Bulletin of Mathematical Biology (2007) 69: 1511–1536
DOI 10.1007/s11538-006-9174-9

ORIGINAL ARTICLE

Epidemiological Models with Non-Exponentially Distributed Disease Stages and Applications to Disease Control

Zhilan Feng[a,*], Dashun Xu[b], Haiyun Zhao[a]

[a] *Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA*

Received: 19 July 2006 / Accepted: 20 September 2006 / Published online: 20 January 2007

Abstract *SEIR* epidemiological models with the inclusion of quarantine and isolation are used to study the control and intervention of infectious diseases. A simple ordinary differential equation (ODE) model that assumes exponential distribution for the latent and infectious stages is shown to be inadequate for assessing disease control strategies. By assuming arbitrarily distributed disease stages, a general integral equation model is developed, of which the simple ODE model is a special case. Analysis of the general model shows that the qualitative disease dynamics are determined by the reproductive number $\mathcal{R}_c$, which is a function of control measures. The integral equation model is shown to reduce to an ODE model when the disease stages are assumed to have a gamma distribution, which is more realistic than the exponential distribution. Outcomes of these models are compared regarding the effectiveness of various intervention policies. Numerical simulations suggest that models that assume exponential and non-exponential stage distribution can produce inconsistent predictions.

Keywords Epidemiological model · Distributed disease stage · Integral equation · Disease control strategies

1. Introduction

The mathematical theory of infectious diseases pioneered by Ross, MacDonald, Kermack, McKendrick and others has played a major role in the study of the control and prevention of infectious diseases (see, for example, Ross, 1911; Kermack and McKendrick, 1927). More recently, mathematical models have been used to investigate how to more effectively control SARS via various disease control measures including vaccination, quarantine, and isolation (see, for

*Corresponding author.
E-mail address: zfeng@math.purdue.edu (Z. Feng).

example, Chowell et al., 2003; Lipsitch et al., 2003; Riley et al., 2003; McLean et al., 2005).

Many of these studies have taken the approach of using simple ODE models to draw conclusions regarding the effectiveness of various disease control programs. The simplicity of these models is often achieved by making the assumption that the disease stages are exponentially distributed. When the models do not include quarantine and/or isolation, or when the isolation is assumed to be perfect (i.e., isolated individuals do not transmit the disease), the exponential distribution assumption (EDA) and the models that use this assumption have been shown to provide valuable information and important insights into the disease dynamics. However, as demonstrated in this article, the EDA may not be appropriate in models for diseases with relatively long latent and/or infectious periods when isolation is not completely effective.

Here, we discuss the problem by considering a simple ODE model that is a commonly used *SEIR*-type model. In a standard *SEIR* model, the whole population N is divided into four sub-classes: susceptible (S), exposed but not yet infectious (E), infectious (I), and immune or recovered (R) individuals. Susceptibles become exposed (latent) at the rate $\lambda_1(t)S(t) = \beta S(t)I(t)/N$ where β is the disease transmission coefficient in the absence of interventions and N is the total population size. Latent individuals progress to the infectious stage at a constant rate α_1 and infectious individuals recover at a constant rate δ_1. A transmission diagram for this case is shown in Fig. 1a. All variables and parameters are listed in Table 1.

To incorporate control measures such as quarantine and isolation, additional sub-classes can be included. For example, quarantine and isolation can be modeled as follows. Let $\lambda(t)$ denote the force of infection (a specific form is given below). Assume that a fraction b of contacts (susceptible individuals who have had contacts with an infectious person) are actually infected, and that the other fraction $(1-b)$ of contacts remain susceptible who will be quarantined (S_Q) and will return to the S class at a rate r (see, e.g., Lipsitch et al., 2003). Among the infected individuals ($b\lambda(t)S$) a fraction γ will be quarantined (Q) at the time of infection (i.e., there is a rate, $\gamma b\lambda(t)S$, from S to Q directly). The fraction $(1-\gamma)$ of the exposed (and infected) individuals (E) who are not quarantined at the time of infection will be quarantined at a constant rate χ throughout the latent period. (We remark that although for most diseases quarantine is not considered as the exposed individuals show no disease symptoms, the situation for SARS is different in which quarantine was implemented in several places including Hong Kong, Taiwan, and China.) The non-quarantined and quarantined (exposed) individuals will progress to the

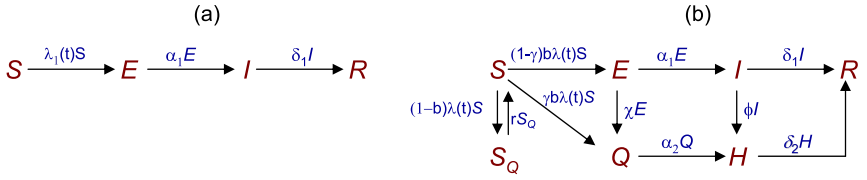

Fig. 1 Disease transmission diagrams (birth and death omitted). (a) No disease control. (b) Quarantine and isolation are included.

Table 1 Definitions of frequently used symbols

Symbol	Definition
$S(t)$	Number of susceptible individuals at time t
$S_Q(t)$	Number of susceptible individuals quarantined at time t
$E(t)$	Number of exposed (not yet infectious) individuals at time t
$Q(t)$	Number of quarantined (exposed) individuals at time t
$I(t)$	Number of susceptible individuals at time t
$H(t)$	Number of isolated (infectious) individuals at time t
$R(t)$	Number of recovered individuals at time t
N	Total population size (constant)
$C(t)$	Number of cumulative new infections at time t
$\lambda(t)$	Force of infection at time t
β	Transmission coefficient
$\alpha_1,\ \alpha_2$	Rate at which non-quarantined, quarantined individuals become infectious
α	Same as α_1
$\delta_1,\ \delta_2$	Rate at which non-isolated, isolated individuals become recovered
δ	Same as δ_1
μ	Natural death rate
χ, ϕ	Rate of quarantine, isolation
ρ	Isolation efficiency ($0 \le \rho \le 1$)
b	Fraction of contacts infected ($b = 1$ in this paper)
γ	Fraction of infecteds quarantined at time of exposure ($\gamma = 0$ in this paper)
$p_i(s),\ P_i(s)$	Probability that disease stage i lasts longer than s time units $(i = E, I)$
$k(s),\ l(s)$	Probability of not being quarantined, isolated at stage age s
$T_E,\ T_I$	Mean of $p_E(s) = \mathrm{e}^{-\alpha s},\ p_I(s) = \mathrm{e}^{-\delta s}$ $(T_E = 1/\alpha,\ T_I = 1/\delta)$
$\mathcal{M}_i(s),\ M$	Expected remaining sojourn at age s: $\int_0^\infty P_i(t\|s)\,\mathrm{d}t$ $(i = E, I)$, $M = \mathcal{M}(0)$
$\mathcal{T}_E$	Probability of surviving and becoming infectious: $\int_0^\infty [-\dot{P}_E(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{T}_{E_k}$	"Quarantine-adjusted" probability (similar to $\mathcal{T}_E$): $\int_0^\infty [-\dot{P}_E(s)k(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{T}_I$	Probability an infectious person survives and recovers: $\int_0^\infty [-\dot{P}_I(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{T}_{I_l}$	"Isolation-adjusted" probability (similar to $\mathcal{T}_I$): $\int_0^\infty [-\dot{P}_I(s)l(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{D}_E$	Mean time in exposed stage (adjusted by death): $\int_0^\infty P_E(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{D}_{E_k}$	"Quarantine-adjusted" mean time in exposed stage: $\int_0^\infty P_E(s)k(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{D}_I$	Mean time in infectious stage (adjusted by death): $\int_0^\infty P_I(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{D}_{I_l}$	"Isolation-adjusted" mean time in infectious stage: $\int_0^\infty P_I(s)l(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}t$
$\mathcal{R}_0$	The basic reproductive number
$\mathcal{R}_c$	The reproductive number under control measures
EDA, GDA	Exponential distribution assumption, Gamma distribution assumption
EDM, GDM	Exponential distribution model, Gamma distribution model

infectious stage at constant rates α_1 and α_2 respectively (the relationship between α_1 and α_2 will be discussed later). Infectious individuals will be isolated (H) at a rate ϕ and individuals in the H class will recover at a rate δ_2 (the relationship between δ_1 and δ_2 will be discussed later). Since we are considering the case of imperfect isolation, the new infections are now produced at the rate $\lambda(t)S(t)$ with

$$\lambda(t) = \beta\left[\frac{I(t) + (1-\rho)H(t)}{N}\right], \tag{1}$$

where $\rho \in [0, 1]$ is the fraction of reduction in the transmission rate of isolated individuals with $\rho = 1$, $\rho = 0$, and $0 < \rho < 1$ representing a completely effective, completely ineffective, and partially effective isolation, respectively. The corresponding transmission diagram is shown in Fig. 1b.

In this paper, we consider only the case that $\gamma = 0$ by assuming that the fraction of infected contacts that can be traced and quarantined at the time of infection is very small, and hence most infected people are quarantined during the latency period. In addition, we assume a large population size N (in comparison with the size of the infected population), in which case the quarantine of susceptibles is unlikely to have a significant impact on the disease transmission dynamics and hence will be ignored. This is equivalent to assuming that $b = 1$. (It may not be appropriate to ignore the S_Q class if one is concerned with the cost associated with quarantine, which is not the case in this paper.) For simplicity, the disease-induced death is ignored and the per-capita birth rate and the natural death rate are assumed to be equal. Hence, the total population size N remains constant. Then the corresponding ODE model is given by the following system

$$\begin{aligned}
S' &= \mu N - \beta S\tfrac{I+(1-\rho)H}{N} - \mu S,\\
E' &= \beta S\tfrac{I+(1-\rho)H}{N} - (\chi + \alpha_1 + \mu)E,\\
Q' &= \chi E - (\alpha_2 + \mu)Q,\\
I' &= \alpha_1 E - (\phi + \delta_1 + \mu)I,\\
H' &= \alpha_2 Q + \phi I - (\delta_2 + \mu)H,\\
R' &= \delta_1 I + \delta_2 H - \mu R.
\end{aligned} \tag{2}$$

" $'$ " denotes the derivative with respect to time t. All involved parameters are nonnegative constants, and all variables and parameters are listed in Table 1.

In the next section, we discuss some of the drawbacks of the simple model (2) when used to evaluate intervention policies. We argue that the main reason for these problems is due to the simplifying assumption of exponential distributions for the disease stages, which is used in the model. This provides a motivation for using more realistic stage distributions. Non-exponential distributions have been considered in epidemiological models (see, for example, Hethcote and Tudor, 1980; Hethcote et al., 1981; Plant and Wilson, 1986; Taylor and Karlin, 1998; Feng and Thieme, 2000a,b; Feng et al., 2001; Lloyd, 2001a,b). However, none of these studies focuses on the evaluation of intervention policies.

In this paper, we develop a general model with arbitrarily distributed disease stages. The general setting allows us to identify new models that are improvements to the simple model (2) while keeping the improved models as simple as possible. We show that in the case of exponential distributions, the general model reduces to the simple model (2) with appropriate constraints on model parameters. We also consider a particular non-exponential stage distribution, the gamma distribution, in which case the general model reduces to another ODE model. Analysis

for both the general model and the model with the gamma distribution assumption (GDA) are provided. We demonstrate that the model under GDA is indeed an improvement on the model under EDA.

2. Drawbacks of EDA and model (2)

One of the main roles of model (2) and its variants is to evaluate various disease control measures. Relevant parameters that represent disease intervention are χ, ϕ, ρ, α_2 and δ_2. It is very important that these parameters have well-defined meanings in order to connect them with epidemiological data and to determine their appropriate values, and to ensure that the model predictions are reasonable regarding the effect of various control strategies. Otherwise, the results obtained from the model might be misleading, as demonstrated below.

One of the quantities that can be used to assess the impact of various control measures is the cumulative number, $C(t)$, of infections determined by the equation

$$C'(t) = \lambda(t)S(t).$$

Let $C(0) = 0$ so that $C(t)$ is the cumulative number of new infections at the end of an epidemic (in the case that the disease is driven into extinction). One would expect that the C value will be reduced if we increase the value of any of the control parameters. However, Fig. 2 (see (a) and (b)) shows that C increases with increasing rates of quarantine (χ) and isolation (ϕ). The parameter values used in Fig. 2 are the following: $\alpha_1 = 0.2$ and $\delta_1 = 0.15$, which correspond to a latency period of $1/\alpha_1 = 5$ days and an infectious period of $1/\delta_1 \approx 7$ days, respectively. These values are in the realistic range of many infectious diseases. The transmission coefficient is chosen to be $\beta = 0.13$ which corresponds to a reproductive number (calculated using the formula (43)) that is equal to approximately 0.9 (so that the disease will die out). The isolation efficiency is $\rho = 0.3$ and other parameters have different values in Figs. 2a–d depending on the assumptions. In Figs. 2a and 2b, no additional constraints are imposed on α_2 and δ_2 which have values 0.17 and 0.1, respectively.

The lack of constraints on α_2 and δ_2 may be responsible for the problem shown in Figs. 2a and 2b. Our simulation results show that the problem can be avoided if the following constraints are imposed

$$\alpha_2 = \alpha_1, \quad \delta_2 = \delta_1. \tag{3}$$

A more rigorous argument for constraint (3) will be provided in Section 3.2. Here, we give only a heuristic argument. The ordinary differential equation model (2) implicitly assumes the exponential distribution for the latent and infectious stages. More precisely, the exponential functions

$$p_E(s) = \mathrm{e}^{-\alpha_1 s} \quad \text{and} \quad p_I(s) = \mathrm{e}^{-\delta_1 s}$$

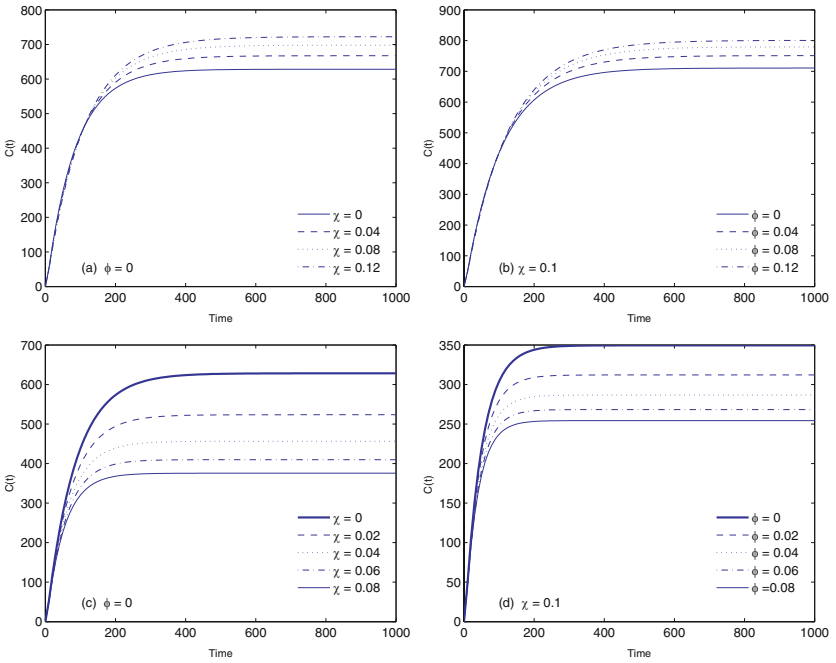

Fig. 2 Numerical simulations of the model (2). The number of cumulative new infections $C(t)$ is plotted for various values of the control parameters χ and ϕ. (a) and (b) are for the case of no constraints on the parameter values. (c) and (d) are for the case when constraint (3) is used.

have been used to describe the probability of remaining in the latent stage and the infectious stage, respectively, and the mean durations of latent and infectious stages are

$$T_E = \int_0^\infty p_E(s)\,\mathrm{d}s = \frac{1}{\alpha_1} \quad \text{and} \quad T_I = \int_0^\infty p_I(s)\,\mathrm{d}s = \frac{1}{\delta_1}. \tag{4}$$

Similarly, the mean sojourn times in the Q and H classes are respectively

$$T_Q = \frac{1}{\alpha_2} \quad \text{and} \quad T_H = \frac{1}{\delta_2}. \tag{5}$$

A fundamental property of the exponential distribution is the memory-less property, which requires that the remaining expected sojourn in the H (or Q) class is independent of the time already elapsed before entering it. This property implies that

$$T_H = T_I \quad \text{and} \quad T_Q = T_E,$$

which is equivalent to the condition given in (3) (see (4) and (5)).

Another argument for the use of (3) is the following. The average time individuals (both isolated and non-isolated individuals) stay in the I class is equal to

$1/(\delta_1+\phi)$, and the average time an isolated individual stays in the H class is $1/\delta_2$. Notice that $\phi/(\delta_1+\phi)$ and $\delta_1/(\delta_1+\phi)$ are fractions of isolated and non-isolated individuals, respectively. Then the weighted average time an individual stays in the infectious stage is

$$\frac{\phi}{\delta_1+\phi}\left(\frac{1}{\delta_1+\phi}+\frac{1}{\delta_2}\right)+\frac{\delta_1}{\delta_1+\phi}\left(\frac{1}{\delta_1+\phi}\right), \tag{6}$$

which is equal to $\frac{1}{\delta_1}$ (i.e., the infectious period) only if we set $\delta_2=\delta_1$ in accordance with constraint (3). It follows from a similar argument that the weighted average time an individual stays in the exposed stage (when quarantine is present) is equal to the latency period only if $\alpha_2=\alpha_1$. Figs. 2c and 2d illustrate that the value of $C(t)$ reduces as the values of control parameters increase, showing the improvement compared to Figs. 2a and 2b. For Figs. 2c and 2d, all parameter values are the same as in Figs. 2a and 2b except that the constraint (3) holds.

Constraint (3) seems to provide a partial solution to the problem exhibited in Figs. 2a and 2b. However, it creates a different problem. Since the isolated or quarantined individuals have already spent some time in the infectious or latent stage before entering the H or the Q class, the condition (3) amounts to allowing for a prolonged period of infectiousness for isolated individuals and a prolonged period of latency for quarantined individuals. In reality, if an infectious person already spent some time in the I class before being isolated, then the expected remaining sojourn in the H class should be shorter than the infectious period. Therefore, the model assumption (EDA) conflicts with biological constraints. A similar argument applies to quarantined individuals. In Section 5.2, we demonstrate how the predictions of the model (2) constrained by (3) may be in disagreement with that of models using more realistic stage distributions.

It should be pointed out that the purpose of this article is not to argue which assumptions/constraints are more appropriate than others or whether they are correct or not. Our goal is to point out the weakness of models that assume exponential distributions for one or both disease stages and to demonstrate possible problems with either constraints of type (3) or no constraint. Therefore, models with more realistic stage distributions may need to be considered.

The above examples demonstrate some of the drawbacks of the EDA and the simple model (2). Epidemiological models with non-exponential distributions such as the gamma distribution have been previously studied (see, for example, Hethcote and Tudor, 1980; Plant and Wilson, 1986; Taylor and Karlin, 1998; Lloyd, 2001a,b). In these studies, the authors discussed other objections to the EDA. For example, it is pointed out that constant recovery is a poor description of real-world infections, and they show that in models with more realistic distributions of disease stages, less stable behavior may be expected and disease persistence may be diminished (see Hethcote et al., 1981; Lloyd, 2001a,b). However, these studies do not focus on the impact of imperfect isolation. In the rest of the paper, we consider models with more realistic disease stage distributions and study the properties of these models.

3. The general model

In this section, we derive an *SEIR* type of model which includes quarantine and isolation and assumes that the disease stages are arbitrarily distributed. We show that this model reduces to the simple model (2) when the disease stages are exponentially distributed and the conditions in (3) are satisfied.

3.1. The model with arbitrarily distributed disease stages

Let $S(t)$, $E(t)$, $Q(t)$, $I(t)$, $H(t)$, $R(t)$, N be as defined in model (2), and let P_E, P_I : $[0,\infty) \to [0,1]$ describe the durations of the exposed (or latent) and infectious stages, respectively. More precisely, $P_i(s)$ $(i = E, I)$ gives the probability that the disease stage i lasts longer than s time units (or the probability of being still in the same stage at stage age s). Then, the derivative $-\dot{P}_i(s)$ $(i = E, I)$ gives the rate of removal from the stage i at stage age s by the natural progression of the disease. These duration functions have the following properties

$$P_i(0) = 1, \quad \dot{P}_i(s) \le 0, \quad \int_0^\infty P_i(s)\,\mathrm{d}s < \infty, \quad i = E, I.$$

Let $k(s), l(s) : [0,\infty) \to [0,1]$ denote, respectively, the probabilities that exposed or infectious individuals have not been quarantined or isolated at stage age s. Hence, $1 - k(s) =: \bar{k}(s)$, $1 - l(s) =: \bar{l}(s)$ give the respective probabilities of being quarantined or isolated before reaching stage age s. Assume that $k(0) = l(0) = 1$, $\dot{k}(s) \le 0$ and $\dot{l}(s) \le 0$. Since we are not focusing on vital dynamics, we use the simplest function $\mathrm{e}^{-\mu t}$ for the probability of surviving natural death. Let the numbers of initial susceptible and removed individuals be $S_0 > 0$ and $R_0 > 0$, respectively. Let $E_0(t)\,\mathrm{e}^{-\mu t}$, $I_0(t)\,\mathrm{e}^{-\mu t}$, $Q_0(t)\,\mathrm{e}^{-\mu t}$, and $H_0(t)\,\mathrm{e}^{-\mu t}$ be the non-increasing functions that represent the numbers of individuals that were initially exposed, infectious, quarantined and isolated, respectively, and are still alive and in the respective classes at time t. For example, in the special case when P_E and P_I are exponentially distributed with mean stage durations α and δ, denoted by $\tilde{P}_E$ and $\tilde{P}_I$, respectively:

$$\tilde{P}_E(s) = \mathrm{e}^{-\alpha s}, \quad \tilde{P}_I(s) = \mathrm{e}^{-\delta s}, \tag{7}$$

and when the survivals from quarantine and isolation are described by the exponential functions

$$k(s) = \mathrm{e}^{-\chi s}, \quad l(s) = \mathrm{e}^{-\phi s} \tag{8}$$

with χ and ϕ being constants, we have

$$E_0(t) = E(0)\,\mathrm{e}^{-(\chi+\alpha)t}, \quad I_0(t) = I(0)\,\mathrm{e}^{-(\phi+\delta)t}, \quad \text{etc.} \tag{9}$$

where $E(0)$ and $I(0)$ are constants representing the number of individuals in the E and I classes, respectively, at time $t = 0$. Let $\tilde{I}_0(t)$, $\tilde{Q}_0(t)$, $\tilde{H}_0(t)$, and $\tilde{R}_0(t)$ denote

those initially infected who have moved into the I, Q, H, and R classes, respectively, and are still alive at time t. For example, if we use (7) and (8), then

$$\tilde{I}_0(t) = \int_0^t \alpha E(0)\, \mathrm{e}^{-(\alpha+\chi+\mu)\tau}\, \mathrm{e}^{-(\delta+\phi+\mu)(t-\tau)}\, \mathrm{d}\tau. \tag{10}$$

The force of infection $\lambda(t)$ is assumed to have the same form as in system (2) and as it is given in (1). Then the number of individuals who became exposed at some time $s \in (0, t)$ and are still alive and in the E class at time t is

$$E(t) = \int_0^t \lambda(s)S(s)P_E(t-s)k(t-s)\, \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s + E_0(t)\, \mathrm{e}^{-\mu t}.$$

Differentiating the above equation,

$$\begin{aligned} E'(t) = {} & \int_0^t \lambda(s)S(s)\dot{P}_E(t-s)k(t-s)\, \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s \\ & + \int_0^t \lambda(s)S(s)P_E(t-s)\dot{k}(t-s)\, \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s \\ & + \lambda(t)S(t) - \mu E(t) + E_0'(t)\, \mathrm{e}^{-\mu t}. \end{aligned} \tag{11}$$

The first and second terms provide inputs for the I and the Q equations, respectively. Hence,

$$\begin{aligned} I(t) = {} & \int_0^t \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau-s)k(\tau-s)]P_I(t-\tau)l(t-\tau) \\ & \times \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s\, \mathrm{d}\tau + I_0(t)\, \mathrm{e}^{-\mu t} + \tilde{I}_0(t), \end{aligned}$$

and

$$\begin{aligned} Q(t) = {} & \int_0^t \int_0^\tau \lambda(s)S(s)[-P_E(\tau-s)\dot{k}(\tau-s)]P_E\big(t-\tau\,\big|\,\tau-s\big) \\ & \times \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s\, \mathrm{d}\tau + Q_0(t)\, \mathrm{e}^{-\mu t} + \tilde{Q}_0(t) \\ = {} & \int_0^t \lambda(s)S(s)P_E(t-s)\bar{k}(t-s)\, \mathrm{e}^{-\mu(t-s)}\, \mathrm{d}s + Q_0(t)\, \mathrm{e}^{-\mu t} + \tilde{Q}_0(t). \end{aligned}$$

Here, $P_E(w|a) = \frac{P_E(w+a)}{P_E(a)}$ is the conditional remaining function which gives the probability that an exposed individual remains non-infectious for w time units longer given that the person was already exposed for a units of time. Individuals in the H class are from two sources, one from the I class via isolation at the rate $-P_I\dot{I}(t)$, and the other one from the Q class via disease progression at the rate

$-\dot{P}_E\bar{k}(t)$. This leads to

$$
\begin{aligned}
H(t) = {} & \int_0^t \int_0^u \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)k(\tau - s)][-P_I(u-\tau)\dot{l}(u-\tau)] \\
& \times P_I\big(t-u\big|u-\tau\big)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau\,\mathrm{d}u \\
& + \int_0^t \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)\bar{k}(\tau - s)]P_I(t-\tau)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau \\
& + H_0(t)\,\mathrm{e}^{-\mu t} + \tilde{H}_0(t),
\end{aligned}
$$

where $P_I(w|a) = \frac{P_I(w+a)}{P_I(a)}$ is the conditional remaining function as $P_E(w|a)$. All individuals described by the three integrals in I and H will eventually enter the recovered class R. Individuals in the R class include those who became exposed at time $s \in (0,t)$, exited the latent stage by becoming infectious at time $\tau \in (s,t)$, exited the infectious stage by recovery at time $u \in (\tau,t)$ at the rate $-\dot{P}_I(u-\tau)$, and are still alive at time t. Adding the corresponding terms we can get the equation for R

$$
\begin{aligned}
R(t) = {} & \int_0^t \int_0^u \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)][-\dot{P}_I(u-\tau)]\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau\,\mathrm{d}u \\
& + R_0(t)\,\mathrm{e}^{-\mu t} + \tilde{R}_0(t) \\
= {} & \int_0^t \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)]\,\mathrm{e}^{-\mu(t-s)} \int_\tau^t [-\dot{P}_I(u-\tau)]\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}\tau \\
& + R_0(t)\,\mathrm{e}^{-\mu t} + \tilde{R}_0(t) \\
= {} & \int_0^t \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)][1 - P_I(t-\tau)]\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau \\
& + R_0(t)\,\mathrm{e}^{-\mu t} + \tilde{R}_0(t).
\end{aligned}
$$

Combining these equations we have the following integral equation model:

$$
\begin{aligned}
S(t) &= \int_0^t \mu N\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s - \int_0^t \lambda(s)S(s)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s + S_0\,\mathrm{e}^{-\mu t}, \\
E(t) &= \int_0^t \lambda(s)S(s)P_E(t-s)k(t-s)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s + E_0(t)\,\mathrm{e}^{-\mu t}, \\
Q(t) &= \int_0^t \int_0^\tau \lambda(s)S(s)[-P_E(\tau - s)\dot{k}(\tau - s)]P_E\big(t-\tau\big|\tau-s\big)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau + \tilde{Q}(t), \\
I(t) &= \int_0^t \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)k(\tau - s)]P_I(t-\tau)l(t-\tau)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau + \tilde{I}(t), \\
H(t) &= \int_0^t \int_0^u \int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau - s)k(\tau - s)][-P_I(u-\tau)\dot{l}(u-\tau)] \\
&\quad \times P_I\big(t-u\big|u-\tau\big)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau\,\mathrm{d}u
\end{aligned}
\tag{12}
$$

$$+\int_0^t\int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau-s)\bar{k}(\tau-s)]P_I(t-\tau)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau+\tilde{H}(t),$$

$$R(t)=\int_0^t\int_0^\tau \lambda(s)S(s)[-\dot{P}_E(\tau-s)][1-P_I(t-\tau)]\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s\,\mathrm{d}\tau+\tilde{R}(t),$$

$$\text{with } \lambda(t)=\beta\frac{I(t)+(1-\rho)H(t)}{N},$$

where $\tilde{X}(t)=X_0(t)\,\mathrm{e}^{-\mu t}+\tilde{X}_0(t)$ $(X=Q,I,H,R)$. Obviously $\tilde{X}(t)\to 0$ as $t\to\infty$. It can be shown that under standard assumptions on initial data and parameter functions the system (12) has a unique nonnegative solution defined for all positive time.

The integral formulation of the model makes it possible to see the role of an important quantity, the *expected remaining sojourn* at stage age s (see Thieme, 2003), which is defined by

$$\mathcal{M}_i(s)=\int_0^\infty P_i(t|s)\,\mathrm{d}t=\int_0^\infty \frac{P_i(t+s)}{P_i(s)}\,\mathrm{d}t,\quad i=E,I \tag{13}$$

with $\mathcal{M}_i(0)=\int_0^\infty P_i(s)\,\mathrm{d}s=M_i$ being the mean sojourn time in stage i (M_i is the latent period if $i=E$ and is the infectious period if $i=I$). This quantity reflects the key difference between various distributions. For example, for the exponential distribution $\mathcal{M}_i(s)=M_i$ for all $s>0$, and for many other distributions $\mathcal{M}_i(s)<M_i, i=E,I$.

3.2. The reduced model of (12) under EDA

In the special case when P_E and P_I are exponentially distributed with mean stage durations α and δ respectively (see (7)), and when the survivals from quarantine and isolation are described by the exponential functions given in (8), we can differentiate the E equation in (12) and get (see (9) and (11))

$$E'(t)=\lambda(t)S(t)-(\chi+\alpha+\mu)E(t).$$

Here we have used the fact from (9) that $E_0'(t)=-(\chi+\alpha)E_0(t)$. To get the $I'(t)$ equation we notice from (9) and (10) that

$$[I_0(t)\,\mathrm{e}^{-\mu t}]'=-(\phi+\delta+\mu)I_0(t)\,\mathrm{e}^{-\mu t},$$

$$\tilde{I}_0'(t)=\alpha E_0(t)\,\mathrm{e}^{-\mu t}-(\phi+\delta+\mu)\tilde{I}_0(t),$$

$$[P_I(t)l(t)\,\mathrm{e}^{-\mu t}]'=-(\phi+\delta+\mu)P_I(t)l(t)\,\mathrm{e}^{-\mu t}.$$

Hence, differentiating the I equation in (12) we get

$$I'(t)=\alpha E(t)-(\phi+\delta+\mu)I(t).$$

Similarly we can differentiate other equations in (12) and get the following system of ordinary differential equations

$$
\begin{aligned}
S' &= \mu N - \beta S\tfrac{I+(1-\rho)H}{N} - \mu S,\\
E' &= \beta S\tfrac{I+(1-\rho)H}{N} - (\chi+\alpha+\mu)E,\\
Q' &= \chi E - (\alpha+\mu)Q,\\
I' &= \alpha E - (\phi+\delta+\mu)I,\\
H' &= \alpha Q + \phi I - (\delta+\mu)H,\\
R' &= \delta I + \delta H - \mu R.
\end{aligned}
\tag{14}
$$

We observe that the simple model (2) is exactly the same as system (14) if we let $\alpha_1 = \alpha_2 = \alpha$ and $\delta_1 = \delta_2 = \delta$. This is in fact a consequence of the memory-less property of the exponential distribution, which implies that $\mathcal{M}_i(s) = M_i$ $(i = E, I)$ for all $s \geq 0$ (see (13)). This also confirms the condition (3) mentioned in the introduction.

In the following section, we will explore the analytical properties of the general model (12). We will also consider a specific non-exponential stage distribution, the gamma distribution, and compare the results of the models that assume different stage distributions.

4. Analysis of model (12)

As in most epidemic models, we compute the reproductive number $\mathcal{R}_c$ (c for control) and show that it determines whether or not the disease can be controlled. We also discuss the relationship between $\mathcal{R}_c$ and the usual basic reproductive number, $\mathcal{R}_0$, which is obtained when no control measures are implemented, i.e., $k(s) = l(s) = 1$.

4.1. The reproductive numbers $\mathcal{R}_c$ and $\mathcal{R}_0$

The expression for $\mathcal{R}_c$ given below is derived from the threshold conditions for the stability of the disease-free equilibrium (see Section 4.2). It also has a clear biological interpretation as the secondary number of infections produced by a typical infectious individual during the entire period of infectiousness. Let

$$
\begin{aligned}
a_1(\tau) &= \mathrm{e}^{-\mu\tau}\int_0^\tau [-\dot{P}_E(\tau-u)k(\tau-u)]P_I(u)l(u)\,\mathrm{d}u,\\
a_2(\tau) &= \mathrm{e}^{-\mu\tau}\int_0^\tau [-\dot{P}_E(\tau-u)k(\tau-u)]P_I(u)\bar{l}(u)\,\mathrm{d}u,\\
a_3(\tau) &= \mathrm{e}^{-\mu\tau}\int_0^\tau [-\dot{P}_E(\tau-u)\bar{k}(\tau-u)]P_I(u)\,\mathrm{d}u,
\end{aligned}
\tag{15}
$$

where $\bar{k}(s) = 1 - k(s)$, $\bar{l}(s) = 1 - l(s)$. Then the reproductive number $\mathcal{R}_c$ is given by

$$
\mathcal{R}_c = \int_0^\infty \beta a_1(\tau)\,\mathrm{d}\tau + \int_0^\infty (1-\rho)\beta[a_2(\tau)+a_3(\tau)]\,\mathrm{d}\tau. \tag{16}
$$

To see the biological meaning of the expression (16) and to simplify the notation in the later sections, we introduce the following quantities:

$$\begin{aligned}
\mathcal{T}_E &= \int_0^\infty [-\dot{P}_E(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}s, \quad \mathcal{T}_{E_k} = \int_0^\infty [-\dot{P}_E(s)k(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}s,\\
\mathcal{T}_I &= \int_0^\infty [-\dot{P}_I(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}s, \quad \mathcal{T}_{I_l} = \int_0^\infty [-\dot{P}_I(s)l(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}s,\\
\mathcal{D}_E &= \int_0^\infty P_E(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}s, \quad \mathcal{D}_{E_k} = \int_0^\infty P_E(s)k(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}s,\\
\mathcal{D}_I &= \int_0^\infty P_I(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}s, \quad \mathcal{D}_{I_l} = \int_0^\infty P_I(s)l(s)\,\mathrm{e}^{-\mu s}\,\mathrm{d}s
\end{aligned} \tag{17}$$

$\mathcal{T}_E$ and $\mathcal{T}_{E_k}$ represent, respectively, the probability and the "quarantine-adjusted" probability that exposed individuals survive and become infectious. $\mathcal{T}_I$ and $\mathcal{T}_{I_l}$ represent, respectively, the probability and the "isolation-adjusted" probability that infectious individuals survive and become recovered. $\mathcal{D}_E$ and $\mathcal{D}_{E_k}$ represent, respectively, the mean sojourn time (death-adjusted) and the "quarantine-adjusted" mean sojourn time (death-adjusted as well) in the exposed stage. $\mathcal{D}_I$ and $\mathcal{D}_{I_l}$ represent, respectively, the mean sojourn time (death-adjusted) and the "isolation-adjusted" mean sojourn time (death-adjusted as well) in the infectious stage. Using (17), we can rewrite $\mathcal{R}_c$ in (16) as

$$\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_{IH} + \mathcal{R}_{QH},$$

where

$$\begin{aligned}
\mathcal{R}_I &= \beta \int_0^\infty a_1(\tau)\,\mathrm{d}\tau = \beta \mathcal{T}_{E_k}\mathcal{D}_{I_l},\\
\mathcal{R}_{IH} &= (1-\rho)\beta \int_0^\infty a_2(\tau)\,\mathrm{d}\tau = (1-\rho)\beta \mathcal{T}_{E_k}(\mathcal{D}_I - \mathcal{D}_{I_l}),\\
\mathcal{R}_{QH} &= (1-\rho)\beta \int_0^\infty a_3(\tau)\,\mathrm{d}\tau = (1-\rho)\beta (\mathcal{T}_E - \mathcal{T}_{E_k})\mathcal{D}_I.
\end{aligned} \tag{18}$$

The three components, $\mathcal{R}_I$, $\mathcal{R}_{IH}$, and $\mathcal{R}_{QH}$ in $\mathcal{R}_c$ represent contributions from the I class and from the H class through isolation and quarantine, respectively. From $0 \le k, l \le 1$, we know that

$$\mathcal{T}_{E_k} \le \mathcal{T}_E, \quad \mathcal{D}_{I_l} \le \mathcal{D}_I. \tag{19}$$

Hence, $\mathcal{R}_{IH}$ and $\mathcal{R}_{QH}$ are both positive. Clearly, each $\mathcal{R}_i$ $(i = I, IH, QH)$ is a product of the transmission rate β (or $(1-\rho)\beta$), the probability of surviving the exposed stage and entering the infectious stage, and the average sojourn time being infectious in the corresponding class (adjusted by natural death).

In the absence of control, i.e., $k(s) = l(s) \equiv 1$, $\mathcal{R}_c$ gives the basic reproductive number:

$$\mathcal{R}_0 = \beta \int_0^\infty [-\dot{P}_E(s)]\,\mathrm{e}^{-\mu s}\,\mathrm{d}s \int_0^\infty P_I(s)\,\mathrm{e}^{-\mu s}\mathrm{d}s = \beta \mathcal{T}_E \mathcal{D}_I. \tag{20}$$

We can express $\mathcal{R}_c$ in terms of $\mathcal{R}_0$ as follows. Notice from (18) that $\mathcal{R}_c$ can be simplified to $(1-\rho)\beta\mathcal{T}_E\mathcal{D}_I + \rho\beta\mathcal{T}_{E_k}\mathcal{D}_{I_l}$. Hence, from (20), we get

$$\mathcal{R}_c = \mathcal{R}_0\left[1-\rho\left(1-\frac{\mathcal{T}_{E_k}\mathcal{D}_{I_l}}{\mathcal{T}_E\mathcal{D}_I}\right)\right]. \tag{21}$$

From (19) it is easy to see that $\mathcal{R}_c < \mathcal{R}_0$. The impact of various (single or combined) control measures represented by ρ, χ, or ϕ on the reduction of $\mathcal{R}_0$ can be evaluated using (21).

4.2. Equilibria and their stability

System (12) always has the disease-free equilibrium (DFE) $U_1 = (S_1, E_1, Q_1, I_1, H_1, R_1) = (N, 0, 0, 0, 0, 0)$. If $\mathcal{R}_c > 1$, then there is a unique endemic equilibrium

$$U^* = (S^*, E^*, Q^*, I^*, H^*, R^*) \tag{22}$$

with

$$S^* = \frac{N}{\mathcal{R}_c}, \quad E^* = \mathcal{D}_{E_k}\lambda^* S^*, \quad Q^* = (\mathcal{D}_E - \mathcal{D}_{E_k})\lambda^* S^*,$$
$$I^* = \mathcal{T}_{E_k}\mathcal{D}_{I_k}\lambda^* S^*, \quad H^* = (\mathcal{T}_E\mathcal{D}_I - \mathcal{T}_{E_k}\mathcal{D}_{I_l})\lambda^* S^*, \quad R^* = \frac{1}{\mu}\mathcal{T}_E\mathcal{T}_I\lambda^* S^*, \tag{23}$$

where $\lambda^* = \mu(\mathcal{R}_c - 1)$. Obviously, U^* exists if and only if $\mathcal{R}_c > 1$.

Theorem 1. *The DFE, U_1, is a global attractor if $\mathcal{R}_c < 1$.*

Proof: We first show that $\lambda(t) \to 0$ as $t \to \infty$. Since we are considering the large time behavior of the solutions, without loss of generality, we let $\tilde{X} = 0$. Let

$$A(\tau) = a_1(\tau) + (1-\rho)\left[a_2(\tau) + a_3(\tau)\right],$$

where a_1, a_2, a_3 are given in (15). Then the reproductive number $\mathcal{R}_c$ can be written as

$$\mathcal{R}_c = \beta\int_0^\infty A(\tau)\,\mathrm{d}\tau, \tag{24}$$

and the I and H equations (for large t) can be rewritten as (see (12))

$$I(t) = \int_0^t \lambda(s)S(s)\int_0^{t-s}\left[-\dot{P}_E(t-s-u)k(t-s-u)\right]P_I(u)l(u)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}u\,\mathrm{d}s$$
$$= \int_0^t \lambda(s)S(s)a_1(t-s)\,\mathrm{d}s, \tag{25}$$

and

$$\begin{aligned} H(t) &= \int_0^t \lambda(s)S(s)\int_0^{t-s}[-\dot{P}_E(t-s-u)k(t-s-u)]P_I(u)\bar{I}(u)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}u\,\mathrm{d}s \\ &\quad + \int_0^t \lambda(s)S(s)\int_0^{t-s}[-\dot{P}_E(t-s-u)\bar{k}(t-s-u)]P_I(u)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}u\,\mathrm{d}s \\ &= \int_0^t \lambda(s)S(s)\,[a_2(t-s)+a_3(t-s)]\,\mathrm{d}s. \end{aligned} \tag{26}$$

Since $S(s) \le N$ for all $s > 0$, using (15), (25) and (26), we know that $\lambda(t)$ satisfies

$$\lambda(t) = \frac{\beta\,(I(t)+(1-\rho)H(t))}{N} \le \beta\int_0^t \lambda(s)A(t-s)\,\mathrm{d}s. \tag{27}$$

Let $\lambda^\infty = \limsup_{t\to\infty}\lambda(t)$, i.e.,

$$\lambda^\infty = \lim_{t\to\infty}\Lambda(t) \quad \text{with} \quad \Lambda(t) = \sup_{s\ge t}\lambda(s). \tag{28}$$

By definition, there exists a sequence $t_n \to \infty$ as $n \to \infty$ such that $\lambda(t_n) \to \lambda^\infty$ as $n \to \infty$. Assume that (otherwise we can choose a subsequence)

$$t_{n+1} - t_n \to \infty \quad \text{as} \quad n \to \infty. \tag{29}$$

Then from (27)

$$\lambda(t_{n+1}) \le \beta\int_0^{t_n}\lambda(s)A(t_{n+1}-s)\,\mathrm{d}s + \beta\int_{t_n}^{t_{n+1}}\lambda(s)A(t_{n+1}-s)\,\mathrm{d}s. \tag{30}$$

Obviously $\lambda(t)$ is bounded by β on $[0,\infty)$. The convergence of $\int_0^\infty A(\tau)\,\mathrm{d}\tau$ and (29) imply that

$$\begin{aligned} \int_0^{t_n}\lambda(s)A(t_{n+1}-s)\,\mathrm{d}s &\le \beta\int_{t_{n+1}-t_n}^{t_{n+1}} A(\tau)\,\mathrm{d}\tau \\ &\le \beta\int_{t_{n+1}-t_n}^{\infty} A(\tau)\,\mathrm{d}\tau \to 0 \quad \text{as } n\to\infty. \end{aligned} \tag{31}$$

Using (24) and (28), we have

$$\int_{t_n}^{t_{n+1}}\lambda(s)A(t_{n+1}-s)\,\mathrm{d}s \le \beta\Lambda(t_n)\int_0^\infty A(\tau)\,\mathrm{d}\tau = \Lambda(t_n)\mathcal{R}_c \tag{32}$$

and (28), (30), (31) and (32) yield that

$$\lambda^\infty \le \lambda^\infty\mathcal{R}_c.$$

Since $\mathcal{R}_c < 1$, the above inequality implies that $\lambda^\infty = 0$, i.e., $\lim_{t\to\infty} \lambda(t) = 0$. Hence,

$$\lim_{t\to\infty} I(t) = 0, \quad \lim_{t\to\infty} H(t) = 0.$$

It also follows from $\lambda^\infty = 0$ that $E(t) \to 0$, $Q(t) \to 0$, and $R(t) \to 0$ as $t \to \infty$. Hence, $S(t) = N - (E(t) + Q(t) + I(t) + H(t) + R(t)) \to N$ as $t \to \infty$. This shows that U_1 is a global attractor. This finishes the proof of Theorem 1.

The following observation allows us to study the disease persistence by considering only two equations. Again, to study the large time behavior of solutions, we ignore those individuals who are initially in the population. From equations (15), (25), and (26), we have

$$\lambda(t) = \frac{\beta}{N}\int_0^t \lambda(s)S(s)A(t-s)\,\mathrm{d}s. \tag{33}$$

Integrating the S equation in (12), we obtain

$$S(t) = N - \int_0^t \lambda(s)S(s)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s. \tag{34}$$

We observe that equations (33) and (34) can be studied independently of other variables, and that the behavior of $\lambda(t)$ determines whether or not the disease will die out. Let (S^*, λ^*) denote an equilibrium of (33) and (34). Then S^* and λ^* are constant solutions of the equations

$$S(t) = N - \int_{-\infty}^t \lambda(t)S(t)\,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s,$$

$$\lambda(t) = \frac{\beta}{N}\int_{-\infty}^t \lambda(t)S(t)A(t-s)\,\mathrm{d}s,$$

or satisfy the equations

$$S^* = N - \int_{-\infty}^0 \lambda^* S^*\,\mathrm{e}^{\mu u}\,\mathrm{d}u,$$

$$\lambda^* = \frac{\beta}{N}\int_{-\infty}^0 \lambda^* S^* A(-u)\,\mathrm{d}u. \tag{35}$$

The equations (35) have two solutions (equilibria)

$$(S_1^*, \lambda_1^*) = (N, 0), \quad (S_2^*, \lambda_2^*) = (N/\mathcal{R}_c, \mu(\mathcal{R}_c - 1)).$$

Clearly, the non-trivial equilibrium (S_2^*, λ_2^*) is feasible only if $\mathcal{R}_c > 1$. To study the stability of (S^*, λ^*), following the approach used in Feng et al. (2001) and Hethcote and Tudor (1980), we translate the equilibrium (S^*, λ^*) to the origin by letting $\hat{S} = S - S^*$, $\hat{\lambda} = \lambda - \lambda^*$. Then the system in terms of $\hat{S}$ and $\hat{\lambda}$ can be written in the following matrix form of a Volterra integral equation

$$X(t) = F(t) + \int_0^t K(t-s)G(X(s))\,\mathrm{d}s \tag{36}$$

with

$$F(t) = \begin{pmatrix} -\int_{-\infty}^{0} \lambda^* S^* \,\mathrm{e}^{-\mu(t-s)}\,\mathrm{d}s \\ \frac{\beta}{N}\int_{-\infty}^{0} \lambda^* S^* A(t-s)\,\mathrm{d}s \end{pmatrix}, \quad K(\tau) = \begin{pmatrix} 0 & -\mathrm{e}^{-\mu(\tau)} \\ 0 & \frac{\beta A(\tau)}{N} \end{pmatrix},$$

$$G(X) = \begin{pmatrix} \hat{S} \\ \hat{\lambda}(\hat{S} + S^*) + \lambda^* \hat{S} \end{pmatrix}, \quad X = \begin{pmatrix} \hat{S} \\ \hat{\lambda} \end{pmatrix}.$$

The following Lemma from Hethcote and Tudor (1980) can be used to study the stability of the origin for the system (36).

Lemma 1. *If solutions of (36) exist on* $[0,\infty)$ *and are bounded,* $F(t) \in C[0,\infty)$, $F(t) \to 0$ *as* $t \to \infty$, $K(t) \in L^1[0,\infty)$, $G(0) = 0$, *the Jacobian* $J = DG(0)$ *of* G *is nonsingular and all roots* w *of the characteristic equation* $\det\left(I - \int_0^\infty \mathrm{e}^{-w\tau} K(\tau) J\,\mathrm{d}\tau\right) = 0$ *have negative real parts, then the origin is locally asymptotically stable (l.a.s.) for (36).*

Theorem 2. *If* $\mathcal{R}_c > 1$, *then the DFE* (S_1^*, λ_1^*) *is unstable and the endemic equilibrium* (S_2^*, λ_2^*) *is l.a.s.*

Proof: The stability of the trivial equilibrium $(\hat{S}, \hat{\lambda}) = (0,0)$ of the system (36) determines the stability of the equilibrium (S^*, λ^*) of the system consisting of (33) and (34). It is easy to verify that $F(t) \to 0$ as $t \to \infty$. Since $S^* \neq 0$, we have

$$\det(J) = \det\begin{pmatrix} 1 & 0 \\ \lambda^* & S^* \end{pmatrix} = S^* \neq 0,$$

and hence J is nonsingular. Noticing that

$$\int_0^\infty \mathrm{e}^{-w\tau} K(\tau) J\,\mathrm{d}\tau = \begin{pmatrix} -\frac{\lambda^*}{w+\mu} & -\frac{S^*}{w+\mu} \\ \frac{\lambda^*}{N} L(w) & \frac{S^*}{N} L(w) \end{pmatrix},$$

where

$$L(w) = \beta \int_0^\infty \mathrm{e}^{-w\tau} A(\tau)\,\mathrm{d}\tau,$$

we obtain the characteristic equation

$$H(w) = \det\left(I - \int_0^\infty \mathrm{e}^{-w\tau} K(\tau) J \,\mathrm{d}\tau\right) = 1 + \frac{\lambda^*}{w+\mu} - \frac{S^*}{N} L(w) = 0. \tag{37}$$

At the DFE $(S_1^*, \lambda_1^*) = (N, 0)$, the characteristic equation (37) reduces to

$$L(w) = 1.$$

From the fact that $L(0) = \mathcal{R}_c$ and that $L'(w) < 0$ for any real number w, we conclude that $H(w) = 0$ has a positive real root if $\mathcal{R}_c > 1$. Therefore, the DFE (S_1^*, λ_1^*) is unstable if $\mathcal{R}_c > 1$.

At the endemic equilibrium (S_2^*, λ_2^*), the characteristic equation (37) simplifies to

$$\frac{w + \mu\mathcal{R}_c}{w+\mu} = \frac{L(w)}{\mathcal{R}_c}. \tag{38}$$

Let $w = a + ib$ with $a \geq 0$. Then the real part of the left-hand side of (38) is

$$\Re\left(\frac{w+\mu\mathcal{R}_c}{w+\mu}\right) = \frac{(a+\mu)(a+\mu\mathcal{R}_c) + b^2}{(a+\mu)^2 + b^2} > 1$$

since $\mathcal{R}_c > 1$. The real part of the right-hand side of (38) is

$$\Re\left(\frac{L(w)}{\mathcal{R}_c}\right) = \frac{\beta\int_0^\infty \mathrm{e}^{-a\tau}\cos b\tau\, A(\tau)\,\mathrm{d}\tau}{\mathcal{R}_c} < \frac{\beta\int_0^\infty A(\tau)\,\mathrm{d}\tau}{\mathcal{R}_c} = 1$$

(see (24)). This shows that w cannot be a root of (38) if $\Re w \geq 0$. Therefore, the endemic equilibrium (S_2^*, λ_2^*) is l.a.s. when $\mathcal{R}_c > 1$. This finishes the proof of Theorem 2.

The asymptotical stability for the system (12) follows immediately from the result of Theorem 2. That is, we have the following result.

Theorem 3. *If $\mathcal{R}_c > 1$, then the endemic equilibrium U^* given by (22) is l.a.s. for the system (12).*

Proof: Note that the components S^* and λ^* of U^* (see (23)) are exactly the same as S_2^* and λ_2^*, the endemic equilibrium of the system (33) and (34). Since $\mathcal{R}_c > 1$, from Theorem 2 we know that $\lambda(t) \to \lambda^*$ and $S(t) \to S^*$ as $t \to \infty$ for all solutions with initial data closed to the endemic equilibrium. By the convergence of the integrals in the E, Q, I, H, and R equations of the system (12), we know that solutions of the system (12) approach U^* as $t \to \infty$. The proof of Theorem 3 is completed.

5. Applications

The integral formulation of the model (12) enables the comparison between both models which differ in their distributions of disease stages and the examination of

their consequent epidemiological implications. In this section, we apply the results obtained for the general model (12) to simplified models that use specific, non-exponential disease stage distributions such as the gamma distribution.

5.1. Models with GDA

One possibility to replace the exponential stage duration function $p(s) = \mathrm{e}^{-\theta s}$, or equivalently the probability density function (PDF) $f(s) = \theta\,\mathrm{e}^{-\theta s}$, is to consider the gamma distribution with parameter θ for which the stage duration function is $p_n(s,\theta) = \mathrm{e}^{-n\theta s}\sum_{k=0}^{n-1}\frac{(n\theta s)^k}{k!}$ where $n \geq 1$. We remark that the exponential distribution is a special case of the gamma distribution when $n = 1$. To see the role of n, the graph of the PDF, $f_n(s)$, is shown in Fig. 3 for $n = 1, 3, 20$. Notice that when $n \to \infty$, it corresponds to a fixed duration. The appropriate value of n may be determined by epidemiological data.

Under the gamma distribution $p_n(s,\theta)$ (or simply denoted by $p_n(s)$) with $n \geq 2$, the expected remaining sojourn at stage age s is from (13)

$$\mathcal{M}_n(s) = \int_0^\infty \frac{p_n(t+s)}{p_n(s)}\,\mathrm{d}t = \frac{1}{p_n(s)}\int_s^\infty p_n(t)\,\mathrm{d}t = \frac{1}{n\theta}\frac{\sum_{k=0}^{n-1}\sum_{j=0}^{k}\frac{(n\theta s)^j}{j!}}{\sum_{k=0}^{n-1}\frac{(n\theta s)^k}{k!}}.$$

After checking $\mathcal{M}_n'(s) < 0$ and $\lim_{s\to\infty}\mathcal{M}_n(s) \to T/n$ where $T = 1/\theta$, we know that $\mathcal{M}_n(s)$ strictly decreases with stage age s, and that when s is large the expected remaining sojourn can be as small as T/n. Hence, the expected remaining sojourn in a stage is indeed dependent on the time already spent in the stage. Therefore, the gamma distribution $p_n(s)$ for $n \geq 2$ provides a more realistic description than the exponential distribution $p_1(s)$ for which $\mathcal{M}_1(s) = T$ for all s.

We now consider the model (12). If both $P_E(s)$ and $P_I(s)$ are chosen to be gamma distributions and $k(s)$ and $l(s)$ are the same as in (8), then the integral equation model (12) can be reduced to an ordinary differential equation model. It has been noted that the use of the gamma distribution $p_n(s,\theta)$ for a disease stage, e.g., the exposed stage, is equivalent to assuming that the entire stage is replaced by a series of n sub-stages, and each of the sub-stages is exponentially distributed with the removal rate $n\theta$ and the mean sojourn time T/n, where $T = 1/\theta$ is the mean sojourn time of the entire stage (see, for example, MacDonald, 1978; Hethcote and

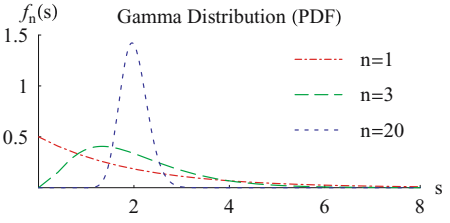

Fig. 3 The probability density function of gamma distribution.

Tudor, 1980; Lloyd, 2001a). This approach of converting a gamma distribution to a sequence of exponential distributions is known as the "linear chain trick."

For the purposes of demonstration and comparison, we adopt the same notation as in (2) and denote $\alpha_2 = \alpha_1$ by α and $\delta_2 = \delta_1$ by δ. Let P_E and P_I be the gamma distributions with the duration functions $P_E(s) = p_m(s, \alpha)$ and $P_I(s) = p_n(s, \delta)$, respectively. Then using the functions $k(s)$ and $l(s)$ given in (8), we can differentiate the equations in system (12) and obtain the following system of ordinary differential equations

$$
\begin{aligned}
S' &= \mu N - \beta S \frac{I + (1-\rho)H}{N} - \mu S,\\
E_1' &= \beta S \frac{I + (1-\rho)H}{N} - (\chi + m\alpha + \mu)E_1,\\
E_j' &= m\alpha E_{j-1} - (\chi + m\alpha + \mu)E_j, \quad j = 2, \ldots, m,\\
Q_1' &= \chi E_1 - (m\alpha + \mu)Q_1,\\
Q_j' &= \chi E_j + m\alpha Q_{j-1} - (m\alpha + \mu)Q_j, \quad j = 2, \ldots, m,\\
I_1' &= m\alpha E_m - (\phi + n\delta + \mu)I_1,\\
I_j' &= n\delta I_{j-1} - (\phi + n\delta + \mu)I_j, \quad j = 2, \ldots, n,\\
H_1' &= m\alpha Q_m + \phi I_1 - (n\delta + \mu)H_1,\\
H_j' &= n\delta H_{j-1} + \phi I_j - (n\delta + \mu)H_j, \quad j = 2, \ldots, n,\\
R' &= n\delta I_n + n\delta H_n - \mu R,\\
&\text{with } I = \sum_{j=1}^{n} I_j, \; H = \sum_{j=1}^{n} H_j.
\end{aligned} \tag{39}
$$

From the formula (21), we get the reproductive number for system (39):

$$
\begin{aligned}
\mathcal{R}_c = &\frac{(m\alpha)^m}{(\mu + m\alpha)^m} \frac{\beta}{\mu + n\delta} \sum_{j=0}^{n-1} \frac{(n\delta)^j}{(\mu + n\delta)^j}\\
&\times \left[1 - \rho \left(1 - \frac{(\mu + m\alpha)^m}{(\mu + m\alpha + \chi)^m} \frac{\mu + n\delta}{\mu + n\delta + \phi} \frac{\sum_{j=0}^{n-1} \frac{(n\delta)^j}{(\mu + n\delta + \phi)^j}}{\sum_{j=0}^{n-1} \frac{(n\delta)^j}{(\mu + n\delta)^j}} \right) \right],
\end{aligned} \tag{40}
$$

with the derivatives

$$
\frac{\partial \mathcal{R}_c}{\partial \chi} = -\beta\rho \frac{m(m\alpha)^m}{(\mu + m\alpha + \chi)^{m+1}} \sum_{j=0}^{n-1} \frac{(n\delta)^j}{(\mu + n\delta + \phi)^{j+1}} < 0, \tag{41}
$$

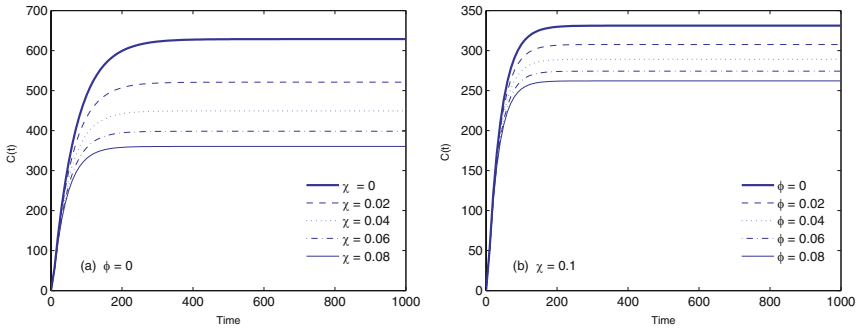

Fig. 4 Numerical simulations of the model (39). It shows that the number of new infections $C(t)$ continuously reduces as the control parameters (χ and ϕ) increase.

$$\frac{\partial \mathcal{R}_c}{\partial \phi} = -\beta\rho \frac{(m\alpha)^m}{(\mu + m\alpha + \chi)^m} \sum_{j=0}^{n-1} \frac{(j+1)(n\delta)^j}{(\mu + n\delta + \phi)^{j+2}} < 0. \tag{42}$$

From (41) and (42), we see that the reproductive number decreases as the disease control parameters χ and ϕ increase. Our simulation results show that the cumulative number $C(t)$ of new infections also decreases with increasing χ and ϕ (see Fig. 4), which appears to be similar to Figs. 2c and 2d. This is not surprising since Figs. 2c and 2d are for model (2) under constraint (3), which is actually a special case of system (39) when $m = n = 1$, whereas Fig. 4 is for system (39) with $m = n = 3$. All parameter values used in Fig. 4 are the same as in Figs. 2c and 2d. However, we observe from Figs. 4b and 4d that for smaller ϕ (e.g., $\phi = 0$) the C value is lower in the GDM than in the EDM ($C = 330$ vs. $C = 350$), whereas for larger ϕ (e.g., $\phi = 0.08$), the C value is higher in the GDM than in the EDM ($C = 262$ vs. $C = 254$). More detailed comparisons between the EDM and the GDM are given in the next section.

5.2. Comparison of EDM and GDM

In this section, we show that when the GDA is used to replace the EDA, model predictions regarding the effectiveness of disease intervention policies may be different both quantitatively and qualitatively. We illustrate this by comparing the two models, EDM (14) and GDM (39). Two criteria are used in the comparison. One is the impact of control measures described by χ and ϕ on the reduction in the magnitude of $\mathcal{R}_c$ and the other one is the reduction in the number of cumulative infections C at the end of an epidemic (the final epidemic size).

From (40)–(42), we know that the reproductive number $\mathcal{R}_c$ for the GDM decreases with increasing χ and ϕ. Similarly, using formula (18), we get the reproductive number for the EDM, $\mathcal{R}_c = \mathcal{R}_I + \mathcal{R}_{IH} + \mathcal{R}_{QH}$, where

$$\mathcal{R}_I = \frac{\beta\alpha}{(\mu + \alpha + \chi)(\mu + \delta + \phi)},$$

$$\mathcal{R}_{IH} = \frac{(1-\rho)\beta\alpha}{\mu+\alpha+\chi}\left(\frac{1}{\mu+\delta} - \frac{1}{\mu+\delta+\phi}\right),$$

$$\mathcal{R}_{QH} = (1-\rho)\beta\left(\frac{\alpha}{\mu+\alpha} - \frac{\alpha}{\mu+\alpha+\chi}\right)\frac{1}{\mu+\delta},$$

which can be written in a simpler form by using (21):

$$\mathcal{R}_c = \frac{\alpha}{\mu+\alpha}\frac{\beta}{\mu+\delta}\left[1-\rho\left(1-\frac{\mu+\alpha}{\mu+\alpha+\chi}\frac{\mu+\delta}{\mu+\delta+\phi}\right)\right]. \tag{43}$$

The derivatives of $\mathcal{R}_c$ with respect to the control parameters are

$$\frac{\partial\mathcal{R}_c}{\partial\chi} = -\beta\rho\frac{\alpha}{(\mu+\alpha+\chi)^2}\frac{1}{\mu+\delta+\phi} < 0,$$

$$\frac{\partial\mathcal{R}_c}{\partial\phi} = -\beta\rho\frac{\alpha}{\mu+\alpha+\chi}\frac{1}{(\mu+\delta+\phi)^2} < 0.$$

Hence, the reproductive number $\mathcal{R}_c$ for the EDM also decreases as the control parameters χ and ϕ increase. Therefore, both models seem to work well when the impact of each individual control measure is considered. When we try to compare model predictions of combined control strategies, however, inconsistent predictions by the two models are observed. For example, in Figs. 5a and 5b, $\mathcal{R}_c$ for both models is plotted either as a function of ϕ for a fixed value of $\chi = 0.1$, or as a function of χ for a fixed value $\phi = 0.1$, or as a function of both χ and ϕ with $\chi = \phi$. For any vertical line except the one at 0.1, the three curves intersect the vertical line at three points that represent three control strategies. The order of these points (from top to bottom) determines the order of effectiveness (from low to high) of the corresponding control strategies since a larger $\mathcal{R}_c$ value will most likely lead to a higher disease prevalence. The order of these three points (labeled by a circle, a triangle, and a square) predicted by the EDM and the GDM is clearly different for the selected parameter sets, suggesting conflicting assessments of interventions between the two models. These conflicting assessments are also shown when we compare the C values. For example, Figs. 5c and 5d plot the function $C(t)$ at the end of epidemics (at which time the number of new infection is zero). Obviously, they show the same problem as that shown in Figs. 5a and 5b in which the $\mathcal{R}_c$ values are compared. The parameter values used in Figs. 5 are $\beta = 0.2$, $\rho = 0.8$, $\alpha = 1/7$, and $\delta = 1/10$, corresponding to a disease with a latency period of $1/\alpha = 7$ days and an infectious period of $1/\delta = 10$ days (e.g., SARS).

To examine in more detail the quantitative differences between the two models, we conducted intensive simulations of the EDM and the GDM for various control measures, some of which are illustrated in Fig. 6. For demonstration purposes, we have used a different α value, $\alpha = 1/10$. All other parameter values are the same as before. Figs. 6a and 6b are for the case in which there is no control ($\chi = \phi = 0$). We observe that both models predict the same value of C (see the C curve). Figs. 6c and 6d are for Strategy I, which implements quarantine alone with $\chi =$

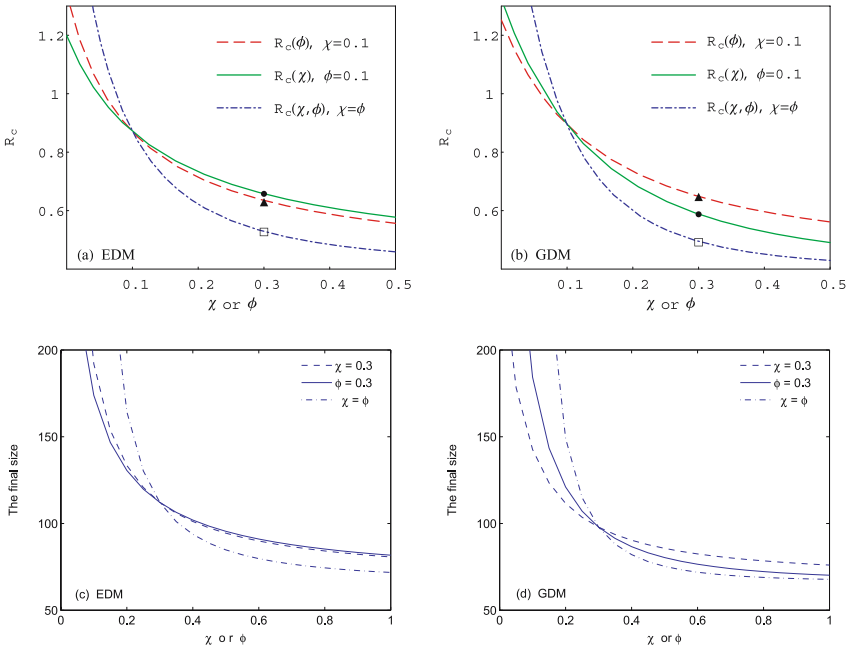

Fig. 5 Comparision of the EDM and the GDM on the impact of various control measures. (a) and (b) are plots of the reproductive number $\mathcal{R}_c$. (c) and (d) are plots of the number of cumulative infections C.

0.08, and Figs. 6e and 6f are for Strategy II, which implements isolation alone with $\phi = 0.1$. The effectiveness of these control measures are reflected by the corresponding $C(t)$ values. According to Figs. 6c and 6e, the EDM predicts that Strategy II, is *more* effective than Strategy I, as the number C of cumulative infections under Strategy II, is 25% lower than the C value under Strategy I, (notice that $C = 2095$ and $C = 1570$ under Strategies I and II, respectively). However, according to Figs. 6d and 6f, the GDM predicts that Strategy II, is *less* effective than Strategy I, as the number C of cumulative infections under Strategy I is 30% lower than the C value under Strategy II (notice that $C = 1540$ and $C = 2270$ under Strategies I and II, respectively). Obviously, in this example, the predictions by the EDM and by the GDM are inconsistent.

6. Conclusion

The goal of this paper is to develop more appropriate epidemiological models for the study of disease control via quarantine and isolation. We first demonstrated that the simple ODE model which assumes exponentially distributed disease stages is not appropriate for the disease control problems under investigation. We then developed a general integral equation model that assumes an arbitrary distribution for each of the disease stages, $P_E(s)$ and $P_I(s)$ (see the model (12)).

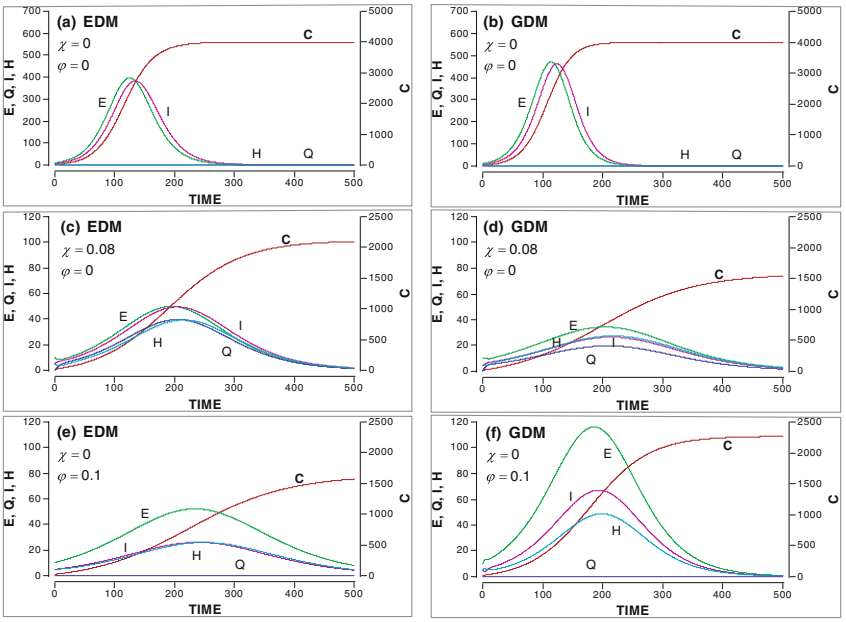

Fig. 6 Numerical simulations of the EDM and the GDM under no control (see (a) and (b)), Strategy I (see (c) and (d)) and Strategy II (see (e) and (f)). The number of cumulative infections $C(t)$ as well as other disease variables $E(t)$, $I(t)$, $Q(t)$, and $H(t)$ are plotted.

One of the key advantages of this general model is that it allows us to compare simplified models obtained by using specific disease stage distributions, including the simple ODE model (14). Consequently, it provides theoretical evidence that simple (and commonly used) ODE models such as (2) or (14) may be generating misleading information regarding disease control strategies. It also suggests that more realistic assumptions on disease stage distribution must be considered when using these models to assess the effectiveness of disease intervention policies.

Many modifications to the simple model (14) can be obtained by using specific and more realistic disease stage distributions. We considered one non-exponential stage distribution, the gamma distribution, and derived the GDM (gamma distribution model) for the case of $n = 3$, which is also an ODE model (see (39)). We argued that this GDM provides a better description of the disease transmission process than the EDM (which is, in fact, a GDM in the extreme case of $n = 1$) by comparing the expected remaining stay at stage age s and the mean sojourn time in the stage for the two stage distributions. It should be pointed out that the GDM is not the complete answer to the problem of discrepancy described in Section 2 for the EDM because the distribution within each sub-stage of the GDM is still exponential. However, the discrepancy is expected to be smaller (especially for larger n values) than it would be in the EDM.

We computed the reproductive number $\mathcal{R}_c$ for the general model (12) and proved that the existence and stability of the equilibria of the model depends on the magnitude of $\mathcal{R}_c$. These results are used to determine another criterion

for the evaluation of control measures. The detailed description of $\mathcal{R}_c$ obtained from the general model (12) also provides important information about the role of the model parameters in the disease transmission dynamics. The comparison of reproductive numbers for the two models under different intervention policies indicate that the EDM again may generate outcomes which conflicts with that of the GDM (see Fig. 5). disease control policies (described by χ and ϕ) using the number of cumulative infections C. Our simulation results from the models (14) and (39) suggest that for many sets of parameter values, the two models predict contradictory outcomes (see Figs. 5 and 6).

To summarize, the results of our study in this article suggest that standard *SEIR*-type ODE models (such as models (2) and (14)), while capable of capturing many essential features of the disease transmission dynamics in the absence of quarantine and isolation, may produce results that are inconsistent with those from models with non-exponentially distributed disease stages. We considered one such model (39) by using the gamma distribution with $n = 3$. Obviously, other types of more realistic stage distributions (P_E and P_I) can be used to derive different improved models. In addition to the choice of P_E and P_I, modifications on the model (14) can be obtained by making different assumptions on $k(s)$ and $l(s)$ as well. These functions obviously play a very important role for the study of disease intervention as these are the parameters (or parameter functions) that reflect the control measures.

Finally, it should be pointed out that the biology of the infection ultimately determines the most appropriate forms for the latent and infectious periods. For some infections it might well be that an n ($n > 1$) stage gamma distribution is a better choice, for others it might even be that an exponential distribution provides a reasonable approximation. In order to determine which model is more appropriate for the problem under investigation, we need to have a deep understanding of the advantages and limitations of these models, as well as the biology of the disease.

Acknowledgements

We thank the Reviewers for very helpful comments and suggestions which gratefully improved the presentation of the paper. Zhilan Feng's research is partially supported by NSF grant DMS-0314575.

References

Anderson, R.M., May, R.M., 1979. Prevalence of schistosome infections within molluscan populations: Observed patterns and theoretical predictions. Parasitology 79, 63–94.

Chowell, G., Fenimorea, P., Castillo-Garsowc, M., Castillo-Chavez, C., 2003. SARS outbreaks in Ontario, Hong Kong and Singapore: The role of diagnosis and isolation as a control mechanism. J. Theo. Bio. 224, 1–8.

Feng, Z., Huang, W., Castillo-Chavez, C., 2001. On the role of variable latent periods in mathematical models for tuberculosis. J. Dyn. Diff. Eqs. 13, 435–452.

Feng, Z., Thieme, H.R., 2000a. Endemic models for the spread of infectious diseases with arbitrarily distributed disease stages I: General theory. SIAM J. Appl. Math. 61(3), 803–833.

Feng, Z., Thieme, H.R., 2000b. Endemic models for the spread of infectious diseases with arbitrarily distributed disease stages II: Fast disease dynamics and permanent recovery. SIAM J. Appl. Math. 61(3), 983–1012.
Hethcote, H., Tudor, D., 1980. Integral equation models for endemic infectious diseases. J. Math. Biol. 9, 37–47.
Hethcote, H., Stech, H.W., van den Driessche, P., 1981. Nonlinear oscillations in epidemic models. SIAM J. Appl. Math. 40(1), 1–9.
Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. A 115, 700–721.
Kermack, W.O., McKendrick, A.G., 1932. Contributions to the mathematical theory of epidemics. II. The problem of endemicity. Proc. Roy. Soc. A 138, 55–83.
Kermack, W.O., McKendrick, A.G., 1933. Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity. Proc. Roy. Soc. A 141, 94–122.
Lipsitch, M., Cohen, T., Copper, B., et al., 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300, 1966–1970.
Lloyd, A., 2001a. Realistic distributions of infectious periods in epidemic models. Theor. Pop. Biol. 60, 59–71.
Lloyd, A., 2001b. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. Proc. R. Soc. Lond. B. 268, 985–993.
MacDonald, N., 1978. Time Lags in Biological Models. Springer-Verlag, New York.
McLean, A.R., Robert, M.M., Pattison, J., Weiss, R.A., 2005. A Case Study in Emerging Infections. Springer-Verlag, Oxford University Press, New York.
Plant, R.E., Wilson, L.T., 1986. Models for age-structured populations with distributed maturation rates. J. Math. Biol. 23, 247–262.
Riley, S., Fraser, C., Donnelly, C.A., Ghani, A.C., et al., 2003. Transmission dynamics of the etiological agent of SARS in Hong Kong—impact of public health interventions. Science 300, 1961–1966.
Ross, R., 1911. The Prevention of Malaria. John Murray, London.
Taylor, H.M., Karlin, S., 1998. An Introduction to Stochastic Modeling, 3rd edition. Academic Press, San Diego.
Thieme, H., 2003. Mathematics in Population Biology. Princeton University Press, Princeton.