THE ROYAL SOCIETY

Received 27 February 2004
Accepted 12 May 2004
Published online 20 October 2004

Modelling strategies for controlling SARS outbreaks

Abba B. Gumel[1], Shigui Ruan[2], Troy Day[3*], James Watmough[4], Fred Brauer[5], P. van den Driessche[6], Dave Gabrielson[1], Chris Bowman[7], Murray E. Alexander[7], Sten Ardal[8], Jianhong Wu[9] and Beni M. Sahai[10]

[1]*Institute of Industrial and Mathematical Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada*
[2]*Department of Mathematics, University of Miami, Coral Gables, FL 33124-4250, USA*
[3]*Departments of Mathematics and Biology, Jeffery Hall, Queen's University, Kingston, Ontario K7L 3N6, Canada* (*tday@mast.queensu.ca*)
[4]*Department of Mathematics and Statistics, University of New Brunswick, PO Box 4400, Fredericton, New Brunswick E3B 5A3, Canada*
[5]*Mathematics Department, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada*
[6]*Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia V8W 3P4, Canada*
[7]*Institute for Biodiagnostics, National Research Council Canada, Winnipeg, Manitoba R3B 1Y6, Canada*
[8]*Central East Health Information Partnership, Box 159, 4950 Yonge Street, Suite 610, Toronto, Ontario M2N 6K1, Canada*
[9]*Laboratory for Industrial and Applied Mathematics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada*
[10]*Cadham Provincial Public Health Laboratory, Winnipeg, Manitoba R3C 3Y1, Canada*

Severe acute respiratory syndrome (SARS), a new, highly contagious, viral disease, emerged in China late in 2002 and quickly spread to 32 countries and regions causing in excess of 774 deaths and 8098 infections worldwide. In the absence of a rapid diagnostic test, therapy or vaccine, isolation of individuals diagnosed with SARS and quarantine of individuals feared exposed to SARS virus were used to control the spread of infection. We examine mathematically the impact of isolation and quarantine on the control of SARS during the outbreaks in Toronto, Hong Kong, Singapore and Beijing using a deterministic model that closely mimics the data for cumulative infected cases and SARS-related deaths in the first three regions but not in Beijing until mid-April, when China started to report data more accurately. The results reveal that achieving a reduction in the contact rate between susceptible and diseased individuals by isolating the latter is a critically important strategy that can control SARS outbreaks with or without quarantine. An optimal isolation programme entails timely implementation under stringent hygienic precautions defined by a critical threshold value. Values below this threshold lead to control, but those above are associated with the incidence of new community outbreaks or nosocomial infections, a known cause for the spread of SARS in each region. Allocation of resources to implement optimal isolation is more effective than to implement sub-optimal isolation and quarantine together. A community-wide eradication of SARS is feasible if optimal isolation is combined with a highly effective screening programme at the points of entry.

Keywords: severe acute respiratory syndrome; modelling; reproduction numbers; quarantine; isolation

1. INTRODUCTION

In March of 2003, the World Health Organization (WHO) reported the emergence of severe acute respiratory syndome (SARS) (WHO 2003*b*), a novel respiratory disease caused by a previously unknown coronavirus, SARS-CoV (Drosten *et al*. 2003; Peiris *et al*. 2003; Marra *et al*. 2003; Rota *et al*. 2003; WHO 2003*c*). The virus apparently evolved from a coronavirus that existed in animals but not in humans (Marra *et al*. 2003; WHO 2003*c*). It is very efficiently transmitted during close person-to-person contact presumably through droplets of respiratory secretions, although its airborne transmission under specific conditions has also been shown to occur (Yu *et al.* 2004). Despite identification and rapid characterization of SARS-CoV, a rapid diagnostic or screening test for SARS is not yet available. The diagnosis of SARS has therefore been based on the presence of clinical symptoms in conjunction with exposure to a known source of SARS-CoV (the epidemiological contact) (WHO 2003*d*). Individuals meeting these criteria have been referred to as 'probable' cases (WHO 2003*d*). The clinical symptoms of SARS include fever, dry cough and dyspnea, often accompanied by radiographic features of pneumonia (Ksiazek *et al.* 2003; Poutanen *et al.* 2003; Tsang *et al.* 2003). The symptoms appear after 3–17 days of incubation and most individuals recover within two weeks of illness (Booth *et al.* 2003; Dwosh *et al.* 2003). The disease nevertheless inflicts a high rate of mortality (*ca*. 15%) especially among the elderly (*ca*. 50%) (Booth *et al*. 2003; Donnelly *et al*. 2003; Ksiazek *et al*. 2003; Poutanen *et al*. 2003; Tsang *et al*. 2003; WHO 2003*a*).

SARS, first reported in November of 2002 in the Guangdong province of China, quickly spread across continents, with the highest prevalence in the Asia Pacific region (WHO 2003*a*). As of 24 June 2003, there were 246 probable cases and 37 deaths in the Greater Toronto Area (GTA; Health Canada 2003*a*); a recent report (Health

* Author for correspondence (tday@mast.queensu.ca).

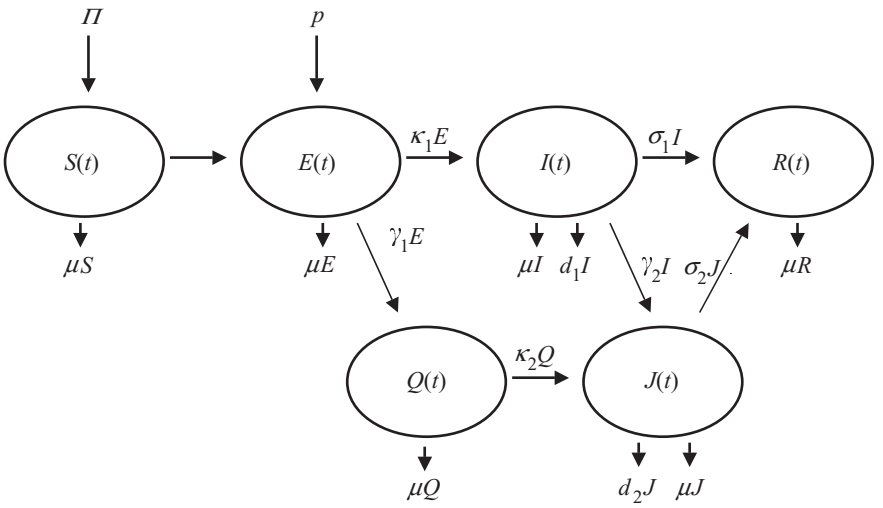

Figure 1. Schematic flow diagram for the SARS model (2.1)–(2.6). The model consists of six sub populations: susceptible ($S(t)$), asymptomatic ($E(t)$), quarantined ($Q(t)$), symptomatic ($I(t)$), isolated ($J(t)$) and recovered ($R(t)$) individuals in a population of $N(t) = S(t) + E(t) + Q(t) + I(t) + J(t) + R(t)$ individuals.

Canada 2003*b*) quoted 44 SARS fatalities for the whole duration of the 2003 SARS outbreaks in Toronto, 1755 probable cases and 296 deaths in Hong Kong (WHO 2003*a*), 206 probable cases and 31 deaths in Singapore (WHO 2003*a*), and 2521 probable cases and 191 deaths in Beijing, the hardest-hit city in the world (Ministry of Health, China 2003; WHO 2003*a*). Overall, SARS accounted for at least 774 deaths and 8098 infections globally (Lingappa *et al.* 2004). Because of its rapid transmissibility across continents and potential to cause large epidemics, the WHO launched a vigorous global campaign to combat SARS. Despite such effort, a definitive treatment, prophylaxis or vaccine against SARS remains elusive. Consequently, isolation of individuals with symptoms of SARS and quarantine of individuals suspected to be exposed to SARS-CoV were adopted worldwide as means to control ongoing and impending outbreaks. Historically, isolation and quarantine of individuals carrying, or suspected to be carrying, infectious pathogens have been effective approaches for containing contagious diseases. However, isolation of individuals with clinical symptoms of SARS, in hospitals or at home, produced mixed results because it was associated with significant new infections in various parts of the world, including GTA where most of all probable cases originated in hospitals involving health-care workers (Booth *et al.* 2003; Donnelly *et al.* 2003; Dwosh *et al.* 2003; MMWR 2003). Notable factors compromising the optimal isolation of symptomatic individuals may include inadequate hygiene precautions (such as the lack of appropriate protective apparel including face masks and other SARS-specific infection control protocols), variability of incubation period, lack of knowledge of the transmissibility of disease early in the outbreak, and, in particular, the vulnerability of health-care workers. Reports of frequent infection of health-care workers while caring for SARS patients in isolation points to inadequate hygiene precautions as a potential major factor in reversing the impact of isolation. This view is further supported by the fact that a major revision of the infection control protocols and a strict adherence to such protocols were needed to reduce, and eventually eliminate, the incidence of new infections in hospitals. Implementation of these measures led to the eventual elimination of all cases of SARS worldwide by August 2003. This provided optimism for global control of SARS, despite the recent re-emergence of a few new cases in China (CDC 2004; WHO 2004*a*,*b*), which have been successfully contained.

Mathematical modelling can be a useful tool for designing strategies to control rapidly spreading infectious diseases in the absence of an effective treatment, vaccine or diagnostic test. Such modelling has recently been used to assess the epidemic potential and control of SARS in Hong Kong (Chowell *et al.* 2003; Lee *et al.* 2003; Riley *et al.* 2003), Singapore (Chowell *et al.* 2003; Lipsitch *et al.* 2003), Beijing (Wang & Ruan 2004), Taiwan (Hsieh *et al.* 2004), China (Zhou *et al.* 2004) and in a community in general (Lloyd-Smith *et al.* 2003). The nosocomial spread of SARS has been studied using models introduced by Lloyd-Smith *et al.* (2003) (discrete, stochastic) and Webb *et al.* (2004) (continuous, deterministic). To gain an insight into the critical factors associated with the control of SARS in a community and the world, we develop a dynamic model to study the outbreak of SARS and its control in four affected regions, namely GTA, Hong Kong, Singapore and Beijing. The model is described in § 2 and the results of simulations are compared with available data in § 3. Finally, the implications for control methods are discussed in § 4.

2. MODEL FORMULATION

General mathematical models for the spread of infectious diseases have been described previously (Anderson & May 1991; Diekmann & Heesterbeek 2000; Hethcote 2000). We develop a variant that reflects some key epidemiological properties of SARS. The model monitors the dynamics of six sub-populations (classes), namely susceptible ($S(t)$), asymptomatic ($E(t)$), quarantined ($Q(t)$), symptomatic ($I(t)$), isolated ($J(t)$) and recovered ($R(t)$) individuals (figure 1). The total population size is $N = N(t) = S(t) + E(t) + Q(t) + I(t) + J(t) + R(t)$. In this model,

quarantine refers to the separation of SARS-CoV-infected individuals from the general population before development of clinical symptoms, whereas *isolation* describes the separation of SARS-CoV-infected individuals with such symptoms.

Our model is a system of ordinary differential equations. In practice, some parameters may change in time as control measures are implemented or changed. The theoretical analysis is performed for systems with constant parameters. In numerical simulations, we use parameters that are piecewise constant to approximate time-dependent parameters.

This model offers a basic description of the transmission dynamics of SARS. A full description of SARS should include social network structure, spatial structure and age structure. At least in the initial stage of the outbreak, the network structure is very important, and a stochastic model is probably essential. Once an epidemic has begun, the assumption of homogeneous mixing made in our model is probably a reasonable approximation. Because our model is focused on the long-term control of SARS, a deterministic formulation seems plausible. Although super-spreading events occurred during the early stages of the epidemic, Riley *et al.* (2003) argue that in Hong Kong the basic reproduction number was large enough that an epidemic would have developed even without super-spreading events. Our model therefore does not incorporate super-spreading events.

Even though a super-spreading event may occur after an epidemic is established, its effect would be a smaller fraction of the epidemic spread. Our approach has been to try to focus on what we believe to be the essentials in formulating a simple model. Because data early in an epidemic are inevitably incomplete and inaccurate, the estimation of many parameters of a detailed model may well produce predictions that are less accurate than those of a simpler model with fewer parameters.

Our model incorporates some demographic effects by assuming a proportional natural death rate $\mu > 0$ in each of the six sub-populations of the model. In addition, our model includes a net inflow of susceptible individuals into the region at a rate Π per unit time. This parameter includes new births, immigration and emigration. Owing to the continuing inflow of travellers into the community during the SARS outbreaks, and the absence of a rapid and effective screening test, it is assumed that some of these travellers are asymptomatically infected and enter the community at a rate p per day. The parameter p can also model cases of animal-to-human transmission of SARS-CoV. Since the estimated value of p is small (see electronic Appendix B), we take $p = 0$ for the theoretical analysis. However, our numerical simulations allow $p > 0$. Brauer & van den Driessche (2001) provide analysis of the role of recruitment of infected individuals ($p > 0$) on disease dynamics in simple models.

(a) *Susceptible individuals:* S(t)

The susceptible population is increased by the net inflow (recruitment) of individuals into the region and is decreased by natural death. The susceptible population also decreases following infection, acquired by contact between a susceptible and an infected individual, who may be symptomatic, asymptomatic, quarantined, or isolated. The transmission coefficients for these four classes of infected individuals are β, $\varepsilon_E\beta$, $\varepsilon_Q\beta$, and $\varepsilon_J\beta$ respectively. Here, we use a basic transmission coefficient, β (which models both the infectiousness of SARS and contact rates), with modification factors for asymptomatic (ε_E), quarantined (ε_Q), and isolated (ε_J) individuals.

Although SARS-CoV is believed to be transmitted exclusively by symptomatic individuals, a very low rate of transmission by asymptomatic individuals cannot yet be ruled out. The model accounts for this possibility by employing the modification parameter ε_E, where $0 \leqslant \varepsilon_E < 1$. The modification parameter $\varepsilon_Q \geqslant 0$ accounts for varying levels of hygiene precautions during quarantine and a similar interpretation is given to the modification parameter ε_J during isolation. Because the quarantine and isolation programmes and hygiene precautions during quarantine and isolation were implemented and enhanced progressively after an outbreak, the transmission coefficients $\varepsilon_Q\beta, \varepsilon_J\beta$, quarantine and isolation rates could be modelled as time-dependent parameters in simulations. The pair $\varepsilon_Q\beta$, γ_1 may be viewed as control parameters in the implementation of a quarantine programme and the pair $\varepsilon_J\beta$, γ_2 may be viewed as control parameters in the implementation of an isolation programme (see parameters γ_1, γ_2 in electronic Appendix B). Furthermore, the interaction between susceptible and infected individuals (asymptomatic, symptomatic, quarantined or isolated) is modelled using 'standard' mixing incidence (Anderson & May 1991; Diekmann & Heesterbeek 2000; Hethcote 2000), in the context of the total population. The rate of change of the susceptible population can be represented by the following equation:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \Pi - \frac{S(\beta I + \varepsilon_E\beta E + \varepsilon_Q\beta Q + \varepsilon_J\beta J)}{N} - \mu S. \qquad (2.1)$$

(b) *Asymptomatic individuals:* E(t)

Asymptomatic individuals have been exposed to the virus, but have not yet developed clinical symptoms of SARS. Such individuals arise as a result of the undetected entry of infected individuals into the community (at a rate p) and new infections of susceptible individuals. The population is diminished by quarantine (at a rate γ_1), development of clinical symptoms (at a rate κ_1) and natural death. The parameter γ_1 represents the rate of quarantining of people who have been in contact with an infected individual. Thus,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = p + \frac{S(\beta I + \varepsilon_E\beta E + \varepsilon_Q\beta Q + \varepsilon_J\beta J)}{N} - (\gamma_1 + \kappa_1 + \mu)E. \qquad (2.2)$$

(c) *Quarantined individuals:* Q(t)

These are asymptomatically infected individuals who are quarantined at a rate γ_1 because of established epidemiological contact with a source of SARS-CoV. For simplicity, we assume that all quarantined individuals are asymptomatic infectives who will go on to develop symptoms and then move to the isolated class. It would be more realistic to assume that some uninfected individuals are also quarantined, but this would complicate the model substantially and would require the addition of several parameters and compartments. In effect, the error introduced by our simplification is to leave in the susceptible population some people who are actually in quarantine and thus are making fewer contacts. The population is diminished by

development of clinical symptoms (at a rate κ_2) with transfer to the isolated class and natural death. Thus,

$$\frac{dQ}{dt} = \gamma_1 E - (\kappa_2 + \mu)Q. \tag{2.3}$$

(**d**) ***Symptomatic individuals:* I*(t)***

The symptomatic population is generated after the development of clinical symptoms of SARS by members of the asymptomatic class. It is diminished by isolation (at a rate γ_2), by disease-induced death (at a rate d_1), recovery (at a rate σ_1) and natural death. It is assumed that γ_2, the rate at which symptomatic individuals seek medical attention and are therefore put into isolation, is greater than γ_1, the rate at which asymptomatic individuals go into quarantine. Thus,

$$\frac{dI}{dt} = \kappa_1 E - (\gamma_2 + d_1 + \sigma_1 + \mu)I. \tag{2.4}$$

(**e**) ***Isolated individuals:* J*(t)***

These are individuals who have developed clinical symptoms and have been isolated, for example, by hospitalization. These come from both the symptomatic class (at rate γ_2) and from the quarantined class (at rate κ_2). This population is diminished by recovery (at a rate σ_2), disease-induced death (at a rate d_2) and natural death. It is assumed that $d_1 > d_2$ and $\sigma_2 > \sigma_1$, because isolated individuals are likely to receive a partly effective, experimental, anti-viral treatment (such as Ribavirin; Booth *et al.* 2003; Lee *et al.* 2003) during hospitalization. Thus,

$$\frac{dJ}{dt} = \gamma_2 I + \kappa_2 Q - (\sigma_2 + d_2 + \mu)J. \tag{2.5}$$

(**f**) ***Recovered individuals:* R*(t)***

It is assumed that recovered individuals possess lasting immunity against SARS although so far this has neither been established nor contradicted. Symptomatic and isolated individuals recover from the disease at rates σ_1 and σ_2; respectively, and this population is diminished by natural death. Thus,

$$\frac{dR}{dt} = \sigma_1 I + \sigma_2 J - \mu R. \tag{2.6}$$

(**g**) ***Reproduction numbers:* $\mathcal{R}_o$, $\mathcal{R}_c$**

It can be shown (by defining an associated positively invariant region) that for non-negative parameters, the model (2.1)–(2.6) with non-negative initial values is well posed both mathematically and epidemiologically. The *basic reproduction number* ($\mathcal{R}_0$) is defined as the expected number of secondary infections produced by an index case (Anderson & May 1991; Diekmann & Heesterbeek 2000; Hethcote 2000). The index case for an outbreak refers to the first infected individual introduced into a population of susceptible individuals. The basic reproduction number is an important epidemiological threshold and can be calculated using the next generation matrix (Diekmann & Heesterbeek 2000; van den Driessche & Watmough 2002). In a deterministic model, if $\mathcal{R}_0 < 1$ an epidemic cannot develop from a small influx of SARS-infected individuals but an epidemic will develop if $\mathcal{R}_0 > 1$. By contrast, an epidemic is not guaranteed in a stochastic model if $\mathcal{R}_0 > 1$, but the probability of an epidemic increases with $\mathcal{R}_0$.

We will follow the convention that the basic reproduction number, $\mathcal{R}_0$, is defined in the absence of control measures and introduce the *control reproduction number*, $\mathcal{R}_c$, to denote the reproduction number when control measures are in place. We calculate $\mathcal{R}_c$ in the same way as we calculate $\mathcal{R}_0$ but using the full model with quarantined and isolated classes. Thus, $\mathcal{R}_0$ is $\mathcal{R}_c$ with $\gamma_1 = \gamma_2 = 0$. For the model (2.1)–(2.6) with $p = 0$, it can be shown (by linearizing the model around the disease-free equilibrium) that

$$\mathcal{R}_0 = \frac{\varepsilon_E \beta}{\kappa_1 + \mu} + \frac{\beta \kappa_1}{(\kappa_1 + \mu)(d_1 + \sigma_1 + \mu)}, \tag{2.7}$$

$$\mathcal{R}_c = \frac{\varepsilon_E \beta}{D_1} + \frac{\beta \kappa_1}{D_1 D_2} + \frac{\varepsilon_Q \beta \gamma_1}{D_1 D_4} + \frac{\varepsilon_J \beta \kappa_1 \gamma_2}{D_1 D_2 D_3} + \frac{\varepsilon_J \beta \gamma_1 \kappa_2}{D_1 D_3 D_4}, \tag{2.8}$$

where $D_1 = \gamma_1 + \kappa_1 + \mu$, $D_2 = \gamma_2 + d_1 + \sigma_1 + \mu$, $D_3 = \sigma_2 + d_2 + \mu$ and $D_4 = \mu + \kappa_2$.

In electronic Appendix A, we provide interpretations of the reproduction numbers, and electronic Appendix B contains estimates of the model parameters.

3. RESULTS

We begin with some estimates of the dependence of the reproduction numbers on parameters. For simulation purposes, we use the following set of parameter values: $\mu = 0.000\,034$, $\kappa_1 = 0.1$, $\kappa_2 = 0.125$ and $\sigma_1 = 0.0337$, $\sigma_2 = 0.0386$, $d_1 = 0.0079$, $d_2 = 0.0068$ as baseline values for all the SARS-affected regions considered here, with the exception of Beijing where we choose $\sigma_1 = 0.0413$, $\sigma_2 = 0.0431$, $d_1 = 0.0055$ and $d_2 = 0.0041$. The parameters γ_1 and γ_2 are initially assumed to be zero until 30 March 2003 when their values are switched to $\gamma_1 = 0.1$ and $\gamma_2 = 0.5$ (to model the late implementation of quarantine and isolation programmes). Furthermore, it is assumed that perfect isolation (ε_J) is attained in all four regions by 20 April 2003. Note that these dates do not correspond to the dates of new policy implementation, but to an average between its start and full effectiveness. The values of the parameters Π, p, ε_J and β generally differ for each of the four regions (see caption to figure 2). It should be noted also that all parameters, except ε_E,ε_Q and ε_J, have units of per day.

With the above choices of parameter values for GTA and $\Pi = 136$, $p = 0.06$, but with β, ε_E, ε_Q and ε_J left unspecified, we obtain $\mathcal{R}_c \approx \beta(0.92 + 5\varepsilon_E + 4\varepsilon_Q + 21.18\varepsilon_J)$. This indicates that $\mathcal{R}_c$ is much more sensitive to changes in the factor ε_J than to changes in $\varepsilon_E, \varepsilon_Q$. Thus, the model predicts that a reduction in the effective contact rate of isolated members of the population (ε_J) has a larger payoff than a reduction in the effective contact rate of quarantined members ($\varepsilon_E, \varepsilon_Q$). For instance, a reduction of ε_J from 0.3 to 0.2 (using $\varepsilon_E = \varepsilon_Q = 0$, $\beta = 0.2$ in both cases) would decrease $\mathcal{R}_c$ from 1.46 to 1.03 (see also figure 3*c*).

Numerical simulations (figure 2) using the aforementioned parameter values, together with the choices of β and ε_J that give the best fit with observed cumulative deaths data, give the following results. The values for $\mathcal{R}_0$ and $\mathcal{R}_c$ for each of the four regions are given below, where the numbers in parenthesis represent $\mathcal{R}_c$ values when isolation measures are enhanced to stringent levels (i.e. $\varepsilon_J = 0$); for GTA, $\beta = 0.2$, $\varepsilon_J = 0.36$, giving $\mathcal{R}_0 = 4.80$, $\mathcal{R}_c = 1.7$ (0.18); for Hong Kong, $\beta = 0.15$, $\varepsilon_J = 0.84$, giving $\mathcal{R}_0 = 3.60$, $\mathcal{R}_c = 2.8$ (0.14); for Singapore, $\beta = 0.21$, $\varepsilon_J = 0.2$, giving $\mathcal{R}_0 = 5.04$, $\mathcal{R}_c = 1.08$ (0.19) and for Beijing, $\beta = 0.23$, $\varepsilon_J = 0.82$, giving $\mathcal{R}_0 = 4.91$, $R_c = 4.03$ (0.21). It is

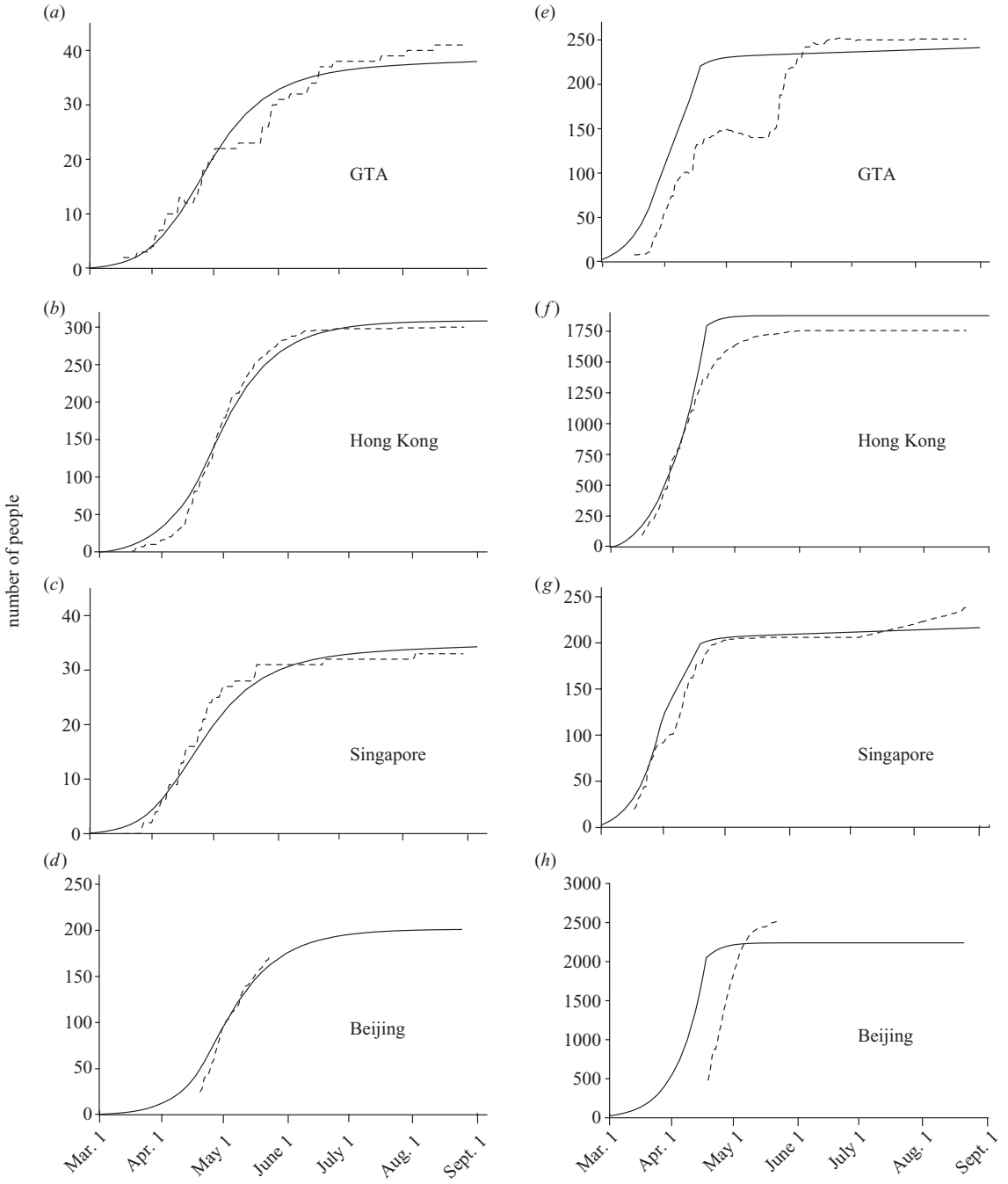

Figure 2. Comparison of simulation results (solid lines) and epidemiological data from the WHO (dotted lines) for (*a–d*) the cumulative number of deaths and probable cases (*e–h*) in GTA, Hong Kong, Singapore and Beijing. We model the late implementation of quarantine and isolation strategies by initially setting γ_1 and γ_2 to zero in all four regions considered, until 30 March, when they are increased to 0.1 and 0.5, respectively. Similarly, ε_J is initially set to the value given in § 3 and then set to zero after 20 April. Unless otherwise stated, the parameter values used are as specified in § 3. (*a*,*e*) GTA, where the outbreak is assumed to begin on 23 February (as per Health Canada 2003*a*). Here, $\Pi = 136$, $p = 0.06$ with initial conditions $S(0) = 4$ million, $E(0) = 6$, $I(0) = 1$, $Q(0) = 0$, $R(0) = 0, J(0) = 0$. The sharp increase in probable cases at the end of May reflects a change in the case definition used in Canada (Health Canada 2003*a*). (*b*,*f*) Hong Kong, where the outbreak consisted of two large clusters, in Amoy Gardens residential complex and Prince of Wales Hospital (Lee *et al.* 2003; Riley *et al.* 2003), which are accounted for in our study by using estimated numbers of cases as initial conditions on 1 March. Although these clusters were reported at the end of March, it is assumed that using the numbers generated during the super-spreading events as initial conditions (on 1 March) would compensate for the absence of explicit formulation for super-spreading events in our model. Other parameter and initial values are $\Pi = 221$, $p = 0$, $S(0) =$ 6.5 million, $E(0) = 124$, $I(0) = 1$, $Q(0) = 0$, $R(0) = 0, J(0) = 0$. (*c*,*g*) Singapore, where the outbreak is assumed to begin on 23 February. Here, $\Pi = 136$, $p = 0.06$, $S(0) = 4$ million, $E(0) = 6$, $I(0) = 1$, $Q(0) = 0$, $R(0) = 0, J(0) = 0$. (*d*,*h*) Beijing, where the outbreak is assumed to begin on 15 February. Other parameter and initial values are $\Pi = 408$, $p = 0$, $S(0) = 12$ million, $E(0) = 10, I(0) = 3$, $Q(0) = 0$, $R(0) = 0$, $J(0) = 0$.The reported case mortality in Beijing is lower than that of the other SARS-affected regions we considered (see Wang & Ruan 2004). Thus, we use $\sigma_1 = 0.0413$, $\sigma_2 = 0.0431$, $d_1 = 0.0055$ and $d_2 = 0.0041$. It is worth pointing out that the increase in the number of cases for Singapore (*g*) is due to our use of linear interpolation of the data owing to the changes in data reporting by the WHO (from providing day-to-day numbers to bi-monthly figures towards the end of July).

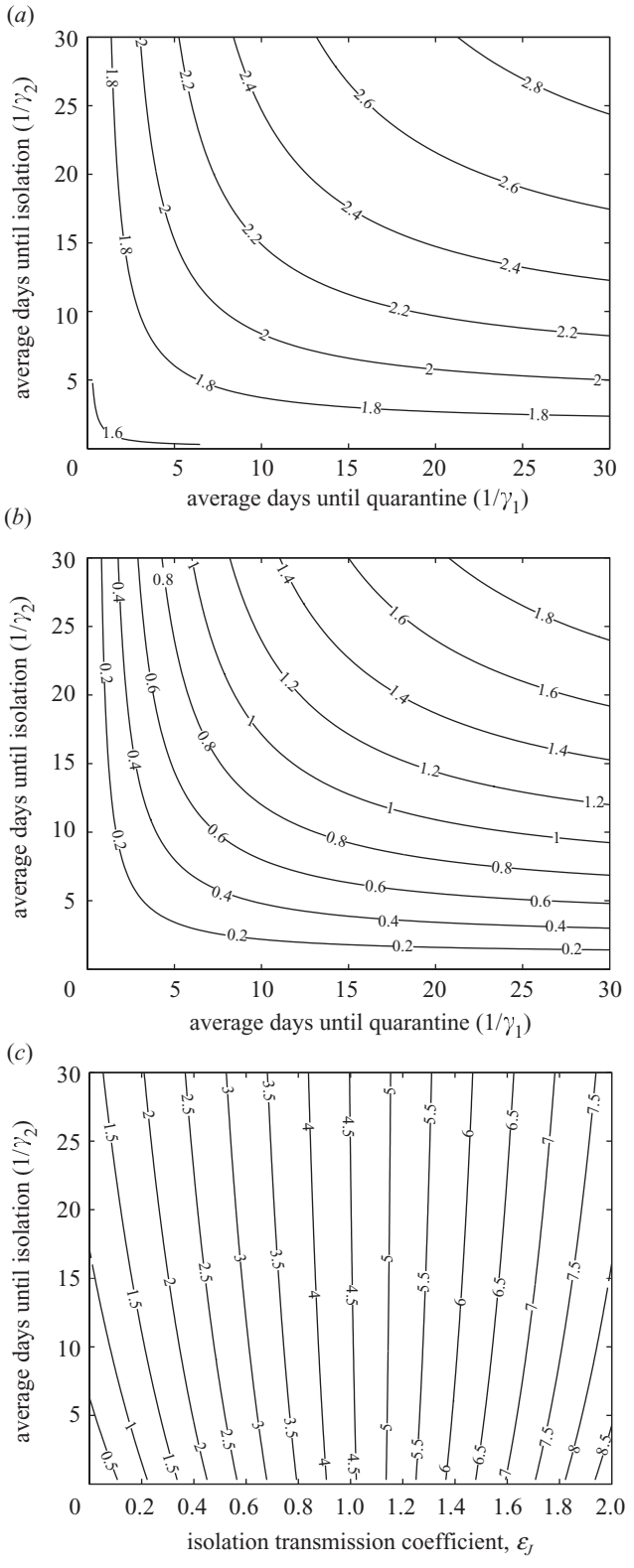

Figure 3. Contour plots of $\mathscr{R}_c$ for the GTA. (*a*,*b*) Plot contours of $\mathscr{R}_c$ versus average days to quarantine ($1/\gamma_1$) and isolation ($1/\gamma_2$), (*a*) in the absence of stringent hygiene precautions ($\varepsilon_J = 0.36$) and (*b*) in the presence of stringent hygiene precautions ($\varepsilon_J = 0$). (*c*) Contour plots of $\mathscr{R}_c$ versus the isolation parameter ε_J and the average days to isolation ($1/\gamma_2$). All parameter values other than γ_1 and γ_2 for (*a*) and (*b*) and ε_J and γ_2 for (*c*) and $p = 0$ are as given in § 3 for the GTA.

worth stating that the estimates we obtain for ε_J and β cannot be reliably used to compare control methods between cities for two reasons: β and ε_J have nearly equal and opposing effects on the incidence rates, and the data do not allow independent estimates of their values. In particular, it is expected that the assumption that $\varepsilon_J = 0$ before 20 April 2003 leads to underestimates of the initial value of ε_J and overestimates of the true value of β for each city. Thus, the differences in our estimates may be more a reflection of the reliability of this assumption for each city.

These $\mathscr{R}_0$ values are slightly higher than some of the reported $\mathscr{R}_0$ values such as $\mathscr{R}_0$ in the range 2.2–3.6 (Lipsitch *et al.* 2003), 2.7–3.4 (Riley *et al.* 2003), and $\mathscr{R}_0 = 2.73$, $\mathscr{R}_c = 1.2$ (Chowell *et al.* 2003). Actually, Chowell *et al*. give $\mathscr{R}_0 = 1.2$, but what they calculate is $\mathscr{R}_c$ in our terminology, and their parameter values give $\mathscr{R}_0 = 2.73$. It should be noted that the estimate for Hong Kong (Riley *et al.* 2003) excludes the contribution of super-spreading events. In any case, our estimates should be viewed as illustrations. This is especially significant for $\mathscr{R}_0$, which is undoubtedly strongly dependent on social networks and stochastic considerations, which are not incorporated in our model. As highlighted by Lipsitch *et al.* (2003), Lloyd-Smith *et al.* (2003) and Riley *et al.* (2003), realistic estimation of $\mathscr{R}_0$ from initial data is difficult.

At the beginning of the outbreak, the values of $\mathscr{R}_0$ for GTA, Singapore, and Beijing are approximately equal, whereas that for Hong Kong is slightly smaller. Overall, this indicates that all four regions displayed similar transmission dynamics and had similar levels of susceptibility to SARS. After the start of quarantine and isolation (but before strict hygienic measures being put in place), our results suggest that the value of $\mathscr{R}_c$ is comparable in the GTA, Singapore and Hong Kong ($\mathscr{R}_c = 1.08$ to 2.8), but our estimate for the value in Beijing is substantially higher ($\mathscr{R}_c = 4.03$). This finding should be interpreted cautiously, however, owing to the limited data available from Beijing, together with possible inaccurate and/or under-reporting of probable cases (see reports cited in Wang & Ruan (2004)). Once an isolation programme of symptomatic individuals is in place, subject to strict hygiene measures (so that $\varepsilon_J = 0$), the estimated values of $\mathscr{R}_c$ for each region indicate that the spread of SARS was rapidly brought under control, with the values from all four regions being essentially the same (where $\mathscr{R}_c$ varies between 0.14 and 0.21).

The contour plots of figure 3*a*,*b* show the dependence of $\mathscr{R}_c$ on the quarantine rate γ_1 and the isolation rate γ_2 for the GTA. The axes for these plots are given as average days from exposure to quarantine ($1/\gamma_1$) and average days from onset of symptoms to isolation ($1/\gamma_2$). In both cases, the contours show that, as expected, the use of anti-SARS control measures (increasing γ_1 and γ_2) reduces the control reproduction number $\mathscr{R}_c$ (and, therefore, SARS cases and mortality). With the remaining parameters set as for the GTA, and with $\varepsilon_J = 0.36$, quarantine and isolation are insufficient to control an outbreak (see figure 3*a*). With these parameter values, as γ_1 increases from 0 to $+\infty$, $\mathscr{R}_c$ decreases from 1.83 to 1.59. Similarly, when γ_2 increases from 0 to $+\infty$, $\mathscr{R}_c$ decreases from 3.19 to 1.58. Thus, $\mathscr{R}_c > 1$ in both cases, and the disease will persist in the population (i.e. the above control measures cannot lead to effective control of the epidemic). By contrast, our study shows that with a perfect isolation programme (so that $\varepsilon_J = 0$),

the outbreak can be controlled with isolation alone (see figure 3*b*). Calculations show that with $\gamma_1 = 0$, $\mathscr{R}_c < 1$ if $\gamma_2 > 0.16$. It follows, that with this high value of $\varepsilon_J = 0.36$, neither the quarantine of asymptomatic individuals nor the isolation of symptomatic individuals can separately control the epidemic. Such control can be achieved by significantly reducing SARS transmission during isolation (i.e. reduced ε_J). For example, by plotting $\mathscr{R}_c$ against ε_J for the GTA, we find that the critical threshold value of ε_J needed to make $\mathscr{R}_c < 1$ is $\varepsilon_J \approx 0.1927$. This shows that unless the isolation programme at GTA is stringent enough to make ε_J less than the above threshold value, its usage will not only fail to offer effective control of the epidemic, it will also result in further (nosocomial) infections of healthcare workers and other caregivers. Lloyd-Smith *et al.* (2003) showed that for $\mathscr{R}_0 \approx 3$ (as reported for SARS in Hong Kong) and the absence of contact precautions in hospitals, a SARS outbreak can be controlled only if isolation reduces transmission by at least a factor of four and the mean onset-to-hospitalization time is less than 3 days. The contour plot of figure 3*c* shows the dependence of the control reproduction number $\mathscr{R}_c$ on the modification factor ε_J and the average days from onset of symptoms to isolation $(1/\gamma_2)$ for the GTA parameter values (note that $\gamma_1 = 0.1$). It is worth noting that since the transmission parameter β is a product of the transmission probability per contact and the rate of contacts, it follows that the modification factor ε_J (for the transmission coefficient of isolated cases) also combines reductions (or increases) in transmission probabilities and contact rates during isolation.

The efficient detection of SARS-infected individuals at points of entry into a community is a daunting task because of the lack of reliable virus- or symptom-based screening tests, human errors and the existence of unregulated points of entry. Thermal screening of passengers at transportation terminals has proven only partly effective. Our results suggest that, unless marked improvements in screening effectiveness were made, changes in protocols at points of entry would probably have had minimal effect. For example, if screening had been perfect throughout the entire duration of the epidemic in the GTA (after the first travellers brought SARS to Canada in February 2003), the cumulative deaths would have decreased by only five individuals; from approximately 38 to 33 (figure 4*a*). Thus, apart from the obvious benefits that strict screening might have had for preventing an outbreak initially, once the outbreak in the GTA began, better screening would have had a relatively minor impact on the epidemic's progression.

The timely implementation of isolation is essential in limiting the impact of SARS outbreaks. Simulations based on no undetected entry of infected individuals into the population $(p = 0)$ show that increasing the time between the onset of an outbreak and the implementation of quarantine and isolation causes an accelerating rise in cumulative deaths (figure 4*b*). For example, using the parameter and initial values for GTA (figure 2*a*,*e*), our simulations show that a delay of only five extra days in the implementation of quarantine and isolation in GTA would have led to approximately 16 additional deaths. Using a stochastic model with $\mathscr{R}_0 \approx 3$, Lloyd-Smith *et al.* (2003) also showed that delays between the onset of an outbreak and the implementation of control measures are detrimental.

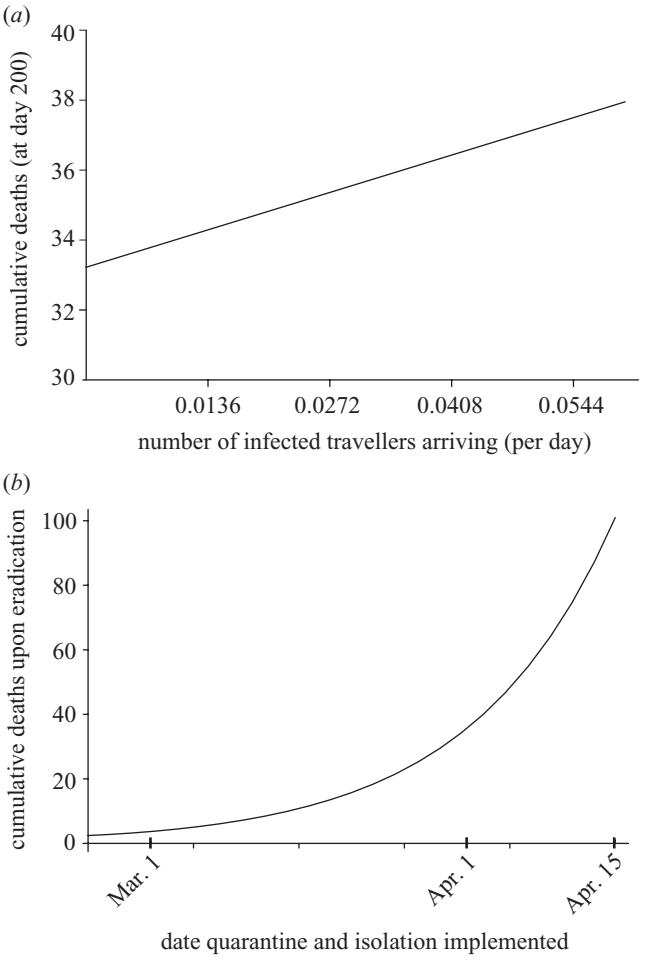

Figure 4. (*a*) The effect of screening at points of entry for GTA. Parameter values are those used for the GTA. Cumulative infections and deaths are shown at day 200, which is chosen arbitrarily because eradication is not possible when $p \neq 0$. (*b*) The effect of the timing of implementation of quarantine and isolation on the cumulative number of deaths once SARS has been eradicated for the GTA. As with other results, we assume that strict hygiene measures were put into place three weeks after implementation of quarantine and isolation.

Finally, we examine how variation in the effectiveness of quarantine and isolation affects the final outcome of the epidemic, by determining the cumulative deaths that would have occurred in the GTA, for different rates of quarantine and isolation (figure 5). It can be seen that both quarantine rate and isolation rate have a pronounced effect on the final outcome of the epidemic. Our simulations that mimic the actual outcome in the GTA are based on an assumption that the average time until isolation of a symptomatic individual is 2 days, whereas the average time until the quarantine of an asymptomatic individual is 10 days (the white dotted lines in figure 5). For these values, it can be seen that alterations in the rate of isolation have a stronger effect than alterations in the rate of quarantine. If the isolation rate were any lower, however (for example, an average of 10 days before isolation) then alterations in the rate of quarantine can have an equivalent or even greater effect than alterations in the rate of isolation. These results point to the conclusion that both isolation and quarantine are important, and quite effective, means for controlling

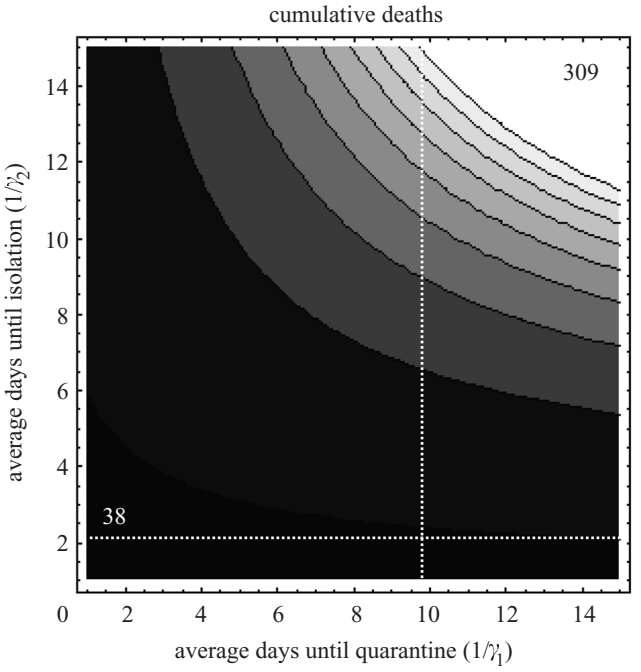

Figure 5. Contour plot of the cumulative number of deaths at day 200 (chosen arbitrarily) as a function of the amount of time asymptomatic and symptomatic individuals remain in the population prior to quarantine or isolation (i.e. the reciprocal of the quarantine and isolation rates, γ_1 and γ_2) for the GTA. White dotted lines are the parameter values used when fitting the model to data.

the spread of SARS. They also illustrate, however that if limited resources are available for investment in these two strategies, then investing all resources in one of them only can yield a better outcome than investing partly in both. This is reflected by the fact that a decrease in either time to quarantine or time to isolation yields the greatest reduction in cumulative deaths when the control parameter in question (i.e. time to quarantine or time to isolation) is already small relative to the other control parameter (figure 5).

4. DISCUSSION

Unlike influenza, which is also transmitted through droplets of respiratory secretion (and easily spread non-nosocomially), the transmission of SARS has largely been nosocomial in every region, despite the evidence for airborne transmission of the disease under some environmental conditions. The key factors contributing to the rapid global spread of SARS include a flurry of nosocomial infections during the isolation of SARS patients despite pre-SARS standard infection control practices, rapid pace of international travel, and the virtual absence of any prior immunity. In keeping with these facts, we develop a relatively simple compartmental deterministic model to study the transmission dynamics and control of SARS. The model incorporates many essential elements of SARS transmission but not the heterogeneity of contacts (caused by spatial considerations and age structure) that are also known to influence disease spread to some extent. The model reasonably mimics the outbreaks in geographically distinct regions, supporting the notion that simple models can be used to provide insights into the dynamics and control of an epidemic in progress.

Our model predicts that isolation of individuals with symptoms of SARS, under stringent hygiene precautions, can lead to effective control in a community and may even eradicate the disease (where all infected components of the model become zero) provided there are no undetected new admissions of SARS-infected individuals. SARS eradication therefore hinges upon (i) implementation of optimal isolation of SARS patients under stringent hygiene precautions; (ii) the ability to detect infected individuals at points of entry; and (iii) the lack of new SARS-CoV transmission to humans from animal sources. The first measure raises the value of γ_2 while decreasing ε_J to values close to zero, thereby halting the spread of SARS during an outbreak. The last two measures, by contrast, reduce the recruitment parameter p to values close to zero.

The study shows that the size and duration of an outbreak can be greatly influenced by the timely implementation of the isolation programme. Although a delay in the implementation of the isolation programme indiscriminately endangers the population at large, its timely implementation with stringent hygiene precautions alone is sufficient to control SARS in a community in the absence of a preventive vaccine or prophylactic drug (figure 4*b*).

The success of an isolation programme is subject to the implementation of stringent hygiene precautions (modelled here by ε_J) to prevent the incidence of nosocomial transmission of SARS to health-care workers and other caregivers. In GTA, for instance, we estimate a value of $\varepsilon_J \approx 0.1927$ to be the critical threshold below which $\mathscr{R}_c$ becomes less than unity, leading to effective control of SARS. By contrast, an approximate twofold rise in the threshold value (to $\varepsilon_J = 0.36$), under the same conditions, increases $\mathscr{R}_c$ to 1.7, resulting in new nosocomial infections during the isolation programme. A secondary deleterious effect of nosocomial infections may result from a reduction in the availability of health-care workers, which has not been considered in the current study.

During the early stages of the outbreak, the isolation of SARS patients in hospitals led to numerous new cases of SARS infection despite the implementation of pre-SARS standard infection control practices that have proven highly effective in controlling the spread of other microbial diseases. As a result, every affected region recorded many more new cases of infection within hospitals than outside hospitals and the new cases were largely confined to health-care workers and other caregivers. The standard pre-SARS infection control practices, although sufficient in controlling other infectious diseases, were unable to control SARS, presumably because of a lack of knowledge of the mode and transmissibility of SARS-CoV and the absence of anti-SARS immunity in the population. A major revision of the infection control guidelines and strict adherence to these guidelines (jointly referred to in this study as stringent hygiene precautions) were necessary to control the incidence of nosocomial infections and eventually control SARS. Timely isolation of SARS patients and the use of stringent hygiene precautions is therefore pivotal to the control of future outbreaks.

A comparison of the relative impact of quarantine of individuals suspected to be exposed to SARS-CoV and of isolation of those diagnosed with the actual disease in GTA reveals that the latter is noticeably more beneficial than the

former provided the isolation programme is implemented rapidly (for example, 2 days after the diagnosis of the index case; see figure 5). Under such conditions, mass quarantine of individuals suspected to be exposed to SARS-CoV offers only marginal benefit. However, if the isolation is delayed (for example, starting 10 days after the initial outbreak), then quarantine of individuals exposed to SARS-CoV becomes equally, or more, important than isolation. Thus, an indirect benefit of timely implementation of isolation is the reduction of economic burden associated with the mass quarantine programme that removes otherwise healthy individuals away from work and other economic activities. The study further shows that timely implementation of optimal isolation (or quarantine, in the case of a delayed isolation programme) would be a far more effective strategy than implementation of sub-optimal isolation and quarantine combined. It follows, therefore, that in cases where health-care resources are limited, investments in optimal isolation of SARS patients *or* quarantine (as the situation requires) would yield the most desirable outcome, as opposed to investing partly in both (figure 5). Because the isolation of symptomatic individuals may be relatively easier to implement than quarantining asymptomatic individuals (owing to the absence of efficient screening tests), allocating resources to optimal isolation at the expense of quarantine might be more desirable (figures 3 and 5).

Owing to the ever-increasing pace of inter-regional travel, the model incorporates a parameter p to account for undetected entry of SARS-infected individuals into a community. A rapid (on-the-spot) screening test capable of detecting such individuals at points of entry, in combination with an optimal isolation programme, might prove useful in preventing outbreaks of SARS. However, once an outbreak is underway, our study shows that even the most effective screening will probably have only a marginal effect on the size of the outbreak (figure 4*a*). By contrast, the eradication of SARS would require the implementation of a reliable and rapid screening test at the entry points in conjunction with optimal isolation, with or without quarantine, provided there is no perinatal or animal-to-human transmission of SARS-CoV.

Since there have been no reported cases of perinatal transmission of SARS, the prospects of global eradication of SARS appear promising provided there are no new cases of animal-to-human transmission of SARS-CoV (thus, effectively rendering $p = 0$ globally). If animal-to-human transmission of SARS-CoV was a one-time event, rather than an ongoing or periodic phenomenon, then the global eradication of SARS is feasible through public health control measures comprising isolation of symptomatic individuals under stringent hygiene precautions, with or without quarantine of asymptomatic individuals. The civet cat has been identified as one of the animal hosts for SARS-CoV-like viruses, possibly including the immediate ancestor of SARS-CoV (Enserink 2003). The factors associated with the evolution and transmission of these viruses to humans remain unknown. Understanding them is fundamentally important for the control and eradication of SARS in communities where interaction between animal hosts of SARS-CoV (or SARS-CoV-like viruses) and humans occurs normally. The urgency of such understanding is highlighted by the re-emergence of a few new cases of SARS in December 2003 in southern China (CDC 2004; WHO 2004*a*). At the present time, neither an epidemiological link between these cases nor their source of exposure has been established. Thus, although we have the ability to control SARS epidemics, its global eradication remains uncertain.

In the wake of an ever-increasing number of travellers between, and within, nations and the emergence of novel infectious diseases, the ability to rapidly contain such diseases in the absence of proven vaccine, diagnostic test, therapy or prior immunity in the population is a challenging but vital public health goal. The current public health threat posed by avian influenza calls for investigation to determine whether or not other novel respiratory diseases can be effectively controlled by basic public health measures such as isolation and quarantine. Owing to possible similarities in the modes of transmission and the absence of prior immunity between SARS and avian influenza viruses, avian influenza may conceivably exhibit transmission dynamics and requirements for control similar to SARS.

This project was supported in part by the Mathematics of Information Technology and Complex Systems (MITACS) and Natural Sciences and Engineering Research Council (NSERC) of Canada, and by the Canada Research Chair Program (T.D. and J.W.). The authors are grateful to Ms Karen Choo (Queen's University, Canada) for data collection, to Dr Ping Yan (Health Canada) for sharing preprints and to Dr Xinxin Lu (Beijing Tongren Hospital) for input on the SARS epidemic in Beijing. They are grateful to the reviewers for their comments, which led to significant improvement in the model and the clarity of the manuscript.

REFERENCES

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans*. London: Oxford University Press.

Booth, C. M. (and 20 others) 2003 Clinical features and short-term outcomes of 144 patients with SARS in the Greater Toronto area. *JAMA* **289**, 1–9.

Brauer, F. & van den Driessche, P. 2001 Models for the transmission of disease with immigration of infectives. *Math. Biosci.* **171**, 143–154.

CDC 2004 Recent SARS cases in China. CDC health update. See http://www.cdc.gov/ncidod/sars/han/sars-fluhanjan1404.htm. (Accessed May 2004.)

Chowell, G., Fenimore, P. W., Castillo-Garsow, M. A. & Castillo-Chavez, C. 2003 SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *J. Theor. Biol.* **224**, 1–8.

Diekmann, O. & Heesterbeek, J. A. P. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. New York: Wiley.

Donnelly, C. A. (and 18 others 2003 Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **361**, 1761–1766.

Drosten, C. (and 24 others) 2003 Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New Engl. J. Med.* **348**, 1967–1976.

Dwosh, H. A., Hong, H. H., Austgarden, D., Herman, S. & Schabas, R. 2003 Identification and containment of an outbreak of SARS in a community hospital. *CMAJ* **168**, 1415–1420.

Enserink, M. 2003 Infectious diseases. Clues to the animal origin of SARS. *Science* **300**, 1351.

Health Canada 2003*a* Information on severe acute respiratory syndrome (SARS). See www.hc-sc.gc.ca/english/protection/warnings/sars/update82.html, 17 June 2003.

Health Canada 2003*b* Learning from SARS—renewal of public health in Canada. See http://www.hc-sc.gc.ca/english/protection/warnings/sars/learning.html. (Accessed May 2004.)

Hethcote, H. W. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653.

Hsieh, Y. H., Chen, C. W. S. & Hsu, S.-B. 2004 SARS outbreak, Taiwan, 2003. *Emerg. Infect. Dis.* **10**, 201–206.

Ksiazek, T. G. (and 25 others) 2003 A novel coronavirus associated with severe acute respiratory syndrome. *New Engl. J. Med.* **348**, 1953–1966.

Lee, N. (and 13 others) 2003 A major outbreak of severe acute respiratory syndrome in Hong Kong. *New Engl. J. Med.* **348**, 1986–1994.

Lingappa, J. R., McDonald, L. C., Simone, P. & Parashar, U. D. 2004 Wresting SARS from uncertainty. *Emerg. Infect. Dis.* **10**, 167–170.

Lipsitch, M. (and 11 others) 2003 Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970.

Lloyd-Smith, J. O., Galvani, A. P. & Getz, W. M. 2003 Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. *Proc. R. Soc. Lond.* B **270**, 1979–1989. (doi:10.1098/rspb.2003.2481)

Marra, M. A. (and 57 others) 2003 The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.

Ministry of Health, China 2003 Available at www.moh.gov.cn/zhgl/yqfb/index.htm.

MMWR 2003 Cluster of severe acute respiratory syndrome cases among protected health-care workers—Toronto, Canada. April 2003, **52**, 433–436. (see http.//www.cdc.gov/mmwr/preview/mmwrhtml/mm5219a1.htm.)

Peiris, J. S. M. (and 15 others) 2003 Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1767–1772.

Poutanen, S. M. (and 19 others) 2003 Identification of severe acute respiratory syndrome in Canada. *New Engl. J. Med.* **348**, 1995–2005.

Riley, S. (and 19 others) 2003 Transmission dynamics of etiological agent of SARS in Hong Kong: the impact of public health interventions. *Science* **300**, 1961–1966.

Rota, P. A. (and 34 others) 2003 Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Sciencep* **300**, 1394–1399.

Tsang, K. W. (and 15 others) 2003 A cluster of cases of severe acute respiratory syndrome in Hong Kong. *New Engl. J. Med.* **348**, 1977–1985.

van den Driessche, P. & Watmough, J. 2002 Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180**, 29–48.

Wang, W. & Ruan, S. 2004 Simulating the SARS outbreak in Beijing with limited data. *J. Theor. Biol.* **227**, 369–379.

Webb, G. F., Blaser, M. J., Zhu, H., Ardal, S. & Wu, J. 2004 Critical role of nosocomial transmission in the Toronto SARS outbreak. *Math. Biosci. Engng* **1**, 1–13.

WHO 2003*a* Cumulative number of reported probable cases of SARS. See www.who.int/csr/sars/country/en. (Accessed May 2004.)

WHO 2003*b* WHO issues a global alert about cases of atypical pneumonia. See www.who.int/sarsarchive/2003_03_12/en, 12 March 2003.

WHO 2003*c* Coronavirus never before seen in humans is the cause of SARS. See www.who.int/mediacentre/release/2003/pr31/en, 16 April 2003.

WHO 2003*d* Case definitions for surveillance of severe acute respiratory syndrome (SARS). See www.who.int/csr/sars/casedefinition/en. (Accessed May 2004.)

WHO 2004*a* New case of laboratory-confirmed SARS in Guangdong, China—update 5. See http://www.who.int/csr/don/2004_01_31/en/. (Accessed May 2004.)

WHO 2004*b* China's latest SARS outbreak has been contained, but biosafety concerns remain—update 7. See http://www.who.int/csr/don/2004_05_18a/en/. (Accessed May 2004.)

Yu, I. T. S. (and 7 others) 2004 Evidence of airborne transmission of SARS virus. *New Engl. J. Med.* **350**, 1731–1739.

Zhou, Y., Ma, Z. & Brauer, F. 2004 A discrete epidemic model for SARS transmission and control in China. *Math. Comput. Model.* (In the press.)

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

Visit www.journals.royalsoc.ac.uk and navigate through to this article in *Proceedings*: Biological Sciences to see the accompanying electronic appendix.