THE ROYAL SOCIETY

Published online 15 June 2004

Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic

Roy M. Anderson[1*], Christophe Fraser[1], Azra C. Ghani[1], Christl A. Donnelly[1], Steven Riley[1], Neil M. Ferguson[1], Gabriel M. Leung[2], T. H. Lam[2] and Anthony J. Hedley[2]

[1]*Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK*

[2]*21 Sassoon Road, Faculty of Medicine Building, University of Hong Kong, Pokfulam, Hong Kong, China*

This paper reviews current understanding of the epidemiology, transmission dynamics and control of the aetiological agent of severe acute respiratory syndrome (SARS). We present analyses of data on key parameters and distributions and discuss the processes of data capture, analysis and public health policy formulation during the SARS epidemic are discussed. The low transmissibility of the virus, combined with the onset of peak infectiousness following the onset of clinical symptoms of disease, transpired to make simple public health measures, such as isolating patients and quarantining their contacts, very effective in the control of the SARS epidemic. We conclude that we were lucky this time round, but may not be so with the next epidemic outbreak of a novel aetiological agent. We present analyses that help to further understanding of what intervention measures are likely to work best with infectious agents of defined biological and epidemiological properties. These lessons learnt from the SARS experience are presented in an epidemiological and public health context.

Keywords: SARS; epidemiology; mathematical models

1. INTRODUCTION

The re-emergence of the viral aetiological agent of SARS in China at the end of 2003 (Paterson 2004), following the epidemic earlier in the year affecting many countries, rang alarm bells in the WHO and elsewhere. Thankfully, prompt action by the Chinese authorities in the isolation of suspect cases and in instigating contact tracing and quarantine measures served to effectively contain the virus. By the end of February 2004, only three confirmed cases and one probable case had been reported with no chains of onward transmission identified (WHO 2004). Where the virus re-emerged from remains uncertain, but the prime suspect as an animal reservoir host remains the civet cat (*Paguma larvata*), and the foci for spread from this host to humans seem to be the animal markets in China, especially those in Guangzhou, in the Guangdong province (Webster 2004). Guangdong province is where the major 2003 outbreak originated that infected more than 8000 people and killed 774 (He *et al.* (2004)).

One measure introduced by the Chinese authorities, following the re-emergence of SARS late in 2003, was a large cull of the civet cat populations in the animal markets and breeding farms, estimated to have involved the removal of over 10 000 animals (Watts 2004). Molecular epidemiological studies suggest that the coronavirus in humans is very closely related to a strain found in civet cats, but differs slightly from the one that caused such devastation earlier in 2003. A clear priority is further surveillance of animals in settings where the human virus spread extensively so as to better understand the origins of the epidemic in humans and the role of animal reservoirs.

In the global response to the 2003 SARS epidemic, orchestrated by the WHO, there were five priority tasks at the start of the outbreak. These were the identification of the aetiological agent of the new disease SARS, the development of diagnostic tests to detect the virus, the development and assessment of treatment protocols to reduce morbidity and mortality, estimation of the key epidemiological parameters that affected spread and persistence, and the formulation and implementation of appropriate public health interventions. Most of these tasks were completed rapidly, including the identification of the viral aetiological agent (Drosten *et al.* 2003; Ksiazek *et al.* 2003; Peiris *et al.* 2003*b*), the delineation of the full genome sequence of the virus (Marra *et al.* 2003), the evaluation of key epidemiological parameters (Donnelly *et al.* 2003) and the impact of different interventions (Lipsitch *et al.* 2003; Riley *et al.* 2003). Problems remain, however, in the areas of quick and accurate diagnosis and effective therapeutic interventions. Detection of the presence of the virus soon after onset of clinical symptoms remains difficult, with the most sophisticated RT–PCR tests still providing only limited sensitivity following the onset of fever. Sensitivity can be improved if samples from the lower respiratory tract can be obtained in the second week following the onset of clinical symptoms. Serological

* Author for correspondence (roy.anderson@imperial.ac.uk).

One contribution of 15 to a Discussion Meeting Issue 'Emerging infections: what have we learnt from SARS?'.

tests are highly sensitive, but after only 21 days following clinical onset (Li *et al.* 2003). With respect to clinical interventions to reduce morbidity and mortality, as yet too little has been done to date to merge clinical patient databases from different settings and countries so as to provide sufficient power to perform rigorous evaluations, taking account of the many confounding factors such as age and various co-morbidities. In any such analyses, a note must be taken of differences among countries in the way clinical symptoms were recorded.

This paper reviews current understanding of the key epidemiological determinants of the transmission dynamics of SARS-CoV, and evaluates what interventions worked best in different settings, using mathematical models to give structure and clarity to the analysis. We also consider gaps in current knowledge and priorities for future research to improve predictions of observed pattern and the evaluation of the impact of interventions for novel infectious agents.

2. DATA NEEDS FOR EPIDEMIOLOGICAL STUDY

The study of the transmission dynamics of an infectious agent is typically based on simple or complex mathematical frameworks (Anderson & May 1991). The goals in model formulation and analysis can be many and varied. They include delineating what needs to be measured to better understand observed pattern, identifying the key determinants of this pattern, and evaluating how different interventions introduced at various stages of the epidemic influence the future incidence of infection and associated disease. For a new pathogen the first of these goals is of central importance both in guiding data collection and analysis, and in the formulation of policies to protect public health. The following sections describe the key steps in data collection and analysis, with the aim of providing robust estimates of the key parameters (and their distributions) that influence transmission and control. Much between-patient variability is often associated with the parameters that influence the typical course of infection in a patient and transmission. Thus, the full distributions must be estimated so that variability in properties such as the incubation period of the disease, the infectious period and how infectiousness changes over time following infection, times from onset of clinical symptoms to isolation in a health care setting, and times from isolation to recovery or death are fully understood.

3. THE CONSTRUCTION OF A PATIENT DATABASE

The creation of a patient database that integrates sociodemographic detail with clinical information, such as treatment and outcome, and epidemiological data, such as contact tracing information and behavioural questionnaire data, is central to the real-time analysis and control of infectious disease outbreaks. Ideally, the electronic database should have one central point of control (such as in a government department or designated research centre), be Web-based with password protection for remote data entry, and be designed to act as a registry and monitoring system to inform policy formulation and allow analyses to be updated daily throughout the epidemic. All patients suspected of having contracted SARS within a country or administrative region should be entered into this database with a unique patient-identifying code. Appropriate measures should be taken to protect patient identity and to conform to data protection legislation. At a global level, such countrywide data should ideally be shared with the WHO, either in its entirety or in a pared down form which includes case incidence over time and basic epidemiological and clinical information, to inform their policy formulation and advisory notices.

Few countries achieved such real-time collection and synthesis of information during the SARS outbreak. Arguably, the best example of good practice occurred in the Hong Kong Special Administrative Region of China, where the Health, Welfare and Food Bureau of the government (in collaboration with university-based public health professionals), created a system called SARSID (SARS Integrated Database) to collect and collate case information. Even in this setting, however, problems were encountered in linking case information with clinical treatment data held in health care settings, and with the results of contact tracing which were held in a separate database managed by the police service. In many countries, the appropriate databases and associated analyses were fully assembled and completed only after the end of the epidemic. As a consequence of the SARS experience (and experiences with other rapidly developing infectious disease outbreaks (see Ferguson *et al.* 2001*a*,*b*)), the development of appropriate software and the training of personnel to use it should be a priority for all government health departments.

4. CASE DEFINITION AND CLINICAL SYMPTOMS

The case definition of SARS has changed since the emergence of the 2002–20003 epidemic. However, the following definition developed by the WHO was used throughout the main period of spread in 2003 (http://www.who.int/csr/sars/casedefinition/en/). Two criteria were used in this definition. The first definition included a person presenting after 1 November 2002 with history of high fever (greater than 38 °C) and cough or breathing difficulty, and one or more exposures during the 10 days prior to onset of symptoms. Exposures are defined as close contact (having cared for, lived with, or had direct contact with respiratory secretions or body fluids) with a person who is a suspect or probable case of SARS; history of travel to an area with recent local transmission of SARS; or residing in an area with recent local transmission of SARS. The second definition included a person with an unexplained acute respiratory illness resulting in death after 1 November 2002, but on whom no autopsy had been performed and who had one or more exposures during the 10 days prior to onset of symptoms. A probable case was defined by the WHO to have three criteria as follows: (i) a suspect case with radiographic evidence of infiltrates consistent with pneumonia or RDS on chest X-ray; (ii) a suspect case of SARS that is positive for SARS-CoV by one or more assays; and (ii) a suspect case with autopsy findings consistent with the pathology of RDS without an identifiable cause.

Because all SARS diagnoses were based on exclusion of other possible causes (e.g. influenza), a case could be excluded if an alternative diagnosis was made at a later

Table 1. Incubation period of SARS.

country/region	mean (days)	number of cases	reference
Hong Kong	4.6	81	Leung *et al.* (2004*a*)
mainland China	4	70	WHO (2003)
Singapore	5.3	46	WHO (2003)
Canada	5	42	Varia *et al.* (2003)

stage following admission to hospital. Throughout the first epidemic and over the months since its termination on 5 July 2003, many cases have been reclassified on the basis of information from virological assays or serological tests based on sera drawn from patients after cessation of clinical symptoms.

Clinical symptoms at admission to hospital, and during hospital stays for confirmed SARS cases (based on virological confirmation) have been recorded in various studies (see Donnelly *et al.* 2003; Tsang *et al.* 2003). At admission, high fever, malaise, cough and headache seem to be the most common symptoms. For confirmed SARS cases, high and prolonged fever and diarrhoea are typical symptoms. Virus can be isolated from sputum, urine and faeces during the mid to late clinical phase of symptoms (Peiris *et al.* 2003*a*).

5. INCUBATION PERIOD

The incubation period is defined as the time from infection to onset of clinical symptoms of disease. This duration is often influenced by factors such as the infecting dose of the virus, the host genetic background, the route of exposure and the age of the patient. The determination of the form of this distribution and its summary statistics at the start of an epidemic of a novel infectious agent are of great importance, given their significance to the duration of quarantine and to contact tracing. The average incubation period also influences the time-scale of the development of the epidemic, by its influence on the rate at which secondary cases are generated.

Infection events cannot be observed, but data on patients with short and well-defined periods of exposure to known SARS cases can be used to estimate the distribution of the incubation period. Estimation of the incubation period is based on sets of patients with a short exposure period of known date to a suspect SARS case. This information may be collated via contact tracing or from questionnaires. The estimates for SARS are based on *ca*. 200 cases from five regions that experienced moderate to severe epidemics (Hong Kong, mainland China (Beijing and Guangdong), Taiwan, Singapore and Canada). The data are summarized in table 1. Donnelly *et al.* (2003) report a mean of 6.37 days (subsequently revised to 4.6 days after the end of the epidemic), with *ca.* 95% of all estimates lying between 2 and 12 days. An observed distribution based on 70 cases in Guangdong and Beijing is shown in figure 1*a*. Figure 1*b* shows the best-fitting gamma distribution to 86 cases with short exposure intervals from Hong Kong.

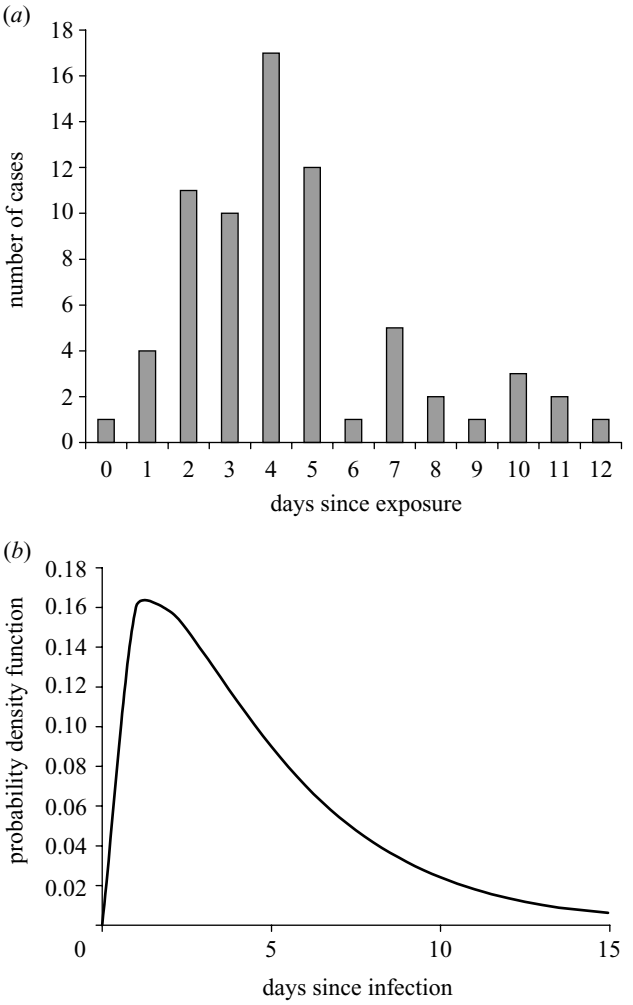


Figure 1. Incubation period distribution of SARS in (*a*) 70 cases in Quangdong (mean, 4.5 days) and (*b*) 86 cases in Hong Kong (mean, 4.6 days), where exposure was known to have occurred over a short interval of time (Donnelly *et al.* 2003).

6. TIME FROM ONSET OF CLINICAL SYMPTOMS TO ADMISSION TO HOSPITAL

Very early on in the SARS epidemic it was understood that reducing the time from onset of clinical symptoms to admission to hospital and subsequent isolation was an important measure to reduce the net rate of transmission within a community or country (Donnelly *et al.* 2003). Most countries compiled statistics on this distribution, with Hong Kong, Taiwan and mainland China having access to such information as the epidemic progressed. Onset and admission times are both observable events.

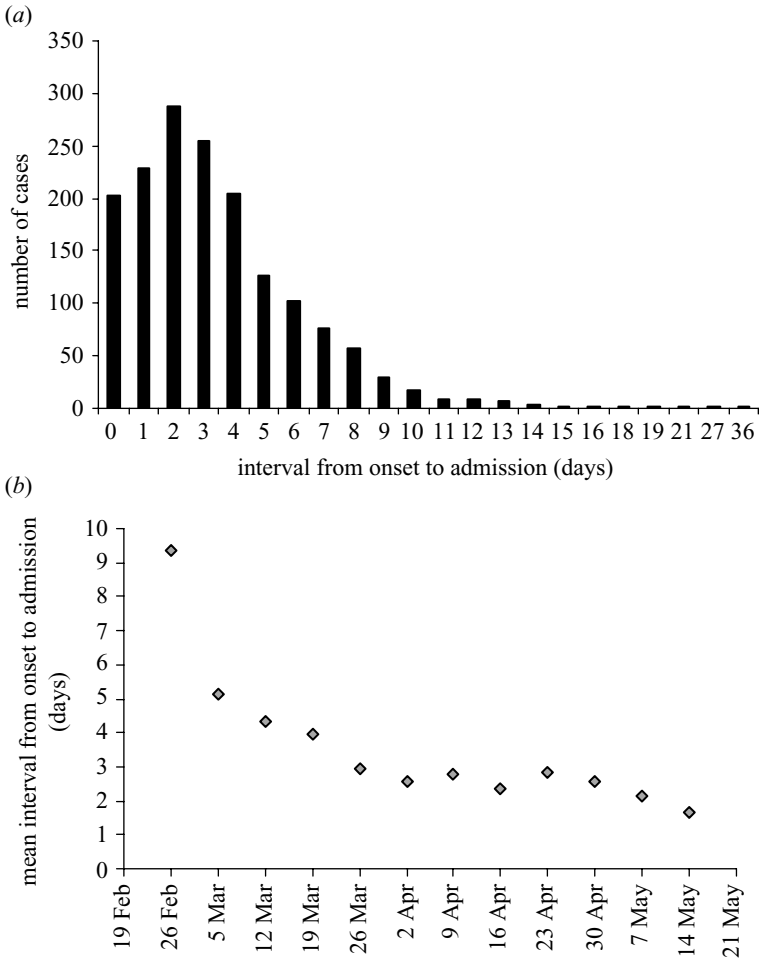


Figure 2. Distribution of interval from onset of clinical symptoms to admission to hospital in Hong Kong. (*a*) Overall distribution; (*b*) mean interval plotted by week of onset of symptoms.

However, in the early phase of a new epidemic, analyses must allow for censoring as a result of incomplete observation. If censoring is not taken into account, the distribution will be biased towards short onset-to-admission times, because patients are only eligible to be included in the hospital-based database on admission to hospital. Patients with recent onsets and long onset-to-admission times are less likely to have been admitted to hospital and thus to be included in any analysis. One such distribution, compiled following the cessation of the epidemic to eliminate the problem of censoring, is recorded in figure 2*a*. As a consequence of public health announcements using the press and media, many regions managed to encourage individuals with symptoms of severe respiratory tract infection to report rapidly to hospitals. For example, in Hong Kong, the mean period shortened greatly over the course of the epidemic (figure 2*b*).

7. TIME FROM ADMISSION TO HOSPITAL TO DISCHARGE

The duration of stay in hospital for those who recovered was typically long, with a mean in excess of 25 days in Hong Kong. Longer durations of stay were associated with older patient age and the presence of co-morbidities. The duration of hospital stay is an important statistic for the effective management of a SARS outbreak, because it describes one aspect of hospital resource use. Ideally, while admitted, patients should be cared for in isolation facilities with the application of rigorous infection control measures.

8. TIME FROM ADMISSION TO DEATH

SARS is a very pathogenic disease with a high CFR. The times from admission to death are again of importance to health care planners in terms of resource use within the hospital setting. In the Hong Kong setting the mean time was 36 days but with a high variance. Age and other co-morbidities greatly increased the mortality rate (Leung *et al.* 2004*a*). This long period generates difficulties in the estimation of the CFR unless methods account for the censoring of data before outcomes are fully known (see figure 3).

A summary of the key distributions derived from data from Hong Kong is presented in table 2.

9. CASE FATALITY RATES

The term CFR is widely used to describe the proportion of those who acquire an infection that eventually die from the disease induced by the aetiological agent. The CFR is not strictly a rate: it is a simple proportion or percentage. The first published study of case fatality in a sample of

Table 2. Summary of the means of four key distributions based on data from Hong Kong (see Donnelly *et al.* 2003).

T_1: exposure to onset (incubation period): mean, 4.6 days, variance 15.9 days^2. Ninety-five per cent of infected individuals onset within 12.5 days. Analysis based on interval-censored exposure data
T_2: onset to admission: reflects rapidity of diagnosis and hence isolation. Decreased from an average of 4.9 days early in epidemic to less than 2 days by mid-May
admission to death (for patients who died): mean of 35.9 days
admission to discharge (for patients who recovered): mean of 23.5 days

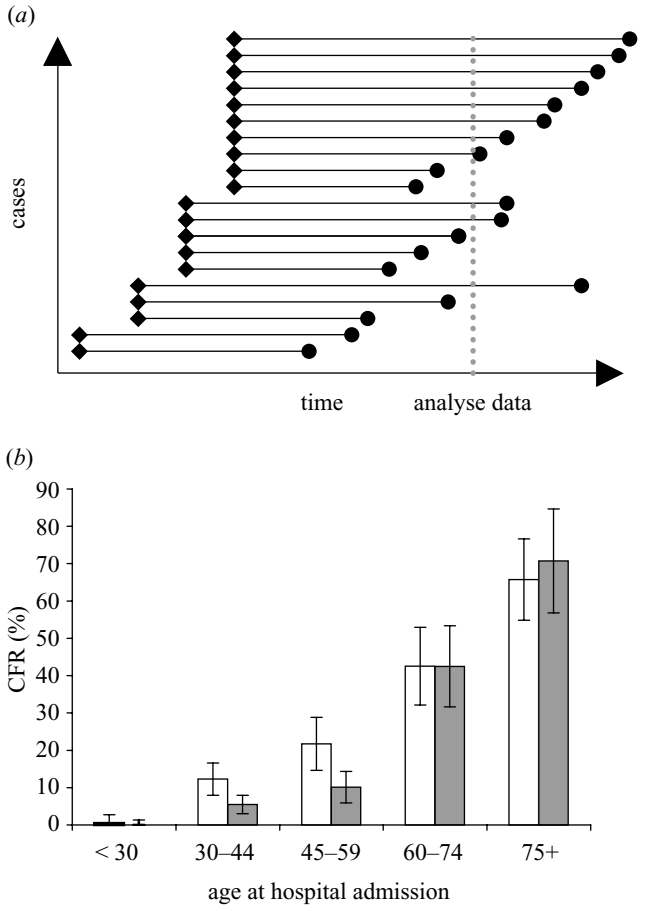


Figure 3. (*a*) A schematic representation of right-hand tail censoring, where estimation underestimates the magnitude of mortality unless account is taken of cases in which the outcome is not known at a defined point in time (t_m). This problem typically arises in the early stage of an epidemic when the duration of stay in hospital is long before either recovery or death occurs (Donnelly *et al.* 2003). (*b*) The final SARS CFR in Hong Kong stratified by gender (white bars, males; grey bars, females) and age at hospital admission.

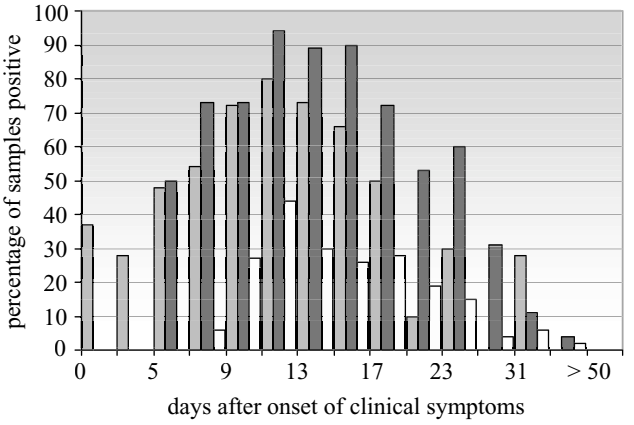


Figure 4. Studies of viral shedding in SARS patients on various days following the onset of clinical symptoms, in stool (dark-grey bars), urine (white bars) and naso-pharangeal aspirate (light-grey bars) (Peiris *et al.* 2003*a*).

SARS patients for which outcome was known, or adjusted for by appropriate statistical methods for censoring of the data (figure 3*a*), was that of Donnelly *et al.* (2003). Using a modified Kaplan–Meier-like non-parametric method this study gave estimates of 6.8% for patients younger than 60 years and 55% for patients older than 60 years. At the end of the first epidemic of SARS-CoV it is possible to examine the mortality associated with the disease in more detail. Overall, a WHO summary suggests a global average of *ca*. 15%. However, this figure hides much variation: there was little mortality in the young and high levels in the elderly. Multivariate analyses identify age effects and the presence of co-morbidities (such as pre-existing heart or respiratory tract disease) as the most important determinants of the outcome (Leung *et al.* 2004*a*). Figure 3*b* shows age- and gender-specific estimates of the CFR from 1628 patients from Hong Kong. Similar patterns have been recorded in Singapore and Taiwan. Lower rates were reported from mainland China (Beijing and Guangdong) but no detail is available as yet concerning laboratory confirmation of reported SARS cases and how these fatality rates relate to patient age and co-morbidity.

In most studies laboratory testing has now become an integral part of mortality assessment to ensure that a reported SARS case, based on clinical features, is confirmed by either virus isolation or serology following recovery.

10. INFECTIOUS PERIOD

A key determinant of the pattern of spread of an infectious agent is the infectiousness distribution, both before and after the onset of clinical symptoms. Two approaches to estimation of the time course of infectiousness within a typical patient are the study of secondary case generation and the measurement of viral load given the assumption that this is proportional to infectiousness to contacts. For SARS-CoV the efficiency of transmission to close contacts appears to be greatest from patients with overt clinical symptoms, usually during the second week following onset. A study reported at the WHO meeting in Geneva in May 2003 but not yet published based on cases in Singapore suggests that peak infectiousness as judged by the generation of secondary cases is *ca.* days 7 to 8 following onset of clinical symptoms. The main conclusion of this

infection	R_0	percentage transmission that must be blocked to prevent epidemic
SARS	2–3	50–67%
HIV	2–5	50–80%
smallpox	5–10	80–90%
pandemic influenza	5–25	80–97.5%

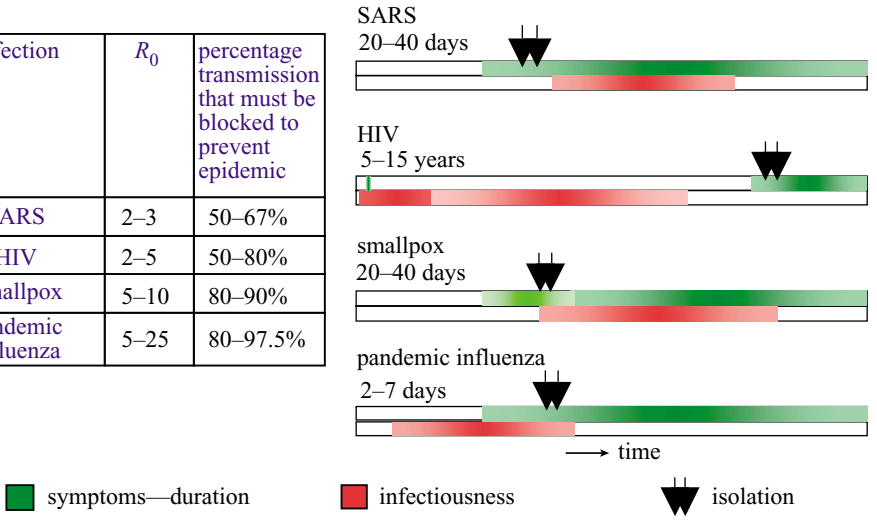


Figure 5. Schematic representation of the incubation and infectious distributions of SARS and other infections, plus when isolation of patients might occur on the basis of onset of clinical symptoms that result in diagnosis (Fraser *et al.* 2004).

R_0 Basic reproduction number
average number of secondary cases generated by one primary case in a susceptible population

R_t Effective reproduction number
number of infections caused by each new case occurring at time t. For illustrative purposes, we use a discrete generation model

- The key determinant of incidence and prevalence of infection is the basic reproductive number R_0
- Many factors determine its magnitude, including those that influence the typical course of infection in the patient and those that determine transmission between people

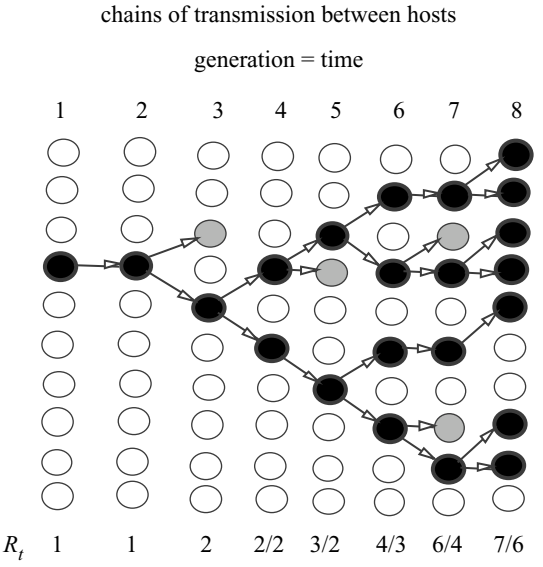


Figure 6. Diagrammatic representation of chains of transmission. The speed of spread is determined by the case reproductive number R_0.

analysis is in good agreement with a study by Peiris *et al.* (2003*a*) of changes over time in viral load (in nasopharyngeal aspirates and nose or throat swabs) following the onset of clinical symptoms of SARS-CoV in 329 patients from Hong Kong. A quantitative RT–PCR was used to determine viral shedding. Maximum virus excretion from the respiratory tract occurs on *ca.* day 10 of illness and declines thereafter to a low level at *ca.* day 23. Virus in stools seems to start later than in respiratory excretions, with a peak between days 12 to 14 and a slower decline thereafter. These results are summarized in figure 4. Virus can also be detected in urine, indicating wide organ involvement in pathogenesis (WHO 2003). Retrospective quantitative studies of viral shedding in patients following recovery and discharge from hospital, based on collected samples of respiratory tract excretions, faeces and urine, are currently underway in a number of countries. For the study of transmission dynamics, quantitative measurements are of great importance to define a distribution of infectiousness before and after onset of clinical symptoms. The viral shedding studies that are available to date suggest that transmission could occur via close contact involving respiratory tract excretions, and via faecal or urine contamination of surfaces. A schematic of the relationship between the average incubation and infectiousness periods is presented in figure 5.

11. ATYPICAL PRESENTATION OF SARS AND ASYMPTOMATIC INFECTION

Atypical presentation of SARS has been documented in a number of papers, where symptoms have included fever

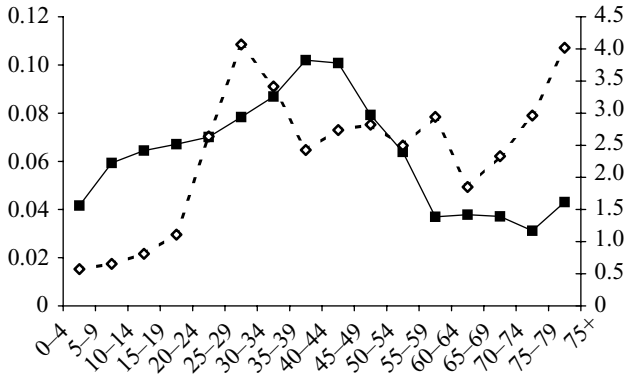
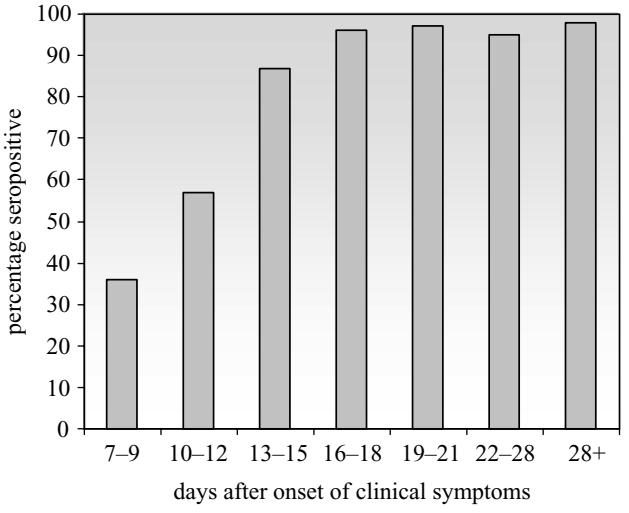


Figure 7. Age distribution of 1628 reported SARS cases and population age distribution in Hong Kong. Filled squares, proportion of population; open diamonds, cases per 10 000 population.



source	no. of samples	no. positive	% positive
clinics	777	10	1.3
hospitals	620	49	7.9
total	1397	59	4.2

Figure 8. Serological studies of SARS patients and contacts with SARS patients in Hong Kong hospital and clinic settings (Li *et al.* 2003). Seroconversion in most patients occurs some 16–18 days after the onset of clinical symptoms (graph), while contacts with such patients by hospital and clinic staff resulted in around a 4% seroconversion rate in individuals who were not diagnosed with SARS.

and diarrhoea, sometimes with bloody stools, but with no respiratory symptoms (e.g. Hsu *et al.* 2003; Chow *et al.* 2004). The incubation periods in such patients ranged from 3 to 8 days, and transmission on to contacts was observed, especially in hospital settings. Such observations might suggest that the recorded cases of SARS, based on clinical criteria, may underestimate the extent of transmission and hence estimates of the case reproductive number, R_0 (figure 6; Lipsitch *et al.* 2003; Riley *et al.* 2003). Further observations relating to the distribution of cases by age, matched with the known age distribution of given populations, revealed a deficit of cases in the young and an excess in the elderly. This pattern could arise from

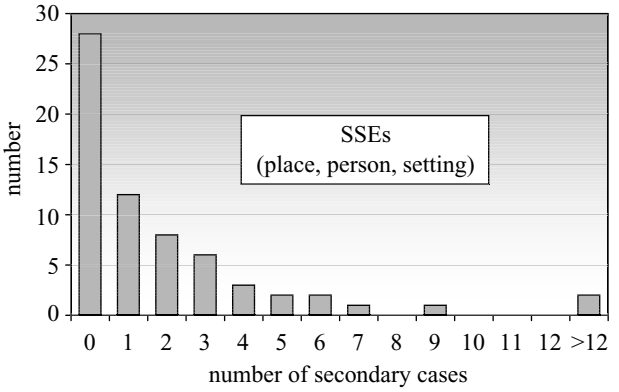


Figure 9. Schematic representation of the distribution of secondary case generation. For SARS, most index cases generated none or a few secondary cases (with an average of *ca.* 2–3), while a very few generated many.

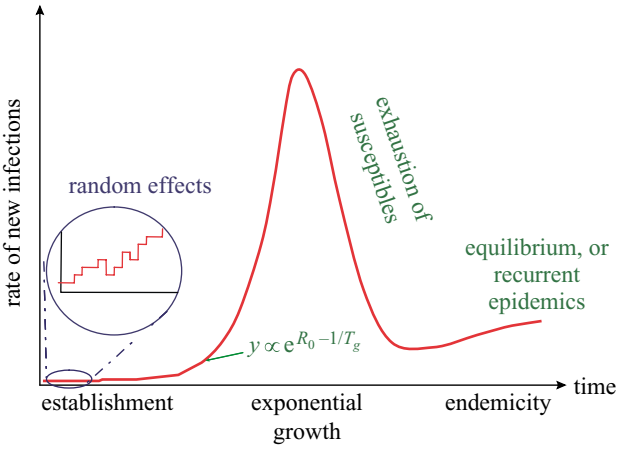


Figure 10. Schematic of the phases of epidemic growth after the introduction of an infectious disease into a virgin population. In the early phase of growth the rate of increase of cases is exponential with growth coefficient $(R_0 - 1)/T_g$, where R_0 is the case reproductive number and T_g is the average generation time (Anderson & May 1991).

frequent asymptomatic infection in the young and more severe infection in the elderly (figure 7).

After the end of the epidemic, a number of studies have been undertaken to assess the extent of transmission from index cases, where no formal identification of SARS was made in such contacts. Serological surveys, for example, have been carried out in Hong Kong using contact tracing data to identify known contacts of SARS cases, to assess the degree of seropositivity in contacts with no recorded symptoms of SARS. In one study of over 1000 serum samples taken from over 3000 contacts of recorded SARS cases, only 0.3% were positive for SARS-CoV immunoglobulin-G antibodies (Leung *et al.* 2004*b*). This result is reassuring, and goes some way to dispelling the doubt that recorded clinical cases of SARS reflect only a small proportion of transmission events. In hospital settings a higher fraction of contacts (who were largely hospital staff) were subsequently found to be seropositive (figure 8).

12. SUPER-SPREADING EVENTS

The epidemics of SARS-CoV in different countries with a moderate to large number of cases were characterized

by a few SSEs, where one case generated large numbers of secondary cases (figure 9). It is important to note that these events were most probably created by different combinations of person-related characteristics (e.g. high viral shedding) and environmental factors (e.g. contamination by fomites or close contact in a health care setting). The occurrence of such events created a distribution of secondary case generation with a high variance and concomitantly a long right-hand tail. Those in the tail of this distribution played a very important role in the emergence of the epidemic and its spread from country to country. One of these events occurred in a hotel in Hong Kong, where an ill traveller from Guangdong infected many other hotel residents, resulting in transmission to many countries from this single case. Others occurred in a hospital setting in Hong Kong, an air flight from Hong Kong to Beijing and in health care settings in Beijing, Singapore and Toronto (Shen *et al.* 2004). A detailed study of the event in a Beijing hospital revealed that one patient with 74 close contacts generated 33 secondary cases. These secondary cases generated a further 43 cases before this chain of transmission petered out (Shen *et al.* 2004).

Evidence from Guangdong province in China, during the very early stages of the global epidemic, suggests that most cases seem to have occurred in food handlers (individuals who kill, handle and sell animals and meat, and those who prepare and serve food) (Breiman *et al.* 2003; WHO 2003). Following this initial stage, the primary mode of person-to-person transmission appears to be via direct contact with respiratory droplets and exposure to fomites in settings where close contact occurred either in a household or a health care facility. In the total epidemic worldwide *ca.* 21% of recorded cases were in health care workers (WHO 2003). Other transmission events occurred in the general population (often unknown in nature) either in workplaces, taxis or aeroplanes. The localized nature of transmission in defined environments suggests an agent of low transmissibility that requires close contact, either with an ill patient or with a recently contaminated surface. There is no direct evidence of faecal–oral transmission but the high and persistent viral shedding in faeces suggests that this could be a route of significance. The exact route of transmission is always difficult to determine and the relative importance of respiratory, faecal and urine excretions remains unclear at present. For the purposes of the analysis of the transmission dynamics of the virus, the most important factors are the settings in which transmission occurs, the incubation and infectious periods, and the distribution of secondary case generation.

13. EPIDEMIC GROWTH

In regions with significant numbers of cases, such as Hong Kong, Taiwan, mainland China and Singapore, the typical pattern of epidemic growth was an initial period of stuttering chains of transmission, interspersed with one or two major SSEs. This was followed by a period of exponential growth slowing to reach a peak and then a period of steady decline with perhaps one or more SSEs leading to temporary resurgence, before the epidemic decayed with a cessation in chains of transmission (figure 10). The decay phases were of varying durations, depending on the efficacy of the interventions introduced in the different regions. The epidemic in mainland China had two phases, with initial spread in Guangdong province followed by a major epidemic in Beijing.

14. TRANSMISSION DYNAMICS

Mathematical models of the transmission dynamics of infectious agents are valuable tools in making assessments of both what needs to be measured to understand spread better and what interventions used alone or in combination are most likely to be effective. They provide a template within which to integrate epidemiological and biological data. In the early to mid stages of the SARS epidemic, a number of research groups formulated mathematical models of viral spread of varying complexity. At one end were the simple deterministic susceptible–infected–recovered frameworks (SIR) (Lipsitch *et al.* 2003), and at the other were more complex stochastic frameworks with more accurate description of disease progression and some representation of spatial structure, SSEs and mixing (Riley *et al.* 2003). Approaches of intermediate complexity were also adopted (Lloyd-Smith *et al.* 2003).

The chain of transmission events within an epidemic is an expanding one if each primary case, on average, generates more than one secondary case. The average number of secondary cases generated in a susceptible population is termed the basic reproductive number, R_0 (figure 6). As the epidemic develops, the effective reproductive number at time t, R_t, which describes the generation of secondary cases in a partly susceptible population, decreases and eventually settles to unity if the disease becomes endemic. The pattern of epidemic growth is governed by two factors: the number of secondary cases generated by one primary case at the start of the epidemic, R_0, and the average time taken for the secondary cases to be infected by a primary case, termed the generation time or serial interval, and denoted T_g (Anderson & May 1991). Essentially, R_0 determines how intensive a policy will need to be to control the epidemic, whereas both T_g and R_0 determine the time available to implement suitably intensive controls. For diseases that are highly infectious with short incubation periods, such as measles ($R_0 > 17$, $T_g < 11$ days), population-wide control requires the long-term reduction of the recruitment of susceptible people, through widespread childhood immunization. By contrast, for less infectious agents (R_0 *ca.* 2–10) that have longer incubation periods ($T_g > 20$ days), if an outbreak is detected in its earliest stages, there is sufficient time for localized control measures to be successfully applied (figure 5). One aim in model development for SARS spread in defined regions was therefore the estimation of both the basic reproductive number and the generation time to assess how difficult it might be to bring the epidemic under control.

Longitudinal monitoring of the magnitude of a further parameter, the effective reproductive number at time t, R_t, also provides crucial information on the success of the interventions that have been implemented to date. For an epidemic to be under control the magnitude of R_t must be less than unity: each primary case generates on average less than one secondary case such that the chains of transmission stutter to extinction.

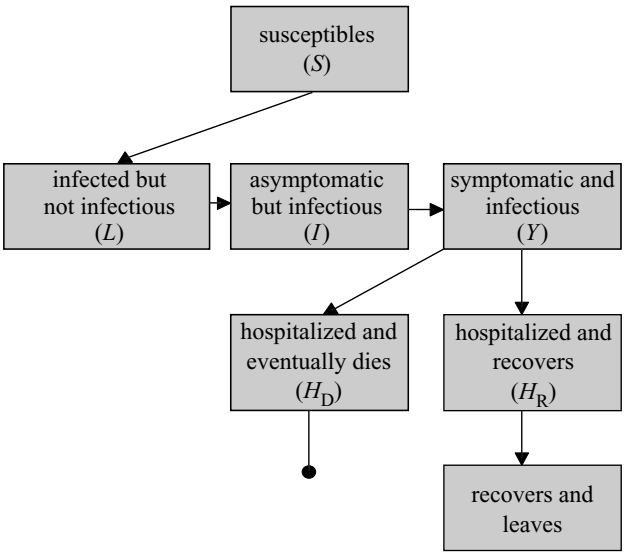


Figure 11. Compartmental framework for a stochastic model of the transmission dynamics of the viral aetiological agent of SARS (Riley *et al.* 2003).

What is the best method to estimate R_0 in an emerging crisis? One simple approach is to estimate the doubling time of the epidemic, t_d, in its early stages, taking account of the stochastic fluctuations that typically occur in the early stages of any epidemic's development (Anderson & May 1991). Estimation of the key parameter, R_0, can be achieved by fitting an exponential growth equation. For simple epidemics of a directly transmitted respiratory or gastro-intestinal pathogen, during early growth in a totally susceptible population, the doubling time t_d is related to the magnitude of R_0 by the simple expression

$$t_d = (\ln 2)D/(R_0 - 1),$$

where D is the average duration of the latent plus infectious periods (essentially the generation time T_g; figure 10). In the absence of knowledge of D, a range of assumptions will have to be made for its value, to get some idea of the magnitude of R_0 (Anderson & May 1991). In the early stages of the SARS epidemic, contact tracing data, especially that based on SSEs, suggested that the value of t_d was somewhere between one and a few weeks (Galvani *et al.* 2003). This method was used by Lipsitch *et al.* (2003) to estimate R_0 based on choosing a few time-points from several separate epidemics. The method was adjusted to allow for a skewed distribution of secondary infections, which amounted to treating SSEs as extreme points from a continuous distribution.

However, the approach based on doubling times is not very robust. Better non-parametric approaches to estimate R_0 directly from time-series can be based on either examining the ratio of incidence to prevalence, or on estimating the likelihood of any person having infected any other. Both approaches are based on the estimated distribution of generation times (Wallinga 2004).The latter approach can be complemented by contact tracing data. The disadvantages are that it does not disaggregate different contributions to R_0 and it is purely descriptive of the data. Another method is to fit a mathematical model of the transmission dynamics of the infectious disease. Although robust, this approach is very parametric, in the sense of being dependent on making a good choice for the framework of the model to fit the known biology and epidemiology.

The simplest models are based on the classification of the population into various disease states such as susceptible, infected but not yet infectious (latent), infectious, and recovered or dead. Equations for each state were constructed denoting rates of flow between the states. These rates may be represented as constants (with an exponential waiting time distribution for movement between states) or by distributed variables given availability of appropriate data (model structures with realistic distributions were used in Riley *et al.* (2003) and Lloyd-Smith *et al.* (2003); see figure 2). For SARS, the simplest structure is represented in figure 11, which pictures in a flow chart an additional class to those outlined above to reflect isolation or quarantine, such that a patient is infectious but unable to transmit. Such models may be stratified by factors such as age, spatial location or environmental setting (i.e. within a health care setting), and may be formulated within a deterministic or stochastic framework.

Population heterogeneity ideally needs to be accounted for, reflecting, for example, variation in infection risk with spatial location, variation in contact rates between groups, and between-case variability in infectiousness (see figure 9; Riley *et al.* 2003). Heterogeneity in R_t between cases appears to be particularly important for SARS because of the occurrence of SSEs. These are defined as rare events where, in a particular setting, an individual may generate many more than the average number of secondary cases. To what extent SSEs are simply extreme values in skewed distributions of infection events (resulting from heterogeneous contact rates) or whether these are special events created by particular settings or host plus virus genetic backgrounds, is uncertain. Some insight can be gained by examining the SIR model with homogeneous mixing: in this case the distribution of secondary infections for each index case is already skew, given by a geometric distribution with mean R_0 and variance $R_0(R_0 + 1)$ rather than the intuitively expected Poisson distribution with mean and variance of R_0 (Ferguson *et al.* 2004). The geometric distribution provides an adequate fit to the observed distribution of secondary infections (see figure 9). The probability of one individual infecting n_c or more people is then

$$p(n \geqslant n_c) = (1 - 1/R_0)^{n_c}.$$

Allowing for heterogeneous contact patterns does not usually greatly change this result. The Hong Kong epidemic, for example, was characterized by two large clusters of cases, together with ongoing transmission to close contacts. In the first cluster, at least 125 people were infected on or soon after 3 March in the Prince of Wales Hospital by the index patient for the Hong Kong epidemic (Lee, N. *et al.* 2003). In the second cluster, an unknown number of people were infected from a probable environmental source in the Amoy Gardens estate (Kwung Tong district). Following mixing with fellow residents, families and friends, over 300 people became infected. Examination of local reports of SARS investigations supports the distinction between these two large SSEs and the other contact-based infections, where many occurred in a hospital setting. Furthermore, our simple analysis, taking the

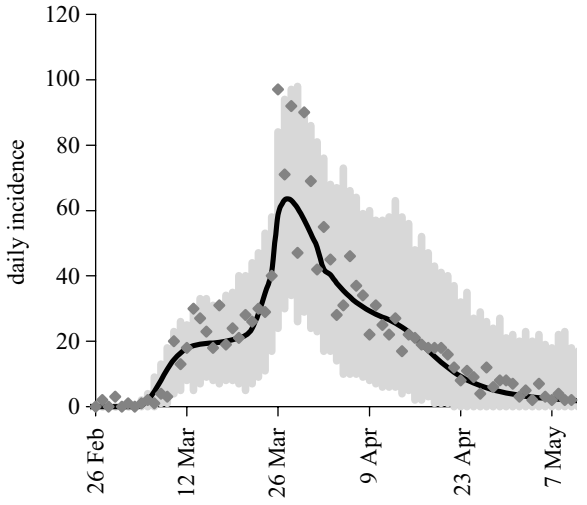


Figure 12. The SARS epidemic in Hong Kong and the fit of a multi-compartment meta-population stochastic model (from Riley *et al.* 2003). The dots are reported SARS cases and the solid line is the best fit model. The vertical grey bars denote 95% prediction intervals.

highest range of R_0 values of five, results in the probability of generating 125 or more secondary infections being $p(n \geqslant 125) < 10^{-12}$, i.e. the probability of such a large cluster arising even once by chance is very, very small. It may be that the distinction between typical infection events and SSEs reflects the two different routes of transmission so far identified as likely for this virus, namely respiratory exudates and faecal–oral contact. However, much still remains uncertain about possible routes of transmission in these SSE settings.

Choice of a suitable model framework must be governed by the degree to which the investigator wishes to capture varying degrees of heterogeneity. A variety of approaches are possible, ranging from a simple deterministic compartmental approach with mixing between patches or settings, to a spatially explicit, individual-based stochastic simulation structure. Riley *et al.* (2003) adopted a stochastic, metapopulation compartmental model, given the quality of the data available in real time from the Hong Kong region. A metapopulation approach was considered to be appropriate because the incidence of SARS varied substantially by geographical district in Hong Kong. A stochastic model was employed because chance fluctuations in case numbers can be large in the early stages of an epidemic. Stochastic models predict both average trends and variability, so that a more robust assessment can be made to examine what changes are likely to be caused by chance and what changes genuinely reflect process and the impact of interventions. They classified the population of each district of Hong Kong into susceptible, latent, infectious, hospitalized, recovered and dead individuals. Data were available to characterize the distributions around these transitions from one compartment to the next. Epidemiological coupling between districts was assumed to depend on their adjacency. Incubating and infectious categories were further divided into multiple stages, chosen in number and duration so as to match accurately the estimated delay distributions determining disease progression and diagnosis (figure 2; Donnelly *et al.* 2003; Riley *et al.* 2003). Multiple realizations of the stochastic model were performed, both for parameter estimation and to generate predicted case incidence time-series. The mean time from the onset of symptoms until hospital admission and subsequent isolation of suspect SARS cases reduced significantly over the course of the epidemic. Changes in the onset-to-hospitalization distribution were treated as an input to the model.

Riley *et al.* (2003) assumed that infectiousness began just before the onset of symptoms and remained constant during the symptomatic phase. With hindsight this was a reasonable assumption. More sophisticated assumptions could now be made given knowledge of how infectiousness changes before and after clinical onset (Peiris *et al.* 2003*a*; figure 4). Simulations were seeded explicitly with the Prince of Wales Hospital and Amoy Gardens clusters. Model fit to the observed case-incidence data in Hong Kong was qualitatively good, both in terms of capturing the temporal development of overall incidence and the pattern of spatial spread (figure 12). Fitting the model to observed trends provided estimates of both R_0, temporal changes in R_t, and the generation time T_g (Riley *et al.* 2003), and as such this approach has many advantages over the use of simpler model constructs given good availability of data in real time. Because the model was seeded explicitly with the two large SSEs, the estimate of R_0 measures the contribution of all non-SSE transmission, including community and hospital transmissions. The contribution of SSEs to R_0 can be calculated only heuristically, but this does not imply that the contribution was unimportant. Indeed, it is possible to construct a hypothetical scenario where the community and hospital contribution to R_0 are below unity, but where SSEs push R_0 above this critical threshold. This would have led to extended periods of epidemic decay between SSEs, until case numbers by chance became sufficient that SSEs began chaining together, causing large and rapid epidemic growth.

The disadvantages of using a more sophisticated and statistically robust approach to estimating R_0 is the need for good computational facilities as well as good data. The development of reliable and efficient parameter estimation methods that do not rely on brute computational force should be a priority.

Estimates for R_0 were based on various methods and on data for a variety of regions. The estimates were in good agreement, despite the different methods employed and the various epidemics in different regions, with an average value of approximately three, independent of setting (table 3). It is important to note that these estimates are based on case reports of individuals with overt clinical symptoms that lead to an initial diagnosis of SARS. To date, however, there are no confirmed cases of transmission from asymptomatic individuals (WHO 2003). The main conclusion to draw from these estimates of R_0 is that SARS-CoV is of low transmissibility by comparison with other directly transmitted viruses such as influenza A (R_0 of *ca.* 7 or more) and the measles virus (R_0 of *ca.* 15–18 prior to widescale immunization (Anderson & May 1991)). Further confirmation of this low level of transmissibility is provided by analyses of SARS cases within those placed in quarantine as a result of close contact (either in health care, family, work or transport settings) with a suspect SARS case. In Taiwan, for example, out of 131 132

Table 3. Estimates of the basic reproductive number, R_0.

reference	region	R_0	comments
Riley *et al.* (2003)	Hong Kong	2–3	excluding SSEs
Lipsitch *et al.* (2003)	Canada and Singapore	3	
J. Wallinga, unpublished (reported in WHO (2003))	Singapore	3.3	

people placed in quarantine, only 45 were recorded as probable SARS cases (Lee, M. L. *et al.* 2003).

The factors that triggered SSEs remain poorly understood at present. The causes probably involve both environmental factors and determinants of the infectiousness of the index patient (the amount of virus excreted in respiratory tract exudates, faeces and urine, and the duration of shedding). The importance of environmental factors is well illustrated by the high proportion of infections worldwide that occurred in health care settings (*ca.* 21% of all reported SARS cases (WHO 2003)). In the estimation of R_0 it is important to recognize that this key epidemiological parameter is a mean drawn from a distribution with high variance. Those events in the right-hand tail of the distribution of secondary case generation may or may not constitute a separate class generated by distinct environmental or case infectiousness factors (figure 9).

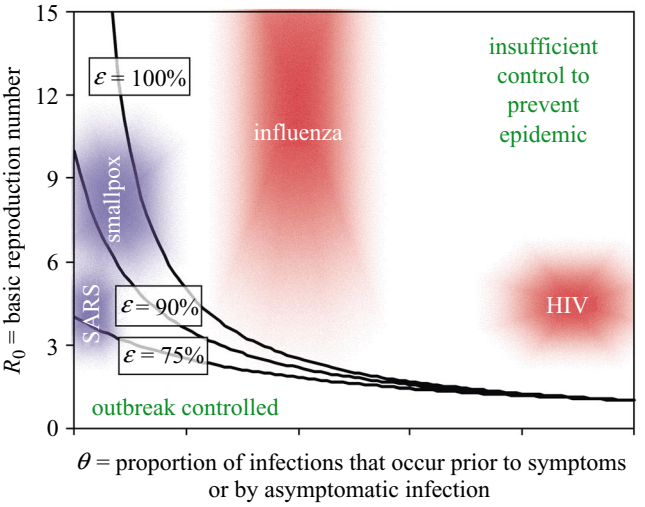


Figure 13. Predicted effect of patient isolation (within 2 days of onset of clinical symptoms) in bringing an epidemic under control (Fraser *et al.* 2004). The solid line denotes a boundary between control and no control, whereas the coloured zones depict ranges for the basic reproductive number, R_0, and the parameter θ. Control can be augmented by other measures such as contact tracing plus isolation of contacts and travel restrictions.

15. HOW WAS THE EPIDEMIC BOUGHT UNDER CONTROL?

To evaluate how different control interventions might impact on any given epidemic during its emergence and early growth, we ideally need to use a mathematical model of transmission that embeds estimates of transmission efficiency and the details concerning the typical course of infection, to explore both what works best and in what combination, and the degree to which a specific intervention must be applied. For example, in the case of SARS, obvious questions are how important is it to reduce the time between onset of clinical symptoms to isolation or quarantine within a health care setting, and what time interval should be the target (i.e. within 1 or 2 days). For international and governmental agencies, the questions may be more complex, such as the value of introducing travel restrictions to and from affected areas, and the effectiveness of screening of passengers at airports for elevated temperature. Such measures may be costly to introduce and may have grave economic consequences for affected regions. It is therefore highly desirable to have available some sort of quantitative template to allow assessments to be made during the heat of the crisis.

For SARS the options for intervention within a country were limited to public health measures, in the absence of a vaccine or effective therapies. There are essentially six intervention categories, namely: (i) restrictions on entry to the country and screening at the point of arrival for fever; (ii) isolation of suspect cases; (iii) the encouragement of rapid reporting to a health care setting following the onset of defined clinical symptoms, with subsequent isolation; (iv) rigorous infection control measures in health care settings; (v) restrictions on movements within a country (restricting travel, limiting congregations such as attendance at school); and (vi) contact tracing and isolation of contacts. To study in quantitative terms their respective impacts on transmission or the impacts of different combinations applied with varying efficiencies, each category must be captured within the mathematical model. One simple illustration is given in the flow chart in figure 10, where a category denoting infectious but isolated is represented. Interventions may reduce the magnitude of R_t but fail to reduce its value to less than unity. In these circumstances the incidence of new infections declines, as if the epidemic is waning, but in reality the trend is to a new endemic state with the infection still persisting. During a crisis, it is difficult to interpret correctly whether or not changes in incidence following the introduction of control measures indicate (i) decay to probable extinction or (ii) continued transmission but at a reduced rate. It is in these circumstances that mathematical models can play an important role as a template for the repeated estimation of R_t through time. Once R_t drops below unity, and stays there, the epidemic is under control provided no relaxation in implementation of the introduced interventions takes place. During the Hong Kong outbreak in 2003, combinations of reductions in onset to hospitalization, in population contact rates (mixing) and with health care setting transmission (improved infection control procedures) reduced the effective reproductive number to *ca.* 1.0 by 21 March, 0.9 by 26 March and then to 0.14 by 10 April (Riley *et al.* 2003).

Teasing out the relative contributions of different interventions is more difficult to achieve if the only quantitative outcome measure is cases of disease reported each day. Ideally other data are required such as changes in travel patterns, the decay in the interval between disease onset and isolation, and the fraction of all contacts of a suspect case traced within a defined time interval. For SARS, aside from case numbers, very few data were available day by day during the course of the epidemic in any given setting, except for information on the temporal change in the distribution of times between onset and isolation (see table 2). Various attempts have been made to dissect the differing impacts of various interventions but with limited success to date. Some of these analyses were based on fitting models to case data to estimate parameters reflecting the efficacy of defined controls (Riley *et al.* 2003), whereas others are more abstract in the sense that parameter values were changed within a model and conclusions were drawn on model predicted trends (Lipsitch *et al.* 2003; Lloyd-Smith *et al.* 2003). Conclusions drawn from both approaches depend on model structure and parameter assignments, and as such should be accepted with caution.

Pooling the results from the published analyses (largely from Lipsitch *et al.* (2003), Lloyd-Smith *et al.* (2003) and Riley *et al.* (2003)), the following conclusions can be drawn. Isolation and quarantine, contact tracing, improved infection control procedures and self-imposed movement and mixing restrictions that limited contact were very effective in combination. They induced the dramatic changes in R_t after the peaks in the epidemic in all regions that were badly affected. Reductions in the onset to admission times were important in most settings owing to the late onset of infectiousness after onset of symptoms of disease. High impact can also be attributed to changes in contact rates (mixing) and better infection control plus isolation and quarantine of symptomatic patients in hospital settings. Spatially explicit models also suggest that restriction on movement between locations within defined communities could have played a useful role in limiting spread. Transmission in most regions was highly localized.

To date, detailed model-based analyses of the effectiveness of contact tracing for SARS suggest that it was far less effective than commonly perceived. More retrospective analyses are required in this area, using large contact-tracing databases. A similarly urgent priority is analysis of the effectiveness of travel directives. Preliminary unpublished studies suggest that they were effective in restricting between country spread. However, this conclusion remains to be confirmed by detailed analyses of travel patterns between major cities before, during and after the SARS epidemic.

Global success in the control of SARS was partly a result of certain epidemiological and biological characteristics of the infectious agent. In the absence of effective vaccines or treatment, understanding the factors that make containment of SARS feasible is important for evaluating how best to control future outbreaks of newly evolved pathogens. Two public health policy options exist for controlling the spread of a novel directly transmitted infectious disease agent: (i) effective isolation of symptomatic individuals (which includes rapid hospitalization after the onset of clinical symptoms) and (ii) tracing and quarantining of the contacts of symptomatic cases. Both measures rely on rapid dissemination of information to facilitate accurate diagnosis of the symptoms of the disease.

For SARS, the timing of the onset of symptoms relative to peak infectivity is probably the most crucial factor in the success of simple public health interventions aimed at reducing transmission. In SARS patients, peak viraemia appears to occur between 5 and 10 days after the onset of symptoms (Peiris *et al.* 2003*a*). Although viraemia does not always predict infectivity, the very low levels measured in the days immediately after the onset of symptoms suggest that peak infectivity occurs somewhat later. Also, no confirmed cases of transmission from asymptomatic patients have been reported to date in detailed epidemiological analyses of clusters of SARS cases (Ksiazek *et al.* 2003; Lee, N. *et al.* 2003). This suggests that for SARS there is a period after symptoms develop during which people can be isolated before their infectiousness increases. It is during this period that transmission can be very effectively interrupted by isolation and quarantine. The second feature is the low transmissibility of SARS-CoV, with its moderately low basic reproductive number, with the exception of the settings in which SSEs occurred.

A recent study (Fraser *et al.* 2004) attempts to analyse why SARS was so effectively contained, using a generic mathematical model for directly transmitted agents. The approach adopted is comparative, and centres on the definition of two key properties of transmission and the typical course of infection, namely the basic reproductive number, R_0, and the proportion of the area under the viral load curve in a typical patient that occurs prior to the onset of easily diagnosable symptoms. This proportion is assumed to reflect infectiousness before and after the onset of symptoms. They show that the proportion of infections that arise prior to the onset of symptoms (or via asymptomatic infection), termed θ, is as strong a predictor of success of the simple public health control measures as the inherent transmissibility of the aetiological agent as measured by R_0. Here, θ is defined as the fraction of transmission (infectiousness) that occurs before the onset of clinical symptoms. Isolation of a proportion ε of symptomatic individuals can control an outbreak if the following expression is satisfied:

$$\varepsilon > (1 - 1/R_0)/(1 - \theta),$$

which can never be reached when $\theta R_0 > 1$. Drawing crisp conclusions is more difficult in the case of contact tracing, or some combination of isolation and contact tracing. In these circumstances a combination of approximate analytical methods and individual-based simulation studies are required to gain insights (figure 13).

From published studies, Fraser *et al.* (2004) estimate these quantities (R_0 and θ) for two moderately transmissible viruses, SARS-CoV and human immunodeficiency virus 1 (HIV-1), and for two highly transmissible viruses, smallpox and pandemic influenza A. They conclude that SARS and smallpox are easier to control using these simple public health measures. This study therefore suggests that in an emerging epidemic of a novel agent, both clinical epidemiological studies of pathogen load and clinical symptoms, plus contact tracing to assess when

transmission occurs during the typical course of infection, should be a priority.

16. DISCUSSION

The evolution, spread and persistence of infectious diseases are facilitated by the mobility of contemporary society, for example through air travel, the continued growth in the world population and the steady rise in the number of densely crowded urban areas (the so-called mega-cities with populations of over 10 million people), especially in Asia. As such, epidemic outbreaks of novel infectious agents are likely to become more common in the twenty-first century than in any previous period of human history. We therefore need to be prepared, and the 'SARS' experience provided many lessons for the future.

What are the lessons to be learnt from the SARS epidemic with respect to the detection and control of novel infectious agents? As an initial task it is important to find the agent that is the cause of observed morbidity and mortality and then to establish whether or not the agent is novel. Careful pathological, microbiological and virological study is the key to discovery of the aetiological agent. With current sequence data on known pathogens, the problem of novelty is relatively easy to solve. The SARS epidemic showed clearly how effective international collaboration (and competition) resulted in the whole genome sequence of the SARS-CoV being available early in the epidemic. In the not too distant future, whole pathogen genome sequencing will be routine in most laboratories in developed countries. Difficulties remain in the poorer regions of the world, with limited surveillance and inadequate laboratory expertise. It is of particular importance in the most populous countries of the world; namely, China, India and Indonesia (Anon. 2004). It is clearly in everyone's interest to greatly enhance global surveillance capabilities, especially in developing regions, and concomitantly to improve basic training in infectious disease and molecular epidemiology.

Effective surveillance structures aim to detect unusual clusters of morbidity and mortality in space and time. What constitutes such a cluster requires robust statistical analysis, and a time history of what normally occurs, taking account of seasonal and longer-term dynamical cycles and chance fluctuations in disease incidence (Anderson & May 1991). It is not too difficult to imagine that in the near future, automated software will be put in place in the richer regions of the world (such as North America and western Europe, which have good data capture systems for infectious disease reports) to analyse reported data on a daily basis to alert authorities of such unusual spatial and temporal clusters. However, it should be noted that the alert physician at the point of contact with sick patients who recognizes an unusual pattern of morbidity and mortality is the foundation of good infectious disease surveillance.

A related problem concerns whether or not the nascent epidemic is indeed about to expand, with R_0 in excess of unity. Several recent studies have started to address this question, taking account of the fact that even if R_0 is less than unity, some long chains of onward transmission will arise by chance before the epidemic stutters to extinction. The key statistics are the average cluster size and the average length of the transmission chains. Statistical and simulation tools are being developed to try and help in such assessments (Ferguson *et al.* 2004), and have been applied recently in the analysis of human cases of bird flu resulting from the major H5N1 epidemic in birds in Asia in the early part of 2004. Detailed contact tracing and clinical investigation could be used early in an outbreak to determine θ, the proportion of infections that occur prior to the expression of symptoms, and thus the likely impact of simple isolation and quarantine measures (see Fraser *et al.* 2004).

More generally, after the surveillance and detection problem, the next set of issues concern data capture, the development of diagnostic tests and treatment algorithms and the identification of public health measures to control epidemic spread. Real-time data capture and associated analysis to reveal how the epidemic is expanding and how interventions are acting to slow spread is essential. Few countries or regions did well on this front during the SARS epidemic, and the WHO struggled to collate detailed data in a timely manner from the affected countries. More must be done in Asia, Europe, North America and the WHO, to improve the information feed into a 'command and control' centre for epidemic outbreak analysis and control. As some countries learnt to their cost, not having a clear command and control structure to manage an outbreak can delay the decision-making process. One central authority needs to collate and analyse data on a day-by-day basis. Furthermore, someone at this centre needs to have the authority to ensure that regional centres and health care settings submit data in a timely manner and that the best available scientific advice forms the basis for policy formulation. A section of this paper describes experiences in Hong Kong where the health authorities performed well in this context in the mid to late phase of the epidemic.

Deciding well before an outbreak emerges what needs to be collected and what sort of analyses should be done day by day is an obvious starting point in the preparation for future outbreaks. Unique patient identifiers should be used to record all socio-demographic, clinical, treatment and epidemiological information (including contact tracing where the nature of contacts and the time period over which they are sought is well defined). Careful thought is needed to define data fields and apply effective data capture across all health care settings. Web-based, password-protected systems need to be ready to be put into action, with information fed daily to one centralized database for analysis and interpretation. No one country or region had such a system in place prior to the emergence of SARS. Ideally some common database and software structure should be used across all countries and regions, such that the WHO could capture information relevant for the global management of an outbreak.

Contact tracing and monitoring travel at points of entry and exit are part of this data capture process. The speedy and accurate collection of contact tracing data serves many purposes. First it acts as a public health measure, in the sense that it gives an opportunity to isolate or quarantine those in close contact with a suspect case. Equally important, however, is the fact that it provides a source of information to estimate key parameters and distributions, and modes of transmission. The incubation period is one

key distribution, as are estimates of the distribution and mean of the effective reproductive number. In most settings that were badly affected by SARS, the percentage of contacts traced within a few days was very limited. We need to learn more about how best to do such tracing for directly transmitted agents, and how best to analyse and interpret the data given uncertain denominators.

The effectiveness of temperature screening at points of entry and exit as a control measure to limit between-country transmission is uncertain at present. This needs more detailed retrospective analysis, as does the controversial measure of travel directives issued by the WHO at various stages of the epidemic. These almost certainly induced substantial changes in business and leisure travel patterns, and had a major impact on the economies of badly affected regions such as Hong Kong, mainland China, Taiwan, Singapore, and Ontario in Canada. More research is needed to establish clear guidelines for the circumstances under which such directives should be issued by the WHO.

The development of rapid diagnostic tests available at the point of patient contact is a clear priority in managing new outbreaks. The particular biology and pathogenesis of SARS-CoV made this task difficult to accomplish during the early and mid phases of the epidemic. Much has been learnt is this context, and the sharing of materials, reagents and viral isolates orchestrated by the WHO was an encouraging aspect of the SARS experience.

A further lesson concerns the topics of clinical epidemiology and treatment algorithms. As noted earlier in this paper, a key property that influences the likelihood of successful control is the typical relationship between the onset of clinical symptoms and viral load as a surrogate marker of infectiousness. With some notable exceptions (Peiris *et al.* 2003*a*), not enough attention was paid to quantifying these properties day by day for large samples of patients. Furthermore, the development of effective treatment algorithms to reduce morbidity and mortality was delayed by poor sharing of patient information. With small numbers of patients in the early stages of a new epidemic, pooling of clinical data in an international database is essential. Randomized clinical trials cannot be conducted easily in the heat of a crisis, and hence careful analysis of observational databases with sufficient patient records to deal with confounders such as age and pre-existing co-morbidities is the only way to decide which pattern of patient management and treatment is most effective.

With respect to the issue of data analysis and the design of the most effective intervention measures for pathogens with specific biological and epidemiological properties, a clear need is the development of generic mathematical models for a wide variety of pathogen types (including vector-transmitted viruses), with differing biological and transmission properties. More than one type of model is needed to deal with problems at a local level (perhaps even within a health care setting), at a community scale and within an international context. Prior exploration of what control option or combination of options works best in defined situations is a clear priority. One approach is that adopted by Fraser *et al.* (2004), but other model templates and methods of analysis would be highly desirable, using more sophisticated stochastic individual-based models. We need to understand more clearly how different interventions impact on agents with given properties, particularly if it is possible to put such measures in a cost-effectiveness framework. The issues of travel directives and airport screening for fever immediately come to mind. We understand very little at present about how effective they were in limiting the spread of SARS-CoV. More important, however, is the issue of isolation, containment and contact tracing. Analyses by Fraser *et al.* (2004) provide a starting point for deciding whether such simple public health tools will work effectively if applied with a given efficacy, to control the spread of specific pathogens with defined biological and epidemiological properties (figure 13). Mathematical models also provide a framework within which different intervention strategies can be evaluated during the course of an epidemic (see Ferguson *et al.* 2001*a*,*b*). Estimation day by day of the effective reproductive number R_t (the average rate of generating secondary cases) provides a quantitative measure of success or failure.

The SARS epidemic caused much suffering, significant mortality, great disruption to social and work activities and considerable economic losses. Draconian public health measures involving the isolation and quarantining of hundreds of thousands of people, and tight restrictions on travel, had to be put in place in some countries. However, it was brought under control—and relatively quickly—with the WHO playing a vital role in coordinating the international response. The quick and effective response of the WHO to the SARS crisis did much to restore faith among the many critics of the effectiveness of international agencies with large bureaucracies and limited resources for action. But it is difficult to escape the conclusion that the world community was very lucky this time round, given the very low transmissibility of the agent, plus the fact that fairly draconian public health measures could be put in place with great efficiency in Asian regions where the epidemic originated. Given the litigious nature of people in North America in particular, and to a lesser degree in western Europe, the control of SSEs in these regions might have presented greater problems if mass quarantining had been required. In the next global epidemic of a directly transmitted short-generation-period infectious agent we may not be so lucky, either in terms of the biology of the agent or the region of its origin. Thus one of the major dangers arising from the effective control of SARS is complacency. Sentiments of the type 'we have been successful once—we will be again' may be far from the truth. Simple public health measures worked well for SARS, but the persistence of the virus (or its close relatives) in animal reservoirs means that re-emergence will occur, as seen at the end of 2003 and the early part of 2004. However, the continuing threat from SARS needs to be kept in perspective, given that influenza A causes many tens of thousands of deaths annually in developed countries. Many informed observers feel that the real threat in the future is an antigenically novel influenza virus, of both high pathogenicity and transmissibility. In these circumstances, simple public health measures are unlikely to be effective (see figure 13), and other options such as more draconian movement restrictions, the greater availability of antiviral drugs and expanded vaccine development and production facilities, will be needed to prevent a devastating impact.

REFERENCES

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford Science Publications.

Anon. 2004 China: towards 'xiaokang', but still living dangerously. *Lancet* **363**, 409.

Breiman, R. F., Evans, M. R., Preiser, W., Maguire, J., Schnur, A., Li, A., Bekedam, H. & MacKenzie, J. S. 2003 Role of China in the quest to define and control severe acute respiratory syndrome. *Emerg. Infect. Dis.* **9**, 1037–1041.

Chow, P. K. H., Ooi, E. E., Tan, H. K., Ong, K. W., Sil, B. K., Teo, M., Ng, T. & Soo, K. C. 2004 Healthcare worker seroconversion in SARS outbreak. *Emerg. Infect. Dis.* **10**, 249–250.

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

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001*a* The foot and mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160.

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. 2001*b* The determinants of transmission intensity and the impact of control policies on the foot and mouth disease (FMD) epidemic in Great Britain. *Nature* **413**, 542–548.

Ferguson, N. M., Fraser, C., Donnelly, C. A., Ghani, A. C. & Anderson, R. M. 2004 Assessing the public health risk posed by the avian H5N1 influenza epidemic. *Science* **304**, 968–969.

Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. 2004 What makes and infectious disease outbreak controllable? *Proc. Natl Acad. Sci. USA* **101**, 6146–6151.

Galvani, A. P., Lei, X. D. & Jewell, N. R. 2003 Severe acute respiratory syndrome: temporal stability and geographic variation in case-fatality rates and doubling times. *Emerg. Infect. Dis.* **9**, 991–994.

He, J. F. (and 52 others) 2004 Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669.

Hsu, L. Y., Lee, C. C., Green, J. A., Ang, B., Paton, N. I., Lee, L., Villacian, J. S., Lim, P. L., Earnest, A. & Leo, Y. S. 2003 Severe acute respiratory syndrome (SARS) in Singapore: clinical features of index patient and initial contacts. *Emerg. Infect. Dis.* **9**, 713–717.

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

Lee, M. L. (and 15 others) 2003 Use of quarantine to prevent transmission of severe acute respiratory syndrome—Taiwan, 2003. *MMWR* **52**, 680–683.

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

Leung, G. M. (and 13 others) 2004*a* The epidemiology of severe acute respiratory syndrome (SARS) in the 2003 Hong Kong epidemic: analysis of all 1755 patients. *Ann. Internal Med.* (In the press.)

Leung, G. M. (and 15 others) 2004*b* Seroprevalence of IgG antibody to SARS coronavirus (SARS-CoV) in a population-based sample of close contacts of all 1755 cases in Hong Kong. *Emerg. Infect. Dis*. (In the press.)

Li, G., Chen, X. J. & Xu, A. L. 2003 Profile of specific antibodies to the SARS-associated coronavirus. *New Engl. J. Med.* **349**, 508–509.

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 58 others) 2003 The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.

Paterson, R. 2004 SARS returns to China. *Lancet Infect. Dis.* **4**, 64.

Peiris, J. S. M. (and 16 others) 2003*a* Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767–1772.

Peiris, J. S. M. (and 15 others) 2003*b* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325.

Riley, S. (and 18 others) 2003 Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966.

Shen, Z., Ning, F., Zhou, W. G., He, X., Lin, C. Y., Chin, D. P., Zhu, Z. H. & Schuchat, A. 2004 Superspreading SARS events, Beijing, 2003. *Emerg. Infect. Dis.* **10**, 256–260.

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.

Varia, M., Wilson, S., Sarwal, S., McGeer, A., Gournis, E., Galanis, E. & Henry, B. 2003 Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada. *Can. Med. Assoc. J.* **169**, 285–292.

Wallinga, J. 2004 Presentation at the Global Meeting on the Epidemiology of SARS, World Health Organization, Geneva, Switzerland, 16–17 May 2003. See http://www.who.int/csr/sars/en/WHOconsensus.pdf.

Watts, J. 2004 China culls wild animals to prevent new SARS threat. *Lancet* **363**, 134.

Webster, R. G. 2004 Wet markets: a continuing source of severe acute respiratory syndrome and influenza? *Lancet* **363**, 234–236.

WHO 2003 Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). See http://www.who.int/csr/sars/en/WHOconsensus.pdf.

WHO 2004 New case of laboratory-confirmed SARS in Guangdong, China–update 5, vol. 2004. Geneva: World Health Organization.

GLOSSARY

CFR: case fatality rate
RDS: respiratory distress syndrome
RT–PCR: reverse transcription–polymerase chain reaction
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus
SSE: super-spreading event
WHO: World Health Organization