantined so that $q_{hj} = 0$ for all j. Deterministic solutions to this SARS model can be generated by directly iterating the above equations for specific sets of parameter values and initial conditions. However, because SARS outbreaks were

invasion scenarios with initially small numbers of cases, stochastic simulations were required to incorporate the important influence of chance on outbreak dynamics.

For a range of R_0 values corresponding to conditions in different cities—but with emphasis on $R_0 \sim 3$ as reported for Hong Kong and Singapore [**35, 49**]—we performed sensitivity analyses (by calculating effective reproductive numbers under different control scenarios) and conducted stochastic simulations to explore the relative merits of different control measures for SARS. We assessed contributions of case management (i.e. isolation and quarantine) and contact precautions (such as masks, gowns, and hand-washing) to containment of a nascent SARS outbreak, and considered the extent to which one measure can compensate for another which is not available in a given setting.

A number of unintuitive and applicable conclusions arose from this analysis. For instance, hospital-wide contact precautions were identified as the single most potent containment measure—an encouraging finding since these are easily implemented and inexpensive. We investigated two types of delay in control efforts. At the individual level, delays of a few days in contact tracing and case identification severely degraded the utility of quarantine and isolation, particularly in high-transmission settings. Still more detrimental were delays at the population level between onset of an outbreak and implementation of control measures: for given control scenarios our model identified windows of opportunity beyond which efficacy of containment efforts was reduced greatly. In settings where hospital-based transmission was continuing, we showed that measures to reduce contact between healthcare workers and the community had dramatic benefits in preventing a widespread epidemic, emphasizing the importance of mixing restrictions (Figure 3) and, hence, considerations of heterogeneity in exposure to disease among individuals in the population.

9. Conclusion

Deterministic and stochastic theory of SIR processes in homogeneous populations has provided a solid foundation for the construction of modern epidemiological theory through the elaboration of the SIR framework. In this chapter we presented on overview of these SIR framework elaborated to include additional disease classes and demographic components in both discrete and continuous time equation formulations. We also demonstrated the extension of the discrete time setting to a population in which susceptible individuals are divided in n homogenous subgroups, with application to the recent SARS epidemics. In the next chapter (Getz et al., 2005), we discuss how elaboration of the SIR framework has much to do with the structuring of populations into homogeneous subclasses of a more general nature than considered here.

Acknowledgements

Grants to WMG from the NSF/NIH Ecology of Infectious Disease Program (DEB-0090323), NIH-NIDA (R01-DA10135), and the James S. McDonnell Foundation 21st Century Science Initiative supported various components of the research reported in this chapter. We are indebted to Philip Johnson for writing the code needed to generate Figure 1 and producing the figure for us. We thank Shirli Bar-David, Paul Cross, Alison Galvani, Philip Johnson, Travis Porco, Maria Snchez

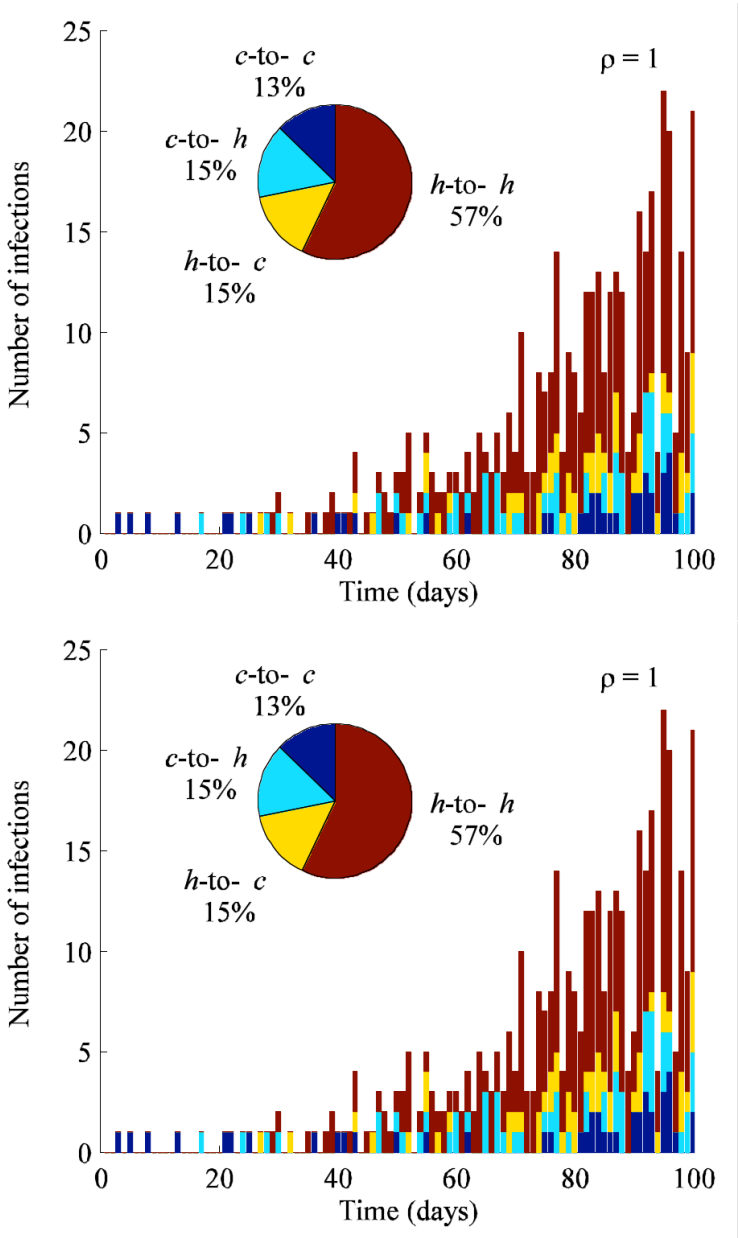

FIGURE 3. Daily incidence of infections are plotted for two stochastic epidemics with identical disease parameters ($R_0 = 3$) and control measures (no quarantine, but daily probability $h_c = 0.3$ of isolating an infectious community member, $h_h = 0.9$ of isolating an infectious HCW) implemented 14 days into the outbreak. Plots differ only in HCW(h) and community (c) mixing precautions (*i.e.*, $\rho = 1$ in **A** and $\rho = 0.1$ in **B**). Inset, pie-charts show average contributions of the different routes of infection for 500 stochastic simulations of each epidemic (after Lloyd-Smith, Galvani, and Getz, 2003).

and an anonymous reviewer for comments that have lead to improvements of this chapter.

References

[1] R. Anderson and R. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991.

[2] D. Anderson and R. Watson, On the Spread of a Disease with Gamma-Distributed Latent and Infectious Periods, *Biometrika* **67** (1980), 191–198.

[3] K. B. Athreya and P. Jagers, *Classical and Modern Branching Processes*, Springer-Verlag, Berlin, 1997.

[4] N. T. J. Bailey, *The Elements of Stochastic Processes*, Wiley, New York, 1964.

[5] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases*, Griffin, London, 1975.

[6] C. T. Bauch, J. O. Lloyd-Smith, M. Coffee, and A. P. Galvani, Dynamically modeling SARS and respiratory EIDs: past, present, future, *Epidemiology* (in review) (2005).

[7] N. E. Beckage, *Parasites and Pathogens: Effects on Host Hormones and Behavior*, Chapman and Hall, New York, 1997.

[8] M. Begon, M. Bennett, R. G. Bowers, N. P. French, S. M. Hazel, and J. Turner, A clarification of transmission terms in host-microparasite models: numbers, densities and areas, *Epidemiol. Infect.* **129** (2002), 147–153.

[9] A. T. Bharucha-Reid, On the Stochastic Theory of Epidemics, in (J. Neyman, ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **4**, 1956, 111–120.

[10] S. N. Busenberg and K. P. Hadeler, Demography and epidemics, *Math. Biosci.* **101** (1990), 63–74.

[11] D. J. Daley and J. Gani, *Epidemic Modeling: An Introduction*, Cambridge University Press, Cambridge, United Kingdom, 1999.

[12] M. C. M. De Jong, A. Bouma, O. Diekmann, and H. Heesterbeek, Modelling transmission: mass action and beyond, *Trends Ecol. Evol.* **17** (2002), 64–64.

[13] M. C. M. De Jong, O. Diekmann, and J. A. P. Heesterbeek, *How Does Transmission of Infection Depend on Population Size? Epidemic Models : Their Structure and Relation to Data*, Cambridge University Press, D. Mollison, 1995, 84–94.

[14] U. Dieckmann, J. A. J. Metz, M. W. Sabelis, and K. Sigmund, eds., *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, Cambridge Series in Adaptive Dynamics, Cambridge University Press, Cambridge, 2002.

[15] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*, Wiley, Chichester, 2000.

[16] K. Dietz and K. P. Hadeler, Epidemiological models for sexually transmitted disease, *Journal of Mathematical Biology* **26** (1988), 1–25.

[17] C. A. Donnelly, A. C. Ghani, G. M. Leung, A. J. Hedley, C. Fraser, S. Riley, L. J. Abu-Raddad, L. M. Ho, T. Q. Thach, P. Chau, K. P. Chan, T. H. Lam, L. Y. Tse, T. Tsang, S. H. Liu, J. H. B. Kong, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong, *Lancet* **361** (2003), 1761–1766.

[18] P. D. Enko, On the course of epidemics of some infectious diseases, *Vrac* **46** (1889), 1008–1010, **47**, 1039–43, **48**, 1061–1063 (For an abridged translation see: *International Journal of Epidemiology* **18**, 749–455).

[19] P. E. M. Fine, The interval between successive cases of an infectious disease, *Am. J. Epidemiol.* **158** (2003), 1039–1047.

[20] C. Fraser, S. Riley, R. M. Anderson, and N. M. Ferguson, Factors that make an infectious disease outbreak controllable, *Proc. Natl. Acad. Sci. U. S. A.* **101** (2004), 6146–6151.

[21] W. M. Getz, A hypothesis regarding the abruptness of density dependence and the growth rate of populations, *Ecology* **77** (1996), 2014–2026.

[22] W. M. Getz and J. Pickering, Epidemic models - thresholds and population regulation, *Am. Nat.* **121** (1983), 892–898.

[23] W. M. Getz, J. O. Lloyd-Smith, P. C. Cross, S. Bar-David, P. L. Johnson, T. C. Porco, M. S. Sánchez, Modeling the invasion and spread of contagious disease in heterogeneous populations, this volume.

[24] B. T. Grenfell and A. P. Dobson, eds., *Ecology of Infectious Diseases in Natural Populations*, Publications of the Newton Institute, Cambridge University Press, Cambridge, England, 1995.
[25] T. E. Harris, *The Theory of Branching Processes*, Springer, Berlin, 1963.
[26] J. A. P. Heesterbeek and J. A. J. Metz, The saturating contact rate in marriage– and epidemic models, *J. Math. Biol.* **31** (1993), 529–539.
[27] J. A. P. Heesterbeek, A brief history of R-0 and a recipe for its calculation, *Acta Biotheor.* **50** (2002), 189–204.
[28] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* **42** (2000), 599–653.
[29] P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek and A. P. Dobson, eds., *The Ecology of Wildlife Diseases*, Oxford University Press, Oxford, 2002.
[30] V. Isham and G. Medley, *Models for Infectious Human Diseases*, Cambridge University Press, 1996.
[31] M. J. Keeling and B. T. Grenfell, Disease extinction and community size: Modeling the persistence of measles, *Science* **275** (1997), 65–67.
[32] M. J. Keeling and B. T. Grenfell, Effect of variability in infection period on the persistence and spatial spread of infectious diseases, *Math. Biosci.* **147** (1998), 207–226.
[33] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Royal Soc. Lond. A* **115** (1927), 700–721.
[34] J. M. Kiesecker, D. K. Skelly, K. H. Beard, and E. Preisser, Behavioral reduction of infection risk, *Proc. Natl. Acad. Sci. U. S. A.* **96** (1999), 9165–9168.
[35] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray, Transmission dynamics and control of severe acute respiratory syndrome, *Science* **300** (2003), 1966–1970.
[36] A. L. Lloyd, Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods, *Proc. Royal Soc. Lond. B* **268** (2001), 985–993.
[37] A. L. Lloyd, Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, *Theor. Popul. Biol.* **60** (2001), 59–71.
[38] J. O. Lloyd-Smith, P. C. Cross, C. J. Briggs, M. Daugherty, W. M. Getz, J. Latto, M. S. Sanchez, A. B. Smith, A. Swei, Population thresholds for disease invasion and persistence in natural populations, to appear in *Trends in Ecology and Evolution* (2005).
[39] J. O. Lloyd-Smith, A. P. Galvani, W. M. Getz, Curtailing SARS transmission within a community and its hospital, *Proc. Royal Soc. Lond. B* **270** (2003), 1979–1989.
[40] J. O. Lloyd-Smith, H. V. Westerhoff, and W. M. Getz, Frequency-dependent incidence in sexually-transmitted disease models: Portrayal of pair-based transmission and effects of illness on contact behaviour, *Proc. Royal Soc. Lond. B* **271** (2004), 625–634.
[41] D. Ludwig, Final size distributions for epidemics, *Mathematical Biosciences* **23** (1975), 33–46.
[42] D. Ludwig, Qualitative behavior of stochastic epidemics, *Mathematical Biosciences* **23** (1975), 47–73.
[43] H. McCallum, N. Barlow, and J. Hone, How should pathogen transmission be modelled?, *Trends in Ecology and Evolution* **16** (2001), 295–300.
[44] H. McCallum, N. Barlow, and J. Hone, Modelling transmission: mass action and beyond – Response from McCallum, Barlow and Hone, *Trends Ecol. Evol.* **17** (2002), 64–65.
[45] D. Mollison, *Epidemic models: their structure and relation to data*, Cambridge University Press, 1995.
[46] I. Nåsell, Stochastic models of some endemic infections, *Mathematical Biosciences* **179** (2002), 1–19.
[47] R. E. Plant and L. T. Wilson, Models for age-structured populations with distributed maturation rates, *Journal of Mathematical Biology* **23** (1986), 247–262.
[48] T. C. Porco, K. A. Holbrook, S. E. Fernyak, D. L. Portnoy, R. Reiter, and T. J. Aragón, Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model, *BMC Public Health* **4** (2004), 34.
[49] S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L.-M. Ho, T.-H. Lam, T. Q. Thach, P. Chau, K.-P. Chan, S.-V. Lo, P.-Y. Leung, T. Tsang, W. Ho, K.-H. Lee, E. M. C. Lau, N. M. Ferguson, and R. M. Anderson, Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions, *Science* **300** (2003), 1961–1966.

[50] M. A. Schiltz and T. G. M. Sandfort, HIV-positive people, risk and sexual behaviour, *Social Sci. Med.* **50** (2000), 1571–1588.
[51] H. E. Soper, The interpretation of periodicity in disease prevalence, *J. Royal Statistical Society* **92** (1929), 34–73.
[52] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, third ed., Academic Press, San Diego, 1998.
[53] H. R. Thieme, *Mathematics in Population Biology*, Princeton University Press, Princeton, New Jersey, 2003.
[54] Y. H. Wang and A. P. Gutierrez, An assessment of the use of stability analysis in population ecology, *J. Animal Ecology* **49** (1980), 435–452.
[55] R. Watson, On the size distribution for some epidemic models, *J. Applied Probability* **17** (1980), 912–921.

Department of Environmental Science, Policy and Management, 140 Mulford Hall #3112, University of California, Berkeley, CA 94720-3112, United States and Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa

E-mail address: getz@nature.berkeley.edu

Biophysics Graduate Group, University of California at Berkeley