



Some properties of a simple stochastic epidemic model of SIR type

Henry C. Tuckwell^{a,*}, Ruth J. Williams^b

^a *Max Planck Institute for Mathematics in the Sciences Inselstr. 22, Leipzig D-04103, Germany*

^b *Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA*

Received 27 May 2005; received in revised form 1 May 2006; accepted 20 September 2006

Available online 11 October 2006

Abstract

We investigate the properties of a simple discrete time stochastic epidemic model. The model is Markovian of the SIR type in which the total population is constant and individuals meet a random number of other individuals at each time step. Individuals remain infectious for R time units, after which they become removed or immune. Individual transition probabilities from susceptible to diseased states are given in terms of the binomial distribution. An expression is given for the probability that any individuals beyond those initially infected become diseased. In the model with a finite recovery time R , simulations reveal large variability in both the total number of infected individuals and in the total duration of the epidemic, even when the variability in number of contacts per day is small. In the case of no recovery, $R = \infty$, a formal diffusion approximation is obtained for the number infected. The mean for the diffusion process can be approximated by a logistic which is more accurate for larger contact rates or faster developing epidemics. For finite R we then proceed mainly by simulation and investigate in the mean the effects of varying the parameters p (the probability of transmission), R , and the number of contacts per day per individual. A scale invariant property is noted for the size of an outbreak in relation to the total population size. Most notable are the existence of maxima in the duration of an epidemic as a function of R and the extremely large differences in the sizes of outbreaks which can occur for small changes in R . These findings have practical applications in controlling the size and duration of epidemics and hence reducing their human and economic costs.

© 2006 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail addresses: tuckwell@mis.mpg.de (H.C. Tuckwell), williams@math.ucsd.edu (R.J. Williams).

Keywords: SIR; Epidemic model

1. Introduction

Infectious diseases are an important and often dramatic cause of human illness and mortality across the globe. New diseases, such as ebola, severe acute respiratory syndrome (SARS), avian or bird ‘flu’, and West Nile virus emerge, and historically significant diseases, such as diphtheria and polio, re-emerge. Smallpox, considered to have been driven to extinction many years ago, has re-emerged as a threat due to the possibility of bioterrorists procuring laboratory samples of the bacterium (see [1], for a quantitative analysis). Human immunodeficiency virus (HIV) is currently threatening to cause more deaths than the great outbreaks of plague in the 14th century (of the order of 25 million deaths or one in four Europeans at the time) and influenza, which caused about 20 million deaths in the early 20th century [2]. Furthermore, epidemics in agricultural animals may have catastrophic economic consequences such as in the recent outbreaks of foot and mouth disease in Britain.

Some well-known classic models of infectious disease population dynamics have been deterministic (see for example [3]). General models, such as the SIR (susceptible, infective, recovered) differential equation model of Kermack and McKendrick [4] have proven useful in ascertaining gross factors affecting rate of growth and final size of an epidemic. However, it seems apparent that the nature of epidemic growth and spread is for the most part stochastic. Probabilistic models have indeed a long and illustrious history going back to Bernoulli [5] and earlier. Reviews of stochastic epidemiological models are contained in [6–8].

It is apparent that some diseases do not fit general simplified schemes and require special consideration of their details as they have characteristic modes of transmission as is the case for malaria. Our approach is in accordance with the views expressed in Isham [8], namely that simple models may be nevertheless useful for understanding underlying principles. In this paper we consider therefore a discrete time, discrete state space stochastic model which includes certain elements of reality, thus extending previous similar models.

There are two classical discrete time stochastic models, both of the so called chain-binomial type. These are the Greenwood [9] model and the Reed-Frost model, which evidently was proposed in 1928 in biostatistics lectures at Johns Hopkins, not published by the proponents but subsequently related in [10]. In these models there are successive generations, indexed by $t = 0, 1, 2, \dots$, of infectives which are only capable of infecting susceptibles for one generation after which they do not participate in the epidemic process. Suppose the population size is n , a constant and let the numbers of susceptibles and (new) infectives of generation t be $X(t)$ and $Y(t)$, respectively. Then the initial condition is $X(0) + Y(0) = n$ and $X(t+1) + Y(t+1) = X(t)$, $t = 0, 1, 2, \dots$, as the infectives and susceptibles of generation $t+1$ are drawn from the susceptibles of generation t . Thus,

$$X(t) + \sum_{j=0}^t Y(j) = n,$$

and the total number infected up to and including generation t is $\sum_{j=0}^t Y(j)$. It is assumed that the number of infectives of generation $t + 1$ is a binomial random variable with parameters $X(t)$ and $p(Y(t))$, the latter being the probability that an existing susceptible will become infected when the number of infectives is $Y(t)$. Thus,

$$P(Y(t+1) = k | X(t) = x, Y(t) = y) = \binom{x}{k} p(y)^k (1 - p(y))^{x-k},$$

for $k = 0, 1, \dots, x$. In the Greenwood model, $p(y) = p$ is a constant not depending on the number y of infectives. In the Reed-Frost model it is supposed that the probability any susceptible escapes being infected when there are y infectives is

$$1 - p(y) = (1 - p)^y,$$

where $p = p(1)$ is the probability a susceptible is infected by one given infective.

The Reed-Frost model has been used to analyze data on meningococcal disease [11]. It has been extensively employed in the analysis of agricultural epidemics such as foot and mouth disease in Japanese beef cattle [12], tuberculosis in Argentinian dairy cows [13,14] and Swedish deer [15]. Despite its apparent simplicity, the Reed-Frost model is not readily analyzed for large n , so that approximations have been sought, such as branching processes in the early stages and a normal approximation for estimating the final size of the epidemic; i.e., the total number infected [16]. However, Ball and O'Neill [19] have succeeded, via a construction of the epidemic process due to Sellke [20], to find the distribution of the final size of an epidemic in the Reed-Frost and other models – see also [21]. See [22] for some generalizations of the Reed-Frost model with application to HIV. A review of continuous time models may be found in [20].

We will explore a mathematical model which incorporates some important features of disease transmission in a discrete time stochastic framework. One of these concerns the group of individuals encountered by a given individual on a particular day. The most realistic situation would make this group consist of a core subgroup which was met almost on a daily basis, such as family members or colleagues, together with a random subgroup whose numbers and composition would change each day, consisting of persons met in travelling or other activities, such as sporting events, shopping or entertainment. Some interesting results have recently been found for models with some such features [17,23] but here it was decided to simplify the model by making the group met by each individual not necessarily the same each day, consisting of a fixed number plus a random number, all being chosen at random from the rest of the population. The second feature consists of a period of R days after the infection of an individual such that only during this period is the individual infected and capable of infecting susceptible individuals. We will mainly be concerned with ascertaining the total number of diseased individuals and how long it takes for the disease to vanish from the population (if ever). The case $R = \infty$ is considerably simpler, so we give some analytical formulas for this case and consider a formal diffusion approximation for the number infected as a random function of time.

2. Description of the model

The model we employ is similar to that of Reed-Frost but has some modifications to make it more realistic and adaptable for different diseases. Because the time scale for data on epidemics is

usually daily at its finest, it is natural to use a discrete time model with a time step which is usually thought of as one day, although in some applications the time step is taken as several days [12].

2.1. Assumptions

We consider a relatively simple stochastic SIR model with assumptions as follows:

- (a) The total population size is fixed at n .
- (b) Time is discrete, with epochs $t = 0, 1, 2, \dots$. The natural unit for the duration of an epoch is one day.
- (c) For individual i , $i = 1, \dots, n$, the random process $Y^i = \{Y^i(t), t = 0, 1, 2, \dots\}$ is such that $Y^i(t) = 1$ if that individual is infected and capable of infecting others (called diseased or infectious) at time t ; otherwise $Y^i(t) = 0$. Thus the total number of diseased and hence infectious individuals at time t is

$$Y(t) = \sum_{i=1}^n Y^i(t), \quad t \geq 0.$$

- (d) Individual i encounters a fixed number (not random) n_i of other individuals each day, drawn randomly from the population. Individual i also meets a randomly chosen and random number $M_i(t)$ of other individuals over $(t, t + 1]$. The variables $M_i(t)$ are mutually independent and independent of the state of the population. These random variables may be, for example, uniformly distributed or have specially tailored discrete distributions to represent as accurately as possible chance meetings in human populations. The total number of individuals met by person i over $(t, t + 1]$ is thus $N_i(t) = n_i + M_i(t)$. An alternative way to view this is that individual i never has less than n_i contacts. We here consider time homogeneous models so that the distribution of $M_i(t)$ is the same for all t .
- (e) If an individual becomes infective, he remains in such a state for R consecutive time points including the initial time point of becoming infected where R is a positive integer constant. (In general R could be a random variable, or even a random process, but this complication is ignored throughout.) Thus, if an individual is diseased for the first time at epoch t , then he is diseased and infectious for the epochs $\{t, t + 1, \dots, t + R - 1\}$. At epoch $t + R$ such an individual recovers but cannot be re-infected. (In real time, if an individual is susceptible at time $t - 1$ and infected at time t then it is assumed that he became infected somewhere in the interval $(t - 1, t]$). For example, if $R = 2$ and individual i becomes infected at some time, then $Y^i(t)$, $t = 0, 1, 2, \dots$ is a string of zeros except for two consecutive time points at which there are ones. We call R the recovery period, although it could equally well be called the infectious period. We also consider the case $R = \infty$ which gives no recovery and hence reduces the model to one of SI type rather than SIR.
- (f) If an individual who has never been diseased up to and including time t encounters an individual in $(t, t + 1]$ who is diseased at time t , then independently of the results of other encounters, this encounter results in transmission of the disease with probability $p \in [0, 1]$, whereupon the individual is infected at epoch $t + 1$.

- (g) Given $Y(t)$, the probability that a randomly chosen individual is diseased at time t is given by $f(t) = \frac{Y(t)}{n}$.

2.2. Description as a Markov chain

If all of the $N_i(t)$ are independent and identically distributed, the model can be construed as an $(R + 1)$ -dimensional Markov chain as follows: for each $t \geq 0$ let

$Y_i(t)$ be the number of individuals who are infected at t and have been infected for exactly i time units, $i = 0, 1, \dots, R - 1$;

$X(t)$ be the number of susceptible individuals at time t ; and let $Z(t)$ be the number of individuals who were previously infected and are recovered at t .

We assume that all of the individuals who are infected at $t = 0$ have just become infected so that $Y(0) = Y_0(0)$ and $Y_i(0) = 0$ for $i = 1, \dots, R - 1$. Also, there are no recovered individuals at $t = 0$ so that $Z(0) = 0$ and $Y_0(0) + X(0) = n$. It is feasible of course that some or all of the initially infected could have been infected prior to $t = 0$. This could be the case if there are infected immigrants who have just entered the population, but we do not consider this possibility here.

Regardless of the initial conditions, it is clear that

$$\mathbf{V}(t) = (X(t), Y_0(t), Y_1(t), \dots, Y_{R-1}(t)), \quad t = 0, 1, 2, \dots \quad (1)$$

is a Markov chain. Note that the value of $Z(t)$ is known if all of the components of $\mathbf{V}(t)$ are known.

There are a number of further constraints on the components as follows:

for $t = 1, 2, \dots$,

$$Y_{k+1}(t) = Y_k(t - 1), \quad k = 0, 1, \dots, R - 2,$$

$$Z(t) = Z(t - 1) + Y_{R-1}(t - 1)$$

and the total number of infectives at t is

$$Y(t) = \sum_{k=0}^{R-1} Y_k(t)$$

so that $(X(t), Y(t), Z(t))$ gives the traditional (S, I, R) description.

In addition to the processes $Y^i = \{Y^i(t), t = 0, 1, 2, \dots\}$, $i = 1, 2, \dots, n$, such that $Y^i(t) = 1$ or 0 depending on whether individual i is infectious or not, it is convenient to introduce the processes X^i which indicate whether individual i is susceptible or not. If then $Z^i(t)$ indicates whether at epoch t individual i has been previously infected and is recovered and incapable of infecting others, then we must have for all i and for all t :

$$X^i(t) + Y^i(t) + Z^i(t) = 1, \quad (2)$$

where two of these variables are zero. Further we must have

$$X(t) = \sum_{i=1}^n X^i(t), \quad Y(t) = \sum_{i=1}^n Y^i(t), \quad \text{and} \quad Z(t) = \sum_{i=1}^n Z^i(t).$$

In general, if the variables $N_i(t)$ are not all identically distributed, in order to give a Markovian state descriptor, we define the processes, $Y_0^i, Y_1^i, \dots, Y_{R-1}^i$ such that $Y_k^i(t) = 1, k = 0, 1, \dots, R - 1$, if individual i was first infected at time $t - k$ and is zero otherwise. Hence

$$Y^i(t) = \sum_{k=0}^{R-1} Y_k^i(t).$$

Then we can use

$$\mathbf{X}(t) = [X^i(t), Y_0^i(t), \dots, Y_{R-1}^i(t), i = 1, \dots, n]$$

as a Markovian state descriptor and the model takes the form of a Markov chain with state space contained in $\{0, 1\}^{(R+1)n}$. Because of its simplicity, this Markov chain will be the one used in our simulations, even when the $N_i(t)$ are independent and identically distributed.

2.3. Transition probabilities

From the above assumptions, the one-step transition probabilities for the Markov chain \mathbf{X} may be written down. However, for an approach through simulation, in which we update the states of individuals at each time step, it is not necessary to catalogue the whole gamut of one-step transition probabilities as many, being deterministic transitions, are taken care of automatically in the simulation program.

At any given general time, t , say, assuming $R \geq 3$, individual i may be in any of $R + 2$ mutually exclusive states so that, one of the variables $X^i(t), Y_0^i(t), \dots, Y_{R-1}^i(t)$ and $Z^i(t)$ is unity whilst the others are zero. If any of the $R + 1$ variables $Y_0^i(t), \dots, Y_{R-1}^i(t), Z^i(t)$ is unity, then the values of all $R + 2$ variables $X^i(t + 1), Y_0^i(t + 1), \dots, Y_{R-1}^i(t + 1)$ and $Z^i(t + 1)$ are determined with probability one. Thus for example, $Z^i(t) = 1 \Rightarrow Z^i(t + 1) = 1$ and $X^i(t + 1) = Y_0^i(t + 1) = \dots = Y_{R-1}^i(t + 1) = 0$; similarly, $Y_0^i(t) = 1 \Rightarrow Y_1^i(t + 1) = 1$ and $X^i(t + 1) = Y_0^i(t + 1) = Y_2^i(t + 1) = \dots = Y_{R-1}^i(t + 1) = Z^i(t + 1) = 0$.

The only individual transition probability (that is not either zero or one) required to simulate the evolution of the process of disease spread is the probability that an individual i susceptible at t becomes infected for the first time at $t + 1$. This probability depends only on the total number $Y(t) = y$ of diseased individuals together with the number $N_i(t)$ of individuals met and the probability p of transmission per contact. Now, assuming n is much greater than $N_i(t)$, so that the binomial approximation may be used, the probability of meeting exactly j infectives is,

$$P_j^i(y, N_i(t); n) \approx \binom{N_i(t)}{j} \left(\frac{y}{n-1}\right)^j \left(1 - \frac{y}{n-1}\right)^{N_i(t)-j} \tag{3}$$

and the probability p_j of becoming infected if j infectives are met is

$$p_j = 1 - (1 - p)^j.$$

Then,

$$P(Y_0^i(t + 1) = 1 | N_i(t), X^i(t) = 1, Y(t) = y, \dots) = \sum_{j=1}^{N_i(t)} p_j P_j^i(y, N_i(t); n), \tag{4}$$

which simplifies (using (3) as an equality) to

$$P(Y_0^i(t+1) = 1 | N_i(t), X^i(t) = 1, Y(t) = y, \dots) = 1 - \left(1 - \frac{py}{n-1}\right)^{N_i(t)} \quad (5)$$

(cf. Eq. (2) of [22]). This also follows because if individual i is susceptible at t , then in $N_i(t)$ independent Bernoulli trials, with a probability of infection on each of $\frac{py}{n-1}$, the probability that the individual does not become infected is $(1 - \frac{py}{n-1})^{N_i(t)}$. Note that this model contains a simplification (which is commonly used), namely, the meeting relationship is not symmetric because if the group randomly chosen to meet individual i contains individual j , the group chosen to meet individual j need not contain individual i .

2.4. Variability of size and duration: simulations

To report results for large ranges of all the parameters in the model would take up much space so the presentation is curtailed by this constraint. In particular, throughout this paper, we report results only for the case in which the $N_i(t)$ are independent and identically distributed. In Section 3, we will consider the simple case of no recovery ($R = \infty$) and in Section 4 we give mean statistics only for finite recovery times. In Figs. 1 and 2 we illustrate the stochastic nature of the epidemic, where we have $N_i(t) = N$, N being a fixed constant. For these simulation results we have chosen the following parameter values. The population size is $n = 200$, the probability of transmission of the disease on contact of a susceptible with an infected is $p = 0.1$, the number of contacts per person per time step (day) is $N = 4$, the fraction of the population initially infected is 0.01 so that $Y(0) = 2$ and the recovery period is $R = 2$ days.

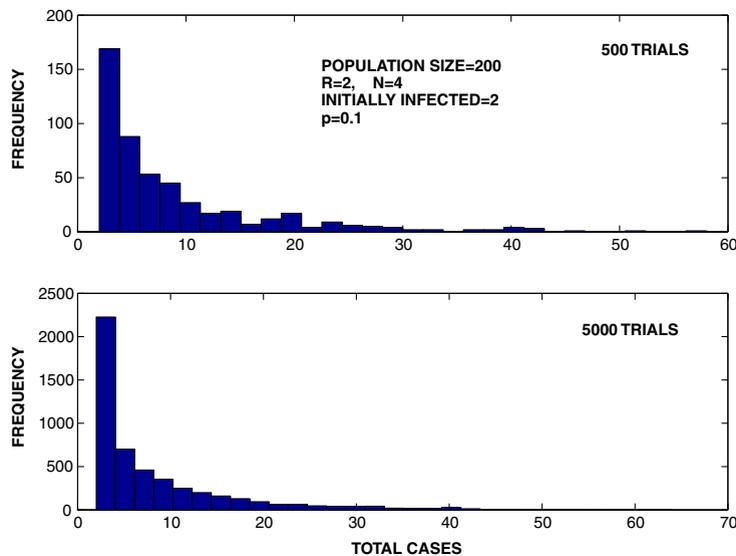


Fig. 1. Histograms of the total number of individuals ever infected for the SIR model; for parameter values see the figure. The upper histogram is obtained with 500 trials whereas the lower one is obtained with 5000 trials.

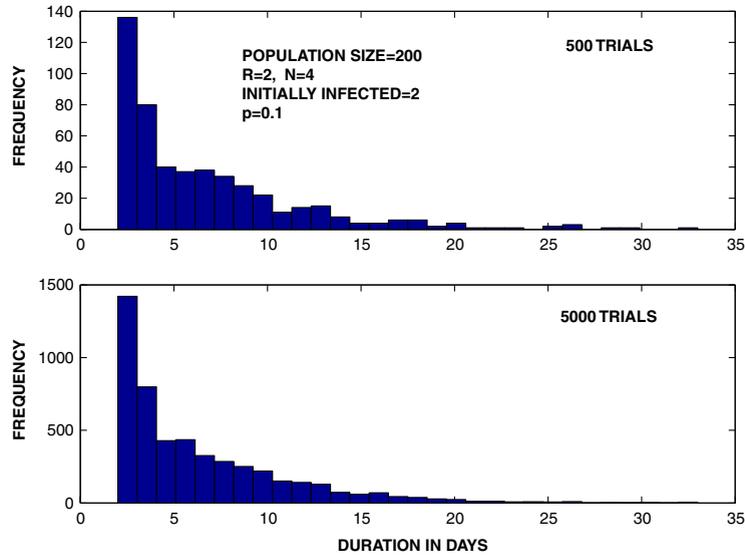


Fig. 2. Histogram of the duration in days of the SIR epidemic model; for parameter values see the figure. The upper histogram is obtained with 500 trials whereas the lower one is obtained with 5000 trials.

Fig. 1 shows the empirical (simulated) distribution (histogram) of the total number of infected individuals at the end of the epidemic. The number of trials for the upper histogram is 500, whereas for the lower histogram it is 5000. We report empirical statistics for the latter case. The maximum number of cases was 64 and the minimum number was 2 (for the latter, there were no new infections beyond the initially infected individuals). The mean number of total cases was 8.39, the standard deviation was 8.44 and the most frequent occurrence was that of no new cases (size 2). The most notable feature is that the same parameter set can lead to either zero or very few new cases or to a large outbreak in which nearly one third of the population becomes infected. This important effect could of course not be discerned with a deterministic model.

Fig. 2 shows the corresponding sets of results (500 and 5000 trials) for the duration of the epidemic, defined as the time required to reach an epoch in which there are no current infectives. The empirical statistics, based on 5000 trials, are as follows. The minimum duration was 2 days (recovery of the initially infected and no new cases) and the maximum duration was 33 days with a mean of 6.52 days and a standard deviation of 4.64 days. The most likely occurrence was a duration of 2 days (no new cases). The variability in the duration is as striking as for the size of the epidemic, especially considering that there is no variability in the number of daily contacts.

In the results shown in Figs. 1 and 2 we have employed samples of 500 and of 5000. The larger sample size is included in order to give a better indication of the underlying distributions, which could be obtained within a theoretical framework. If we consider the process $\{(X(t), Y_0(t), Y_1(t)), t = 0, 1, 2, \dots\}$ with initial value $(n - 1, 1, 0)$ and state space $\{(x, y_0, y_1) \in \mathbb{Z}_+^3 : x + y_0 + y_1 \leq n\}$, then the duration is the time T to absorption of the process on the x -axis, with $Y_0(T) = Y_1(T) = 0$, and the total number infected is $n - X(T)$. Although an analytical approach via Markov chain theory is potentially feasible to find the distributions of T and $X(T)$, the formulas are so unwieldy that we restrict our attention to estimation by simulation.

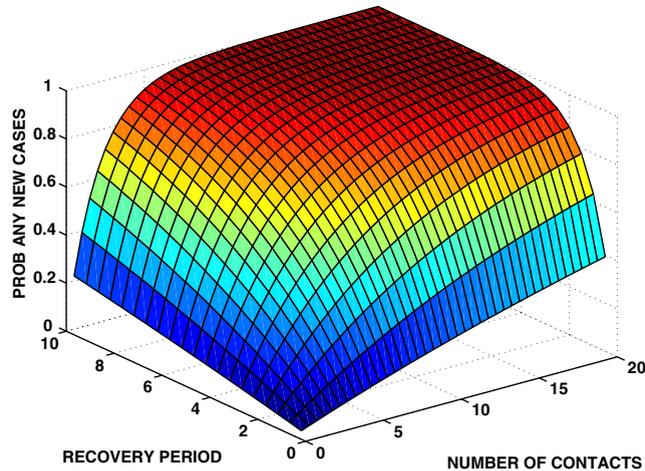


Fig. 3. A 3-D plot of $1 - Q_0$ against the number of contacts per day (n_i) and the recovery period (R). Q_0 is given by formula (6).

2.5. The probability of infections beyond those initially infected

Given that there are y_0 initially infected and a recovery period of length R , from the previous expressions we can readily determine an expression for the probability Q_0 that the disease does not spread to any new individuals other than those initially infected. This must be, in the case of fixed numbers $N_i(t) = n_i$ of contacts for the i th susceptible,

$$Q_0 = \left[\prod_{i=1}^{n-y_0} \left(1 - \frac{py_0}{n-1} \right)^{n_i} \right]^R. \quad (6)$$

For the parameter values $y_0 = 2$, $R = 2$, $p = 0.1$, $n = 200$ as in Figs. 1 and 2, and all $n_i = 4$ this gives $Q_0 = 0.203$ which compares favorably with the fraction of trials, namely 20.02% in the data of Figs. 1 and 2 (5000 trials) in which there were no new cases or the duration was 2 days. (Note that the bin widths in Figs. 1 and 2 are not unity.) As a further illustration, we have plotted in Fig. 3, as a function of number of contacts and the length of the recovery period, the probability (computed from (6)) of having any new cases (that is $1 - Q_0$) at fixed values of $p = 0.05$, $Y(0) = 1$ and $n = 200$.

3. The model without recovery ($R = \infty$)

A simplifying assumption is that infectious individuals remain infectious throughout the course of the epidemic. Such a situation can arise when a disease causing agent has a long life as with tuberculosis in deer [15] or with HIV in humans [23], especially with life-prolonging drug therapies. Without recovery the number of infectives at time t is a classical discrete time, discrete state space Markov chain, for which the one-step transition probabilities can be written down explicitly. We assume that at time t there are $Y(t) = y$ infectious individuals which, as there are no

recovered individuals, implies that there are $n - y$ susceptibles. There are two cases we wish to consider: (i) the number of individuals met is random; and (ii) the number of individuals met is constant, with no random component.

3.1. The number of meetings for each individual is random

We remind the reader that we are assuming that the $N_i(t)$ are independent and identically distributed. With $Y(t) = y$ infectives, given the value of $N_i(t)$, the probability that the i th susceptible, $i = 1, 2, \dots, n - y$, meets exactly j infectives is given by (3) and the probability that this individual is newly infected by $t + 1$ is given by (5). Thus, the probability of no new infectives at $t + 1$ is

$$P(Y(t + 1) = y | N_i(t), i = 1, \dots, n - y; Y(t) = y) = \prod_{i=1}^{n-y} \left(1 - \frac{py}{n-1}\right)^{N_i(t)} \tag{7}$$

and the probability of one new infective is

$$\begin{aligned} P(Y(t + 1) = y + 1 | N_i(t), i = 1, \dots, n - y; Y(t) = y) \\ = \sum_{i=1}^{n-y} \left[1 - \left(1 - \frac{py}{n-1}\right)^{N_i(t)}\right] \prod_{j=1, j \neq i}^{n-y} \left(1 - \frac{py}{n-1}\right)^{N_j(t)}. \end{aligned} \tag{8}$$

Expressions can be written down for the chances of larger increments in the number of infectives, $Y(t + 1) - Y(t) = \tilde{y}$, $\tilde{y} = 2, 3, \dots$ but they are unwieldy – see below for a more manageable case.

3.2. The number of meetings per individual is constant

If all susceptibles meet the same constant (non-random) number of individuals N per epoch then each susceptible has the same chance to become infected. This is equivalent to a Reed-Frost model, modified so that the number of individuals met by an infective is not the group of all susceptibles but a subset of them [23].

Using (5) with $N_i(t) = N$, we find that the probability that any individual who is susceptible at t is infected at $t + 1$, when the number of infectives at time t is y , is

$$\tilde{p} = 1 - \left(1 - \frac{py}{n-1}\right)^N.$$

Thus, the distribution of the increment in the number of infectives must be

$$P(Y(t + 1) = y + \bar{y} | Y(t) = y) = \binom{n-y}{\bar{y}} \left(1 - \left\{1 - \frac{py}{n-1}\right\}^N\right)^{\bar{y}} \left(1 - \frac{py}{n-1}\right)^{N(n-y-\bar{y})}, \tag{9}$$

where $\bar{y} = 0, 1, 2, \dots$. Then the increment in the number of infectives has a mean given by

$$E[Y(t + 1) - Y(t) | Y(t) = y] = (n - y) \left(1 - \left\{1 - \frac{py}{n-1}\right\}^N\right) \tag{10}$$

and its variance is

$$\text{Var}[Y(t + 1) - Y(t) | Y(t) = y] = (n - y) \left(1 - \left\{1 - \frac{py}{n-1}\right\}^N\right) \left\{1 - \frac{py}{n-1}\right\}^N. \tag{11}$$

3.2.1. *A diffusion approximation*

The computations (10) and (11) for the mean and variance of the one-step increments of Y suggest that for a large population size n and small transmission probability p such that nNp is of moderate size, one might approximate a suitably rescaled version of Y by a diffusion process. More precisely, if we speed up time and rescale the state to define

$$\hat{Y}^n(t) = \frac{Y([nt])}{n} \quad \text{for all } t \geq 0,$$

where $[\cdot]$ denotes the greatest integer part, then $\hat{Y}^n(t)$ is the *fraction* of the population that has been infected by the time $[nt]$ in the original time scale of Y . From (10) and (11), for large n and small p such that $\theta = nNp$ is of moderate size, using the approximation $1 - (1 - x)^N \approx Nx$ for small x , we see that with $\Delta t = \frac{1}{n}$ and $t = 0, \frac{1}{n}, \frac{2}{n}, \dots$,

$$E[\hat{Y}^n(t + \Delta t) - \hat{Y}^n(t) | \hat{Y}^n(t) = \hat{y}] \approx (1 - \hat{y})Np\hat{y} = \theta\hat{y}(1 - \hat{y})\Delta t$$

and

$$\text{Var}[\hat{Y}^n(t + \Delta t) - \hat{Y}^n(t) | \hat{Y}^n(t) = \hat{y}] \approx (1 - \hat{y})Np\hat{y}(1 - Np\hat{y})\Delta t \approx \frac{\theta}{n}\hat{y}(1 - \hat{y})\Delta t. \tag{12}$$

This suggests that one might approximate \hat{Y}^n by a diffusion process \hat{Y} that lives in $[0,1]$ and satisfies the stochastic differential equation

$$d\hat{Y}(t) = \theta\hat{Y}(t)(1 - \hat{Y}(t))dt + \sqrt{\frac{\theta}{n}\hat{Y}(t)(1 - \hat{Y}(t))}dW(t), \tag{13}$$

where $W = \{W(t), t \geq 0\}$ is a standard Wiener process with $W(0) = 0$, $E(W(t)) = 0$ and $\text{Var}(W(t)) = t$. (For proofs of similar approximations for continuous time Markov chains, see [18, Chapter 11].) To Eq. (13) there corresponds a forward (and backward) Kolmogorov equation satisfied by the transition probability density function $p(\hat{y}, t | \hat{y}_0)$ (see for example [24]):

$$\frac{\partial p}{\partial t} = -\theta \frac{\partial}{\partial \hat{y}} (\hat{y}(1 - \hat{y})p) + \frac{\theta}{2n} \frac{\partial^2}{\partial \hat{y}^2} (\hat{y}(1 - \hat{y})p). \tag{14}$$

In Fig. 4, statistical aspects of some simulations of the diffusion process \hat{Y} are compared with those for simulations of the original process. The parameter values are $n = 200$, $N = 4$, and $p = 0.05$, with 2 individuals initially infected. The figure shows the (empirical) stochastic means \pm one standard deviation (computed from 50 trials) for the number of infectives as a function of time for both Y and \hat{Y} . As noted in the caption to the figure, for ease of comparison, we have rescaled the diffusion plot so that what is shown is a graph corresponding to values of $n\hat{Y}(t/n)$ for $t = 0, 1, 2, \dots$

If the variability is small, so that the noise term in (13) has little effect, then it is natural to conjecture that the mean of $\hat{Y}^n(t)$ can be approximated by $\hat{m}(t)$ where \hat{m} satisfies the logistic equation

$$\frac{d\hat{m}}{dt} = \theta\hat{m}(1 - \hat{m}) \tag{15}$$

with solution

$$\hat{m}(t) = \frac{1}{1 + \frac{1-\hat{m}_0}{\hat{m}_0} \exp(-\theta t)}, \quad t \geq 0, \tag{16}$$

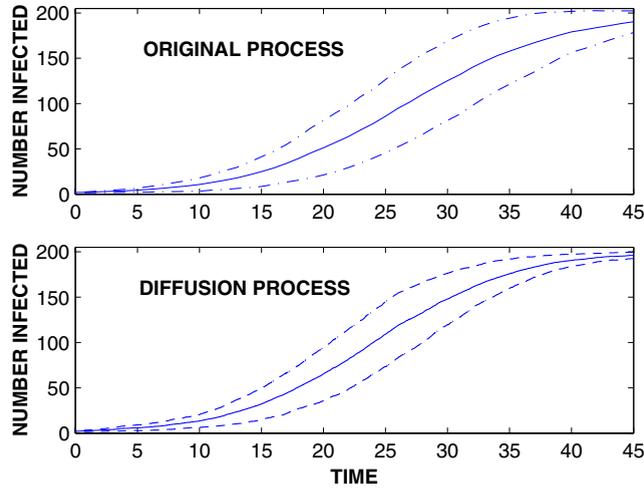


Fig. 4. A comparison of statistics for the diffusion approximation (13) and the original discrete Markov chain model, based on 50 trials, with a population size of $n = 200$, probability $p = 0.05$ of transmission of the disease on contact, $N = 4$ contacts per individual per time period and 2 initial infectives. The stochastic means \pm one standard deviation are plotted against time. To facilitate direct comparison of the two plots, in the diffusion plot the state value of the diffusion has been scaled up by n and the time has been rescaled by the same factor.

where $\hat{m}_0 = E[\hat{Y}^n(0)] = \frac{1}{n}E[Y(0)]$. This suggests that for $t = 0, 1, 2, \dots$ (with attendant scaling up of error terms),

$$E[Y(t)] = nE\left[\hat{Y}^n\left(\frac{t}{n}\right)\right] \approx n\hat{m}\left(\frac{t}{n}\right) = \frac{n}{1 + \frac{n-y_0}{y_0} \exp(-pNt)}, \tag{17}$$

where $y_0 = E[Y(0)]$.

Fig. 5 shows a comparison of values of the above logistic approximation with values of the stochastic mean computed from simulations of the discrete process Y for three values of N , the number of contacts per day. Here the parameters are $n = 500$, $p = 0.1$, the initial number infected is one, and there are 10 trials for each parameter set for the stochastic model.

3.2.2. Time to reach a given fraction of infectives

Having seen that the logistic can give a reasonably accurate estimate of the expected number of infected individuals as a function of time, it is interesting to ascertain roughly the dependence on the parameters p , n and N of the time taken for the number infected to reach a given fraction of the population. That is, we ask for the time t_α such that

$$\hat{m}\left(\frac{t_\alpha}{n}\right) = \alpha,$$

where $0 < \alpha < 1$ and where $\hat{m}(t)$ is given by (16). Substitution leads to an explicit solution

$$t_\alpha = \frac{1}{pN} \ln \left[\frac{\frac{n}{y_0} - 1}{\frac{1}{\alpha} - 1} \right]. \tag{18}$$

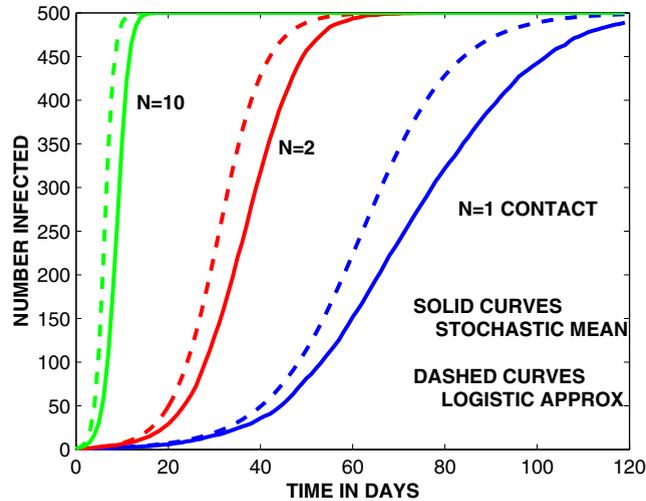


Fig. 5. Logistic curves for various numbers of contacts per day and the corresponding means obtained from simulations for the stochastic epidemic model without recovery.

We see therefore that under the logistic approximation:

- For a given population size, number of contacts per day per individual and number initially infected, the time for a fraction of the population to become infected is inversely proportional to the probability of infection on contact between a susceptible and an infective.
- For a given population size, probability of infection on contact between a susceptible and an infective and number initially infected, the time for a fraction of the population size to become infected is inversely proportional to the number of contacts per day. Furthermore, if it can be assumed that the ratio n/y_0 of total population size to the number initially infected is much larger than one, then we have approximately

$$t_\alpha \approx \frac{1}{pN} \left[\ln \left(\frac{n}{y_0} \right) - \ln \left(\frac{1}{\alpha} - 1 \right) \right]. \quad (19)$$

Then we also have

- For fixed N and p , the time taken for 50% of the population to become infected ($\alpha = 1/2$) is proportional to the logarithm of the reciprocal of the initial fraction of the population that is infected.

4. The model with recovery ($R < \infty$)

For $R < \infty$ the Markov chain model described in Section 2 is rather complicated for an analytical approach and hence results for it have been obtained by simulations, some of which are now described. We will describe results for the mean computed over several trials, based on simulation

of the population on an individual by individual basis using a Matlab program. Results showing the variability of the population response to the introduction of a few infected individuals have been given in Section 2.4.

4.1. Effects of various numbers of contacts

It is interesting to first examine the effects of varying the number of contacts per individual per day. In this section the number of contacts per day is held constant and denoted by N . Results are shown in Figs. 6 and 7 for two values of the recovery period, namely $R = 2$ and $R = 4$. In these figures, the mean numbers (over 25 trials) of infected individuals at time t are plotted against time t , assumed to be measured in days. The population size was chosen as 500 and the probability of transmission set at $p = 0.1$. There is initially just one diseased individual. Referring to Fig. 6, for a recovery period of $R = 2$, $E(Y(t))$ does not grow much past the initial number and diminishes to zero within several days for $N = 1$ and $N = 2$. For $R = 4$ and $N = 1$, the expected number of infected individuals becomes zero after about 15 days; for $R = 4$ and $N = 2$ the duration of the epidemic is prolonged substantially to as long as 30 days.

In Fig. 7, corresponding results for the larger contact rates $N = 5$ and $N = 10$ are shown. Here the results are somewhat unexpected as the times taken for $E(Y(t))$ to vanish are longer for fewer contacts per day $N = 5$ for both values of the recovery period. When $N = 5$ doubling the recovery period from 2 to 4 days has a very large effect on the maximum number of expected cases, taking it from a few to over 80. Similarly when $N = 10$, doubling the recovery period increases the maximum number of cases by a factor of about 4 but does not significantly change the time taken for $E(Y(t))$ to vanish. For these larger values of N it is seen that the larger N leads to a larger but shorter epidemic.

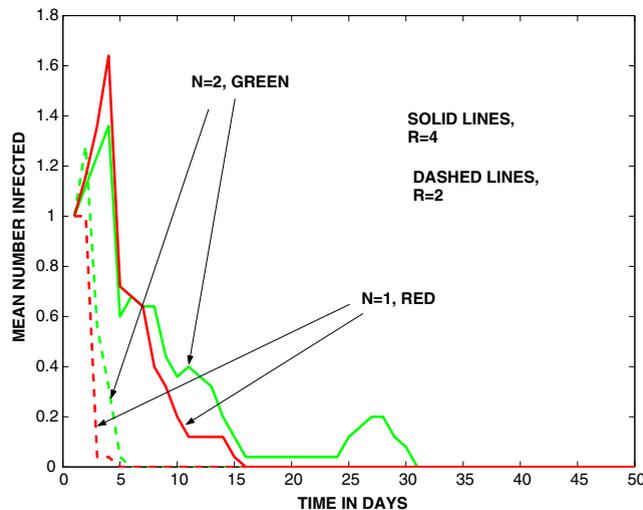


Fig. 6. A plot showing how the time course of the SIR epidemic depends on the number of contacts per day, here $N = 1$ and $N = 2$, and the recovery period which takes values $R = 2$ and $R = 4$. The mean number infected at time t is plotted against t .

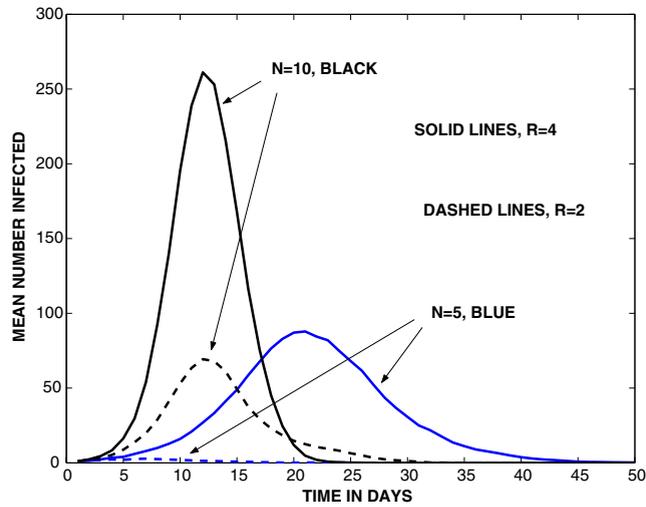


Fig. 7. A plot showing how the time course of the SIR epidemic depends on the number of contacts per day, here $N = 5$ and $N = 10$, and the recovery period which takes values $R = 2$ and $R = 4$. The mean number infected at time t is plotted against t .

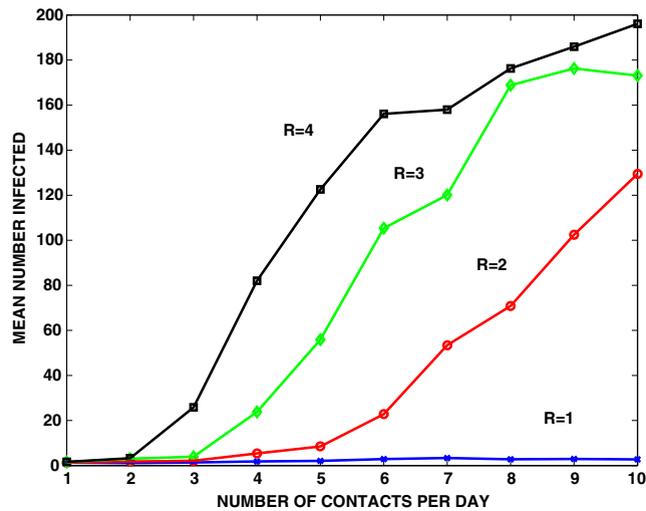


Fig. 8. A plot illustrating how the mean total size of the SIR epidemic depends on the number of contacts N , for recovery periods of $R = 1, 2, 3$ and 4 days.

Figs. 8 and 9 show the effects of increasing the number of contacts per day on the mean total number of cases and the mean total duration of the epidemic. The same values of n, p and $Y(0)$ were employed as for Figs. 6 and 7, and the averages are over 25 trials. Fig. 8 shows the steady increase in total number of cases at each of the values of the recovery period. Most noticeable, however, is the enormous difference between the sizes of the epidemic for intermediate values of N (4, 5 and 6) as the recovery period changes from 2 to 4. For example, when there are 5

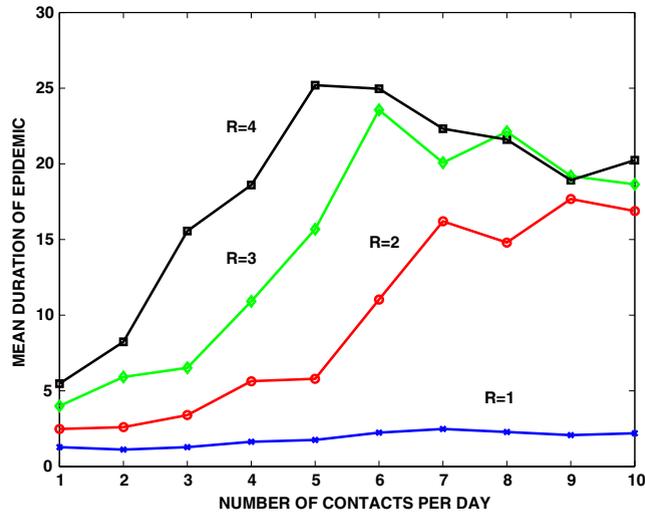


Fig. 9. A plot showing the dependence of the mean duration of the SIR epidemic as the number of contacts N varies for recovery periods of $R = 1, 2, 3$ and 4 days.

contacts per day, the mean total number afflicted is about 5 with $R = 2$ but is about 120 for $R = 4$; similarly, with $N = 6$, there is a mean total number of cases of just less than 20 with $R = 2$ but this becomes nearly 160 if $R = 4$. Fig. 9 shows the duration of epidemics corresponding to the results of Fig. 8. For $R = 1$ there is little change in the duration as the number of contacts per day increases. When $R = 2$ the duration increases quite rapidly and achieves a maximum (indicated in the figure as being at about $N = 9$) before declining at large values of N . When $R = 4$, a maximum is apparently achieved at about only 5 contacts per individual per day.

The results of Fig. 8 suggest that when the number of contacts is small (less than 3 per day per individual) there is little benefit in reducing R . A similar conclusion might be drawn when N is large (greater than 9 per day per individual). For intermediate numbers of contacts per day, large reductions in total number of infected individuals can be effected by reducing the duration of the recovery (infectious) period. This has implications for both pharmacological intervention and other treatments that accelerate recovery or for social policy in which afflicted individuals are taken out of circulation when sick, possibly on a volunteer basis, thus effectively reducing R and/or N .

4.2. Scale invariance

An important aspect of the model that we wished to consider was how the development of an epidemic might differ qualitatively and quantitatively as the total population size varied. Although population size may be quite small in isolated animal herds, or even isolated human settlements, urban populations often involve much greater numbers. The simulation of such large populations with the present model, and probably any reasonably accurate model, is very time consuming so it is important to know whether the behavior of solutions for relatively small populations is a reliable indicator of that for larger ones. Fig. 10 shows results for the final fraction of the population that is infected for populations of sizes $n = 100, 500$ and 1000 for various initial fractions of the

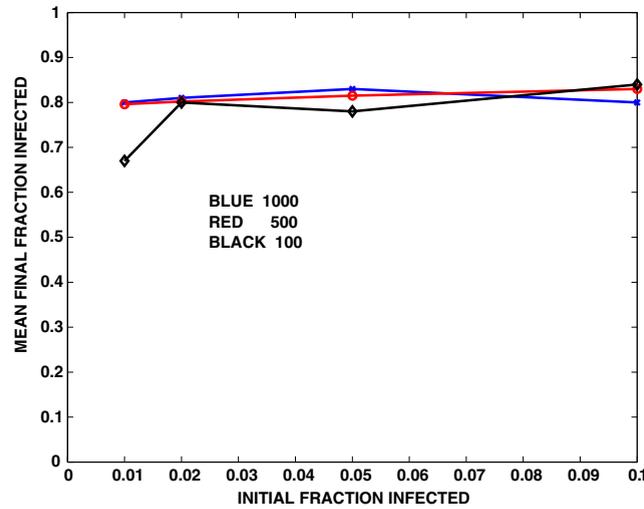


Fig. 10. Results for the stochastic SIR model showing the relative invariance of the mean final fraction infected with regard to both the initial fraction infected and the population size. For parameter values see text.

population infected. In obtaining these results the number of contacts made by each individual per time unit was random, with a distribution as specified below.

The remaining parameters for these trials were recovery period $R = 2$ days (epochs), probability of transmission of disease on contact between a susceptible and an infective, $p = 0.1$; and the numbers of contacts per individual were all $5 + U$ where U is uniform on $[0, 1, \dots, 10]$. The results are the means for 50 trials.

Here it is seen that for different populations there are significant differences in the final fraction infected for small initial fractions (< 0.02) but only when the population is less than 500. Otherwise, the final fraction infected is practically the same (for 500, 1000) and increases gently as the initial fraction infected increases. Beyond an initial infected fraction of 0.02 the final fraction infected is practically independent of population size for all values considered.

The mean duration of an epidemic is shown in Fig. 11 as a function of the initial fraction of infected individuals for the same set of population sizes and remaining parameters as for Fig. 10. A stronger dependence on population size is found for the duration than for the final fraction infected. When the initial fractions infected are the same, as the population size increases, the mean duration of the epidemic increases. An explanation is sought in terms of the number of effective contact operations required. Suppose the population size is n , the initial fraction infected is ρ_1 and the final fraction infected is ρ_2 , which is assumed to be about the same for various n . The number of new cases is thus $n(\rho_2 - \rho_1)$ which is much larger for $n = 1000$ than for $n = 100$ and of course is larger the smaller ρ_1 is for fixed n .

4.3. Dependence on transmission probability p

In this subsection we report some results from our investigations on how certain properties of the epidemic depend on the probability p of the development of disease in a susceptible on contact

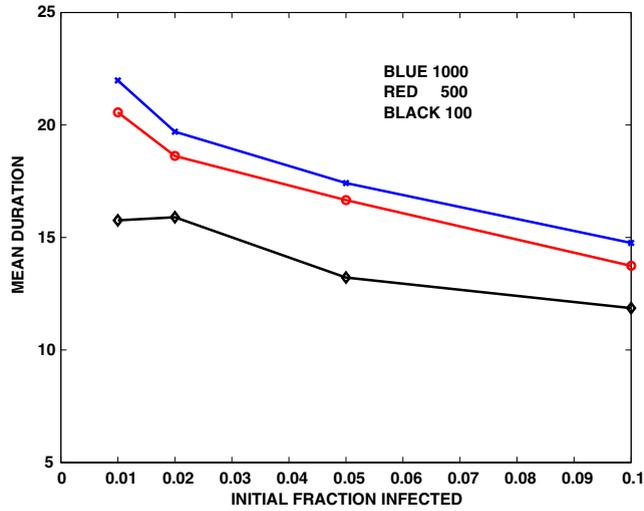


Fig. 11. A plot showing the dependence of the mean duration of an epidemic in the discrete stochastic SIR model on the initial fraction of infected individuals for various population sizes. Remaining parameter values are as for the previous figure.

with an infective. In Fig. 12, plots are shown of the mean final number of cases, averaged over 50 trials, for a population of size 500 of whom 5 are initially infected. The four sets of results are for recovery periods of $R = 1, 2, 3$ and 4. The number of contacts per day per individual is uniformly distributed on the integers 5–15.

When the recovery period is $R = 1$ (day), (see the blue curve), so that diseased individuals only have the capability to spread infection for a very limited time, the mean fraction of the population

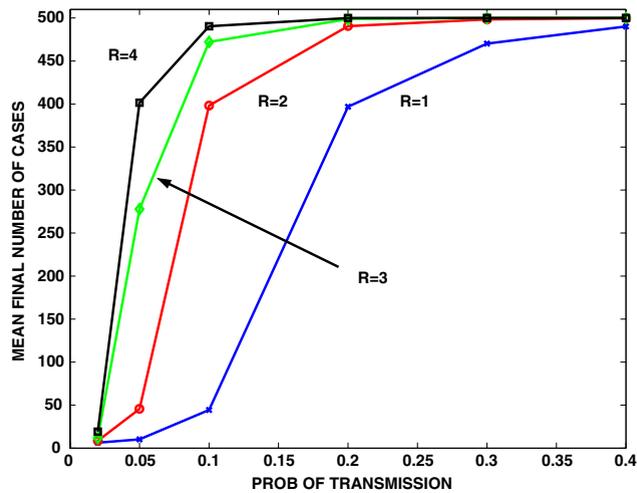


Fig. 12. A plot showing the dependence on p of the mean final size of an epidemic in a population of 500 with 5 initially infected individuals for various recovery periods $R = 1, 2, 3$ and 4, which label the curves.

who become infected is less than 50% until p reaches the high and somewhat unlikely value of 0.15, whereafter it climbs to almost 100%. When susceptibles are potentially exposed to infectives for 3 or 4 days, the entire population is infected whenever $p > 0.2$ and there is over 50% penetration for p as small as 0.05.

An aspect of special interest is the change in the mean final number of cases when R changes. For example, when $p = 0.1$, for a recovery period of $R = 2$, the mean final number infected is nearly 400. In contrast, when $R = 1$, the mean final number infected is less than 50, so there are eight times as many cases, on average, when the recovery period is 2 days as for a recovery period of 1 day. This observation is of great interest in reducing the burden of an epidemic, which is measured not only in human suffering and inconvenience, but also economic cost. The mean number of cases can be reduced not only by reducing the transmission probability p but also, and quite dramatically, by reducing R . In practice this reduction in R could be effected by either making sure that diseased individuals are prevented from circulating in the population as soon as possible after they become infective, or possibly by the use of medication or treatment which accelerates recovery.

In Fig. 13 we show the dependence of the mean duration on p for the same parameters as in Fig. 12. For each of the four values of R considered, there is a maximum mean duration at a particular value of p , which is about 0.05 for $R = 2, 3$ and 4 and about 0.2 for $R = 1$. Mean duration seems to depend significantly on p when p varies from 0 to 0.2, particularly when $R = 3$ and $R = 4$, but not when p is greater than 0.2.

4.4. Effects of changing R

In Fig. 14 we illustrate the dependence of the mean total number of cases on the recovery period R for four different values of the transmission probability p . The data are the same as for Figs. 12

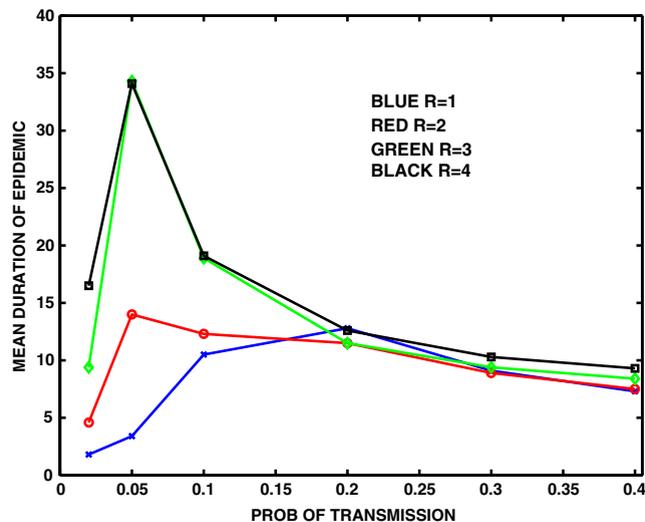


Fig. 13. A plot showing the dependence on the probability, p , of transmission of the disease on contact with an infected individual, of the mean duration of an epidemic in a population size 500 with 5 initially infected individuals for recovery periods of $R = 1, 2, 3$ and 4.

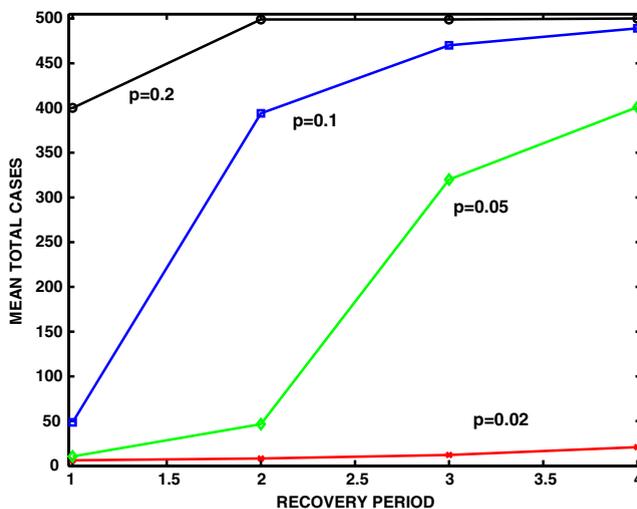


Fig. 14. A plot showing the mean total numbers of infected individuals versus recovery period for various values of the probability of transmission of the disease on contact with an infected individual.

and 13 but are plotted differently. When transmission is fairly likely, $p \geq 0.2$, nearly the whole population is infected regardless of the length of the recovery period. Furthermore, when transmission is very unlikely at $p = 0.02$ only small numbers are infected even when the recovery period is as large as 4 days. The dependence on R is thus not severe at very small or relatively large transmission probabilities. By contrast, at intermediate values of p , as R increases there is a rapid increase of mean epidemic size. Hence at such values of p , reducing the infectious period can have an extremely beneficial effect on the containment of the spread of the disease.

5. Discussion

We have formulated a simple stochastic model for the spread of disease throughout a homogeneous community of a fixed size. Time is discrete and individuals may meet a fixed number plus a random number of other individuals per day. The model is Markovian and offers somewhat more realistic characterizations of epidemics than classical discrete time models such as the Reed-Frost model which has often been employed for analyzing agricultural epidemics.

We have been concerned with situations where there are initially a few diseased individuals. By analytical and simulation methods we considered how the parameters of the model affected the time course of the spread of the disease and the final outcome in terms of total cases and total duration. Apart from the initial condition there are four variable elements: n , the total population size; p , the probability of transmission from diseased to susceptible; R , the number of days an individual remains infective; and the set of N_i , $i = 1, 2, \dots, n$, where N_i is the random number of contacts per day made by individual i ; N_i may have fixed and random components. A degree of scale invariance was noted in the sense that the fraction of the population ultimately infected depended on the fraction initially infected rather than the absolute size of the population. This possibly can be interpreted by considering that each initially infected individual more or less starts his own epidemic independently of other initially infected individuals.

An approximate expression was easily obtained for the probability of any or zero new cases after $t = 0$. We found that there was considerable variability of response as a small number of initially infected individuals have the capacity to give rise to either very small outbreaks or, for the parameters considered, with small probability, very large outbreaks, as seen in Fig. 1. The corresponding durations also exhibited great variability. For the case $R = \infty$ a formal diffusion approximation was obtained for the number of cases as a function of time, which also leads to a useful approximate logistic equation for the mean number of cases.

In Sections 4.1, 4.3 and 4.4, using simulation, we have examined, in the mean, the effects of varying the contacts per day, the probability of transmission and the length of the recovery period. Most noteworthy were the maxima in the mean duration that occur as the contact rate and the probability of transmission increase, and the drastic reductions in the size of the outbreak as the recovery period is reduced for certain, but not all, ranges of the other parameters. These findings may have practical applications as they shed light on some of the factors controlling the size and duration of epidemics and hence their human and economic costs. More simulations will be required to do a thorough investigation of the factors involved. In addition, asymptotic analysis may be useful for large n as well as a comparison of results from the present model with those for differential equation models and the original Reed-Frost model. These aspects will be the subjects of future articles.

References

- [1] E.H. Kaplan, D.L. Craft, L.M. Wein, Emergency response to a smallpox attack: the case for mass vaccination, *Proc. Natl. Acad. Sci. USA* 99 (2002) 10935.
- [2] R.M. Anderson, R.M. May, *Infectious Diseases of Humans*, OUP, Oxford, 1992.
- [3] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* 42 (2000) 599.
- [4] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics part I, *Proc. Roy. Soc. Lond.* A115 (1927) 700.
- [5] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'Inoculation pour la prévenir, *Mém. Math. Phys. Acad. Roy. Sci.* 6 (1760) 1.
- [6] N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, Griffin, London, 1975.
- [7] O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*, Wiley, Chichester, 2000.
- [8] V. Isham, Stochastic models for epidemics: current issues and developments, in: *Celebrating Statistics: Papers in honor of Sir David Cox on his 80th birthday*, Oxford University Press, Oxford, 2005.
- [9] M. Greenwood, On the statistical measure of infectiousness, *J. Hyg. Camb.* 31 (1931) 336.
- [10] H. Abbey, An examination of the Reed-Frost theory of epidemics, *Hum. Biol.* 24 (1952) 201.
- [11] J. Ranta, P.H. Makela, A. Takala, E. Arjas, Predicting the course of meningococcal disease outbreaks in closed subpopulations, *Epidemiol. Infect.* 123 (1999) 359.
- [12] T. Tsutsui, N. Minamib, M. Koiwai, T. Hamaokaa, I. Yamanea, K. Shimura, A stochastic-modeling evaluation of the foot-and-mouth-disease survey conducted after the outbreak in Miyazaki, Japan in 2000, *Prev. Vet. Med.* 61 (2003) 45.
- [13] A.M. Perez, M.P. Ward, V. Ritacco, Simulation-model evaluation of bovine tuberculosis-eradication strategies in Argentine dairy herds, *Prev. Vet. Med.* 54 (2002) 351.
- [14] A.M. Perez, M.P. Ward, A. Charmandarian, V. Ritacco, Simulation model of within-herd transmission of bovine tuberculosis in Argentine dairy herds, *Prev. Vet. Med.* 54 (2002) 361.
- [15] H. Wahlstrom, L. Englund, T. Carpenter, U. Emanuelson, A. Engvall, I. Vagsholm, A Reed-Frost model of the spread of tuberculosis within seven Swedish extensive farmed fallow deer herds, *Prev. Vet. Med.* 35 (1998) 181.

- [16] A.D. Barbour, S. Utev, Approximating the Reed-Frost epidemic process, *Stoch. Proc. Appl.* 113 (2004) 173.
- [17] F.G. Ball, O.D. Lyne, Optimal vaccination policies for stochastic epidemics among a population of households, *Math. Biosci.* 177,178 (2002) 333.
- [18] S.N. Ethier, T.G. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley, New York, 1986.
- [19] F.G. Ball, P. O’Neill, The distribution of general final state random variables for stochastic epidemic models, *J. Appl. Prob.* 36 (1999) 473.
- [20] T. Sellke, On the asymptotic distribution of the size of a stochastic epidemic, *J. Appl. Prob.* 20 (1983) 390.
- [21] C. Lefevre, P. Picard, A non-standard family of polynomials and the final-size distribution of Reed-Frost epidemic processes, *Adv. Appl. Prob.* 22 (1990) 25.
- [22] J. Ng, E.J. Orav, A generalized chain-binomial model with application to HIV infection, *Math. Biosci.* 101 (1990) 99.
- [23] P. Neal, Compound Poisson limits for household epidemics, *J. Appl. Prob.* 42 (2005) 334.
- [24] H.C. Tuckwell, *Elementary Applications of Probability Theory: An Introduction to Stochastic Differential Equations*, Chapman Hall, London, 1995.