The Annals of Applied Probability
1997, Vol. 7, No. 1, 90–109

AN EPIDEMIC MODEL WITH REMOVAL-DEPENDENT INFECTION RATE

BY PHILIP O'NEILL

University of Bradford

This paper is concerned with a model for the spread of an epidemic in a closed, homogeneously mixed population in which new infections occur at rate $\beta(z)xy/(x+y)$, where x, y and z denote, respectively, the numbers of susceptible, infective and removed individuals. Thus the infection mechanism depends upon the number of removals to date, reflecting behavior change in response to the progress of the epidemic. For a deterministic version of the model, a recurrent solution is obtained when $\beta(z)$ is piecewise constant. Equations for the total size distribution of the stochastic model are derived. Stochastic comparison results are obtained using a coupling method. Strong convergence of a sequence of epidemics to an unusual birth-and-death process is exhibited, and the behavior of the limiting birth-and-death process is considered. An epidemic model featuring sudden behavior change is studied as an example, and a stochastic threshold result analagous to that of Whittle is derived.

1. Introduction. Mathematical models for the spread of an infectious disease in a population typically assume that the parameters governing infection remain unaltered throughout an epidemic. However, when modelling certain diseases it may be necessary to consider the effects of behavior changes. In particular, individuals may take steps to reduce the risk of infection if they are sufficiently aware of the presence of the disease in the population. This kind of feature occurs in real-life epidemics; for example, there is evidence of changing sexual practices among certain homosexual groups in the United Kingdom in response to AIDS awareness advertising in the 1980's [see BMRB (1987), Anderson, Blythe, Gupta and Konings (1989)].

The purpose of this paper is to examine a simple closed-population model for an epidemic in which the infection rate depends upon the number of removals that have occurred. In a modelling context, this infection mechanism reflects behavior change in response to the perceived state of the epidemic. In particular, this approach caters to epidemics where the current number of infectives is unknown, but where some information about the number of removed individuals is available. Data concerning sexually transmitted diseases are often of this type. Deterministic models for sexually transmitted diseases that consider behavior change have been proposed in Blythe, Brauer, Castillo-Chavez, and Velasco-Hernandez (1992), Brauer,

Received November 1995; revised June 1996.
AMS 1991 *subject classifications*. 92D30, 60J27.
Key words and phrases. Epidemics, size of epidemic, deterministic and stochastic models, threshold theorems, coupling, strong convergence, birth-and-death process.

Blythe, and Castillo-Chavez (1992) and Hadeler and Castillo-Chavez (1995). In particular, these models allow the rate of recruitment into the sexually active part of the population to depend upon the current perceived state of the epidemic.

The model that we shall consider is a generalization of the modified stochastic epidemic [see Gleissner (1988) and Ball and O'Neill (1993)]. Specifically, our model is defined as follows. Consider a population consisting of initially $n > 0$ susceptible and $a > 0$ infective individuals, and for $t \geq 0$ let $X(t)$, $Y(t)$ and $Z(t)$ denote, respectively, the numbers of susceptible, infective and removed individuals at time t. The epidemic is a continuous-time Markov chain, with transition probabilities

$$\begin{aligned}
&\Pr\{(X(t+\delta t), Y(t+\delta t), Z(t+\delta t)) = (x-1, y+1, z) \mid \\
&\qquad\qquad (X(t), Y(t), Z(t)) = (x, y, z)\} \\
&\quad = \beta(z)\, xy\, \delta t/(x+y) + o(\delta t), \\
&\Pr\{(X(t+\delta t), Y(t+\delta t), Z(t+\delta)) = (x, y-1, z+1) \mid \\
&\qquad\qquad (X(t), Y(t), Z(t)) = (x, y, z)\} \\
&\quad = \gamma y\, \delta t + o(\delta t),
\end{aligned}$$

where $\beta(z) > 0$ and $\gamma > 0$ are known as the infection and removal rates, respectively. The two transitions defined above describe, respectively, an infection and a removal. Notice that since $X(t) + Y(t) + Z(t) = n + a$ for all $t \geq 0$, it is sufficient to describe the epidemic in terms of $(X(t), Y(t))$ only. The epidemic ceases as soon as there are no infectives remaining in the population. Note that by setting $\beta(Z(t)) = n + a - Z(t) = X(t) + Y(t)$, the epidemic reduces to the widely studied general stochastic epidemic model [see, e.g., Bailey (1975), page 88].

The above definition of our model implies that when an individual becomes infective, it remains so for a period of time having a negative exponential distribution of mean γ^{-1}, and is then removed. This period of time is known as the infectious period. It is possible to generalize our model so as to allow the infectious period to be any nonnegative random variable, and we shall consider this in Sections 3.2 and 3.3.

The paper is organized as follows. In Section 2 we consider a deterministic version of the model and obtain results concerning the total size and temporal behavior. The stochastic model is examined in Section 3. We show that the methods of Picard and Lefèvre (1993) can be applied to our model to yield a system of equations for the total size distribution. Comparison results for epidemics with different $\beta(Z)$ functions are considered by using the coupling methods of O'Neill (1995b). By extending the methods of Ball and Donnelly (1995), we show that the epidemic can be approximated by an unusual birth-and-death process. Section 4 is concerned with a specific example, namely a model for a sudden change in behavior. Again, deterministic and stochastic models are considered, and for the latter we derive a Whittle-style threshold result [see Whittle (1955)].

2. Deterministic model. For $t \geq 0$ let $x(t)$, $y(t)$ and $z(t)$ denote, respectively, the numbers of susceptible, infective and removed individuals in the population at time t. The epidemic, with initial condition $(x(0), y(0), z(0)) = (n, a, 0)$, is described by the differential equations

$$\frac{dx}{dt} = \frac{-\beta(z)xy}{x+y}, \tag{2.1}$$

$$\frac{dy}{dt} = \frac{\beta(z)xy}{x+y} - \gamma y, \tag{2.2}$$

$$\frac{dz}{dt} = \gamma y, \tag{2.3}$$

where $\beta(z)$, $\gamma > 0$.

2.1. *Total size.* The total size of an epidemic is defined as the total number of susceptibles who ultimately become infected. It is not hard to see that $y(t) \to 0$ as $t \to \infty$, and thus, writing T for total size, $T = n - x(\infty) = z(\infty) - a$, where $x(\infty) = \lim_{t \to \infty} x(t)$. We proceed by following Kendall (1956, 1965). Dividing (2.1) by (2.3) we obtain

$$\frac{dx}{dz} = \frac{-\beta(z)x}{\gamma(n+a-z)},$$

which upon integration yields that

$$x(t) = n \exp\left\{-\gamma^{-1}\int_0^{z(t)} (N-u)^{-1}\beta(u)\,du\right\},$$

where $N = n + a$. Now since $dz/dt \to 0$ as $t \to \infty$, it follows from (2.3) that $z(\infty)$ must be a root in $[a, N]$ of the equation

$$N - \xi - n \exp\left\{-\int_0^{\xi} (N-u)^{-1}\sigma(u)\,du\right\} = 0, \tag{2.4}$$

where $\sigma(u) = \beta(u)/\gamma$.

LEMMA 2.1. *Equation* (2.4) *always has at least one root in* $(a, N]$, *and* $z(\infty)$ *is given by the least such root.*

PROOF. Let $H(\xi)$ denote the left-hand side of (2.4), and define

$$I_\sigma = \int_0^N (N-u)^{-1}\sigma(u)\,du.$$

Then if $I_\sigma = \infty$, it is immediate from (2.4) that $H(N) = 0$. Conversely, if $I_\sigma < \infty$, then $H(N) < 0$, which, combined with the facts that $H(a) > 0$ and H is continuous in ξ, implies that (2.4) has a root in (a, N). Suppose now that (2.4) has two roots, ξ_1 and ξ_2, where $a < \xi_1 < \xi_2 \leq N$, and for contradiction assume that $z(\infty) = \xi_2$. Now since $z(0) = 0$ and $z(t)$ is nondecreasing in t,

the continuity of z ensures that there must exist an $s > 0$ such that $z(s) = \xi_1$. However, since ξ_1 is a root of (2.4), we have that $(dz/dt)_s = 0$ and thus $y(s) = 0$. It follows that $z(\infty) = \xi_1$, providing the required contradiction.

Consider now two epidemics with infection rate functions $\beta_1(z)$ and $\beta_2(z)$. Suppose that, for all $\xi \in [a, N]$,

$$\int_0^\infty (N-u)^{-1}\sigma_1(u)\,du \geq \int_0^\xi (N-u)^{-1}\sigma_2(u)\,du,$$

using the obvious notation. It follows that $H_1(\xi) \geq H_2(\xi)$ for $\xi \in [a, N]$, and thus $z_1(\infty) \geq z_2(\infty)$. In particular, note that it is not necessary that $\beta_1(z) \geq \beta_2(z)$ for all z in order to obtain this inequality.

2.2. *Temporal solution.* The set of equations (2.1)–(2.3) does not, in general, appear to have a closed-form solution. However, by setting $\beta(z) = \beta > 0$ for all z, the epidemic reduces to a modified epidemic, for which explicit formulas are available for $x(t)$, $y(t)$ and $z(t)$ [see Ball and O'Neill (1993), Section 2.1]. By exploiting this fact, it is straightforward to obtain a recursive solution for our more general model whenever $\beta(\cdot)$ is piecewise constant. Specifically, let $0 = \delta_0 < \delta_1 < \delta_2 < \cdots < \delta_m = N$, and suppose that $\beta(z) = \beta_k > 0$ for $z \in [\delta_k, \delta_{k+1})$ $(k = 0, 1, \ldots, m-1)$. In the following, let $\rho_k = \gamma\beta_k^{-1}$, $\tau_k = \inf\{t \geq 0\colon z(t) = \delta_k\}$, $n_k = x(\tau_k)$, $a_k = y(\tau_k)$ and $N_k = n_k + a_k$. Then for $t \in [\tau_k, \tau_{k+1})$ $(k = 0, 1, \ldots, m-1)$, if $\rho_k = 1$ we obtain

$$\begin{aligned} x(t) &= n_k \exp(-a_k\,\beta_k(t-\tau_k)/N_k), \\ y(t) &= a_k \exp(-a_k\,\beta_k(t-\tau_k)/N_k), \\ z(t) &= N_k\bigl(1-\exp(-a_k\,\beta_k(t-\tau_k)/N_k)\bigr) + \delta_k \end{aligned}$$

and thus

$$\tau_{k+1} = \frac{N_k}{a_k\,\beta_k}\log\left(\frac{N_k}{N_k - (\delta_{k+1}-\delta_k)}\right) + \tau_k,$$

while if $\rho_k \neq 1$, we obtain

$$\begin{aligned} x(t) &= n_k\bigl\{N_k^{-1}\bigl[n_k + a_k\exp((\beta_k-\gamma)(t-\tau_k))\bigr]\bigr\}^{1/(\rho_k-1)}, \\ y(t) &= a_k\bigl\{N_k^{-1}\bigl[n_k + a_k\exp((\beta_k-\gamma)(t-\tau_k))\bigr]\bigr\}^{1/(\rho_k-1)} \\ &\quad\times\exp((\beta_k-\gamma)(t-\tau_k)), \\ z(t) &= N_k\Bigl\{1-\bigl[N_k^{-1}\bigl(n_k + a_k\exp((\beta_k-\gamma)(t-\tau_k))\bigr)\bigr]^{\rho_k/(\rho_k-1)}\Bigr\} + \delta_k \end{aligned}$$

and

$$\tau_{k+1} = (\beta_k-\gamma)^{-1}\log\Bigl\{a_k^{-1}\Bigl[N_k\bigl(1-(\delta_{k+1}-\delta_k)/N_k\bigr)^{(\rho_k-1)/\rho_k} - n_k\Bigr]\Bigr\} + \tau_k.$$

By successive evaluation and substitution of the k-dependent quantities τ_k, n_k, a_k and N_k, the above equations provide a complete temporal description of the epidemic. The total size may also be calculated using this method, although in practice it will be easier to proceed directly from equation (2.4).

3. Stochastic model.

3.1. *Total size.* The total size distribution of the modified stochastic epidemic has been studied by Picard and Lefèvre (1993), using the martingale methods of Picard (1980). It transpires that this approach is also suitable, with appropriate modifications, for our epidemic. We proceed as follows. For $t \geq 0$, let $\mathscr{F}_t$ denote the σ-algebra generated by $\{(X(u), Y(u), Z(u))\colon 0 \leq u \leq t\}$, and for $k = 0, 1, \ldots, n$ and $\theta \geq 0$ define

$$\alpha_{k,j}(\theta) = \gamma(k+j)\big[k\tilde{\beta}(k+j) + (\gamma+\theta)(k+j)\big]^{-1}, \qquad j = 1, 2, \ldots, n+a-k,$$

where $\tilde{\beta}(z) = \beta(n+a-z)$. For $i, j \in \mathbb{N}$, define $i_{[j]} = i(i-1)\cdots(i-j+1)$, and set $i_{[0]} = 1$. The following is a generalization of Picard and Lefèvre's (1993) Proposition 5.1.

LEMMA 3.1. *For $k = 0, 1, \ldots, n$ and $\theta \geq 0$,*

$$\left\{\left(X(t)_{[k]}\left(\prod_{j=1}^{X(t)+Y(t)-k}\alpha_{k,j}(\theta)\right)\exp\left(-\theta\int_0^t Y(u)\,du\right), \mathscr{F}_t\right)\colon t \geq 0\right\}$$

is a martingale.

PROOF. The method of proof is essentially the same of that used by Picard and Lefèvre, which we now briefly outline. Now $\{(M(t), \mathscr{F}_t)\colon t \geq 0\}$ will be a martingale provided that, for all $t \geq 0$,

$$\frac{dE[M(t)]}{dt} = 0. \tag{3.1}$$

Let $M(t) = f(X(t), V(t))\exp(-\theta\int_0^t Y(u)\,du)$, where $V(t) = X(t) + Y(t)$. Evaluation of the left-hand side of (3.1) leads us to deduce that $\{(M(t), \mathscr{F}_t)\colon t \geq 0\}$ will be a martingale if

$$\begin{aligned}\tilde{\beta}(v)xv^{-1}[f(x-1, v) - f(x, v)] + \gamma[f(x, v-1) - f(x, v)] \\ = \theta f(x, v),\end{aligned} \tag{3.2}$$

where $(x, v) \in \mathscr{D} = \{(i, j) \in \mathbb{N}^2\colon\ 0 \leq n,\ \ i < j \leq n + a\}$. Writing $f(x, v) = x_{[k]}b(v)$, where $k \in \{0, 1, \ldots, n\}$, it follows from (3.2) that

$$b(v) = \gamma v b(v-1)\big[k\tilde{\beta}(v) + (\gamma+\theta)v\big]^{-1}.$$

The result now follows upon setting

$$b(v) = \prod_{j=1}^{v-k}\alpha_{k,j}(\theta).$$

Let S denote the time at which the epidemic ceases, so that $S = \inf\{t \geq 0\colon Y(t) = 0\}$. It follows at once from Lemma 3.1 that for $k = 0, 1, \ldots, n$ and

for $\theta \geq 0$,

$$(3.3)\quad E\left\{X(S)_{[k]}\left(\prod_{j=1}^{X(S)-k}\alpha_{k,j}(\theta)\right)\exp(-\theta A(S))\right\} = n_{[k]}\prod_{j=1}^{n+a-k}\alpha_{k,j}(\theta),$$

where $A(S) = \int_0^S Y(u)\,du$, the final severity of the epidemic. By taking $\theta = 0$ and $k = n, n-1, \ldots, 0$ in (3.3) we thus obtain a triangular system of linear equations in the probabilities $\Pr(T = j)$ $(j = n, n-1, \ldots, 1)$, whence $\Pr(T = 0)$ may also be calculated.

3.2. *Comparison results.* In Section 2, we found that the comparison of the total sizes of two deterministic epidemics with different $\beta(\cdot)$ functions depended upon the relative values of the integral $\int_0^\xi (N-u)^{-1}\beta(u)\,du$. For the stochastic case, a corresponding relationship seems far harder to establish. However, we can make progress by imposing a rather stronger condition on the two $\beta(\cdot)$ functions, and this is the subject of the following paragraphs.

We shall proceed by using a coupling method described in O'Neill (1995b). This approach enables us to generalize our model slightly in two ways. First, we shall no longer require that the infectious periods have a negative exponential distribution, and second, the infectious periods of different individuals need not be identically distributed. In practice, this second feature can be used to model any discrepancy between the infectious periods of initially infective individuals (these being measured from time $t = 0$) and the infectious periods of those susceptibles who subsequently become infected. More precisely, the model is now defined as follows. There are initially n susceptibles and a infectives. An infective individual, j say, remains so for a period of time L_j where L_j is distributed according to some nonnegative random variable. During its infectious period, j infects susceptibles according to a Poisson process of intensity $\beta(Z(t))X(t)(X(t)+Y(t))^{-1}$. It follows that the probability of an infection occurring in $[t, t+\delta t)$ is the same as for our original model, namely $\beta(Z(t))X(t)Y(t)(X(t)+Y(t))^{-1}\,\delta t + o(\delta t)$. The epidemic ends as soon as there are no more infectives left in the population.

We now describe a method of construction for the above epidemic model. Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ on which are defined a Poisson process Q of rate 1, and an independent sequence of random variables $L_{-(a-1)}$, $L_{-(a-2)}, \ldots, L_n$, where each L_i $(i = -(a-1), -(a-2), \ldots, n)$ has some arbitrary distribution. Let the initial infectives be numbered $-(a-1), -(a-2), \ldots, 0$. For the Poisson process Q we denote the points of Q by $\tilde{Q}_1, \tilde{Q}_2, \ldots$ and the number of points in $[0, s]$ by $Q(s)$. We define

$$I(t) = Q\left(\int_0^t f(X(u), Z(u))\,du\right),$$

$$\tau_j = \begin{cases} 0, & \text{if } j \leq 0, \\ \tilde{I}_j, & \text{if } j \geq 1, \end{cases}$$

$$R(t) = \sum_{j=-(a-1)}^{I(t)} \mathscr{I}_{\{\tau_j + L_j \leq t\}},$$

where $f(x, z) = \beta(z)xy/(x + y) = \beta(z)x(n + a - x - z)/(n + a - z)$,

$$X(t) = n - I(t),$$
$$Y(t) = a + I(t) - R(t),$$
$$Z(t) = R(t)$$

and where $\mathscr{I}_A$ denotes the indicator function of the set A. In the above, $I(t)$ and $R(t)$ are, respectively, the number of infections and the number of removals to have occurred by time t. For $j \geq 1$, τ_j denotes the time of the jth infection, while for $j \leq 0$, $\tau_j = 0$ is the time that the jth initial infective begins infecting. Similarly, for $j \geq 1$, L_j is the infectious period of the jth susceptible to be infected, while for $j \leq 0$, L_j is the infectious period of initial infective j. It is straightforward to verify that the above construction does indeed yield the required epidemic model [see O'Neill (1995b)].

Consider now two epidemics, $E^{(1)}$ and $E^{(2)}$, say, constructed on $(\Omega, \mathscr{F}, \mathbb{P})$ using the common random variables $L_{-(a-1)}, L_{-(a-2)}, \ldots, L_n$ and the common Poisson process Q, but with infection rate functions $\beta^{(1)}$ and $\beta^{(2)}$, where $\beta^{(1)}(z) \leq \beta^{(2)}(\tilde{z})$ for all $z \geq \tilde{z}$. Note that these conditions on $\beta^{(1)}$ and $\beta^{(2)}$ may be equivalently expressed by the requirement that, in the obvious notation, $f^{(1)}(x, z) \leq f^{(2)}(x, \tilde{z})$ for all x, where $z \geq \tilde{z}$.

For $l = 1, 2$ let $\tau_j^{(l)}$ denote the time of the jth infection in $E^{(l)}$. If there are a total of m infections in $E^{(l)}$, then set $\tau_j^{(l)} = \infty$ for $j > m$. The following result shows that corresponding infections will always occur first in $E^{(2)}$, and furthermore that the time lag between such corresponding infections is weakly increasing.

LEMMA 3.2. *The sequence*

$$\left(\tau_1^{(1)} - \tau_1^{(2)}, \tau_2^{(1)} - \tau_2^{(2)}, \ldots\right)$$

is nonnegative and nondecreasing.

The proof of the above lemma is similar to that of Lemma 4.2 in O'Neill (1995b); full details are available in O'Neill (1995a).

For $l = 1, 2$ let $X^{(l)}$ and $Z^{(l)}$ denote, respectively, the numbers of susceptibles and removed individuals in $E^{(l)}$. It follows at once from Lemma 3.2 that for all $t \geq 0$ and $\omega \in \Omega$, $X^{(1)}(\omega, t) \geq X^{(2)}(\omega, t)$. Further, since the infectious periods are common in the two epidemics, we also obtain that $Z^{(1)}(\omega, t) \leq Z^{(2)}(\omega, t)$. This in turn immediately implies that for all $t \geq 0$, $X^{(1)}(t) \geq_{\text{st}} X^{(2)}(t)$ [i.e., $X^{(1)}(t)$ is stochastically greater than $X^{(2)}(t)$] and $Z^{(1)}(t) \leq_{\text{st}} Z^{(2)}(t)$. Thus, as expected, there is a faster rate of disease spread in $E^{(2)}$.

It should be noted that our results extend those found in Ball and O'Neill (1993), Section 3.4), which compare the general and modified stochastic epidemics, both of which are special cases of our more general model. In particular, our approach provides additional temporal information, since as

well as establishing that infections occur sooner in the more severe epidemic, we have also shown that the time discrepancy between corresponding infections increases as the two epidemics progress.

Finally, our results can also be used to obtain corresponding results for the deterministic model. Specifically, we can apply the methods of Kendall and Saunders (1983) to obtain that, in the obvious notation, $x^{(1)}(t) \geq x^{(2)}(t)$ and $z^{(1)}(t) \leq z^{(2)}(t)$ for all $t \geq 0$. The details of the method are explained in O'Neill (1995b), and for brevity we do not include them here.

3.3. *Birth-and-death process approximation.* It is frequently possible to linearize an epidemic model in order to create a simpler approximating process. The best-known example of this is found in the fact that the early stages of the trajectory of infectives in a general stochastic epidemic may be well approximated by a linear birth-and-death process. This in turn gives rise to threshold results, which, broadly speaking, describe conditions under which the epidemic is likely either to die out quickly or take hold. In recent years, a number of papers have addressed this area in a broader context, using coupling arguments to prove strong convergence of a wide range of epidemic models to branching processes and birth-and-death processes. The convergence occurs as the initial number of susceptibles increases to infinity, while the initial number of infectives is held constant. For further details, see Ball (1983), Ball and O'Neill (1994), Ball and Donnelly (1995) and O'Neill (1996).

In the following, we shall assume that the infection rate function can also depend upon the initial population size, so that $\beta = \beta(Z, n)$. This assumption reflects the idea that behavior changes in a population are likely to depend upon the population size as well as the number of removals.For example, it seems reasonable to suppose that behavior changes occur in response to the proportion of individuals in a population that are removed, rather than the absolute number. Notice that the results of the previous sections remain unaltered if β is n-dependent, since we have only considered the situation where the population size is fixed. Our convergence results in this section will apply in the case where $\beta(Z, n) \to \beta(Z)$ as $n \to \infty$. Before stating the results, we first quickly outline some key ideas; full details are available in the papers cited.

Ball and Donnelly's (1995) paper makes use of the following construction. A branching process is given, and at each birth time a member of the set of initially susceptible individuals is chosen uniformly and at random. If this individual is still susceptible, then it becomes infected and lives for the same length of time as the corresponding individual that has just been born in the branching process. If however the individual has previously been infected, then no infection occurs, and the birth in the branching process is ignored. This ignored individual is called a ghost. The salient feature of this construction is that as the number of susceptibles increases, the probability of a ghost being created tends to zero, yielding a convergence result.

The above construction has two features which need to be adjusted in order to proceed. Firstly, the form of the construction implies that the branching process acts as an upper bound for the number of infectives created, since infections can only occur as a result of births. As such, if the infection rate of the epidemic can potentially exceed the birth rate, it becomes necessary to modify the construction. This can be achieved by generating extra infections by using suitably time-transformed Poisson processes, as described in O'Neill (1996). The second difficulty, which is somewhat more technical, is the presence of the $(X + Y)^{-1}$ term in the modified epidemic infection mechanism. This will be dealt with by increasing the infection probability when a birth occurs in an appropriate manner.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which are defined the following.

1. A birth-and-death process W, with a initial individuals. Individuals live for a time distributed according to some arbitrary nonnegative random variable L, during which they give birth to new individuals according to a Poisson process of rate $\beta(D(t))$, where $D(t)$ denotes the number of deaths during $[0, t]$. All individuals behave independently of one another, and the size of W at time t will be denoted by $W(t)$.
2. Two independent and identically distributed sequences of Uniform$(0, 1)$ random variables $U_1, U_2, \ldots$ and $V_1, V_2, \ldots$.
3. A Poisson process P of rate 1.
4. An independent and identically distributed sequence of random variables $L_1, L_2, \ldots,$ where L_j is distributed according to L in (1) above.

We now describe the construction of a sequence of epidemic processes indexed by n. All random variables in the following are deemed to be evaluated at some fixed $\omega \in \Omega$, but we shall suppress explicit reference to this fact. For $x \in \mathbb{R}$, $\lceil x \rceil$ shall denote the least integer greater than or equal to x. Let the initial susceptibles be labeled $1, 2, \ldots, n$, and the initial infectives $(n + 1), (n + 2), \ldots, (n + 1)$. For $j \in \mathbb{N}$ let χ_j denote the jth individual to be born in W, and denote its lifetime by $L(\chi_j)$. As χ_j is born, consider the $\lceil (n + a)U_j \rceil$th individual in the initial population. If it is a susceptible that has not yet been infected, then that individual immediately becomes infective. Conversely, suppose that the chosen individual is either an initial infective or an initial susceptible that has already been infected. Define

$$\begin{aligned} p &= p(X, Y, Z, D, n) \\ &= \min\left\{1, \left(\frac{X}{X + Y}\right)\left(\frac{(n + a)\beta(Z, n) - (X + Y)\beta(D)}{(n + a - X)\beta(D)}\right)\right\}, \end{aligned}$$

where X, Y, Z and D denote, respectively, the current numbers of susceptibles, infectives, removed individuals, and deaths to have occurred in W. Then an uninfected initial susceptible becomes infected if $V_l < p$, where V_l is any previously unused member of the $V_1, V_2, \ldots$ sequence. The susceptible infected in this way is chosen uniformly and at random from those still available, using another previously unused member of the V sequence. Notice

that since $p = 0$ whenever $X = 0$, an infection cannot occur in this manner if there are no more susceptibles. If, however, $V_l \geq p$, then no infection occurs, and the births of χ_j and all its subsequent offspring are ignored. In this case, we refer to χ_j as a ghost. A susceptible that becomes infective as a consequence of the birth of χ_j remains so for a time $L(\chi_j)$ before becoming removed.

Infections also arise at the times given by the points of the process

$$P\left(\int_0^t \phi(u)\, du\right),$$

where

$$\phi(u) = \max\left\{\frac{\beta(Z(u), n)X(u)}{X(u) + Y(u)} - \beta(D(u)), 0\right\}Y(u).$$

When an infection occurs in this way, then a susceptible is chosen at random from those available by using a fresh member of the V sequence. Following O'Neill (1996), such infectives are referred to as shadow infectives. The jth shadow infective created remains infectious for a time L_j and is then removed. The epidemic ceases as soon as there are no more infectives present in the population.

We may verify that the above construction gives rise to our epidemic as follows. The infectious periods are clearly of the desired form, and thus it only remains to show that the infection mechanism is correct. So let δt be some fixed small time interval and suppose that $Y(t) > 0$, so that the epidemic has not yet terminated. For $\phi(t) > 0$ it is easily verified that $p(X(t), Y(t), Z(t), D(t), n) = 1$, and so an infection can occur either as the result of a non-ghost individual giving birth in W, or a shadow infection. Combining these possibilities yields that the probability of an infection occurring during $[t, t + \delta t)$ is given by

$$\beta(D(t))Y(t)\,\delta t + \left(\frac{\beta(Z(t), n)X(t)}{X(t) + Y(t)} - \beta(D(t))\right)Y(t)\,\delta t + o(\delta t)$$
$$= \frac{\beta(Z(t), n)X(t)Y(t)}{X(t) + Y(t)}\delta t + o(\delta t),$$

as required. Alternatively, if $\phi(t) \leq 0$, then no shadow infections can occur, so that an infection can only arise as a result of a non-ghost individual having a non-ghost birth in W. The required probability is thus

$$\beta(D(t))Y(t)\,\delta t\left(\frac{X(t)}{n + a} + \left(1 - \frac{X(t)}{n + a}\right)p(X(t), Y(t), Z(t), D(t), n)\right)$$
$$= \frac{\beta(Z(t), n)X(t)Y(t)}{X(t) + Y(t)}\,\delta t + o(\delta t).$$

The key feature of our construction is that the epidemic process and the birth-and-death process are identical up until the appearance of either the

first ghost or the first shadow infective. However, the probability of either of these events occurring during a finite time interval tends to zero as n tends to infinity, which leads to the following theorem. We shall denote the number of infectives in the nth epidemic at time t by $Y_n(t)$.

THEOREM 3.3. *Let d be any metric on the space of sample paths of W, and let A be that subset of Ω on which W ultimately becomes extinct.*

(i) *For $\mathbb{P}$-almost all $\omega \in A$,*

$$\lim_{n\to\infty} \sup_{0\le t\le\infty} d(W(\omega,t), Y_n(\omega,t)) = 0.$$

(ii) *For $\mathbb{P}$-almost all $\omega \in A^c$,*

$$\lim_{n\to\infty} \sup_{0\le t\le T} d(W(\omega,t), Y_n(\omega,t)) = 0$$

for all $T > 0$.

PROOF. (i) Let Ω_1 be that subset of Ω on which $U_i \neq U_j$ for all $i, j \in \mathbb{N}$ such that $i \neq j$, and note that $\mathbb{P}(\Omega_1) = 1$. Let $\omega \in \Omega_1 \cap A$, $T(\omega) < \infty$ be the duration of W and $B(\omega) < \infty$ be the number of births to occur during $[0, T(\omega)]$. Then for all sufficiently large n, $[(n + a)U_i] \neq [(n + a)U_j]$ for $1 \le i, j \le B(\omega)$, $i \neq j$, from which it follows that no ghosts can be produced during $[0, T(\omega)]$.

It only remains to show that no shadow infections can occur, so suppose for contradiction that the first shadow infective appears at time $s \in [0, T(\omega)]$. Then if the first point of P occurs at τ, we have

$$\int_0^s \phi(u)\, du = \tau. \tag{3.4}$$

Let $t \in [0, s)$, and consider the time interval $[0, t]$. No shadow infections occur, and for all sufficiently large n, no ghosts are created. It follows that during this time interval the number of deaths in W and the number of removals in the epidemic are identical, and further that the number of infections and births are the same, the latter being bounded by $B(\omega)$. Combining these observations with the fact that $\beta(Z, n) \to \beta(Z)$ as $n \to \infty$ yields that we can make $\phi(t)$ arbitrarily small by taking n sufficiently large. In particular, we obtain that

$$\int_0^t \phi(u)\, du < \tau/2,$$

for each $t \in (0, s)$, which contradicts (3.4).

(ii) Let $T \ge 0$, and notice that, with probability 1, only finitely many births occur in W during $[0, T]$. The result now follows by an identical argument found in (i) above.

3.4. *Properties of the limiting birth-and-death process.* We shall briefly consider the behavior of the W process introduced in the previous section. For

$t \geq 0$, let $W(t)$ and $D(t)$ denote, respectively, the number of individuals in the population and the total number of deaths to have occurred at time t. We suppose that each individual has a lifetime distributed according to some nonnegative random variable L, and that during its lifetime an individual reproduces according to a Poisson process of intensity $\beta(D(t))$.

First, it is straightforward to show that W will go extinct with probability 1 if, for all sufficiently large d, $\beta(d) \leq \gamma$. However, if this does not occur, then the process will have some nonzero probability of never becoming extinct. It is not clear how this probability could be evaluated in general. In the special case where L is exponentially distributed, so that W is a Markov process, we can make a little more progress, as follows.

For $m = 1, 2, \ldots$ let $\tilde{W}_m$ denote the size of the population immediately after the mth death has occurred, and set $\tilde{W}_0 = a$. Then

$$\tilde{W}_{m+1} = \tilde{W}_m + R_m - 1,$$

where R_m denotes the number of births that occur between the mth and $(m + 1)$th deaths. Thus for $k = 0, 1, \ldots,$

$$\Pr(R_m = k) = \left(\frac{\beta(m)}{\gamma + \beta(m)}\right)^k \left(\frac{\gamma}{\gamma + \beta(m)}\right),$$

so that $\{\tilde{W}_m : m \geq 0\}$ is a time-inhomogeneous Markov chain. The process terminates as soon as $\tilde{W}_m = 0$, and we define M^* to be the least integer m such that $\tilde{W}_m = 0$. The structure of the process enables us to write down an expression for $\Pr(M^* = k)$ $(k = a, a + 1, \ldots)$, which we achieve in the following manner. First, note that if the process terminates upon the event of the kth death, then $\tilde{W}_{k-1} = 1$. It follows that R_m can only be nonzero for $m = 0, 1, 2, \ldots, k - 2$. For these values of m, let $b_m \geq 0$ denote the value of R_m. It follows that for $m = 0, 1, \ldots, k - 2$ we must have

$$\tilde{W}_m = a + b_0 + b_1 + \cdots + b_{m-1} - m > 0, \tag{3.5}$$

since the process cannot terminate until the kth death. Further, the total number of births to occur during the lifetime of the process is $k - a$, and thus

$$\sum_{m=0}^{k-2} b_m = k - a. \tag{3.6}$$

Denote by B the set of all possible vectors with nonnegative integer components $\mathbf{b} = (b_0, b_1, \ldots, b_{k-2})$ such that $b_0, b_1, \ldots, b_{k-2}$ satisfy (3.5) and (3.6). It follows that

$$\Pr(M^* = k) = \left\{\prod_{m=0}^{k-1} \left(\frac{\gamma}{\gamma + \beta(m)}\right)\right\} \sum_{\mathbf{b} \in B} \prod_{m=0}^{k-2} \left(\frac{\beta(m)}{\gamma + \beta(m)}\right)^{b_m}.$$

Finally, the probability that the process ever terminates is given by

$$\sum_{k=a}^{\infty} \Pr(M^* = k),$$

which does not appear to have a simpler form.

4. Example: a sudden change model. In this section we shall consider a specific choice of $\beta(Z)$, namely that $\beta(Z) = \beta_1$ if $Z < z_0$, and $\beta(Z) = \beta_2$ otherwise, where $\beta_1 \neq \beta_2$. This fairly simple model is based upon the idea that no change of behavior occurs until the number of removals has reached some critical value, which seems a plausible feature of real-life epidemics.

Since the results of the previous two sections apply to the sudden change model, we shall in the following only draw attention to particular points of interest. The deterministic model admits an explicit closed form solution, from which the total size is also available. For the stochastic model, we shall derive a threshold result based on the approach of Whittle (1955).

4.1. *Deterministic model.* For $j = 1, 2$ let $\rho_j = \gamma/\beta_j$, and let $\tau = \inf\{t \geq 0: z(t) = z_0\}$. The following results are easily verified.

(i) $\rho_1 = 1$. From Gleissner (1988) we have that for $t \in [0, \tau]$,

$$\begin{aligned} x(t) &= n \exp(-a\beta_1 t/N), \\ y(t) &= a \exp(-a\beta_1 t/N), \\ z(t) &= N(1 - \exp(-a\beta_1 t/N)), \end{aligned} \tag{4.1}$$

where $N = n + a$. It follows from (4.1) that

$$\tau = (N/a\beta_1)\log[N/(N - z_0)],$$

and so

$$\begin{aligned} x(\tau) &= n(1 - z_0/N), \\ y(\tau) &= a(1 - z_0/N), \\ z(\tau) &= z_0. \end{aligned}$$

Since a behavior change occurs at time τ, it follows that $\rho_1 \neq \rho_2$, so that $\rho_2 \neq 1$. From Gleissner (1988) it follows that for $t > \tau$,

$$x(t) = x(\tau)\left(\frac{x(\tau) + y(\tau)\exp[(\beta_2 - \gamma)(t - \tau)]}{x(\tau) + y(\tau)}\right)^{1/(\rho_2 - 1)},$$

with similar expressions holding for $y(\tau)$ and $z(\tau)$. After a little manipulation we obtain that

$$x(t) = n(1 - z_0/N)\left(\frac{n + a\exp[(\beta_2 - \gamma)(t - \tau)]}{n + a}\right)^{1/(\rho_2 - 1)},$$

$$y(t) = a(1 - z_0/N)\left(\frac{n + a\exp[(\beta_2 - \gamma)(t - \tau)]}{n + a}\right)^{1/(\rho_2 - 1)} \times \exp[(\beta_2 - \gamma)(t - \tau)],$$

$$z(t) = N(1 - z_0/N)\left[1 - \left(\frac{n + a\exp[(\beta_2 - \gamma)(t - \tau)]}{n + a}\right)^{\rho_2/(\rho_2 - 1)}\right] + z_0.$$

(ii) $\rho_1 \neq 1$. Proceeding as before, we have that for $t \in [0, \tau]$,

$$x(t) = n\{N^{-1}[n + a\exp((\beta_1 - \gamma)t)]\}^{1/(\rho_1 - 1)},$$

$$y(t) = a\{N^{-1}[n + a\exp((\beta_1 - \gamma)t)]\}^{1/(\rho_1 - 1)}\exp((\beta_1 - \gamma)t),$$

$$z(t) = N\{1 - [N^{-1}(n + a\exp((\beta_1 - \gamma)t))]^{\rho_1/(\rho_1 - 1)}\}.$$

By direct calculation,

$$\tau = (\beta_1 - \gamma)^{-1}\log a^{-1}[N(1 - z_0/N)^{(\rho_1 - 1)/\rho_1} - n],$$

provided that $N(1 - z_0/N)^{(\rho_1 - 1)/\rho_1} - n > 0$. However, this last condition is always true if $\rho_1 < 1$, whilst if $\rho_1 > 1$ then the condition is equivalent to the statement $z_0 < N[1 - (n/N)^{\rho_1/(\rho_1 - 1)}] = z(\infty)$, which says that the epidemic does not terminate before the behavior change occurs. Thus when a behavior change does occur,

$$x(\tau) = n(1 - z_0/N)^{1/\rho_1},$$

$$y(\tau) = N - z_0 - n(1 - z_0/N)^{1/\rho_1},$$

$$z(\tau) = z_0.$$

Suppose that $\rho_2 \neq 1$; then, for $t > \tau$,

$$x(t) = n(1 - z_0/N)^{1/\rho_1}A^{1/(\rho_2 - 1)},$$

$$y(t) = [N - z_0 - n(1 - z_0/N)^{1/\rho_1}]A^{1/(\rho_2 - 1)}\exp[(\beta_2 - \gamma)(t - \tau)],$$

$$z(t) = (N - z_0)[1 - A^{\rho_2/(\rho_2 - 1)}] + z_0$$

where

$$A = \left\{\frac{n + [N(1 - z_0/N)^{(\rho_1 - 1)/\rho_1} - n]\exp[(\beta_2 - \gamma)(t - \tau)]}{N(1 - z_0/N)^{(\rho_1 - 1)/\rho_1}}\right\}.$$

Conversely if $\rho_2 = 1$ then we obtain that for $t > \tau$,

$$x(t) = n(1 - z_0/N)^{1/\rho_1} \exp\Big[\big(nN^{-1}(1 - z_0/N)^{(1-\rho_1)/\rho_1} - 1\big)\beta_2(t - \tau)\Big],$$

$$y(t) = (1 - z_0/N)^{1/\rho_1} \times \Big[N(1 - z_0/N)^{(\rho_1-1)/\rho_1} - n\Big] \times \exp\Big[\big(nN^{-1}(1 - z_0/N)^{(1-\rho_1)/\rho_1} - 1\big)\beta_2(t - \tau)\Big],$$

$$z(t) = (N - z_0)\Big\{1 - \exp\Big[\big(nN^{-1}(1 - z_0/N)^{(1-\rho_1)/\rho_1} - 1\big)\beta_2(t - \tau)\Big]\Big\} + z_0.$$

We may evaluate the total size of the deterministic model without difficulty. First, if $\rho_1 = 1$, then

$$T(z_0) = \begin{cases} n, & \text{if } \rho_2 < 1, \\ n - (N - z_0)(n/N)^{\rho_2/(\rho_2-1)}, & \text{if } \rho_2 > 1. \end{cases} \tag{4.2}$$

If $\rho_1 \neq 1$, then it is possible for the epidemic to terminate before a behavior change occurs. This will happen if the final number of removed individuals in a modified epidemic of relative removal rate ρ_1 is not greater than z_0, that is, $N(1 - (n/N)^{\rho_1/(\rho_1-1)}) \leq z_0$. We can rewrite this condition in terms of ρ_1, so that a behavior change does not occur unless $\rho_1 < \zeta(z_0)$, where

$$\zeta(z_0) = \frac{\log[(N - z_0)/N]}{\log[(N - z_0)/n]} > 1.$$

We thus obtain that for $\rho_1 \neq 1$,

$$T(z_0) = \begin{cases} n, & \text{if } \rho_1 < \zeta(z_0),\ \rho_2 \leq 1, \\ F(z_0), & \text{if } \rho_1 < \zeta(z_0),\ \rho_2 > 1, \\ n\big(1 - (n/N)^{1/(\rho_1-1)}\big), & \text{if } \rho_1 \geq \zeta(z_0), \end{cases} \tag{4.3}$$

where

$$F(z_0) = n\Big\{1 - \Big[(n/N)(1 - z_0/N)^{(\rho_2-\rho_1)/\rho_1}\Big]^{1/(\rho_2-1)}\Big\}.$$

REMARK. In the following, we suppose, for realism, that $\rho_1 < \rho_2$. Suppose that a behavior change actually occurs, so that $z_0 < N(1 - (n/N)^{\rho_1/(\rho_1-1)}) = z_m$, say. We may investigate the impact of this change upon the progress of the epidemic by considering $Q(z_0) = \tilde{T} - T(z_0)$, where $\tilde{T}$ is the total size of a modified epidemic with relative removal rate ρ_1, and $T(z_0)$ is given by (4.2) or (4.3). If $\rho_2 \leq 1$, then $Q = 0$, so that no benefit occurs as a result of the behavior change. So suppose that $\rho_2 > 1$. If $\rho_1 = 1$, then it is easily seen that Q is linear in z_0. If however $\rho_1 \neq 1$, then Q contains the term $(1 - z_0/N)^{\eta}$, where $\eta = (\rho_2 - \rho_1)/\rho_1(\rho_2 - 1) > 0$. If $\rho_1 < 1$, then $\eta - 1 > 0$, while if $\rho_1 > 1$, we find that $\eta - 1 < 0$. It follows that d^2Q/dz_0^2 is positive when $\rho_1 < 1$ and negative if $\rho_1 > 1$, yielding that $Q(z_0)$ is, respectively, convex or concave

on $[0, z_m)$. These results can be interpreted as giving an indication of the effectiveness of altering z_0, in the following broad sense. Suppose for example that $\rho_1 < 1$; then the smaller z_0 is, the greater the impact of any reduction of z_0 upon the epidemic. Conversely, as z_0 approaches z_m, reductions of z_0 are of less benefit.

4.2. *Stochastic threshold result.* In Section 3.3 we noted the fact that when considering results that depend upon the initial number of susceptibles increasing towards infinity, $\beta(Z)$ may be regarded as a function of n. For the sudden change model there are essentially two possibilities of interest: as $n \to \infty$, either $z_0(n) \to l < \infty$ or $z_0(n) \to \infty$. In the former case, the limiting process has birth rate β_1 until l deaths have occurred, and thereafter it has birth rate β_2, with the death rate being γ at all times. Consider now the latter case, in which the limiting birth-and-death process is simply linear and homogeneous, with birth rate β_1 and death rate γ. In particular, the behavior change aspect of the epidemic disappears in the limit as $n \to \infty$. In the case where $z_0 = O(n)$, an alternative method of analysis is given by the methods of Whittle (1955), and this is the subject of this section. We shall proceed by following Whittle's arguments; see also Ball and O'Neill (1993) for the corresponding modified stochastic epidemic case.

Let P_w denote the probability of an epidemic of total size w ($w = 0, 1, 2, \ldots, n$), and recall that the *intensity* of an epidemic is defined as the proportion of the initial susceptibles who ultimately contract the disease. We define π_i as the probability that an epidemic of intensity i or less occurs, so that

$$\pi_i = \sum_{w=0}^{\lceil ni \rceil} P_w .$$

Suppose further that $z_0 = z_0(n) = c(n + a) = cN$, where $c \in (0, 1)$. This assumption means that a behavior change occurs when some fixed proportion c of the population becomes removed. In order to ensure that there is some positive probability of a behavior change actually occurring, we set $c < i$. Let τ_n denote the time of behavior change in the nth epidemic, so that $\tau_n = \inf\{t \geq 0\colon\ Z(t) \geq z_0\}$. Then if a behavior change occurs in an epidemic of intensity i or less we obtain that for all $t \geq \tau_n$,

$$n(1 - i) \leq X(t) \leq N(1 - c),$$

and

$$cN \leq Z(t) \leq a + ni.$$

It follows that

$$\frac{\beta_2 n(1 - i)Y(t)}{N(1 - c)} \leq \frac{\beta_2 X(t)Y(t)}{X(t) + Y(t)} \leq \frac{\beta_2 N(1 - c)Y(t)}{n(1 - i)},$$

which when n is large can be approximated by

$$\frac{\beta_2(1-i)Y(t)}{(1-c)} \le \frac{\beta_2 X(t)Y(t)}{X(t)+Y(t)} \le \frac{\beta_2(1-c)Y(t)}{(1-i)}. \tag{4.4}$$

It follows that for $t \ge \tau_n$ we can think of the epidemic as being sandwiched between two linear birth-and-death processes, with birth rates given by (4.4) and common death rate γ. We shall denote the number of individuals ever born into the upper-bound process by T_U and the number born into the lower-bound process by T_L.

Now

$$\begin{aligned}\Pr(\text{a behavior change ever occurs}) &= \Pr(Z(\infty) \ge \lceil cN \rceil) \\ &= 1 - \Pr(Z(\infty) \le \lceil cN \rceil - 1) \\ &= 1 - \Pr(\text{intensity} \le j),\end{aligned}$$

where by direct calculation $j = (\lceil cN \rceil - a - 1)/n \approx c$ for large n. Thus the probability of a behavior change occurring is approximately $1 - \pi_c(\rho_1)$, where $\pi_c(\rho_1)$ is the probability that the intensity of a modified stochastic epidemic with relative removal rate ρ_1 is not greater than c. From Ball and O'Neill (1993), we have approximate bounds on this quantity when n is large, namely that

$$\{\min[\rho_1(1-c), 1]\}^a \le \pi_c(\rho_1) \le \left\{\min\left[\frac{\rho_1 N}{n(1-c)}, 1\right]\right\}^a.$$

Let π_i^* denote the probability of an epidemic of intensity i or less, conditional upon the occurrence of a behavior change. Then,

$$\pi_i = (1 - \pi_c(\rho))\pi_i^* + \pi_c(\rho_1), \tag{4.5}$$

the point here being that if a behavior change does not occur, then the intensity must automatically be less than i. Let r_n denote the number of infections that occur during (τ_n, ∞). Then

$$\Pr(T_U \le r_n) \le \pi_i^* \le \Pr(T_L \le r_n).$$

However, $\Pr(T_L \le r_n) \le \Pr(T_L < \infty)$, while $\Pr(T_U \le r_n) = \Pr(T_U < \infty) - \Pr(T_U \in (r_n, \infty))$. A difficulty now arises, since all that can be said about r_n is that

$$0 \le r_n \le N(i-c) + a(1-i).$$

However, we may proceed by making use of a further approximation, namely that $X(\tau_n) \approx x(\tilde{\tau}_n)$, where $x(\tilde{\tau}_n)$ is the number of susceptibles in the deterministic model at the time of behavior change. Our justification for this is as follows. One of the main differences between different realizations of a modified stochastic epidemic is the time taken until the epidemic "takes off."

Once this has happened however, the epidemic is well approximated by the corresponding deterministic trajectory. Indeed for the general stochastic epidemic, Metz (1978) has conjectured that the stochastic model is well modeled by a random time translation of the deterministic model. Thus in some sense, the stochastic and deterministic models will look similar if compared at corresponding points of progress, although they might be quite different if compared at some fixed time point. Regarding the actual time taken until take-off, Ball and O'Neill (1993) indicate that the modified stochastic epidemic is well approximated by a linear birth-and-death process until around $\sqrt{n}$ of the initial susceptibles become infected. After this time, the deterministic approximation becomes more appropriate [see Barbour (1980) for a discussion of similar piecewise approximations for Markov processes]. In the present case, however, since at least $cN - a$ infections must have occurred by the time of the behavior change, the epidemic is certainly at a stage where the deterministic approximation is reasonable. As further justification for the present case, Table 1 illustrates deterministic and stochastic numbers of susceptibles and infectives for various values of ρ_1, with $n = 999$ and $a = 1$. As can be seen from the table, the approximations are quite good.

Returning to our argument, suppose that the actual intensity of the epidemic is μ, where $\mu \in (c, i]$. Then $r_n = X(\tau_n) - (1 - \mu)n \approx x(\tilde{\tau}_n) - (1 - \mu)n = O(n)$, since from Section 4.1, $x(\tilde{\tau}_n) = n(1 - c)$ if $\rho_1 = 1$ and $n(1 - c)^{1/\rho_1}$ otherwise. It follows that $\Pr(T_U \in (r_n, \infty)) \to 0$ as $n \to \infty$, so that using the standard result for the probability of extinction for a linear birth-and-death process we obtain the approximate result

$$(4.6) \qquad \left[\min\left(\rho_2\left(\frac{1-i}{1-c}\right), 1\right)\right]^{y(\tilde{\tau}_n)} \le \pi_i^* \le \left[\min\left(\rho_2\left(\frac{1-c}{1-i}\right), 1\right)\right]^{y(\tilde{\tau}_n)},$$

where $y(\tilde{\tau}_n) \approx Y(\tau_n)$, the latter being the initial population size of both the birth-and-death processes. If $\rho_1 = 1$, then $y(\tilde{\tau}_n) = a(1 - c)$. Otherwise,

TABLE 1
$[X(\tau), Y(\tau)]$ *for modified stochastic epidemics where* $n = 999$ *and* $a = 1$

Simulation number	**1**	**2**	**3**	**4**	**5**	$[x(\tilde{\tau}), y(\tilde{\tau})]$
			(i) $\beta_1 = 2.0, \gamma = 1.0$			
$z_0 = 250$	(593, 157)	(563, 187)	(556, 194)	(572, 178)	(535, 215)	(562, 188)
$z_0 = 500$	(264, 236)	(285, 215)	(247, 253)	(234, 266)	(286, 214)	(250, 250)
$z_0 = 750$	(71, 179)	(59, 191)	(66, 184)	(58, 192)	(84, 166)	(62, 188)
			(ii) $\beta_1 = 0.8, \gamma = 0.2$			
$z_0 = 250$	(298, 452)	(332, 418)	(302, 448)	(347, 403)	(318, 432)	(316, 434)
$z_0 = 500$	(71, 429)	(56, 444)	(62, 438)	(77, 423)	(51, 449)	(62, 438)
$z_0 = 750$	(8, 242)	(4, 246)	(3, 247)	(4, 246)	(3, 247)	(4, 246)

$y(\tilde{\tau}_n) = O(n)$, so in this case for large n we find that, approximately,

$$\pi_i^* = \begin{cases} 1, & \text{if } \rho_2 \geq \left(\dfrac{1-c}{1-i}\right), \\ 0, & \text{if } \rho_2 < \left(\dfrac{1-i}{1-c}\right), \end{cases}$$

while if $\rho_2(1-i)/(1-c) < 1 < \rho_2(1-c)/(1-i)$ then (4.6) does not yield any useful bounds. Upon substitution into (4.5) we find that

$$\pi_i = \begin{cases} 1, & \text{if } \rho_2 \geq \left(\dfrac{1-c}{1-i}\right), \\ \pi_c(\rho_1), & \text{if } \rho_2 < \left(\dfrac{1-i}{1-c}\right). \end{cases}$$

We may interpret these results as follows. For large n, if $\rho_1 \geq 1$ or $\rho_2 \geq (1-c)/(1-i)$ then there is zero probability of an epidemic occurring. However if $\rho_1 < 1$ and $\rho_2 < (1-i)/(1-c)$ then the probability of an epidemic is approximately $1 - \rho_1^a$, for small values of c. Notice that this probability does not depend upon the value of ρ_2, which is essentially for the following reason: once the epidemic has taken off sufficiently for a behavior change to occur, then there will be a large number of infectives present in the population at the time of the change, and thus the epidemic is extremely unlikely to die out suddenly.

REFERENCES

ANDERSON, R. M., BLYTHE, S. P., GUPTA, S. and KONINGS, E. (1989). The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behaviour. *Philos. Trans. Roy. Soc. London Ser. B* **325** 45–98.

BAILEY, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd ed. Griffin, London.

BALL, F. G. (1983). The threshold behaviour of epidemic models. *J. Appl. Probab.* **20** 227–241.

BALL, F. G. and DONNELLY, P. J. (1995). Strong approximations for epidemic models. *Stochastic Process. Appl.* **55** 1–21.

BALL, F. G. and O'NEILL, P. D. (1993). A modification of the general stochastic epidemic motivated by AIDS modelling. *Adv. in Appl. Probab.* **25** 39–62.

BALL, F. G. and O'NEILL, P. D. (1994). Strong convergence of epidemic models. *Adv. in Appl. Probab.* **26** 629–655.

BARBOUR, A. D. (1980). Density dependent Markov population processes. *Lecture Notes in Biomath.* **38** 36–49. Springer, Berlin.

BLYTHE, S. P., BRAUER, F., CASTILLO-CHAVEZ, C. and VELASCO-HERNANDEZ, J. X. (1992). Models for sexually transmitted diseases with recruitment. In *Proceedings of the Third International Conference on Population Dynamics*. Univ. Pau, France.

BRMB (1987). *AIDS advertising campaign, report on four surveys during the first year of advertising, 1986–7*. British Market Research Bureau Limited.

BRAUER, F., BLYTHE, S. P. and CASTILLO-CHAVEZ, C. (1992). Demographic recruitment in sexually transmitted disease models. Technical Report BU-1154-M, Biometrics Unit, Cornell Univ.

GLEISSNER, W. (1988). The spread of epidemics. *Appl. Math. Comput.* **27** 167–171.
HADELER, K. P. and CASTILLO-CHAVEZ, C. (1995). A core group model for disease transmission. *Math. Biosci.* **128** 41–55.
KENDALL, D. G. (1956). Deterministic and stochastic epidemics in closed populations. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **4** 149–165. Univ. California Press, Berkeley.
KENDALL, D. G. (1965). Mathematical models for the spread of infection. In *Mathematics and Computer Science in Biology and Medicine* 213–225. MRC, HMSO, London.
KENDALL, W. S. and SAUNDERS, I. W. (1983). Epidemics in competition II: the general epidemic. *J. Roy. Statist. Soc. Ser. B* **45** 238–244.
METZ, J. A. J. (1978). The epidemic in a closed population with all susceptibles equally vulnerable; some results for large susceptible populations and small initial infections. *Acta Biotheoretica* **27** 75–123.
O'NEILL, P. D. (1995a). An epidemic model with removal-dependent infection rate. Report SOR 95-71, Dept. Mathematics, Univ. Bradford.
O'NEILL, P. D. (1995b). Epidemic models featuring behaviour change. *Adv. in Appl. Probab.* **27** 960–979.
O'NEILL, P. D. (1996). Strong approximations for some open population epidemic models. *J. Appl. Probab.* **33** 448–457.
PICARD, P. (1980). Applications of martingale theory to some epidemic models. *J. Appl. Probab.* **17** 583–599.
PICARD, P. and LEF ÈVRE, C. (1993). Distribution of the final state and severity of epidemics with fatal risk. *Stochastic Process. Appl.* **48** 277–294.
WHITTLE, P. (1955). The outcome of a stochastic epidemic: a note on Bailey's paper. *Biometrika* **42** 116–122.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRADFORD
BRADFORD, BD7 1DP
ENGLAND
E-MAIL: p.oneill@bradford.ac.uk