Available online at www.sciencedirect.com
SCIENCE DIRECT®
Mathematical Biosciences 198 (2005) 119–131
Mathematical Biosciences
www.elsevier.com/locate/mbs

The Kermack–McKendrick epidemic model revisited ☆

Fred Brauer *

Department of Mathematics, University of British Columbia, Vancouver, B.C. V6T 1Z2, Canada

Received 19 November 2004; received in revised form 6 April 2005; accepted 20 July 2005
Available online 30 August 2005

Abstract

The Kermack–McKendrick epidemic model of 1927 is an age of infection model, that is, a model in which the infectivity of an individual depends on the time since the individual became infective. A special case, which is formulated as a two-dimensional system of ordinary differential ordinary differential equations, has often been called the Kermack–McKendrick model. One of the products of the SARS epidemic of 2002–2003 was a variety of epidemic models including general contact rates, quarantine, and isolation. These models can be viewed as age of infection epidemic models and analyzed using the approach of the full Kermack–McKendrick model. All these models share the basic properties that there is a threshold between disappearance of the disease and an epidemic outbreak, and that an epidemic will die out without infecting the entire population.

Keywords: Epidemic; Age of infection model; SARS models

1. Introduction

Almost since the beginning of recorded history there have been epidemics, sudden outbreaks of a disease that infects a substantial portion of the population in a region before it disappears. One

☆ This research was supported by MITACS and an NSERC Research Grant.
* Tel.: +604 733 1860; fax: +604 822 6074.
E-mail address: brauer@math.ubc.ca

0025-5564/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2005.07.006

of the early triumphs of mathematical epidemiology [1] was the formulation of a simple model that predicted just such behaviour. What has become known as the Kermack–McKendrick epidemic model is a special case of the model described in [1]. This model is a system of two ordinary differential equations

$$\begin{aligned} S' &= -\beta SI \\ I' &= (\beta S - \alpha)I. \end{aligned} \tag{1}$$

The population being studied is divided into compartments, namely a susceptible class S, an infective class I, and a removed class R. The model is based on the following assumptions:

(1) An average member of the population makes contact sufficient to transmit infection with βN others per unit time, where N represents total population size and is a function of time if there are disease deaths.
(2) Infectives leave the infective class at rate α per unit time.
(3) There is no entry into or departure from the population, except possibly through death from the disease.

The system (1) has been used successfully to fit data from many epidemics, and it is very easily analyzed qualitatively; see for example, [2, Section 7.2]. Two fundamental properties of this model are (i) that there is a basic reproduction number determining whether the disease will die out without spreading or whether there will be an epidemic, and (ii) that the epidemic will die out leaving some members of the population uninfected.

The basic reproduction number R_0 is defined to be the number of secondary infections caused by a single infective introduced into a wholly susceptible population of initial size K over the course of the infection of this single infective. In the situation described by (1) an infective makes βK contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period is $1/\alpha$; thus the basic reproduction number is $\beta K/\alpha$.

Our definition of an epidemic is a situation in which the number of infectives grows initially. Because the equation for I may be approximated for t close to zero by

$$I' = (\beta K - \alpha)I$$

with initial exponential growth rate

$$r = \beta K - \alpha = \alpha(R_0 - 1)$$

we see that there is an epidemic if and only if $R_0 > 1$.

The orbits in the (S, I)-plane of (1) are the curves

$$I = -S + \frac{\alpha}{\beta}\log S + c, \tag{2}$$

with c an arbitrary constant of integration. The constant c is determined by the initial values $S(0)$ and $I(0)$ of S and I, respectively. With $S(0) + I(0) = K$ we have

$$c = K - \frac{\alpha}{\beta}\log S(0).$$

If we use the fact that $\lim_{t\to\infty} I(t) = 0$, and let $S_\infty = \lim_{t\to\infty} S(t)$, then we obtain

$$K - \frac{\alpha}{\beta}\log S(0) = S_\infty - \frac{\alpha}{\beta}\log S_\infty, \tag{3}$$

the *final size equation*. Since no orbit reaches the I-axis, $S > 0$ for all times. In particular, $S_\infty = \lim_{t\to\infty} S(t) > 0$, which implies that part of the population escapes infection. The final size equation makes it possible to calculate the fraction S_∞/K of the population that escapes the epidemic.

The SARS epidemic of 2002–2003 renewed interest in the modelling of epidemics. Until the SARS epidemic, most of the work in mathematical epidemiology concentrated on studies of specific diseases or on the interplay between epidemiological and demographic effects. Although the AIDS epidemic renewed interest in epidemic modelling, the long epidemiological time scale for AIDS meant that it was necessary to include demographic effects (recruitment into and departure from a sexually active population). The study of short-term models which did not incorporate demographic effects was neglected, with the notable exception of [3]. Thus the studies of the SARS epidemic began with the original so-called Kermack–McKendrick model but then incorporated a variety of factors to make models more realistic; some dynamic models may be found in [4,5]. While these more recent extensions formulated for SARS included calculations for the basic reproductive number they did not establish the property (ii) formally. Even for the simple model (1) almost all descriptions of its analysis except [6] have shown graphically that the infection dies out but have omitted an analytic proof.

The purpose of this work is to show that the properties (i) and (ii) hold for an age of infection model closely related to the original Kermack–McKendrick model, of which (1) was a special case and thence to conclude that these properties hold for the extended models which have come out of the SARS epidemic. Our age of infection model is more general than the Kermack–McKendrick age of infection model in that it allows general contact rates. Thus our contribution is to round out the general theory. We remark that since S is a monotone decreasing function of time there is no possibility of periodic solutions, and since I approaches zero there is no possibility of multiple endemic equilibria or backward bifurcations. The general theory of epidemics is thus much simpler than the corresponding theory for models which include demographic effects, that admit possibilities such as endemic situations and periodic solutions.

2. General contact rates

The assumption in the model (1) of a rate of contacts per infective which is proportional to population size N, called *bilinear incidence*, was used in all the early epidemic models. It is more realistic to assume a contact rate which is a non-increasing function of total population size. For example, a situation in which the number of contacts per infective in unit time is constant, called *standard incidence*, is probably a more accurate description for sexually transmitted diseases [7–9].

We will assume that on the average a member of the population makes $C(N)$ contacts in unit time with $C'(N) \geqslant 0$, [10,11] and we define

$$\beta(N) = \frac{C(N)}{N}.$$

It is reasonable to assume $\beta'(N) \leqslant 0$ to express the idea of saturation in the number of contacts. Then bilinear incidence corresponds to the choice $C(N) = \beta N$, $\beta(N) = \beta$, and standard incidence corresponds to the choice $C(N) = \lambda$, $\beta(N) = \lambda/N$. The assumptions $C(N) = N\beta(N)$, $C'(N) \geqslant 0$ imply that

$$\beta(N) + N\beta'(N) \geqslant 0. \tag{4}$$

Various other forms, satisfying (4) have been used for the contact function $C(N)$; see for example, [11–14]. If there are no disease deaths, so that the total population size is constant, all incidence assumptions involving total population size are equivalent, but if there are disease deaths the behaviour of a model may depend on the form of the incidence.

An epidemic model in which the incidence is assumed to depend on total population size must include an equation for total population size. This forces us to make a distinction between members of the population who die of the disease and members of the population who recover with immunity against reinfection. We assume that a fraction f of the members leaving the infective class at time t recover and the remaining fraction $(1 - f)$ die of disease. For emerging diseases it is not immediately clear that recovery brings permanent immunity, but the short time scale being considered suggests that it is reasonable to assume immunity at least for the duration of a single epidemic outbreak.

We generalize (1) to include a general contact rate. We use S, I, and N as variables, with $N = S + I + R$. It will be convenient to use N as a model variable in place of R because the rate of new infections is now $\beta(N)SI$. We obtain the three-dimensional model

$$\begin{aligned} S' &= -\beta(N)SI, \\ I' &= \beta(N)SI - \alpha I, \\ N' &= -(1 - f)\alpha I. \end{aligned} \tag{5}$$

We also have the equation $R' = f\alpha I$, but we need not include it in the model since R is determined when S, I, and N are known.

3. An age of infection epidemic model

The general model described by Kermack and McKendrick included a dependence of infectivity on the time since becoming infected (age of infection). We will describe an age of infection epidemic model similar to the Kermack–McKendrick model but even more general in allowing a general contact rate. We continue to let $S(t)$ denote the number of susceptibles at time t and $R(t)$ the number of members who have recovered with immunity, but now we let $I^*(t)$ denote the number of infected (but not necessarily infective) members. We also let $\phi(t)$ be the total infectivity at time t.

We make the following assumptions:

(1) An average member of the population makes $C(N)$ contacts in unit time. We define $\beta(N) = C(N)/N$ and assume $\beta'(N) \leqslant 0$, $C'(N) \geqslant 0$.
(2) $B(\tau)$ is the fraction of infecteds remaining infected at infection age τ; we assume an infected period of finite length, $\int_0^\infty B(\tau)\mathrm{d}\tau < \infty$.

(3) A fraction $f(\tau)$ of infected members who leave the infected class when they have age of infection τ recovers with immunity and a fraction $(1-f(\tau))$ dies of disease. In all our examples f will be constant.
(4) $\pi(\tau)$ with $0 \leqslant \pi(\tau) \leqslant 1$ is the infectivity at infection age τ; let $A(\tau) = \pi(\tau)B(\tau)$, so that
$$\int_0^\infty A(\tau)\mathrm{d}\tau \leqslant \int_0^\infty B(\tau)\mathrm{d}\tau < \infty.$$
(5) Initial data is $S(s) = N(s) = K$, $\phi(s) = 0$ for $-\infty < s < 0$.

We let $i_0(t)$ be the density of new infecteds at time t, and $i(t,\tau)$ the density of infecteds at time t with infection age τ. Then
$$i_0(t) = S(t)\beta(N(t))\phi(t), \quad i(t,\tau) = i_0(t-\tau)B(\tau), \ \ 0 \leqslant \tau \leqslant t$$
and
$$\begin{aligned} S'(t) &= -\beta(N(t))S(t)\phi(t), \\ I^*(t) &= \int_0^\infty i(t,\tau)\mathrm{d}\tau = \int_0^\infty i_0(t-\tau)B(\tau)\mathrm{d}\tau = \int_0^\infty \beta(N(t-\tau))S(t-\tau)\phi(t-\tau)B(\tau)\mathrm{d}\tau, \\ \phi(t) &= \int_0^\infty i_0(t-\tau)A(\tau)\mathrm{d}\tau = \int_0^\infty \beta(N(t-\tau))S(t-\tau)\phi(t-\tau)A(\tau)\mathrm{d}\tau. \end{aligned}$$
Differentiation of the equation for I^* shows that the rate of recovery plus the rate of disease death is
$$-\int_0^\infty \beta(N(t-\tau))S(t-\tau)\phi(t-\tau)B'(\tau)\mathrm{d}\tau.$$
Thus the SI^*R model is
$$\begin{aligned} S' &= -\beta(N)S\phi, \\ \phi(t) &= \int_0^\infty \beta(N(t-\tau))S(t-\tau)\phi(t-\tau)A(\tau)\mathrm{d}\tau, \\ N'(t) &= \int_0^\infty (1-f(\tau))\beta(N(t-\tau))S(t-\tau)\phi(t-\tau)B'(\tau)\mathrm{d}\tau. \end{aligned} \tag{6}$$
Since I^* is determined when S, ϕ, N are known we have dropped the equation for I^* from the model, but it will be convenient to recall
$$I^*(t) = \int_0^\infty \beta(N(t-\tau))S(t-\tau)\phi(t-\tau)B(\tau)\mathrm{d}\tau.$$
If $f(\tau) \equiv 1$ then $N(t)$ is a constant K, and the dimension of the model may be reduced, replacing N by the constant K. In this case, since $\beta(N)$ is replaced by the constant $\beta(K)$ the contact rate may be taken to be bilinear.

If a single infective is introduced into a wholly susceptible population, making $K\beta(K)$ contacts in unit time, the fraction still infective at infection age τ is $B(\tau)$ and the infectivity at infection age τ is $A(\tau)$. Thus R_0, the total number of secondary infections caused, is

$$\int_0^\infty K\beta(K)A(\tau)\mathrm{d}\tau = K\beta(K)\int_0^\infty A(\tau)\mathrm{d}\tau.$$

At the beginning of a disease outbreak modelled by (6) we have

$$S \approx K, \quad N \approx K$$

and the linear approximation to the model is

$$u'(t) = -K\beta(K)\phi(t)$$
$$\phi(t) = \int_0^\infty K\beta(K)\phi(t-\tau)A(\tau)\mathrm{d}\tau$$
$$v'(t) = \int_0^\infty (1-f(\tau))K\beta(K)\phi(t-\tau)B(\tau)\mathrm{d}\tau.$$

The corresponding characteristic equation is

$$K\beta(K)\hat{A}(\lambda) = 1,$$

where $\hat{A}(\lambda)$ is the Laplace transform of $A(\tau)$. But if $\Re\lambda \geqslant 0$ we have

$$|K\beta(K)\hat{A}(\lambda)| \leqslant K\beta(K)\hat{A}(0) = R_0.$$

Thus all roots of the characteristic equation have negative real part if $R_0 < 1$, so that the solutions of (6) die out exponentially initially. If $R_0 > 1$, the characteristic equation has a positive real root and solutions of (6) grow exponentially initially. This shows that if $R_0 < 1$ the disease dies out while if $R_0 > 1$ there is an epidemic. We see also from this analysis that at an equilibrium $(S_\infty, 0, N_\infty)$ of (6) with $\phi = 0$ the corresponding characteristic equation is

$$S_\infty\beta(N_\infty)\hat{A}(\lambda) = 1.$$

Thus the equilibrium is asymptotically stable if and only if

$$S_\infty\beta(N_\infty)\hat{A}(0) < 1. \tag{7}$$

The model (1) is the special case of mass action incidence and

$$A(\tau) = B(\tau) = \mathrm{e}^{-\alpha\tau}.$$

Because of the assumption of bilinear incidence in (1) it is not necessary to distinguish between infectives who die of disease and infectives who recover. Of course, in order to track disease deaths in modelling an epidemic it is necessary to make this distinction.

Returning to the general case (6), we note that $S(t)$ and $N(t)$ are non-negative, monotone, non-increasing functions and therefore, have non-negative limits, S_∞ and N_∞, respectively; if there are no disease deaths then $N(t)$ is the constant K. We will show shortly that $S_\infty > 0$. Since $S'(t) \to 0$ as $t \to \infty$, we see from the first equation of (6) that $\phi(t) \to 0$ as $t \to \infty$. Thus the model approaches an equilibrium $(S_\infty, 0, N_\infty)$, and since this equilibrium must be asymptotically stable, the relation (7) must be satisfied.

In fact, we have some information on the rate at which $\phi(t)$ approaches zero. From the first equation of (6) we see that

$$S(t) = S(0)\mathrm{e}^{-\int_0^t \beta(N(s))\phi(s)\mathrm{d}s}$$

so that

$$S_\infty = S(0)\mathrm{e}^{-\int_0^\infty \beta(N(s))\phi(s)\mathrm{d}s}.$$

Since $S_\infty > 0$ we must have

$$\int_0^\infty \beta(N(s))\phi(s)\mathrm{d}s < \infty.$$

Because $\beta(N(s)) \geqslant \beta(K)$ we obtain the estimate

$$\int_0^\infty \phi(s)\mathrm{d}s < \infty.$$

In order to obtain a relation between N_∞ and S_∞ we note that

$$-S'(t-\tau) = S(t-\tau)\beta(N(t-\tau))\phi(t-\tau)$$

and integrate the equation for N in (6) with respect to t from 0 to ∞, obtaining

$$\begin{aligned} K - N_\infty &= \int_0^\infty \left[\int_0^\infty (1-f(\tau))[-S'(t-\tau)]B'(\tau)\mathrm{d}\tau\right]\mathrm{d}t \\ &= \int_0^\infty (1-f(\tau))\int_0^\infty [-S'(t-\tau)]\mathrm{d}t\,B'(\tau)\mathrm{d}\tau \\ &= \int_0^\infty (1-f(\tau)[S(-\tau)-S_\infty]B'(\tau)\mathrm{d}\tau \\ &= (K-S_\infty)\int_0^\infty (1-f(\tau))B'(\tau)\mathrm{d}\tau. \end{aligned} \tag{8}$$

Here, $K - N_\infty$ is the number of disease deaths over the course of the epidemic, $K - S_\infty$ is the total number of disease cases over the course of the epidemic, and $\int_0^\infty (1-f(\tau))B'(\tau)\mathrm{d}\tau$ is the mean disease mortality rate. We note that this result is independent of the form of the equation for S in the model (6).

To show that $S_\infty > 0$ for the model (6) we write

$$-\frac{S'(t)}{S(t)} = \beta(N(t))\int_0^\infty [-S'(t-\tau)]A(\tau)\mathrm{d}\tau.$$

Integration with respect to t from 0 to ∞ gives

$$\begin{aligned} \log\frac{K}{S_\infty} &= \int_0^\infty \beta(N(t))\int_0^\infty [-S'(t-\tau)]A(\tau)\mathrm{d}\tau\,\mathrm{d}t \\ &= \int_0^\infty A(\tau)\int_0^\infty \beta(N(t))[-S'(t-\tau)]\mathrm{d}t\,\mathrm{d}\tau \\ &\leqslant \beta(0)\int_0^\infty A(\tau)\int_0^\infty [-S'(t-\tau)]\mathrm{d}t\,\mathrm{d}\tau \\ &= \beta(0)\int_0^\infty A(\tau)[S(-\tau)-S_\infty]\mathrm{d}\tau \\ &= \beta(0)(K-S_\infty)\int_0^\infty A(\tau)\mathrm{d}\tau \end{aligned}$$

and this shows that $S_\infty > 0$. It is assumed here that $\beta(0)$ is finite. However, an assumption that $\beta(N)$ is unbounded as $N \to 0$ is biologically unreasonable. In particular, standard incidence is not realistic for small population sizes. A more realistic assumption would be that the number of contacts per infective in unit time is linear for small population size and saturates for larger population sizes [10–14].

4. Some examples

The models (1) and (5) are age of infection models with $\pi(\tau) = 1$ and $A(\tau) = B(\tau) = e^{-\alpha\tau}$. An obvious extension adds an exponentially distributed exposed period. We may view a model with an exposed period as an age of infection model. We let $u(\tau)$ be the fraction of infected members with infection age τ who are not yet infective and $v(\tau)$ the fraction of infected members who are infective. Then the rate at which members become infective at infection age τ is $\kappa u(\tau)$, and we have

$$\begin{aligned} u'(\tau) &= -\kappa u(\tau), \quad u(0) = 1 \\ v'(\tau) &= \kappa u(\tau) - \alpha v(\tau), \quad v(0) = 0. \end{aligned} \tag{9}$$

The solution of the first of the equations of (9) is

$$u(\tau) = e^{-\kappa\tau}$$

and substitution of this into the second equation gives

$$v'(\tau) = \kappa e^{-\kappa\tau} - \alpha v(\tau).$$

When we multiply this equation by the integrating factor $e^{\alpha\tau}$ and integrate, we obtain the solution

$$v(\tau) = \frac{\kappa}{\kappa - \alpha}\left[e^{-\alpha\tau} - e^{-\kappa\tau}\right]$$

and this is the term $A(\tau)$ in the general model. The term $B(\tau)$ is $u(\tau) + v(\tau)$. Thus we have

$$\begin{aligned} A(\tau) &= \frac{\kappa}{\kappa - \alpha}\left[e^{-\alpha\tau} - e^{-\kappa\tau}\right] \\ B(\tau) &= \frac{\kappa}{\kappa - \alpha}e^{-\alpha\tau} - \frac{\alpha}{\kappa - \alpha}e^{-\kappa\tau} \\ B'(\tau) &= -\frac{\alpha\kappa}{\kappa - \alpha}\left[e^{-(\mu+\alpha)\tau} - e^{-(\mu+\kappa)\tau}\right]. \end{aligned}$$

With these choices and the identifications

$$I = \phi, \quad E = I^* - \phi$$

we may verify that the system (6) reduces to the SEIR model

$$\begin{aligned} S' &= -\beta(N)SI, \\ E' &= \beta(N)SI - \kappa E, \\ I' &= \kappa E - \alpha I, \\ N' &= -(1 - f)\alpha I. \end{aligned} \tag{10}$$

If there is an asymptomatic period during which individuals have infectivity reduced by a factor ϵ, then the model can be described by the system

$$\begin{aligned} S' &= -\beta(N)S(I+\epsilon E),\\ E' &= \beta(N)S(I+\epsilon E)-\kappa E,\\ I' &= \kappa E-\alpha I,\\ N' &= -(1-f)\alpha I. \end{aligned} \tag{11}$$

This may be considered as an age of infection model with the same identifications of the variables and the same choice of $u(\tau)$, $v(\tau)$ but with $A(\tau)=\epsilon u(\tau)+v(\tau)$.

Since the models (10) and (11) may be viewed as age of infection models they have the same qualitative behaviour as the model (1), namely that there is a basic reproduction number which distinguishes between disappearance of the disease and an epidemic outbreak, and that if there is an epidemic some members of the population are left untouched.

5. Models incorporating reactions to an epidemic

An actual epidemic differs considerably from the idealized models (1), (5) and (11), as was shown by the SARS epidemic of 2002–2003. Some notable differences are:

(1) If a vaccine is available for the disease which has broken out, public health measures will include vaccination of part of the population. Various vaccination strategies are possible, including vaccination of health care workers and other first line responders to the epidemic, vaccination of members of the population who have been in contact with diagnosed infectives, or vaccination of members of the population who live in close proximity to diagnosed infectives.
(2) Diagnosed infectives may be hospitalized, both for treatment and to isolate them from the rest of the population.
(3) Contact tracing of diagnosed infectives may identify people at risk of becoming infective, who may be quarantined (instructed to remain at home and avoid contacts) and monitored so that they may be isolated immediately if and when they become infective.
(4) Isolation may be imperfect; in-hospital transmission of infection was a major problem in the SARS epidemic.

In the SARS epidemic of 2002–2003 in-hospital transmission of disease from patients to health care workers or visitors because of imperfect isolation accounted for many of the cases. This points to an essential heterogeneity in disease transmission which must be included whenever there is any risk of such transmission.

All these generalizations have been considered in studies of the SARS epidemic of 2002–2003. While the ideas were suggested in SARS modelling, they are in fact relevant to any epidemic. One beneficial effect of the SARS epidemic has been to draw attention to epidemic modelling which may be of great value in coping with future disease outbreaks.

If a vaccine is available for a disease which threatens an epidemic outbreak, a vaccinated class which is protected at least partially against infection should be included in a model. While this is

not relevant for an outbreak of a new disease, it would be an important aspect to be considered in modelling a bioterrorist outbreak of smallpox.

For an outbreak of a new disease, where no vaccine is available, isolation and quarantine are the main control measures available. Let us formulate a model for an epidemic once control measures have been started. Thus, we assume that an epidemic has started, but that the number of infectives is small and almost all members of the population are still susceptible.

We introduce a class Q of quarantined members and a class J of isolated members and we formulate a general model to describe the course of an epidemic of a disease for which no vaccine is available when control measures are begun under the assumptions:

(1) Exposed members may be infective with infectivity reduced by a factor ϵ_E, $0 \leqslant \epsilon_E < 1$.
(2) Exposed members who are not quarantined become infective at rate κ_1.
(3) Exposed members are quarantined at rate γ_1 per unit time (in practice, a quarantine will also be applied to many susceptibles, but we ignore this in the model). Quarantine is not perfect, but reduces the contact rate by a factor ϵ_Q. The effect of this assumption is that some susceptibles make fewer contacts than the model assumes.
(4) Infectives are diagnosed at rate γ_2 per unit time and isolated. In addition, quarantined members are monitored and isolated immediately when they develop symptoms at rate κ_2.
(5) There may be transmission of disease by isolated members, with an infectivity factor of ϵ_J.
(6) Infectives who are not isolated leave the infective class at rate α_1 with a fraction f_1 recovering, and isolated members leave the isolated class at rate α_2 with a fraction f_2 recovering.

These assumptions lead to the SEQIJR model

$$\begin{aligned} S' &= -\beta(N)S[\epsilon_E E + \epsilon_E\epsilon_Q Q + I + \epsilon_J J] \\ E' &= \beta(N)S[\epsilon_E E + \epsilon_E\epsilon_Q Q + I + \epsilon_J J] - (\kappa_1 + \gamma_1)E \\ Q' &= \gamma_1 E - \kappa_2 Q \\ I' &= \kappa_1 E - (\alpha_1 + \gamma_2)I \\ J' &= \kappa_2 Q + \gamma_2 I - \alpha_2 J \\ N' &= -(1 - f_1)\alpha_1 I - (1 - f_2)\alpha_2 J. \end{aligned} \tag{12}$$

Here, we have used an equation for N to replace the equation

$$R' = f_1\alpha_1 I + f_2\alpha_2 J.$$

The model before control measures are begun is the special case

$$\gamma_1 = \gamma_2 = \kappa_2 = \alpha_2 = f_2 = 0$$

of (12). The model (12) is equivalent to the SARS model of [5] except for the extension from standard incidence to a general contact rate and the omission of immigration and natural death rate terms. The model of [4] has two susceptible classes with different susceptibilities, which does not affect the age of infection structure of the model, and includes an isolated class but no quarantined class. From an age of infection point of view it is a special case of (12).

We define the *control reproduction number* R_c to be the number of secondary infections caused by a single infective in a population consisting essentially only of susceptibles with the control

measures in place. It is analogous to the basic reproduction number but instead of describing the very beginning of the disease outbreak it describes the beginning of the recognition of the epidemic. The basic reproduction number is the value of the control reproduction number with

$$\gamma_1 = \gamma_2 = \kappa_2 = \alpha_2 = f_2 = 0.$$

We may calculate R_c in the same way as we calculate R_0 but using the full model with quarantined and isolated classes. We obtain

$$R_0 = \frac{\epsilon_E K\beta(K)}{\kappa_1} + \frac{K\beta(K)}{\alpha_1}$$
$$R_c = \frac{\epsilon_Q\epsilon_E K\beta(K)}{D_1} + \frac{K\beta(K)\kappa_1}{D_1D_2} + \frac{\epsilon_Q K\beta(K)\gamma_1}{D_1\kappa_2} + \frac{\epsilon_J K\beta(K)\kappa_1\gamma_2}{\alpha_2 D_1 D_2} + \frac{\epsilon_J K\beta(K)\gamma_1}{\alpha_2 D_1},$$

where $D_1 = \gamma_1 + \kappa_1$, $D_2 = \gamma_2 + \alpha_1$.

Each term of R_c has an epidemiological interpretation. The mean duration in E is $1/D_1$ with contact rate $\epsilon_E\beta$, giving a contribution to R_c of $\epsilon_E K\beta(K)/D_1$. A fraction κ_1/D_1 goes from E to I, with contact rate β and mean duration $1/D_2$, giving a contribution of $K\beta(K)\kappa_1/D_1D_2$. A fraction γ_1/D_1 goes from E to Q, with contact rate $\epsilon_Q\epsilon_Q\beta$ and mean duration $1/\kappa_2$, giving a contribution of $\epsilon_Q K\beta(K)\gamma_1/D_1\kappa_2$. A fraction $\kappa_1\gamma_2/D_1D_2$ goes from E to I to J, with a contact rate of $\epsilon_J\beta$ and a mean duration of $1/\alpha_2$, giving a contribution of $\epsilon_J K\beta(K)\kappa_1\gamma_2/\alpha_2 D_1 D_2$. Finally, a fraction γ_1/D_1 goes from E to Q to J with a contact rate of $\epsilon_J\beta$ and a mean duration of $1/\alpha_2$ giving a contribution of $\epsilon_J K\beta(K)\gamma_1/D_1\alpha_2$. The sum of these individual contributions gives R_c.

We may view the epidemic management model (12) as an age of infection model at least if $f_1 = f_2$. We define $I^* = E + Q + I + J$, and we need only calculate the kernels $A(\tau)$, $B(\tau)$. We let $u(\tau)$ denote the number of members of infection age τ in E, $v(\tau)$ the number of members of infection age τ in Q, $w(\tau)$ the number of members of infection age τ in I, and $z(\tau)$ the number of members of infection age τ in J. Then (u, v, w, z) satisfies the linear homogeneous system with constant coefficients

$$u'(\tau) = -(\kappa_1 + \gamma_1)u(\tau)$$
$$v'(\tau) = \gamma_1 u(\tau) - \kappa_2 v(\tau)$$
$$w'(\tau) = \kappa_1 u(\tau) - \alpha_1 w(\tau) - \gamma_2 w(\tau)$$
$$z'(\tau) = \gamma_2 w(\tau) + \kappa_2 v(\tau) - \alpha_2 z(\tau)$$

with initial conditions $u(0) = 1$, $v(0) = 0$, $w(0) = 0$, $z(0) = 0$. This system is easily solved recursively, and then the system (12) is an age of infection epidemic model with

$$A(\tau) = \epsilon_E u(\tau) + \epsilon_E\epsilon_Q v(\tau) + w(\tau) + \epsilon_J z(\tau)$$
$$B(\tau) = u(\tau) + v(\tau) + w(\tau) + z(\tau).$$

It follows that the the properties (i) and (ii) and the relation (8) hold for the management model (12).

The model (12) is not included in (6) if $f_1 \neq f_2$ because of the assumption in (12) that the disease death fraction may not be the same in the infective and isolated classes, but it is possible to obtain an analogue of the relation (8) directly, as is done in [15].

The asymptotic behaviour of the treatment model (12) is the same as that of the simpler model (1). If the control reproduction number $R_c < 1$ the disease dies out and if $R_c > 1$ there is an epidemic which will pass leaving some members of the population untouched.

6. Discussion

We have established that general epidemic models behave in the same way in the sense that there is a basic reproduction number which determines whether there will be an epidemic and that an epidemic will pass through a population leaving some members untouched. We conjecture that this remains true for more complicated models with more compartments and more stages, including models with heterogeneity of mixing. This fact would be established by showing how to interpret such a model as an age of infection models. Thus the age of infection approach started by Kermack and McKendrick was 75 years ahead of its time and is a unifying way to view epidemic models. It is important to know that all epidemic models have the same basic properties even though these properties may appear to be 'obvious'. Of course, our models assume that the course of the epidemic is rapid enough that demographic effects may be ignored. If this is not true, then it would be possible for a disease to become endemic.

The management model (12) is viewed as a two stage model, with the imposition of control measures marking the beginning of the second stage. The initial stage of an epidemic is actually more complicated than is suggested by our model. For example, stochastic effects are certainly important early in an epidemic and a continuous model is not appropriate when the number of infectives is very small. Such considerations are discussed in [3]. The beginning of an epidemic is often marked by 'superspreading events' in which a single source may produce a large number of infections. Examples indicate that the probability of an epidemic depends strongly on the contact network at the beginning of a disease outbreak. The study of complex networks is a field which is developing very rapidly. Some basic references are [16,17]. Also, [18] investigates the beginning of a disease outbreak from a network perspective.

We have been considering an epidemic in a single location, ignoring travel between locations of individuals who may be infective. Modern transportation has permitted the rapid transfer of infectious diseases over great distances, and an aspect of epidemic control that has become important is the screening of travellers who may be infective. Epidemic models which include some movement into and out of populations are a natural extension of the models considered here.

Our model does not take into account behavioural changes in a population when an epidemic breaks out. If the disease is debilitating, infectives make fewer contacts because of their weakness, and the total number of contacts would depend on the number of infectives as well as the total population size. Another aspect of epidemics which we have ignored is that when an epidemic occurs some members of the population will undoubtedly avoid crowds and change their behaviour in other ways to reduce the number of contacts they make and will also adopt hygienic measures to reduce the probability that a contact will transmit infection. This may occur because of personal decisions or because of government instructions. Inclusion of these effects in an epidemic model is an important problem to be explored.

References

[1] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. London 115 (1927) 700.
[2] F. Brauer, C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, Springer, New York, 2001.
[3] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases, Wiley, Chichester, 2000.
[4] G. Chowell, P.W. Fenimore, M.A. Castillo-Garsow, C. Castillo-Chavez, SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, J. Theor. Biol. 224 (2003) 1.
[5] A. Gumel, S. Ruan, T. Day, J. Watmough, F. Brauer, P. van den Driessche, D. Gabrielson, C. Bowman, M.E. Alexander, S. Ardal, J. Wu, B.M. Sahai, Modeling strategies for controlling SARS outbreaks in Toronto, Hong Kong, Singapore and Beijing, Proc. R. Soc. London, Ser. B 271 (2004) 2223.
[6] P. Waltman, Deterministic threshold models in the theory of epidemics, Lecture Notes in Biomathematics, 1, Springer, Berlin, Heidelberg, New York, 1974.
[7] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (2000) 599.
[8] H.W. Hethcote, J.W. Van Ark, Modeling HIV transmission and AIDS in the United States, Lecture Notes in Biomathematics, 95, Springer, Berlin, Heidelberg, New York, 1992.
[9] H.W. Hethcote, J.A. Yorke, Gonorrhea transmission and control, Lecture Notes in Biomathematics, 56, Springer, Berlin, Heidelberg, New York, 1984.
[10] C. Castillo-Chavez, K. Cooke, W. Huang, S.A. Levin, The role of long incubation periods in the dynamics of HIV/AIDS. Part 1: Single populations models, J. Math. Biol. 27 (1989) 373.
[11] K. Dietz, Overall patterns in the transmission cycle of infectious disease agents, in: Population Biology of Infectious Diseases, in: R.M. Anderson, R.M. May (Eds.), Life Sciences Research Report, 25, Springer, Berlin, Heidelberg, New York, 1982, p. 87.
[12] J.A.P. Heesterbeek, J.A.J. Metz, The saturating contact rate in marriage and epidemic models, J. Math. Biol. 31 (1993) 529.
[13] C.M. Kribs-Zaleta, To switch or taper off: The dynamics of saturation, Math. Biosci. 192 (2004) 137.
[14] J. Mena-Lorca, H.W. Hethcote, Dynamic models of infectious diseases as regulators of population size, J. Math. Biol. 30 (1992) 693.
[15] F. Brauer, Some simple epidemic models, Math. Biosci. Eng., in press.
[16] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167.
[17] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268.
[18] L.A. Meyers, B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, R.C. Brunham, Network theory and SARS: predicting outbreak diversity, J. Theor. Biol. 232 (2005) 71.