The Annals of Applied Probability
1993, Vol. 3, No. 1, 1–27

SPECIAL INVITED PAPER

STOCHASTIC MODELS FOR EPIDEMICS WITH SPECIAL REFERENCE TO AIDS[1]

BY VALERIE ISHAM

University College London

This paper gives a review of some of the recent work on stochastic epidemic models and their deterministic counterparts. In particular it focusses on models for the spread of HIV infection and AIDS. The variability between realisations of an epidemic is discussed in some detail, and methods of assessing this variation are described. Numerical examples are given to illustrate various aspects of the models considered.

1. Introduction. In recent years there has been a great upsurge of interest in models for epidemics, partly generated by concern over the AIDS epidemic. Much of the work on building mathematical models for the transmission dynamics of HIV infection is deterministic, although allowance for many sources of variability can be made in the models [see, e.g., the review given by Isham (1988) and Anderson, Blythe, Gupta and Konings (1989)]. However, the models are almost always described in stochastic terms even when a deterministic analysis is envisaged. This raises the question as to the connection between the stochastic models and their deterministic analogues, the correspondence being many-to-one since a variety of stochastic models will have the same deterministic counterpart. A deterministic model for the (multivariate) state of a system will consist of a set of differential equations which, given assumed parameter values, can be solved numerically if not analytically to give a fixed temporal evolution of the system. Realisations of a corresponding stochastic model can be simulated and will in general, and in contrast, exhibit considerable interrealisation variability. It will often be difficult or impossible to obtain explicit expressions for the properties of a stochastic model.

The aim of this paper is to review some of the recent work on stochastic models for epidemics and their deterministic versions. It will focus on the AIDS epidemic which has stimulated much progress in this area, although the results have more general implications and applications. The paper will start in Section 2 by describing properties and results for a very simple and well-understood model, before going on in Section 3 to consider more general

Received August 1991; revised February 1992.

[1]IMS Special Invited Lecture given to the Biometric Society ENAR meeting in Houston, Texas in March 1991.

AMS 1991 *subject classification*. 60K99.

Key words and phrases. AIDS, HIV infection, epidemic model, general stochastic epidemic, variability of epidemic realisations.

and realistic models for AIDS. Some areas of current interest and activity will be outlined in Section 4. In Section 5 we discuss one specific aspect of stochastic epidemics, that of the variability between realisations, in more detail and use numerical examples to illustrate some of the ideas touched upon in the paper. Finally, some summary remarks are given in Section 6.

2. A simple epidemic model.

The general stochastic epidemic. Perhaps the simplest interesting stochastic epidemic model is the *general stochastic epidemic*, which is a Markov process in which a closed population of size n is subdivided into three classes consisting of those susceptible to the infection, those who are infected and those who have recovered and are immune to reinfection. We denote the numbers of individuals in these classes at time t by $X(t)$, $Y(t)$ and $Z(t)$, respectively, so that $X(t) + Y(t) + Z(t) \equiv n$. Infections occur at a rate proportional to the current values of both X and Y, that is, we shall assume

$$P(X(t+dt) = i-1, Y(t+dt) = j+1, Z(t+dt) = k | X(t) = i, Y(t) = j, Z(t) = k) = \alpha n^{-1} ij\, dt + o(dt) \tag{2.1}$$

while the periods of infection of distinct individuals are independent and exponentially distributed with parameter ν, so that

$$P(X(t+dt) = i, Y(t+dt) = j-1, Z(t+dt) = k+1 | X(t) = i, Y(t) = j, Z(t) = k) = \nu j\, dt + o(dt). \tag{2.2}$$

This model is also known as an S-I-R model, which alludes to the susceptible-infected-recovered progression of the disease. An illustration of the course of this epidemic is given in Figure 1. This will be discussed further in Section 5.

It follows from (2.1) and (2.2) that the increments in the variables satisfy the equations

$$\begin{aligned} E(dX(t) | X(t), Y(t)) &= -\alpha n^{-1} X(t) Y(t)\, dt + o(dt), \\ E(dY(t) | X(t), Y(t)) &= \left[\alpha n^{-1} X(t) - \nu\right] Y(t)\, dt + o(dt), \end{aligned} \tag{2.3}$$

where we shall concentrate on the variables X and Y, Z being then completely determined. The corresponding deterministic model is obtained by equating these conditional expected increments to the actual increments, thus it is given by the differential equations

$$\frac{dx(t)}{dt} = -\alpha n^{-1} x(t) y(t), \qquad \frac{dy(t)}{dt} = \left[\alpha n^{-1} x(t) - \nu\right] y(t). \tag{2.4}$$

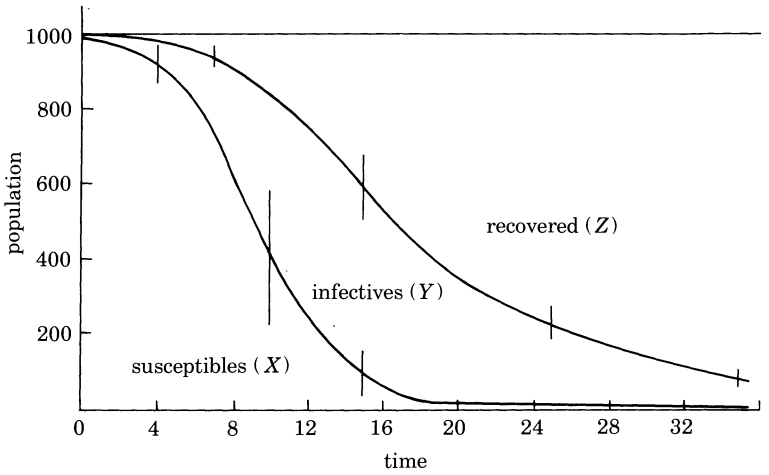

FIG. 1. *The expected numbers of susceptibles* [$E(X)$, *lower curve*] *and susceptibles + infectives* [$E(X+Y)$, *upper curve*] *for the general stochastic epidemic with* $n = 1000$, $X(0) = 990$, $Y(0) = 10$, $\alpha = 0.6$, $\nu = 0.1$. *The bars indicate* ± 2 (*standard deviation*) *for* X *or* $X+Y$ *as appropriate at particular time points.*

Since the means of the stochastic system satisfy

$$(2.5)\qquad \begin{aligned} \frac{dE(X(t))}{dt} &= -\alpha n^{-1} E(X(t))E(Y(t)) - \alpha n^{-1}\operatorname{cov}(X(t),Y(t)), \\ \frac{dE(Y(t))}{dt} &= \left[\alpha n^{-1} E(X(t)) - \nu\right] E(Y(t)) + \alpha n^{-1}\operatorname{cov}(X(t),Y(t)), \end{aligned}$$

it is immediately clear that because the model is nonlinear, the deterministic model (2.4) does not give the mean curves for the stochastic system. If, at $t = 0$, $X(0)$ is close to n and $Y(0)$ is small, then for small t, $Y(t)$ will increase as $X(t)$ decreases so that their covariance is negative. Thus the stochastic mean number infected will increase more slowly than the corresponding deterministic curve at the start of the epidemic.

Sometimes the constant of proportionality α/n in (2.1) is replaced by α', say. The distinction only really becomes important when the population size n is regarded as variable; for example, as when the limit $n \to \infty$ is considered. However, the former parameterisation is more appropriate in many cases including that of AIDS, where it is natural to assume that each individual "chooses" partners with whom to come into contact, at random from the population. A common parameterisation for HIV transmission assumes that individuals form new partnerships at rate κ and that there is a probability β that infection will be transmitted from an infective to a susceptible partner. This results in an infection rate as in (2.1) with $\alpha = \beta\kappa$. Note that in order to use the S-I-R model as a simple model for AIDS we must reinterpret "infected"

as being infected with HIV but not yet having full AIDS and "recovered" as having progressed to a full AIDS diagnosis. With these definitions, the S-I-R model is appropriate if we assume that those who have been infected with HIV cease to transmit the infection once they have been diagnosed as having AIDS.

Properties of the model. It follows from the deterministic equations (2.4) that if $x(0)$ is close to the population size n, then an epidemic will only occur (in the sense that the number of infectives initially increases) if $\alpha/\nu > 1$. Corresponding threshold effects exist in the stochastic case also, where if $\alpha/\nu > 1$, then an epidemic occurs with probability $1 - (\nu/\alpha)^{Y(0)}$. The critical ratio $R_0 = \alpha/\nu$ is the *reproductive ratio* of the epidemic, which represents the expected total number of individuals who would be infected by a single infective, assuming all partners are susceptible.

The properties of the general stochastic epidemic have been considered by many authors. Bailey (1975, Chapter 6) gives a thorough and authoritative account, while Lefèvre (1990) provides a short survey of the considerable volume of work on the model published in the last fifteen years. The special case, when $\nu = 0$, is known as the *simple stochastic epidemic*. This provides a model of a situation where individuals are assumed to remain infectious indefinitely. In many applications this will be a gross oversimplification. Nevertheless it will often be an appropriate approximation to the start of an epidemic. For example, in the AIDS context, it is well known that the incubation periods of individuals between infection with HIV and diagnosis with AIDS are typically many years and therefore there will be few transitions from the "infected" (interpreted as infected with HIV but not yet having full AIDS) to the "recovered" (interpreted as full AIDS) state during the early years of the epidemic. Thus the simple stochastic epidemic model can give a good idea of the spread of HIV infection at the beginning of the AIDS epidemic. The advantage of using the simple stochastic epidemic model is that its properties are much easier to obtain by algebraic means.

The solution of the deterministic equations (2.4) when $\nu = 0$ is simply the logistic curve given by

$$y(t) = \frac{ny(0)e^{\alpha t}}{x(0) + y(0)e^{\alpha t}}, \tag{2.6}$$

where $x(t) + y(t) \equiv n$. If t is small, then $y(t) \simeq y(0)e^{\alpha t}$, which underpins the common assumption that an epidemic grows exponentially in its early stages. Of course as $t \to \infty$, $y(t) \to n$ and, in this model, the infection spreads through the whole population. For the stochastic model with $\nu = 0$, it is easy to show that $Y(t)$ has moment generating function $M_Y(\theta; t)$ satisfying the forward equation

$$\frac{\partial M_Y(\theta;t)}{\partial t} = \alpha n^{-1}(e^{\theta} - 1)\left\{\frac{n\partial}{\partial\theta} - \frac{\partial^2}{\partial\theta^2}\right\}M_Y(\theta;t). \tag{2.7}$$

The solution of this equation can be expressed in terms of a series of hypergeo-

metric functions which leads naturally to the moments of $Y(t)$ and approximations to these via a variety of expansions [Bailey (1975, Chapter 5)].

In the general case, with $\nu > 0$, things are more difficult. Recursive solutions to the forward equations for the (joint) probability distribution of the state of the epidemic have been investigated by several authors; for a discussion see Bailey (1975, Chapter 6.3) and Kryscio (1975). For the most part these solutions are computationally very intensive although the scheme by Billard and Zhao (1991, 1992), which exploits a technique involving the transformation of the variables first introduced by Severo (1967, 1969), looks promising. Another approach, taken by Ludwig (1973), is to replace the forward equations of the process by an approximating system. In contrast with the case when $\nu = 0$, the epidemic can die out without spreading through the whole population and the threshold results previously mentioned address this point.

3. Models for the AIDS epidemic. In the previous section we saw that the general stochastic epidemic could be regarded as a very simple model for the spread of AIDS if we interpret the three states as *susceptible*, *infected with HIV but not yet diagnosed as having (full) AIDS* and *diagnosed with AIDS*, and take the constant of proportionality α to be the product ($\alpha = \beta\kappa$) of the rate κ of partner change and the probability β of transmission of infection. There are, however, many unsatisfactory features of this model which need to be modified to make the model more realistic. Realism is of course not the only goal, as we want a simple parsimonious model which can be interpreted easily. Nevertheless it is vital to include sources of variation which are known to have an important influence on the course of an epidemic.

Models for the incubation period. In the general stochastic epidemic, infected individuals are assumed to have a constant hazard of recovery. In the AIDS context this would mean that the incubation periods between infection with HIV and the progression to full AIDS are exponentially distributed. In fact it is known that short incubation periods (of 2–3 years, say) are, relatively, very unlikely; see, for example, Bacchetti and Moss (1989). If the hazard of developing AIDS is taken to be a function $\nu(\tau)$ of the time τ since infection, then it is straightforward to write down equations for a Markov model in which the state of the system must now be $\{X(t), Y(t;\tau)\ \tau > 0,\ Z(t)\}$ where $Y(t;\tau)\,d\tau$ represents the number of infected individuals in the population at time t who became infected during $(t-\tau, t-\tau+d\tau)$. For example, the deterministic equations [cf. (2.4)] in this case are

$$\begin{aligned} \frac{dx(t)}{dt} &= -\alpha n^{-1} x(t) \int_0^\infty y(t;\tau)\,d\tau, \qquad y(t;0) = -\frac{dx(t)}{dt}, \\ &\left(\frac{\partial}{\partial\tau} + \frac{\partial}{\partial t}\right) y(t;\tau) = -\nu(\tau) y(t;\tau). \end{aligned} \tag{3.1}$$

The Weibull distribution, for which $\nu(\tau) \propto \tau^{\gamma-1}$ (some $\gamma > 0$), is a convenient parametric distribution, quite often assumed for the AIDS incubation

period. If the index γ is greater than 1, then the incubation period has an increasing hazard, and fitted values of γ for AIDS are typically in the range 2.0–2.5 [Billard, Medley and Anderson (1990)]. However, for the purposes of simulation of the stochastic process or numerical solution of the deterministic equations, it is simpler to use a gamma distribution with integer index (k, say), exploiting the method of stages in which the incubation period is represented as a series of k stages, the durations of stay in the stages being independent, identically distributed exponential variables. In this way, we need only the numbers $Y^{(i)}(t)$, $i = 1, \ldots, k$, of infectives in the k stages rather than the density $Y(t; \tau)$, and within a stage the hazard is constant. In this case, for a gamma distribution with mean k/ν, the deterministic equations corresponding to (2.4) and (3.1) would be

$$\frac{dx(t)}{dt} = -\alpha n^{-1} x(t) y(t), \qquad \frac{dy^{(1)}(t)}{dt} = \alpha n^{-1} x(t) y(t) - \nu y^{(1)}(t),$$

$$\frac{dy^{(i)}(t)}{dt} = \nu y^{(i-1)}(t) - \nu y^{(i)}(t), \qquad i = 2, \ldots, k, \tag{3.2}$$

where $y(t) = y^{(1)}(t) + \cdots + y^{(k)}(t)$. When fitted to AIDS incubation period data, the gamma distribution has been found to have an index of around 2.6; Billard, Medley and Anderson (1990).

The method of stages uses independent, exponential variables but there is no requirement that these have the same parameter ν (i.e., that the hazards are constant between as well as within stages), so that a wider class of distributions than the gamma class can be considered by letting the ith stage have hazard ν_i. This then allows the possibility of modelling explicitly the clinical progression of HIV infection through a series of clinically defined stages, as long as the durations of these stages can be taken as independent exponential (or gamma) variables. This approach has been used by several authors, including Anderson, Blythe, Gupta and Konings (1989), Bailey (1990), Brookmeyer and Liao (1990) and Longini, Clark, Byers, Ward, Darrow, Lemp and Hethcote (1989). In particular in the latter paper, the data for the incubation period was fitted using three exponential stages, the latent period, an asymptomatic phase until certain specified symptoms developed and finally the symptomatic phase until progression to AIDS.

Note that, as the epidemic progresses, it will be necessary to incorporate into the model changes in the incubation period distribution in real time to reflect advances in symptomatic management and treatment with drugs like zidovudine.

Variable infectiousness. Modelling the incubation period in terms of the time τ since infection means that the transmission probability β can also be allowed to depend on τ, which is an advantage since it has been suggested [Blythe and Anderson (1988)] that an infective individual is more infectious soon after infection and again later during the symptomatic part of the

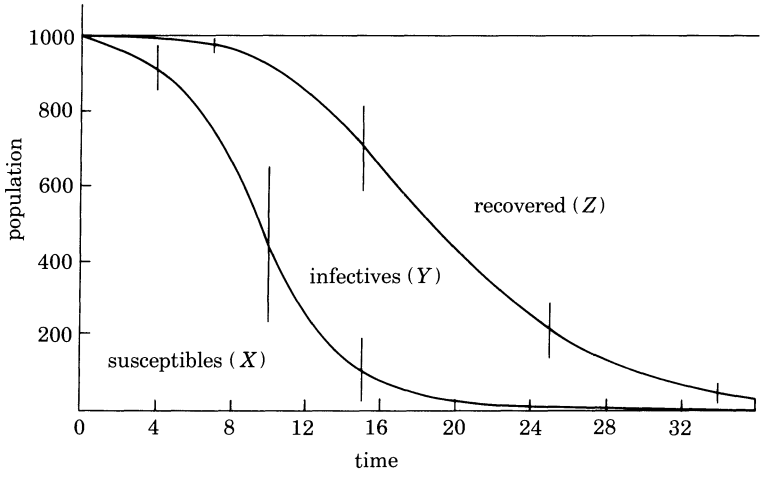

FIG. 2. *The expected numbers of susceptibles* [$E(X)$, *lower curve*] *and susceptibles + infectives* [$E(X + Y)$, *upper curve*] *for the stochastic epidemic with staged incubation period and* $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$, $\boldsymbol{\alpha} = (1, 0, 1)$, $\boldsymbol{\nu} = (1, 0.25, 0.2)$. *The bars indicate* ± 2 (*standard deviation*) *for* X *or* $X + Y$ *as appropriate at particular time points.*

incubation period than during the long asymptomatic phase. When the incubation period is divided into stages with constant hazard functions (ν_i for the ith stage), then it is most natural to let the transmission probability also be a constant (β_i) within stage i but vary with i. Note that in this simple model no allowance is made for variable susceptibility of individuals to infection by an infected partner.

In Figure 2, the course of a stochastic epidemic with a three-stage incubation period and varying transmission probability is illustrated. This can be compared with Figure 1, where an epidemic with a single-stage incubation period and constant transmission probability is represented. The initial conditions, mean incubation period and reproductive ratio are the same in both cases. The effects of the staged incubation period and of the varying transmission probability can be seen separately in Figure 3(a) and (b). As before, the initial conditions, the mean incubation period and the reproductive ratio (R_0) are kept fixed. These three figures will be discussed in Section 5.

Heterogeneity in sexual activity. Another important source of heterogeneity between individuals is that of sexual activity, represented in our model by the rate κ of partner change, which can be incorporated by dividing the population into subgroups with differing activity rates. Let $X_i(t)$ and $Y_i(t)$ be the numbers of susceptibles and infectives in the ith population subgroup having rate κ_i of partner change [$Y_i(t)$ will be further divided by stage of incubation period] and assume that there is a mixing matrix (p_{ij}) which specifies the probability that an individual in activity subgroup i will choose a partner in subgroup j. Then infections in subgroup i occur at a rate

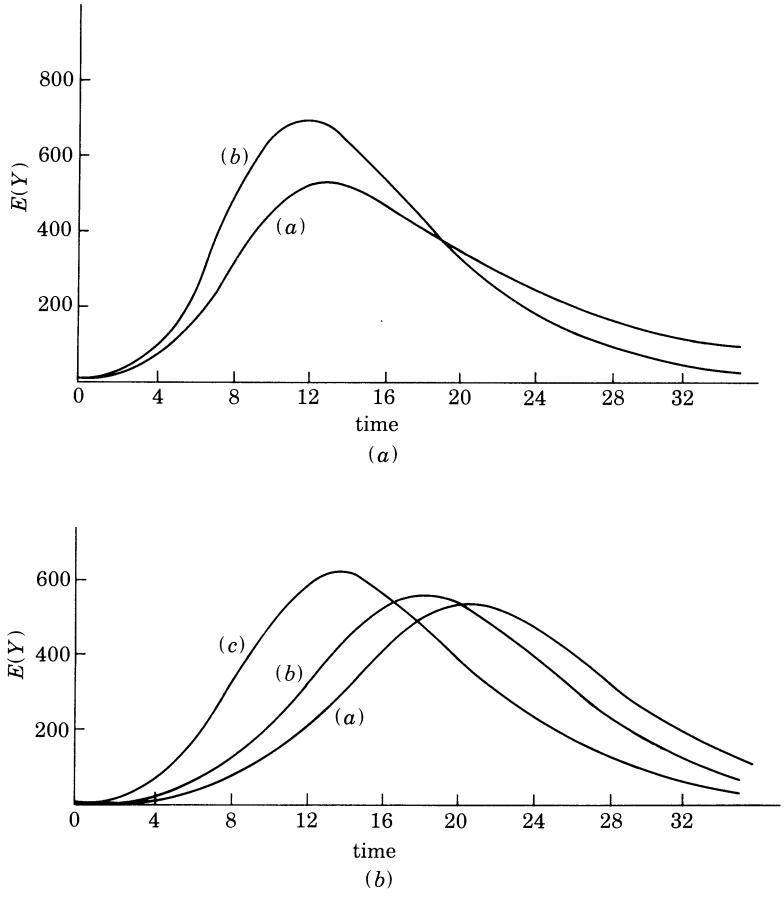

FIG. 3. *Comparison of expected numbers of infectives, $E(Y)$, for models with staged incubation periods and* $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$, $R_0 = 6$. (i) *Curve* (a) *single stage* $\nu = 1$, $\alpha = 0.6$; *Curve* (b) *three stages* $\boldsymbol{\nu} = (1, 0.25, 0.2)$ $\boldsymbol{\alpha} = (0.6, 0.6, 0.6)$. (ii) *Three stages* $\boldsymbol{\nu} = (1, 0.25, 0.2)$ *with Curve* (a) $\boldsymbol{\alpha} = (0.25, 0, 1.15)$; *Curve* (b) $\boldsymbol{\alpha} = (0.5, 0, 1.1)$; *Curve* (c) $\boldsymbol{\alpha} = (1, 0, 1)$.

$\beta X_i \kappa_i \Sigma_j p_{ij} Y_j / n_j$, where n_i is the total number of individuals in the ith subgroup. The effects of heterogeneity of sexual activity will be discussed further in Section 4.

Demographic factors. Our model can also be made more realistic by allowing the population and subgroup sizes to vary. In particular, because of the very long time scale applying to AIDS, we should allow for immigration of new susceptibles into the population (subgroups), for natural mortality of all individuals and for excess mortality due to HIV infection and AIDS. The population should also be stratified by age since it is known that the length of the incubation period tends to decrease with increasing age in adults and also

that young children have short incubation periods [e.g., Billard, Medley and Anderson (1990)].

Changes in behaviour. Changes in behaviour can be incorporated into the model by letting individuals move from one sexual activity subgroup to another. The effects of such behaviour changes are the subject of a recent paper by Scalia-Tomba (1991). Another sort of behaviour change is the adoption of safer sexual practices as the epidemic progresses or publicity campaigns are mounted. This could be introduced by taking the transmission probability β to be a function of real time t (as well as the time τ since infection).

Other risk groups. So far the model has been formulated with a rather promiscuous homosexual community in mind, but it is also important to widen it to include other risk groups. In modelling heterosexuals we must yet further divide the population into male and female. Then, in including less promiscuous individuals it will be necessary carefully to model pair formation and separation, as in the papers by Dietz; see Dietz (1987, 1988) and Dietz and Hadeler (1988). We should consider models appropriate for transmission amongst drug users and those which allow for vertical transmission to children, too.

All the modifications of the general stochastic epidemic described in this section can be made in a totally straightforward manner, although the dimension of the resulting state space will be high if all these features are incorporated simultaneously. Much recent work has concentrated on investigating the effects of allowing for these various sorts of heterogeneity. Considerable effort is also being expended on data analysis, involving the estimation of model parameters from biological or behavioural data, model fitting and prediction. In the next section, we consider a few of the issues of current interest concerning models and their properties in a little more detail.

4. The effect of population heterogeneity.

Heterogeneity of sexual activity and mixing. In the previous section, many modifications of the general stochastic epidemic were described. The purpose of these is to allow for the inherent variability within individuals over time and between individuals, both in terms of biological characteristics (e.g., length of incubation period) and behavioural ones (e.g., choice of sexual partner). That incorporating variable rates of partner change into the models has a crucial effect on the outcome of the epidemic is well known [Andersen, Medley, May and Johnson (1986)]. Perhaps the clearest demonstration of this effect is given by Jacquez and Simon (1990). If we make the *proportional mixing* assumption that the elements of the mixing matrix (as defined in Section 3) are given by

$$p_{ij} = \kappa_j n_j \Big/ \sum_l \kappa_l n_l$$

for all activity subgroups i, j, then the deterministic equation for the number

of infectives in the ith subgroup is

$$\frac{dy_i(t)}{dt} = \frac{\beta\kappa_i x_i(t)\Sigma_j \kappa_j y_j(t)}{\Sigma_l \kappa_l n_l} - \nu y_i(t). \tag{4.1}$$

Now the quantity which drives the epidemic is the total rate $V(t) = \Sigma\kappa_j Y_j(t)$ at which contacts are made by infectives, and the corresponding deterministic equation for V is

$$\frac{dv(t)}{dt} = \left[\frac{\beta\Sigma_i \kappa_i^2 x_i(t)}{\Sigma_l \kappa_l n_l} - \nu\right] v(t). \tag{4.2}$$

If we write μ_κ and σ_κ^2 for the population mean and variance of the rate κ of partner change, that is, $\mu_\kappa = \Sigma\kappa_i n_i/n$ and $\sigma_\kappa^2 = \Sigma\kappa_i^2 n_i/n - \mu_\kappa^2$, and assume that at the beginning of the epidemic almost all individuals are susceptible so that $x_i(t) \simeq n_i$, then (4.2) can be approximated by

$$\frac{dv(t)}{dt} \simeq \left[\beta\left(\frac{\sigma_\kappa^2}{\mu_\kappa} + \mu_\kappa\right) - \nu\right] v(t), \tag{4.3}$$

for small t. Thus we see that it is not μ_κ but $\mu_\kappa + \sigma_\kappa^2/\mu_k$ which determines the rate of growth at the start of the epidemic. In particular the reproductive ratio for this model is $R_0 = (\beta/\nu)(\sigma_\kappa^2/\mu_k + \mu_\kappa)$, where we need $R_0 > 1$ for epidemic growth.

Modelling the mixing matrix. It is clear from the preceding argument that heterogeneity of sexual activity must be incorporated into the model. Then the question arises as to the most appropriate form for the mixing matrix. This cannot be an entirely arbitrary stochastic matrix because its entries must satisfy the basic constraint

$$n_i\kappa_i p_{ij} = n_j\kappa_j p_{ji} \tag{4.4}$$

for all activity subgroups i, j, which says that pairs in which a subgroup i individual chooses a subgroup j partner must balance those with i and j interchanged. It is easy to verify that the proportional mixing probabilities do obey this constraint. In that case, p_{ij} does not depend on i and individuals choose partners in a particular subgroup in proportion to the total rate of activity of members of that subgroup. Another possibility is that of *restricted mixing*, where individuals choose partners only from within their own subgroups ($p_{ij} = \delta_{ij}$, where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$). A less extreme form of preference (*preferred mixing*) is to take

$$p_{ij} = \varepsilon\delta_{ij} + (1 - \varepsilon)n_j\kappa_j \Big/ \sum_l n_l\kappa_l, \tag{4.5}$$

in which a proportion ε of all choices are reserved for like-with-like within-group partners and the rest are spread proportionately through the other subgroups. Jacquez, Simon and Koopman (1989) and Koopman, Simon, Jacquez and Park (1989) devised a mixing scheme (*structured mixing*) which

allows the proportions of choices to be specified over all the subgroups. In this scheme, choices are made in "contact classes" which can be thought of as locations (e.g., geographical or social) in which the partner choices take place. Alternatively, perhaps they could be determined by type of sexual activity or age group. A matrix must be specified giving the probabilities f_{ik} that an individual in activity subgroup i chooses a partner in contact class (location) k. Then for each contact class k, a stochastic matrix, with elements $p_{ij}(k)$, must be specified giving the probabilities that an individual from activity subgroup i who is choosing a partner in contact class k, chooses a partner from subgroup j. For example, we could assume proportional mixing within each contact class, in which case

$$p_{ij}(k) = n_j \kappa_j f_{jk} \Big/ \sum_l n_l \kappa_l f_{lk}. \tag{4.6}$$

Of course it is necessary that the balance equations (4.4) are satisfied within each contact class.

In Section 3, the modification of the basic model was mentioned in which the population and subgroup sizes vary in time because of demographic factors, and this will be a very necessary feature of a model for AIDS. Then n_i is a function of time $n_i(t)$ and the constraint (4.4) must hold for all t. Thus we cannot take a fixed set of activity rates κ_i and a fixed mixing matrix (p_{ij}); either the rates or the mixing matrix or both must vary with time. Note that all of the mixing schemes described above automatically satisfy (4.4) for all t, when the activity rates κ_i are kept constant. However, this point becomes important if the entries in the mixing matrix are to be determined from empirical data rather than being given a priori.

Some authors have considered the estimation of $m_{ij} = n_i \kappa_i p_{ij}$ using as data the number x_{ij} in some sampled population, of subgroup i individuals choosing subgroup j partners. Since it is unlikely that the sampled population will be closed under partner choice, the x_{ij} will not usually be symmetric in i and j. Thus, for example, Pugliese (1990) fits the "nearest" symmetric matrix (m_{ij}) to the observed data (x_{ij}) which preserves the zeros in the data matrix, where two different metrics are suggested to define "nearest." Morris (1991a, b) discusses the use of log-linear models for the x_{ij}, which offer the possibility of using more or less parsimonious models as appropriate. She notes that a proportional mixing model is equivalent to a log-linear model with no $i \times j$ interaction. While the former author is concerned with the situation at a single time point, Morris takes account of the fact that m_{ij} must be symmetric in continuous time by putting forward the model

$$m_{ij}(t) = n_i(t) n_j(t) \alpha_{ij}, \tag{4.7}$$

where α_{ij} is to be symmetric and estimated from data, and suggests taking $\alpha_{ij} \propto s_{ij} s_{ji}$ where s_{ij} is the probability that a subgroup i individual signals "yes" to a subgroup j individual.

The reproductive ratio. In Section 2 we saw that the reproductive ratio R_0 for the general stochastic epidemic is given by $R_0 = \beta \kappa D$, where $D = 1/\nu$ is

the mean incubation period, and earlier in this section we showed that when the rate κ varies over the population, we must replace κ not by μ_κ but by $\mu_\kappa + \sigma_\kappa^2/\mu_\kappa$ in this expression for R_0. If the incubation period has a gamma distribution $\Gamma(k, \nu)$ with mean k/ν, then we substitute k/ν for D in R_0. If, however, we allow deaths to occur, at a constant rate μ, say, over the incubation period, then things become a little more complicated because the individual may die before completing all stages of his incubation period. In this case D should be the mean infectious period between infection and whichever is the sooner of development of AIDS and death. Within each stage of the incubation period which is attained, the hazard for the combined "AIDS or death" event is $\nu + \mu$ and hence D is given by

$$D = \frac{1}{\nu + \mu} \sum_{i=1}^{k} \left(\frac{\nu}{\nu + \mu} \right)^{i-1}. \tag{4.8}$$

If the transmission probability β varies from stage to stage of the incubation period, then in the expression for R_0, β should be replaced by the time-weighted average value $\bar{\beta}$ given by

$$\bar{\beta} = \frac{1}{\nu + \mu} \sum_{i=1}^{k} \beta_i \left(\frac{\nu}{\nu + \mu} \right)^{i-1}. \tag{4.9}$$

A recent paper by Jacquez, Simon and Koopman (1991) discusses the appropriate form for the reproductive ratio in their important structured mixing model. The essential idea is that for a deterministic model with recruitment of susceptibles at a constant rate, there will be global stability of the endemic state if $R_0 > 1$ and of the disease-free equilibrium if $R_0 \leq 1$. If all the activity subgroups have within-group reproductive ratios of at most one, then it is intuitively clear that there will be a disease-free equilibrium, while if all the subgroups have ratios exceeding one, there will be a persistent epidemic. The interest is therefore in detailed results in between these two extremes, which are described in Jacquez, Simon and Koopman (1991). There is also the question as to the stochastic counterparts of these threshold results, in the same way that there are corresponding threshold theorems for the general stochastic epidemic and its deterministic analogue. Jacquez and O'Neill (1991) address this question for homogeneous populations.

5. Variability of stochastic epidemics.

A multivariate normal approximation. In general it is not easy to get exact explicit expressions for the properties of the general stochastic epidemic, let alone the more complicated modifications of the model described in earlier sections of this paper. There is interest, therefore, in any methods which claim to give reasonable approximations to properties and which are straightforward to implement. In Isham (1991) a means of obtaining a bivariate normal approximation to the joint distribution of $X(t)$ and $Y(t)$ for the general stochastic epidemic is described and discussed. The method goes back to a

paper of Whittle (1957) and, in essence, consists of taking the forward equations for the means, variances and covariances of the variables and replacing any higher-order moments in these equations by the functions of the first- and second-order moments appropriate in the case of a multivariate normal distribution. Thus for the general stochastic epidemic, a set of five simultaneous differential equations is obtained, which are simple to solve numerically using particular choices for the parameter values. The first two equations, for the means, are the exact equations (2.5). Nevertheless the solution of the set of equations will not give exact stochastic means because the equations for the second-order moments are only approximate.

Theoretical justification for the method is provided by a series of papers [Daley and Kendall (1965), Kurtz (1970, 1971) and Barbour (1972, 1974)] obtaining Gaussian diffusion processes about the deterministic curve, as limiting processes for density-dependent Markov processes on a lattice, where the limit is taken as the population size tends to infinity but the initial proportions of susceptibles and infectives are kept fixed. Numerical illustrations of this convergence are given later in this section, in Table 6 and Figure 6 and related discussion. An equivalent approximation obtained using perturbation techniques is given by Daniels (1991).

The importance of the multivariate normal approximation is that it provides a practicable means of obtaining a better approximation to the stochastic mean than the solution of the deterministic equations, and in addition gives a way of assessing the variability likely to occur between realisations of the stochastic epidemic. This is particularly relevant in the context of prediction of the AIDS epidemic where often predictions are given, based on the deterministic approximation, and where allowance is made for uncertainty over parameter values but not for the fact that a single realisation of the epidemic will be observed. In Isham (1991), the multivariate normal approximations are compared with estimates of the true moments determined by simulation of the stochastic process, and are found to work well for the chosen parameter values. In the remaining part of this paper, the multivariate normal approximation will be used to investigate and illustrate properties of some of the more general models described earlier.

Comparison of the multivariate normal approximation and the deterministic curve with a stochastic epidemic.

(a) Exponential incubation period. In Table 1 we compare the multivariate normal approximation and deterministic curve with simulation estimates for the general stochastic epidemic in a population of size $n = 1000$ in which 990 are susceptible and 10 infected at $t = 0$. The incubation period is assumed to be exponentially distributed with a mean of 10 years ($\nu = 0.1$) and the reproductive ratio is chosen to be 6 ($\alpha = \beta\kappa = 0.6$), the values being chosen as those which might be appropriate for AIDS. Estimates of α of approximately 0.6 are reported by Bailey (1990) for Swiss data. In the table, values are given for the means and standard deviations of the numbers of susceptibles (X), infectives (Y) and diagnosed (AIDS, Z) individuals in the population; for ease of display,

TABLE 1
Comparison of results for the general stochastic epidemic with $n = 1000$, $X(0) = 990$, $Y(0) = 10$, $\alpha = 0.6$ *and* $\nu = 0.1$[1]

Time (Years)	X		Y		Z	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$t = 0$	990		10		0	
$t = 2$	970	8.1	27	7.9	3.1	1.7
	970	8.0	27	7.8	3.4	1.8
	970		27		3.4	
$t = 4$	920	25	68	22	12	4.2
	920	25	68	22	12	4.2
	920		68		12	
$t = 7$	727	72	221	59	52	15
	727	74	220	61	52	15
	720		227		53	
$t = 10$	406	90	442	61	152	34
	407	90	441	59	152	34
	387		456		157	
$t = 15$	90	29	503	24	407	43
	90	26	504	25	407	42
	82		503		415	
$t = 25$	10	3.8	225	22	765	23
	10	3.8	225	21	765	22
	9.8		221		769	
$t = 35$	4.3	2.2	86	11	910	12
	4.3	2.2	86	11	910	11
	4.2		84		911	

[1]The three lines in each band of results correspond to those from 10^4 simulations of the general stochastic epidemic, the multivariate normal approximation and the deterministic approximation.

correlations between the variables have not been given. The estimates obtained by simulation are based on 10^4 realisations of the stochastic epidemic, the normal approximation values are found by numerical solution of (3.5) from Isham (1991) while the deterministic approximation is obtained by solution of (2.4). The simulation estimates are also illustrated in Figure 1. It is clear that, as expected on theoretical grounds, the deterministic epidemic curve $y(t)$ increases too quickly over the first part of the epidemic. We also see that the normal approximation gives mean values which are very close to the simulation estimates. The standard deviations, while also close, are such that the normal approximation values for the standard deviations lie outside confidence intervals based on the simulations more often than would be expected if the variables were truly multivariate normal. Nevertheless the normal approximation values give a good idea of the true standard deviations and the approximation would improve with a larger population. Note that, for small t,

$X(t) + Y(t) \simeq n$ so that var $X(t) \simeq$ var $Y(t)$. Similarly, when the infection process has largely ceased, var $Y(t) \simeq$ var $Z(t)$. It is noticeable that the standard deviations grow relatively quite large as the epidemic progresses indicating that, given any particular initial conditions, there will be substantial variations between distinct realisations. Of course for the parameter values chosen, the probability $(\nu/\alpha)^{Y(0)} \simeq 1.65 \times 10^{-8}$, so that the chance of obtaining a realisation in which the epidemic dies out without all the susceptibles being infected is negligible in these simulations. Therefore the standard deviations shrink to zero as the epidemic dies away.

(b) Gamma incubation period. A second comparison of approximations and simulation results is given in Table 2, where we examine the effect of replacing the (single stage) exponential incubation period with one made up of three independent, exponentially distributed stages with means 1, 4 and 5 years, respectively ($\nu_1 = 1$, $\nu_2 = 0.25$, $\nu_3 = 0.2$). We then have the problem of

TABLE 2

Comparison of results for the stochastic epidemic with staged incubation period and $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$, $\boldsymbol{\alpha} = (1, 0, 1)$ *and* $\boldsymbol{\nu} = (1, 0.25, 0.2)$[1]

Time (Years)	X		Y		Z	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$t = 0$	990		10		0	
$t = 2$	967	11	33	11	0.54	0.73
	967	11	33	11	0.54	0.73
	967		33		0.54	
$t = 4$	919	31	77	30	3.8	2.1
	919	31	77	31	3.8	2.1
	918		78		3.8	
$t = 7$	739	83	239	76	21	8.2
	740	87	238	80	21	8.3
	730		249		22	
$t = 10$	443	106	482	85	75	24
	446	108	480	86	74	25
	418		505		77	
$t = 15$	106	44	603	26	292	56
	106	38	604	27	291	58
	92		606		302	
$t = 25$	7.6	3.5	211	35	782	37
	7.6	3.4	211	34	781	36
	7.0		202		791	
$t = 35$	3.2	1.9	41	13	956	13
	3.2	1.9	43	10	954	10
	3.1		40		957	

[1]The three lines in each band of results correspond to those from 10^4 simulations of the general stochastic epidemic, the multivariate normal approximation and the deterministic approximation.

distributing the 10 individuals already infected at $t = 0$ between the three incubation stages. Here we make the simple, but entirely arbitrary, assumption that all 10 are in the first of the three stages. We also assume that individuals are not infectious during the middle stage of the incubation period ($\beta_2 = 0$) but we keep infectiousness constant over the first and third stages ($\beta_1 = \beta_3$) in such a way that the reproductive ratio is again 6. Although the model distinguishes the numbers $Y^{(i)}(t)$, $i = 1, 2, 3$, in the three stages of the incubation period, for simplicity only their total $Y(t)$ is tabulated. The comments made above concerning Table 1 apply equally to this case also. The equations used to derive the normal approximation for this model are given in the Appendix [(A.3)–(A.6)]. The simulation results are shown graphically in Figure 2.

The effect of varying infectivity. If we try to compare Tables 1 and 2, it is hard to disentangle the effect of the staged incubation period from that of allowing the transmission probability to vary from stage to stage. In Table 3 and subsequent tables and figures, results will be given for the normal

TABLE 3
Comparison of results for staged incubation periods with $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$ *and* $R_0 = 6$[1]

Time (Years)	Single stage, $\nu = 1$ $\alpha = 0.6$ (see Table 1)		Three exponential stages, $\nu = (1, 0.25, 0.2)$ $\alpha = (0.6, 0.6, 0.6)$		$\alpha = (0.25, 0, 1.15)$		$\alpha = (0.5, 0, 1.1)$		$\alpha = (1, 0, 1)$	
	Y	Z	Y	Z	Y	Z	Y	Z	Y	Z
$t = 0$	10	0	10	0	10	0	10	0	10	0
$t = 2$	27	3.4	32	0.49	15	0.38	19	0.43	33	0.54
	7.8	1.8	8.4	0.70	3.1	0.61	4.8	0.65	11	0.73
$t = 4$	68	12	92	3.7	24	2.0	34	2.4	77	3.8
	22	4.2	27	2.0	7.7	1.4	11	1.5	31	2.1
$t = 7$	220	52	338	25	57	7.8	88	10	238	21
	61	15	74	7.6	21	3.1	32	4.0	80	8.3
$t = 10$	441	152	637	98	126	21	203	31	480	74
	59	34	47	23	45	7.5	66	11	86	25
$t = 15$	504	407	592	372	354	87	488	133	604	291
	25	42	34	41	87	29	71	40	27	58
$t = 25$	225	765	158	837	448	504	363	615	211	781
	21	22	19	20	56	76	55	62	34	36
$t = 35$	86	910	29	968	129	866	88	908	43	954
	11	11	6.3	6.4	31	32	21	21	10	10

[1]The two lines in each band give the means and standard deviations of the numbers of infectives (Y) and diagnoses (Z) for each of the chosen parameter sets.

approximation only (we assume that the case for this approximation has been made by Tables 1 and 2; as in those cases the corresponding simulation results for subsequent tables are very similar to those from the normal approximation). In Table 3, we give only the total number of infectives (Y) and the number of diagnosed (Z) (the numbers of susceptibles can be obtained by subtraction), and we consider a variety of parameter sets. The three stages of the incubation period will be as before but now several different sets of α_i, $\alpha_i = \beta_i \kappa$, are used, varying from β_i constant across all three stages, sets where $\beta_2 = 0$ but $\beta_1 \neq \beta_3$, to a set with $\beta_2 = 0$, $\beta_1 = \beta_3$ (as in Table 2). In each case the reproductive ratio is fixed at 6 as in Tables 1 and 2. For comparison, we repeat, from Table 1, the results obtained using a single stage incubation period. The evolution of the number of infectives over the epidemic, for each parameter set considered, is shown graphically in Figure 3(a), (b).

Compare first in Table 3 and Figure 3(a) the results obtained when the transmission probability stays fixed and where the incubation period consists of a single exponential stage, with those when the incubation period has three exponential stages (the first two pairs of columns). Although the mean incubation period is fixed at 10 years in both cases, its smaller variance in the latter case with fewer very short and very long incubation periods has the effect that there are more infectives and fewer diagnosed cases in the early part of the epidemic, but fewer infectives and more diagnosed in the later part. This observation confirms results in a recent paper by Malice and Kryscio (1989), which show that the expected cumulative number $[E(Y(t) + Z(t))]$ infected by time t is greater for the less variable incubation period distribution. Note that the proof of this result uses a general branching process approximation and therefore applies for small t in the finite population case.

Now consider the other results in Table 3 and Figure 3(b) which illustrate the effect of varying the transmission probability over the incubation period, while the reproductive ratio is held constant. It is clear that for fixed R_0 and $\beta_2 = \alpha_2 = 0$, the larger β_1 the faster the epidemic spreads through the population. On the other hand if we compare the $\alpha_i = 0.6$, $i = 1, 2, 3$, results with those when $\boldsymbol{\alpha} = (1, 0, 1)$, we see that after an initial spurt, the epidemic proceeds more slowly in the latter case (reflecting the fact that in that case the average α in the first stage of incubation is 1, but over the first two stages is only 0.2, as compared with 0.6 both times in the former case). Thus we see that changing the incubation period distribution and the transmission probability over the incubation period both have profound effects on the course of the epidemic. It is therefore important to fit well the tail behaviour of the incubation period distribution. Data is of course only available now on the left-hand tail although information on the right-hand tail is gradually increasing as the epidemic progresses. Unfortunately this information will inevitably be confounded with treatment effects as new methods of treating HIV infection are put into practice. Similarly, it is vital that understanding is achieved of how the infectiousness of HIV infected individuals varies over the incubation period.

Varying sexual activity. Note that in all these numerical comparisons we have assumed a single fixed value for the rate κ of partner change. It is a straightforward though rather tedious matter to write down the equations for the multivariate normal approximation when there are several sexual activity subgroups characterised by distinct values of κ. In the case so far considered with a three-stage incubation period, there are essentially four variables (the fifth being constrained by the fixed population size) which results in the set of 14 simultaneous differential equations given in the Appendix. Even if there were just two activity subgroups giving eight variables (assuming fixed numbers in each subgroup) we would need a set of 44 equations to determine all their second order properties. If individuals were allowed to change subgroup, then the number of equations would increase further.

Further modifications to the basic model. In the models for AIDS described so far, it has been assumed that infected individuals will take steps not to transmit HIV infection once they have progressed to full AIDS. However, it seems likely that most such individuals will be aware of their infected status and cease transmission of infection at an earlier stage. To investigate the effect of this we suppose that the third infectious stage is shortened to have a mean of three years ($\nu_3 = 0.\dot{3}$) rather than 5 ($\nu_3 = 0.2$) even though full AIDS will not develop for another two years on average. In Table 4 figures are tabulated showing the effect of this, when $\boldsymbol{\alpha} = (1, 0, 1)$. Note that this choice means that, for the first set of columns (repeated from Table 3), the mean total infectious period is 10 years and the reproductive ratio is 6 while in the second set the mean total infectious period is eight years and the reproductive ratio therefore drops to 4. In the third set of columns we increase $\boldsymbol{\alpha}$ to $(1.5, 0, 1.5)$ so that the reproductive ratio is kept at 6 even though the mean total infectious period is only eight years. In all three cases we tabulate the numbers of susceptibles X, those who are infectious (transmitting infection) Y^* and those who are infected but no longer transmitting infection Z^*. In the first case, these numbers are the variables X, Y, Z as before, but in the second and third cases Z^* includes those with AIDS but also those in the final part of their incubation period to AIDS. These latter are of course excluded from Y^*. The evolution of the number Y^* of infectives for the three cases is illustrated in Figure 4.

From the table we see that when the infectious period is shortened (mean eight years) so that R_0 drops to 4, the infection spreads through the population a little more slowly (as is to be expected). If however the reproductive ratio is kept fixed ($R_0 = 6$) while the incubation period is shortened, the increased initial infectiousness ($\alpha_1 = 1.5$) raises dramatically the speed at which the epidemic spreads, illustrating again the vital influence of the transmission probability (β_1) at the beginning of the incubation period on the rate at which the epidemic affects the population.

Another modification to the basic model which we might wish to make is the following. The argument for taking the infection rate in the AIDS interpretation of the general stochastic epidemic to be $\beta\kappa X(t)Y(t)/n$ is that new partnerships involving susceptibles occur at a total rate $\kappa X(t)$ and it is

TABLE 4

Comparison of results when transmission of infection ceases before the end of the incubation period, with $n = 1000$, $X(0) = 990$ *and* $Y(0) = Y^{(1)}(0) = 10$[1]

Time (Years)	$\nu = (1, 0.25, 0.2)$, $\alpha = (1, 0, 1)$, $R_0 = 6$			$\nu = (1, 0.25, 0.\dot{3})$, $\alpha = (1, 0, 1)$, $R_0 = 4$			$\nu = (1, 0.25, 0.\dot{3})$, $\alpha = (1.5, 0, 1.5)$, $R_0 = 6$		
	X	$Y = Y^*$	$Z = Z^*$	X	Y^*	Z^*	X	Y^*	Z^*
$t = 0$	990	10	0	90	10	0	990	10	0
$t = 2$	967	33	0.54	967	32	0.84	934	65	1.1
	11	11	0.73	11	11	0.91	24	24	1.1
$t = 4$	919	77	3.8	923	72	5.6	758	232	11
	31	31	2.1	31	29	2.6	84	81	4.5
$t = 7$	740	238	21	769	203	29	314	604	81
	87	80	8.3	83	74	11	86	64	25
$t = 10$	446	480	74	516	391	93	89	648	264
	108	86	25	113	83	33	28	27	49
$t = 15$	106	604	291	178	491	330	14	356	630
	38	27	58	55	28	71	5.1	40	43
$t = 25$	7.6	211	781	33	152	815	3.2	50	947
	3.4	34	36	8.5	33	38	1.9	10	10
$t = 35$	3.2	4.3	954	22	25	954	2.6	5.2	992
	1.9	10	10	5.7	8.0	10	1.7	2.5	3.0

[1]The two lines in each band give the means and standard deviations of the numbers of susceptibles (X), those transmitting infection (Y^*) and those who have ceased to transmit infection (Z^*) for each of the given parameter sets.

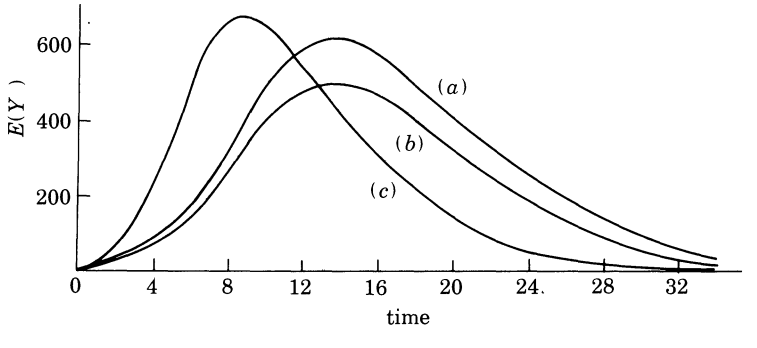

FIG. 4. *Comparison of expected numbers of infectives* $E(Y^*)$ *when transmission of infection ceases before the end of the third stage of the incubation period*: (*a*) *third stage of infectious period has mean 5 years, with* $\nu = (1, 0.25, 0.2)$, $\alpha = (1, 0, 1)$, $R_0 = 6$; (*b*) *Third stage of infectious period has mean 3 years, with* $\nu = (1, 0.25, 0.\dot{3})$, $\alpha = (1, 0, 1)$, $R_0 = 4$; (*c*) *Third stage of infectious period has mean 3 years, with* $\nu(1, 0.25, 0.\dot{3})$, $\alpha = (1.5, 0, 1.5)$, $R_0 = 6$.

TABLE 5

Comparison of results for models with standard and modified infection rates, with $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$, $\boldsymbol{\nu} = (1, 0.25, 0.2)$, $\boldsymbol{\alpha} = (1, 0, 1)$ *and* $R_0 = 6$[1]

Time (Years)	Standard model			Modified model		
	X	Y	Z	X	Y	Z
$t = 0$	990	10	0	990	10	0
$t = 2$	967	33	0.54	967	33	0.54
	11	11	0.73	11	11	0.73
$t = 4$	919	77	3.8	919	77	3.8
	31	31	2.1	31	31	2.1
$t = 7$	740	238	21	737	242	21
	87	80	8.3	89	82	8.4
$t = 10$	446	480	74	426	498	75
	108	86	25	113	90	26
$t = 15$	106	604	291	70	630	300
	38	27	58	34	30	60
$t = 25$	7.6	211	781	0.12	197	802
	3.4	34	36	0.35	34	34
$t = 35$	3.2	4.3	954	0.00	36	964
	1.9	10	10	0.01	8.8	8.8

[1]The two lines in each band give the means and standard deviations of the numbers X, Y, Z in each of the three main classes. The infection rate for the standard model is $(\alpha_1 Y^{(1)} + \alpha_2 Y^{(2)} + \alpha_3 Y^{(3)})X/n$ while that for the modified model is $(\alpha_1 Y^{(1)} + \alpha_2 Y^{(2)} + \alpha_3 Y^{(3)})X/(X + Y^{(1)} + Y^{(2)} + Y^{(3)})$.

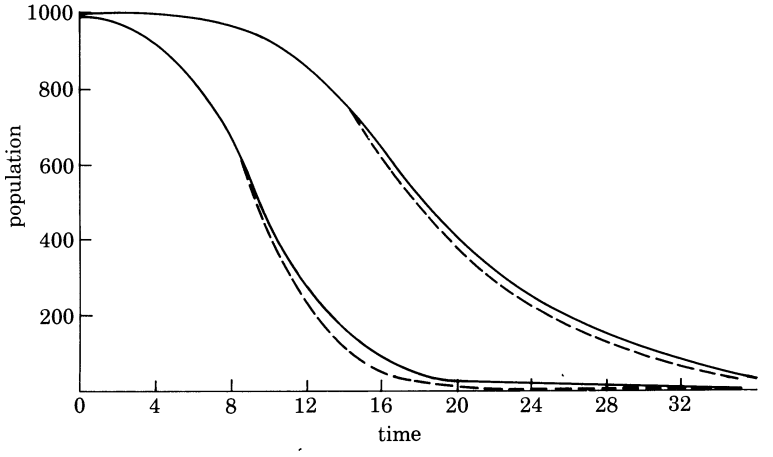

FIG. 5. *The expected numbers of susceptibles* $[E(X)$, *lower curve*$]$ *and susceptibles + infectives* $[E(X + Y)$, *upper curve*$]$ *for the stochastic model with three-stage incubation period and* $n = 1000$, $X(0) = 990$, $Y(0) = Y^{(1)}(0) = 10$, $\boldsymbol{\nu} = (1, 0.25, 0.2)$, $\boldsymbol{\nu} = (1, 0, 1)$, $R_0 = 6$: (i) (*unbroken curves*) *standard model with infection rate* $(\alpha_1 Y^{(1)} + \alpha_2 Y^{(2)} + \alpha_3 Y^{(3)})X/n$; (ii) (*broken curves*) *modified model with infection rate* $(\alpha_1 Y^{(1)} + \alpha_2 Y^{(2)} + \alpha_3 Y^{(3)})X/(X + Y^{(1)} + Y^{(2)} + Y^{(3)})$.

assumed that each susceptible chooses a partner at random from the population so that the probability that the chosen partner is an infective is $Y(t)/n$. Implicit in this is the assumption that those who are infected and have developed AIDS [of whom there are $Z(t)$] will not transmit infection to any partners although they will still form new partnerships. But it could be more realistic to assume that these individuals play no further part in the epidemic, that is, that they cease to form new partnerships. In this case the appropriate infection rate is $\beta\kappa X(t)Y(t)/[X(t) + Y(t)]$, subject to whatever other modifications are made to the model. Thus, for example, if the incubation period

TABLE 6

Comparison of results with increasing population size n and with $X(0) = 0.99n$, $Y(0) = Y^{(1)}(0) = 0.01n$, $\nu = (1, 0.25, 0.2)$, $\alpha = (1, 0, 1)$ *and* $R_0 = 6$[1]

Time (Years)	$1000X/n$		$1000Y/n$		$1000Z/n$	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$t = 0$	990		10		0	
$t = 2$	967	11	33	11	0.54	0.73
	967	5.1	33	4.9	0.54	0.33
	967	3.4	33	3.5	0.54	0.23
	967	1.6	33	1.5	0.54	0.10
$t = 4$	919	31	77	31	3.8	2.1
	918	14	78	14	3.8	0.94
	918	9.9	78	9.7	3.8	0.66
	918	4.4	78	4.3	3.8	0.30
$t = 7$	740	87	238	80	21	8.3
	732	38	247	35	21	3.7
	731	27	248	25	21	2.6
	730	12	249	11	21	1.2
$t = 10$	446	108	480	86	74	25
	423	46	501	36	76	11
	420	32	503	25	77	7.9
	418	14	505	11	77	3.5
$t = 15$	106	38	604	27	291	58
	94	15	606	13	300	25
	93	11	606	9.1	301	18
	92	4.7	606	4.1	302	7.8
$t = 25$	7.6	3.4	211	34	781	36
	7.1	1.4	204	14	789	15
	7.1	0.99	203	10	790	10
	7.1	0.44	202	4.5	791	4.6
$t = 35$	3.2	1.9	36	8.8	961	9.1
	3.1	1.4	41	4.2	956	4.3
	3.1	0.58	40	3.0	956	3.0
	3.1	0.26	40	1.3	957	1.4

[1]The four lines in each band give the results when the population size n takes values 1000, 5000, 10000 and 50000 respectively.

consists of three exponential stages with corresponding transmission probabilities $\beta_1, \beta_2, \beta_3$, infections will occur at rate

$$\kappa\big(\beta_1 Y^{(1)}(t) + \beta_2 Y^{(2)}(t) + \beta_3 Y^{(3)}(t)\big)/ \big(X(t) + Y^{(1)}(t) + Y^{(2)}(t) + Y^{(3)}(t)\big). \tag{5.1}$$

The equations which determine the normal approximation to the model are given in the Appendix [(A.3)–(A.5), (A.7)].

In Table 5 we see the effect of this modification for the case already illustrated in Table 2, with $\boldsymbol{\nu} = (1, 0.25, 0.2)$, $\boldsymbol{\alpha} = (1, 0, 1)$ so that $R_0 = 6$, with $n = 1000$ [$X(0) = 990$, $Y^{(1)}(0) = 10$]. It is clear that initially the effect of using $n - Z(t)$ rather than n as the denominator in (5.1) is very small, but gradually the infection rate in the former case gains on that for the latter case and the corresponding epidemic spreads more quickly. However, at least with these parameter values, the effect of the modification to the model is relatively small. This effect is shown graphically in Figure 5.

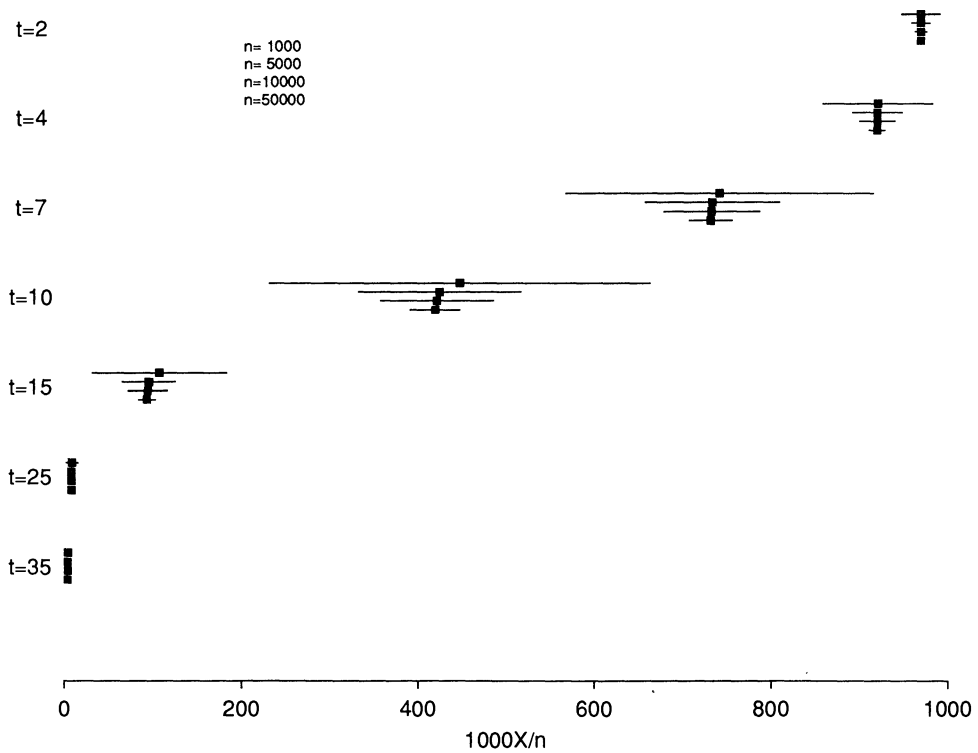

FIG. 6. *Comparison of results with increasing population size n and with $X(0) = 0.99n$, $Y(0) = Y^{(1)}(0) = 0.01n$, $\boldsymbol{\nu} = (1, 0.25, 0.2)$, $\boldsymbol{\alpha} = (1, 0, 1)$, $R_0 = 6$. At each time point the square ■ gives the expected value of $1000X/n$ and the bar shows the interval ± 2(standard deviation of $1000X/n$) for four values of n*: 1000, 5000, 10000 *and* 50000.

The effect of increasing the population size. Finally, in Table 6 we illustrate the effect of the diffusion approximation by increasing the population size n. To make the comparison straightforward we tabulate the proportions X/n, Y/n, Z/n each multiplied by 1000, for n taking values 1000, 5000, 10,000 and 50,000 with the initial proportions $X(0)/n = 0.99$, $Y(0)/n = Y^{(1)}(0)/n = 0.01$ fixed and the other parameters as before (Table 2, Table 5). By comparison with Table 2 it is clear that, to the degree of accuracy given, the stochastic mean when $n = 50{,}000$ is the same as the solution of the corresponding deterministic model. While the stochastic means increase in direct proportion to n, the standard deviations increase in proportion to $\sqrt{n}$ and therefore the coefficient of variation decreases as $1/\sqrt{n}$. The convergence is represented diagramatically in Figure 6.

6. Discussion. In this paper, some of the recent work on stochastic models for epidemics and their deterministic counterparts has been reviewed. We have concentrated particularly on the development of simple models for the AIDS epidemic and the sort of behaviour these exhibit. In particular, the method of approximating the distribution of the state of an epidemic by a multivariate normal observation based on a diffusion approximation is a valuable one. It makes it possible to improve on the solution of the corresponding deterministic model as an approximation to the mean of the stochastic process. This may be particularly important when the population or the many subpopulations into which the population is divided are small. But we also obtain approximations to the second moments of the state of the epidemic which, at least for the sorts of parameter sets investigated so far, appear to be a reliable guide to the true moments. These approximations are straightforward to determine and can be found with relatively little computational effort. It is apparent that the variability between realisations of the same epidemic model is in general high, and this has implications for prediction. In particular, this provides another reason (in addition to that of uncertainty about model assumptions and parameter values) why prediction more than a very short while ahead is probably inadvisable.

APPENDIX

Suppose that the incubation period consists of three independent exponentially distributed stages with hazards, ν_i, and corresponding transition probabilities β_i, $i = 1, 2, 3$. Denote the numbers of infectives in these stages by $\mathbf{Y} = (Y^{(1)}, Y^{(2)}, Y^{(3)})$ and the total infection rate by $\lambda_{X,Y}(t)$. We shall consider here two possible forms for $\lambda_{X,Y}(t)$:

$$\lambda_{X,Y}(t) = \left[\beta_1 Y^{(1)}(t) + \beta_2 Y^{(2)}(t) + \beta_3 Y^{(3)}(t)\right]\kappa X(t)/n \tag{A.1}$$

and

$$\lambda_{X,Y}(t) = \left[\beta_1 Y^{(1)}(t) + \beta_2 Y^{(2)}(t) + \beta_3 Y^{(3)}(t)\right]\kappa X(t)/\left[X(t) + Y^{(1)}(t) + Y^{(2)}(t) + Y^{(3)}(t)\right]. \tag{A.2}$$

The forward equations for the first- and second-order moments of the epidemic model are as follows, where we use the abbreviated notation

$$E(X(t)) = \mu_X, \qquad \operatorname{var}(X(t)) = \sigma_{XX},$$

$$E(Y^{(i)}(t)) = \mu_i, \qquad \operatorname{cov}(X(t), Y^{(i)}(t)) = \sigma_{Xi}, \qquad \operatorname{cov}(Y^{(i)}(t), Y^{(j)}(t)) = \sigma_{ij}$$

for $i, j = 1, 2, 3$ and the dependence on time is not represented explicitly:

$$\begin{aligned} d\mu_X/dt &= -E(\lambda_{X,Y}), \\ d\mu_1/dt &= E(\lambda_{X,Y}) - \nu_1\mu_1, \\ d\mu_2/dt &= \nu_1\mu_1 - \nu_2\mu_2, \\ d\mu_3/dt &= \nu_2\mu_2 - \nu_3\mu_3, \end{aligned} \tag{A.3}$$

$$\begin{aligned} d\sigma_{XX}/dt &= E(\lambda_{X,Y}) - 2\operatorname{cov}(X, \lambda_{X,Y}), \\ d\sigma_{11}/dt &= E(\lambda_{X,Y}) + 2\operatorname{cov}(Y^{(1)}, \lambda_{X,Y}) - 2\nu_1\sigma_{11} + \nu_1\mu_1, \\ d\sigma_{22}/dt &= 2\nu_1\sigma_{12} - 2\nu_2\sigma_{22} + \nu_1\mu_1 + \nu_2\mu_2, \\ d\sigma_{33}/dt &= 2\nu_2\sigma_{23} - 2\nu_3\sigma_{33} + \nu_2\mu_2 + \nu_3\mu_3, \end{aligned} \tag{A.4}$$

$$\begin{aligned} d\sigma_{X1}/dt &= \operatorname{cov}(X, \lambda_{X,Y}) - \operatorname{cov}(Y^{(1)}, \lambda_{X,Y}) - \nu_1\sigma_{X1} - E(\lambda_{X,Y}), \\ d\sigma_{X2}/dt &= -\operatorname{cov}(Y^{(2)}, \lambda_{X,Y}) + \nu_1\sigma_{X1} - \nu_2\sigma_{X2}, \\ d\sigma_{X3}/dt &= -\operatorname{cov}(Y^{(3)}, \lambda_{X,Y}) + \nu_2\sigma_{X2} - \nu_3\sigma_{X3}, \\ d\sigma_{12}/dt &= \operatorname{cov}(Y^{(2)}, \lambda_{X,Y}) - (\nu_1 + \nu_2)\sigma_{12} + \nu_1\sigma_{11} - \nu_1\mu_1, \\ d\sigma_{13}/dt &= \operatorname{cov}(Y^{(3)}, \lambda_{X,Y}) - \nu_1\sigma_{13} + \nu_2\sigma_{12} - \nu_3\sigma_{13}, \\ d\sigma_{23}/dt &= \nu_1\sigma_{13} + \nu_2(\sigma_{22} - \sigma_{23}) - \nu_3\sigma_{23} - \nu_2\mu_2. \end{aligned} \tag{A.5}$$

When $\lambda_{X,Y}$ is given by (A.1), the normal approximation to the joint distribution of X and $\mathbf{Y}$ is obtained by solving equations (A.3)–(A.5), substituting the approximations

$$\begin{aligned} E(\lambda_{X,Y}) &= \kappa(\beta_1\sigma_{X1} + \beta_2\sigma_{X2} + \beta_3\sigma_{X3})/n + \kappa\mu_X\mu_B/n, \\ \operatorname{cov}(X, \lambda_{X,Y}) &\simeq \kappa(\beta_1\sigma_{X1} + \beta_2\sigma_{X2} + \beta_3\sigma_{X3})\mu_X/n + \kappa\sigma_{XX}\mu_B/n, \\ \operatorname{cov}(Y^{(1)}, \lambda_{X,Y}) &\simeq \kappa(\beta_1\sigma_{11} + \beta_2\sigma_{12} + \beta_3\sigma_{13})\mu_X/n + \kappa\sigma_{X1}\mu_B/n, \\ \operatorname{cov}(Y^{(2)}, \lambda_{X,Y}) &\simeq \kappa(\beta_1\sigma_{12} + \beta_2\sigma_{22} + \beta_3\sigma_{23})\mu_X/n + \kappa\sigma_{X2}\mu_B/n, \\ \operatorname{cov}(Y^{(3)}, \lambda_{X,Y}) &\simeq \kappa(\beta_1\sigma_{13} + \beta_2\sigma_{23} + \beta_3\sigma_{33})\mu_X/n + \kappa\sigma_{X3}\mu_B/n, \end{aligned} \tag{A.6}$$

where $\mu_B = \beta_1\mu_1 + \beta_2\mu_2 + \beta_3\mu_3$, and which can be deduced using

$$\operatorname{cov}(U, VW) \simeq \sigma_{UV}\mu_W + \sigma_{UW}\mu_V,$$

which is exact when U, V, W have a multivariate normal distribution.

When $\lambda_{X,Y}$ is given by (A.2) the normal approximation is obtained by solving (A.3–A.5), in conjunction with approximations for $E(\lambda_{X,Y})$, $\operatorname{cov}(X, \lambda_{X,Y})$ and $\operatorname{cov}(Y^{(i)}, \lambda_{X,Y})$, $i = 1, 2, 3$, obtained by using Taylor expansions to second order, namely

$$\begin{aligned} E(\lambda_{X,Y}) &\simeq \kappa\mu_X\mu_B/\mu_{\cdot} + \kappa\mu_X\mu_B(\sigma_{X\cdot} + \sigma_{1\cdot} + \sigma_{2\cdot} + \sigma_{3\cdot})/\mu_{\cdot}^3 \\ &\quad - \kappa\{\sigma_{X\cdot}\mu_B + \mu_X(\beta_1\sigma_{1\cdot} + \beta_2\sigma_{2\cdot} + \beta_3\sigma_{3\cdot})\}/\mu_{\cdot}^2 \\ &\quad + \kappa(\beta_1\sigma_{X1} + \beta_2\sigma_{X2} + \beta_3\sigma_{X3})/\mu_{\cdot}, \\ \text{(A.7)} \qquad \operatorname{cov}(X, \lambda_{X,Y}) &\simeq \kappa\{\sigma_{XX}\mu_B + \mu_X(\beta_1\sigma_{X1} + \beta_2\sigma_{X2} + \beta_3\sigma_{X3})\}/\mu_{\cdot} \\ &\quad - \kappa\mu_X\mu_B\sigma_{X\cdot}/\mu_{\cdot}^2, \\ \operatorname{cov}(Y^{(i)}, \lambda_{X,Y}) &\simeq \kappa\{\sigma_{Xi}\mu_B + \mu_X(\beta_1\sigma_{1i} + \beta_2\sigma_{2i} + \beta_3\sigma_{3i})\}/\mu_{\cdot} \\ &\quad - \kappa\mu_X\mu_B\sigma_{i\cdot}/\mu_{\cdot}^2, \end{aligned}$$

where $\mu_{\cdot} = \mu_X + \mu_1 + \mu_2 + \mu_3$, $\mu_B = \beta_1\mu_1 + \beta_2\mu_2 + \beta_3\mu_3$, $\sigma_{X\cdot} = \sigma_{XX} + \sigma_{X1} + \sigma_{X2} + \sigma_{X3}$ and $\sigma_{i\cdot} = \sigma_{Xi} + \sigma_{i1} + \sigma_{i2} + \sigma_{i3}$, $i = 1, 2, 3$.

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.

ANDERSON, R. M., MEDLEY, G. F., MAY, R. M. and JOHNSON, A. M. (1986). A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA J. Math. Appl. Med. Biol.* **3** 229–263.

BACCHETTI, P. and MOSS, A. R. (1989). Incubation period of AIDS in San Francisco. *Nature* **328** 251–253.

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

BAILEY, N. T. J. (1990). Applications of stochastic epidemic modelling in the public health control of HIV/AIDS. In *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomath.* (J.-P. Gabriel, C. Lefèvre and P. Picard, eds.) **86** 14–20. Springer, Berlin.

BARBOUR, A. D. (1972). The principle of diffusion of arbitrary constants. *J. Appl. Probab.* **9** 519–541.

BARBOUR, A. D. (1974). On a functional central limit theorem for Markov population processes. *Adv. in Appl. Probab.* **6** 21–39.

BILLARD, L., MEDLEY, G. F. and ANDERSON, R. M. (1990). The incubation period for the AIDS virus. In *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomath.* (J.-P. Gabriel, C. Lefèvre and P. Picard, eds.) **86** 21–35. Springer, Berlin.

BILLARD, L. and ZHAO, Z. (1991). Three-stage stochastic epidemic model: an application to AIDS. *Math. Biosci.* **107** 431–449.

BILLARD, L. and ZHAO, Z. (1992). The stochastic general epidemic model revisited and a generalisation. To appear.

BLYTHE, S. P. and ANDERSON, R. M. (1988). Variable infectiousness in HIV transmission models. *IMA J. Math. Appl. Med. Biol.* **5** 181–200.

BROOKMEYER, R. and LIAO, J. (1990). Statistical modelling for the AIDS epidemic for forecasting health care needs. *Biometrics* **46** 1151–1163.

DALEY, D. J. and KENDALL, D. G. (1965). Stochastic rumours. *J. Inst. Math. Appl.* **1** 42–55.

DANIELS, H. E. (1991). A look at perturbation approximations for epidemics. In *Selected Proc. Sheffield Symp. Applied Probability. IMS Lecture Notes Monograph Ser.* (I. V. Basawa and R. L. Taylor, eds.) **18**. IMS, Hayward, CA.

DIETZ, K. (1987). Epidemiological models for sexually transmitted infections. In *Proc. 1st World Congress of Bernoulli Society (Tashkent, 1986)* **2** 539–542. VNU Science Press, Utrecht.

DIETZ, K. (1988). On the transmission dynamics of HIV. *Math. Biosci.* **90** 397–414.

DIETZ, K. and HADELER K. P. (1988). Epidemiological models for sexually transmitted diseases. *J. Math. Biol.* **26** 1–25.

ISHAM, V. (1988). Mathematical modelling of the transmission dynamics of HIV infection and AIDS: a review. *J. Roy. Statist. Soc. Ser. A* **151** 5–30.

ISHAM, V. (1991). Assessing the variability of stochastic epidemics. *Math. Biosci.* **107** 209–224.

JACQUEZ, J. A. and O'NEILL, P. (1991). Reproduction numbers and thresholds in stochastic epidemic models I. Homogeneous populations. *Math. Biosci.* **107** 161–186.

JACQUEZ, J. A. and SIMON, C. P. (1990). AIDS: the epidemiological significance of two different mean rates of partner change. *IMA J. Math. Appl. Med. Biol.* **7** 27–32.

JACQUEZ, J. A., SIMON, C. P. and KOOPMAN, J. (1989). Structured mixing: heterogeneous mixing by the definition of activity groups. In *Mathematical and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomath.* (C. Castillo-Chavez, ed.) **83** 301–315. Springer, Berlin.

JACQUEZ, J. A., SIMON, C. P. and KOOPMAN, J. (1991). The reproduction number in deterministic models of contagious diseases. *Comm. Theor. Biol.* **2** 159–209.

KOOPMAN, J. S., SIMON, C. P., JACQUEZ, J. A. and PARK, T. S. (1989). Selective contact within structured mixing with an application to HIV transmission risk from oral and anal sex. In *Mathematical and Statistical Approaches to AIDS Epidemiology. Lecture Notes in Biomath.* (C. Castillo-Chavez, ed.) **83** 316–349. Springer, Berlin.

KRYSCIO, R. J. (1975). The transition probabilities of the general stochastic epidemic model. *J. Appl. Probab.* **12** 415–424.

KURTZ, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* **7** 49–58.

KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8** 344–356.

LEFÈVRE, C. (1990). Stochastic epidemic models for S-I-R infectious diseases: a brief survey of the recent general theory. In *Stochastic Processes in Epidemic Theory. Lecture Notes in Biomath.* (J.-P. Gabriel, C. Lefèvre and P. Picard, eds.) **86** 1–12. Springer, Berlin.

LONGINI, I. M., CLARK, W. S., BYERS, R. H., WARD, J. W., DARROW, W. W., LEMP, G. F. and HETHCOTE, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine* **8** 831–43.

LUDWIG, D. (1973). Stochastic approximation for the general epidemic. *J. Appl. Probab.* **10** 263–276.

MALICE, M. P. and KRYSCIO, R. J. (1989). On the role of variable incubation periods in simple epidemic models. *IMA J. Math. Appl. Med. Biol.* **6** 233–242.

MORRIS, M. (1991a). Networks and diffusion: modelling the effects of selective mixing on the spread of disease. Unpublished manuscript.

MORRIS, M. (1991b). A log-linear modelling framework for selective mixing. *Math. Biosci.* **107** 349–377.

PUGLIESE, A. (1990). Contact matrices for multipopulation epidemic models: How to build a consistent matrix close to data? Technical report, Univ. Trento, Italy.

SCALIA-TOMBA, G. (1991). The effect of structural behaviour change on the spread of HIV in a one-sex population. *Math. Biosci.* **107** 547–555.

SEVERO, N. C. (1967). Two theorems on solutions of differential-difference equations and applications to epidemic theory. *J. Appl. Probab.* **4** 271–280.

SEVERO, N. C. (1969). A recursion theorem in solving differential-difference equations and applications to some stochastic processes. *J. Appl. Probab.* **6** 673–681.

WHITTLE, P. (1957). On the use of the normal approximation in the treatment of stochastic processes. *J. Roy. Statist. Soc. Ser. B* **19** 268–281.

DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON WC1E 6BT
ENGLAND