Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Epidemiological and genetic analysis of severe acute respiratory syndrome

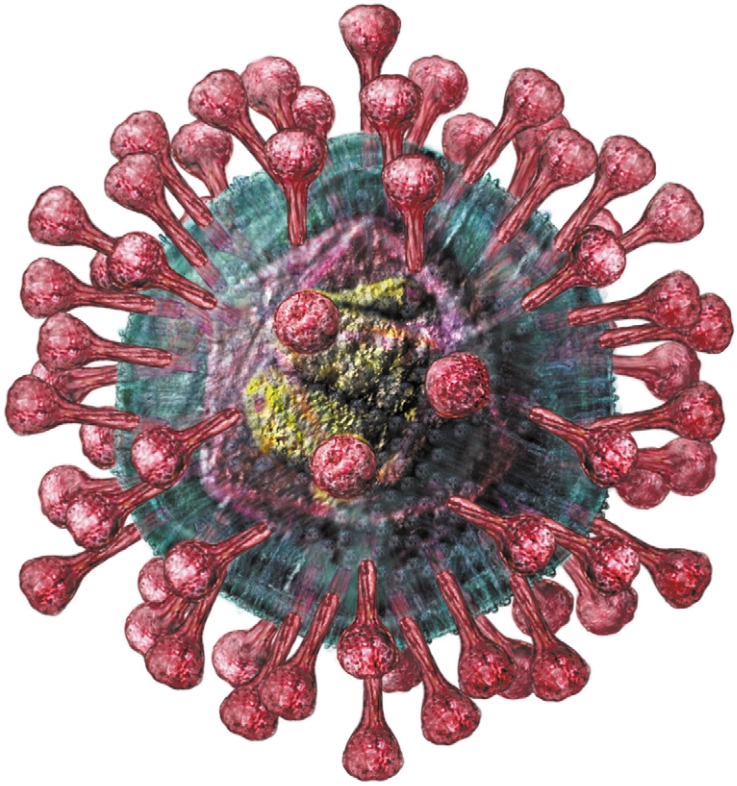
Christl A Donnelly, Matthew C Fisher, Christophe Fraser, Azra C Ghani, Steven Riley, Neil M Ferguson, and Roy M Anderson

The severe acute respiratory syndrome (SARS) epidemics in 2002–2003 showed how quickly a novel infectious disease can spread both within communities and internationally. We have reviewed the epidemiological and genetic analyses that have been published both during and since these epidemics, and show how quickly data were collected and analyses undertaken. Key factors that determine the speed and scale of transmission of an infectious disease were estimated using statistical and mathematical modelling approaches, and phylogenetic analyses provided insights into the origin and evolution of the SARS-associated coronavirus. The SARS literature continues to grow, and it is hoped that international collaboration in the analysis of epidemiological and contact-network databases will provide further insights into the spread of this newly emergent infectious disease.

Lancet Infect Dis 2004; **4:** 672–83

The rapid worldwide spread of the coronavirus (figure) that causes severe acute respiratory syndrome (SARS) led to 29 countries reporting cases in 2003.[1] The first human case was identified in Guangdong Province, China on Nov 16, 2002, and the last known case of the initial epidemic experienced the onset of symptoms on June 15, 2003, in Taiwan.[1] (However, due to differences in case definitions,[2,3] the USA has reported probable cases of SARS with onset of illness after July 5, 2003[1]) Subsequent cases have arisen in Singapore,[4] Taiwan,[5] and China[6] (most recently in April, 2004[7]) because of laboratory-related infections and onward transmission.[8] Worldwide surveillance, coordinated by WHO, resulted in the identification of 8098 clinically affected SARS cases, of whom 774 died.[1]

SARS is believed to be zoonotic in origin, with the palm civet cat (*Paguma larvata*) being implicated as an important animal reservoir, although evidence of infection has been identified in other species (the raccoon dog, *Nyctereutes procyonoides*, and the Chinese ferret-badger, *Melogale moschata*).[9] Close human–animal contact was associated with many early SARS cases,[10] indicating that the SARS coronavirus had jumped host, although most of the infections over the course of the epidemic were due to human–human transmission. The spread of infectious diseases, such as SARS, within human hosts is facilitated by our increasingly mixed and densely packed global society with its high degree of connectedness through increased long-distance air travel, continued growth in world population,[11] and increasing number of densely



Russell Kightley Media

Computer generated image of SARS coronavirus.

CAD, NMF, and RMA are professors, MCF is a lecturer, ACG and CF are research fellows, and SR is a postdoctoral researcher at the Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College, London, UK.

Correspondence: Professor Christl Donnelly, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK. Tel +44 (0)20 7594 3394; fax +44 (0)20 7262 3180; email c.donnelly@imperial.ac.uk

inhabited urban areas, which are particularly common in Asia.[12]

Although the emergence of SARS was recent, a wide range of epidemiological studies have been published both during and since the 2003 epidemics. Increasingly sophisticated techniques and more powerful computers now permit rigorous epidemiological and genetic analysis of epidemics as they unfold. The goals of such analyses include the following: understanding the origin of the disease with a view to preventing subsequent outbreaks; estimation of key biological and epidemiological parameters; identification of risk factors for susceptibility, infectiousness, and mortality; prediction of future trends in infection and case incidence; and assessment of the effectiveness of public-health control measures. Some of these goals can be achieved through the application of widely-used epidemiological methods, such as case-control studies,[13] which are applicable to a wide range of medical conditions. However, others require techniques specifically designed for the analysis of infectious diseases,[14] which, although less widely used, have a long history.[15–19]

Here we review the epidemiological literature on SARS, as an example of an important novel infectious disease, and consider the contributions of the various approaches. In doing so, we highlight the benefits that were realised from epidemiological analyses and to suggest how, in future outbreaks, such methods might be used even more effectively.

Evolution of the SARS coronavirus genome

Unprecedented levels of international cooperation led to the sequencing of two SARS viral genomes within 6 weeks of the identification of atypical pneumonia in Hong Kong.[20,21] Alignment to the genomes of known groups 1, 2, and 3 coronaviruses showed that SARS coronavirus is phylogenetically distinct, and only distantly related to the other coronavirus clades. These early analyses showed that SARS coronavirus is not a recently evolved pathogen, and group 4 was proposed within which to classify the novel coronavirus.[20,21] Subsequent alignment of the SARS coronavirus replicase open reading frame (ORF) 1b (about 5500 bp) genome segment, using the genus Torovirus (order Nidovirales) to root the tree, suggested that SARS coronavirus represents an early split-off from the coronavirus group 2 lineage[22] and should be thought of as a group 2 subgroup. This result has found support from other studies using alternative outgroups[23,24] and methods.[23–27]

However, as clearly shown by analyses of other coronaviruses,[28,29] single-stranded RNA viruses are prone to recombination within and between lineages. This greatly complicates phylogenetic analyses, as different regions of the genome will have different evolutionary histories. Marra and colleagues[20] proposed that SARS coronavirus may have undergone intergroup recombination after noting that the SARS coronavirus S2m motif is found in the group 3 avian coronaviruses, as well as the more distantly related equine rhinovirus (Picornaviridae). Later studies focused on the SARS coronavirus spike (S), matrix (M), nucleocapsid (N), replicase polyprotein (PP1ab), and RNA-dependent RNA polymerase (RDRP) proteins with Bayesian techniques. Here, phylogenetic incongruence was reported to occur across the SARS coronavirus genome, by showing that the *M* and *N* genes were most likely to have originated from an ancestral bird (group 3) coronavirus, whereas *PP1ab* showed homology with a mammalian-like (group 2) coronavirus.[24] Sliding-window approaches suggested that the *S* and *RDRP* genes are candidate mosaic sequences and identified the possible site of the original interlineage recombination events.[24,30] It has been proposed that recombination in the crucial S protein may have generated a virus with modified host specificities, leading to the contemporary emergence in human populations.[24] An event of similar nature may have led to the 1918 Spanish influenza pandemic. However, the bootstrap methods used[24,30] support an independent, and genetically distant, SARS coronavirus clade for all genes. This suggests that any recombination events, if they have occurred, are evidently ancient and therefore not implicated in the current emergence of SARS coronavirus in human populations.

Origins and dating the emergence of human SARS coronavirus

Retrospective assessments of case reports identified 11 index cases from the Guangdong Province, China, the earliest of which was recorded from the city of Foshan on Nov 16, 2002.[10] These index cases were unlinked, except for the epidemiological observation that seven of the 11 patients worked with animals in the food industry.[31] Within live-animal markets in Guangdong Province, 13–40% of wild-animal traders and slaughterers were seropositive for SARS;[9,32] these findings led to the speculation that SARS is a zoonosis from an unidentified animal source. Suspicion focused on palm civet cats because 73% of the traders primarily trading in masked palm civet cats tested seropositive for the virus.[32]

PCR and serological surveys of 25 animals from the live-animal market found serological evidence for infection in five Himalayan palm civet cats, a raccoon dog, and a Chinese ferret-badger. Coronaviruses were successfully isolated from the palm civet cats and raccoon dog, yielding two full-length genomes with 99·8% homology to human SARS coronavirus.[9] Further surveys of civet cats farmed in Hubei Province have shown that these too are infected with a SARS-like coronavirus.[31] Comparative analyses of the Shenzhen civet cat sequences against those from 11 human isolates showed that the human and animal isolates are phylogenetically distinct.[33] The genetic distance between civet cat coronavirus isolates is greater than that observed between geographically (China, Hong Kong, Canada) and temporally separated (early and mid-epidemic) human SARS coronavirus isolates.[9] These data suggest that SARS originated from an animal reservoir, and that the ultimate source of the coronavirus that caused the emergence of the human SARS coronavirus genotype remains unclear. However, sequence data from a recent case (GD03T0013, isolated Dec 16, 2003) has found closer grouping with the civet cat coronaviruses than was previously observed,[31] suggesting that civet cats may indeed be the source of SARS coronavirus. Further attempts to isolate the virus from

market animals has met with little success.[34–36] Final confirmation of the animal reservoir of SARS coronavirus therefore awaits a systematic survey of Chinese fauna.

By May 9, 2003, 14 genomes of SARS coronavirus had been sequenced.[37] This total has risen to 100 GenBank depositions by May, 2004. Molecular analyses have shown that the early phase of the epidemic was characterised by two genotypes. The first (cluster A) is composed of 10 isolates corresponding to the very early cases from Guangdong Province and three separate introductions to Hong Kong.[33] The second (cluster B) corresponds to the so-called super-spreading event (SSE) triggered by the arrival of patient 1 (HKU-33) in Hong Kong;[31,33] this is the genotype that ultimately became pan-global. The finding that most of the SARS coronavirus genetic diversity occurs within cluster A is consistent with the epidemiological observations that Guangdong Province, China, is the geographical point of origin for the emergence of the virus.

If we assume that SARS coronavirus had a single emergence within human populations, the most recent common ancestor of SARS coronavirus will correspond to our best approximation for the emergence of the virus. Efforts to date the most recent common ancestor of SARS coronavirus have mostly followed the rationale that was used to date the emergence of HIV-1 M group viruses.[38] This relies on building a phylogenetic tree of isolates to find the most deeply branched sequences, then assuming neutral clock-like evolution to date the root of the tree. Using the divergence of *S*-gene sequences from 139 patients,[33] linear regression dated the emergence of SARS coronavirus to mid-December, 2002 (95% CI late September, 2002, to mid-January, 2003).[39] A recent study by the Chinese SARS Molecular Epidemiology Consortium[31] attempted to correct for the potential effects of selection by only using synonymous (*Ks*) substitutions. They dated the ancestral sequence with the deeply rooted isolate GZ02 as an outgroup, and estimated an origin of mid-November, 2002 (95% CI early June, 2002, to late December, 2002). If correct, these data suggest that the earliest known SARS case, in November, 2002, was not far removed from the theoretical origin of the epidemic.

However, although promising, these studies necessarily rely on isolates that are collected over a short timescale and are probably rapidly evolving. Whereas contemporary isolates are rare, there have been recent infections that are not associated with laboratory escapes.[31] Due to the observation that the isolate GD03T0013 is the most deeply rooted yet seen, use of this sequence to date the most recent common ancestor of SARS coronavirus will push back the epidemic's origin, perhaps significantly. If done, such an analysis would suggest that SARS coronavirus has been circulating, undetected, in China for longer than was previously expected. It is also evident from comparisons of non-synonymous to synonymous substitution rates (*Ka*/*Ks*) in the S protein that the SARS genome is under strong directional selection.[31] The use of techniques that account for variation in the rates of evolution over the course of the epidemic (ie, Bayesian evolutionary analysis sampling trees, BEAST v1·0·2, available from http://evolve.zoo.ox.ac.uk/beast/ [accessed Oct 4, 2004]) would therefore be appropriate.

Incubation period

A key factor in the epidemiology of an infectious disease is the incubation period, which is defined as the time from infection to onset of clinical symptoms of disease.[40] The distribution of the incubation period has important implications for contact tracing and quarantine strategies, so accurate estimates of the distribution are an important goal for early epidemiological investigations of a novel disease. Furthermore, the average incubation period influences the timescale of the development of the epidemic, as it partly determines the time interval between a case and the infections that the case subsequently generates. Identification of determinants of the incubation period, such as age, infectious dose, and host genetics, can provide insights into the mechanisms of disease progression.

Although infection events cannot be observed directly, some patients retrospectively reported well-defined periods of exposure to one or more known SARS cases. When an event (ie, infection) is only known to have occurred within a defined period, the data are said to be interval censored.[41] Patients with long periods of exposure are uninformative; however, patients with short and well-defined periods of exposure are informative, even though the exact date of infection is unknown. These data, when analysed appropriately, can be used to estimate the distribution of incubation periods in the patient population.

Summaries and analyses of incubation period data have been published for various populations of patients (table 1). In many cases, the difficulties posed by interval censoring led to researchers presenting descriptive summary statistics without further analysis. Other work corrected for the interval censoring by use of both parametric[47] and non-parametric[57] approaches. However, it should be noted that naive analyses that assume patients were equally likely to have been infected throughout their reported interval[57] overestimate the variance in the distribution and could also bias the estimates of the mean incubation period, with the size of these problems depending on the width of the reported exposure intervals.

Given the difficulties inherent in the interpretation of interval-censored data, the central estimates (means and medians; table 1) are remarkably consistent in patients in China, Hong Kong, Singapore, and Canada (with central estimates ranging from 4 to 6). The mean from European cases, 7·2 days, was somewhat higher, but the estimate is uncertain because it is based on only five cases.[54]

The maximum incubation period is less clear, with a number of reports of incubation periods exceeding WHO's maximum incubation period of 10 days.[45,50,54] The WHO consensus document on the epidemiology of SARS, published in October, 2003, noted the existence of incubation period outliers of more than 10 days, but suggested they had "not necessarily been subjected to rigorous and standardised investigation".[54] Furthermore, interval censoring causes particular difficulties in assessing the maximum incubation period, and if midpoints in large

Table 1. Published data on the incubation period of SARS

First author	Publication date	Location	Number of patients	Interval censoring (IC) or multiple exposure	Other comments	Estimates (days)
Tsang[42]	March 31, 2003	Hong Kong	9	IC present in 5 of 9,	Individual data published	..
WHO[43]	March 21, 2003	Worldwide	..	..	Range and maximum reported	Range 2–7
Poutanen[44]	March 31, 2003	Canada	10	IC present	Individual data published	..
Lee[45]	April 7, 2003	Hong Kong	..	Unclear, described as "the interval between exposure to the index patient or ward and the onset of fever"	Range and median reported	Median 6, range 2–16
Booth[46]	May 6, 2003	Canada	144	Some multiple exposures	Reported median and IQR from earliest self-reported exposure to onset of symptoms (caution urged)	Median 6, IQR 3–10
Donnelly[47]	May 7, 2003	Hong Kong	57	IC present, estimates based on patients with single exposure	Maximum likelihood allowing for IC	Mean 3·8, variance 8·3
WHO[48]	May 7, 2003	Worldwide	..	..		Maximum 10
Leo[49]	May 9, 2003	Singapore	21 patients with point exposures; 94 with "well-defined exposures"	IC present	Mean, median, 95th percentile reported; gave separate estimates for those with "well-defined point exposures"; mid-points used for IC data.	Mean 5·2, median 5 Mean 5, median 4·3
Wu[50]	June, 2003	Guangzhou, China	..	..	Mean and range reported	Mean 5·9, range 1–20
Avendano[51]	June 24, 2003	Canada	14*	4 with single exposure and 10 with multiple exposure	Mean and SD reported separately for patients with single and multiple exposures	Mean 4, SD 3 (single exposure); mean 3·5, SD 3 (multiple exposure)
Varia[52]	July 29, 2003	Canada	42	..	Mean and range reported	Mean 5, median 4, range 2–10
Choi[53]	Oct 1, 2003	Canada	..	..	..	Median 5
WHO[54]	Oct 17, 2003	Singapore;	46	Single exposure, IC not mentioned	Mean, median, range reported	Mean 5·3, median 5, range 1–10
		Guangdong, China;	70	IC not mentioned		Mean 4, median 4, range 1–12
		WHO European Region	5	IC not mentioned		Mean 7·2, median 7, range 5–10
Olsen[55]	Dec 18, 2003	In-flight transmission	22	IC and multiple exposure not present due to limited (in-flight) exposure	Mean and range reported, full data given in figure	Mean 4, range 2–8
Chow[56]	Jan 15, 2004	Singapore	15	Multiple exposures present: "complex"	Range reported and full data given in figure	Mean 4·3, median 4, variance 2·2, range 3–8
Meltzer[57]	Feb, 2004	Hong Kong, Canada, USA	20†	Present; all data published	Assumed uniform distribution to allow for IC to estimate distribution	Median 4, range 1–18

*These 14 patients were among the 144 SARS patients previously described by Poutanen and colleagues;[44] however, this study only reported incubation period information for patients with single exposures, whereas Avendano and colleagues[51] reports data for both singly and multiply exposed patients. †The data analysed include those published by Tsang and colleagues[42] and Poutanen and colleagues[44] in addition to previously unpublished data on two USA patients. ..=not specified.

exposure intervals are relied on then errors can result. For example, an incubation period in a patient exposed to a SARS case 5–25 days before symptom onset might be naively (and most likely incorrectly) reported as a 15-day incubation period.

Three studies reported somewhat higher mean incubation periods (8, 7·3, and 7·6 days).[58–60] Similarly, another group[61] reported that, although the index case had an incubation period of 4 days, the secondary and tertiary generations had incubation periods of 7 and 8 days, respectively, noting that shorter incubation periods were associated with longer fevers and greater clinical severity. By contrast, He and colleagues[62] reported a mean incubation period of 4·5 days in patients from Guangdong Province, China, and Li and colleagues[63] reported a median incubation period of 3 days (range 1–10 days) in Beijing, China. However, we were unable to obtain full translations of these papers so were not able to determine how these estimates were obtained.

Since only a small proportion of the SARS cases will have data suitable for estimation of the incubation period distribution, international collaboration would be particularly valuable. (WHO has suggested that such data exist on only 200 cases worldwide.[54]) Such efforts would, however, need to go well beyond straightforward analysis of a merged dataset, due to the care and precision required to define periods of exposure to SARS infection accurately. If an international dataset were systematically compiled, then rigorous overall estimates of the incubation period distribution could be obtained, and any dependence of the

incubation period distribution on patients' characteristics (ie, age, sex, stage of the epidemic) could be determined.

Infectiousness and disease progression

Following the appearance of symptoms, virtually all SARS patients were either admitted to hospital or placed under home quarantine. Disease progression was best summarised by Peiris and colleagues[64] who followed 75 SARS patients for 3 weeks after admission to hospital in Hong Kong. Patients experienced a recurrence of symptoms after a mean of 8·9 days, a peak in viral load at approximately 10 days after admission, and 60% of patients had seroconverted by 20 days. The rates of admission to intensive care units (ICUs) in cohorts worldwide ranged from 23% to 40%, with a high proportion of those admitted with acute respiratory distress syndrome and requiring mechanical ventilation.[45,65–69]

Transmission of SARS in most cases could be linked to direct close contact with another SARS case. Many of these contacts were nosocomial. Roughly half of cases were healthcare workers, in studies in Hong Kong[45] and Guangzhou, China,[70] and 77% of cases were exposed in hospital, in a study from the Toronto area, Canada.[46] Although these studies are convenience samples drawn from particular hospitals, the levels reported are similar to those reported from the national database of SARS cases in Hong Kong (with 49% of the SARS cases resulting from infections occurring in clinics, hospitals, or elderly or nursing homes; G M Leung, Department of Community Medicine, University of Hong Kong, personal communication).

Within the community, rates of transmission were generally low, with 8% of household contacts infected in one study in Hong Kong[71] and 6·7% in a study in Singapore.[72] The closeness of contact also seems to be important. For example, in a large retrospective examination of case notes of quarantined close contacts of SARS cases in Beijing, China, the overall attack rate was 6·3%, with the highest rates (15·4%) in spouses and lowest rates in work and school contacts (0·36%).[73] These figures are further supported by large-scale screening for SARS coronavirus antibodies in direct contacts of SARS cases, which have very low rates (0·2% were positive for SARS coronavirus IgG antibodies[74]). Nosocomial transmission therefore seems to have been the major source of new infections, with higher attack rates reported in this setting. For example, in a study of nosocomial outbreak in Toronto, rates in nurses working in the emergency department, ICU, and coronary care unit ranged from 10% to 60%.[52] Good barrier protection was essential to prevent transmission: in a case-control study in Hong Kong, inconsistent use of goggles, gowns, gloves, and caps was significantly associated with a higher risk of SARS.[75]

Case fatality rate

Early in the epidemic, with little known about the pathogenicity of SARS, there was substantial concern about the increasing rates of morbidity and mortality that were being reported through individual case reports. Estimation of the case fatality rate (CFR; the percentage of people diagnosed as having a specific disease who die as a result of that disease) during an outbreak is complicated because the eventual outcomes of patients still in hospital are unknown at the time of analysis. The duration of hospital stay depended on the severity of illness, but for most patients this was in the region of 14 days to 1 month.[47,76] Naive estimates of the CFR, based simply on the cumulative number of deaths divided by the cumulative number of cases,[77] were therefore particularly misleading because they yielded underestimates of the true mortality. This bias was reduced as the epidemic progressed (and as the outcome of a greater percentage of patients was known), which is why reported estimates of CFR seemed to indicate that mortality was increasing over time, causing some to incorrectly conclude that SARS coronavirus was evolving to be more lethal.[78,79] The earliest estimates that used appropriate statistical methodology were published in May, 2003, with data from Hong Kong[47] and worldwide,[48] and gave final CFR estimates of between 14% and 18%. Table 2 shows the estimates obtained in various cohorts, with CFRs at 21 days after hospital admission of 6·5% and 10%, at 28 days of 10%, and at 3 months of 12%.[46,65–67] CFRs for those admitted to ICUs were significantly higher, with estimates at 28 days of 26% and 34% in Hong Kong and Toronto cohorts, respectively.[80,82]

Several cohorts have consistently described the course of disease for SARS patients, using these data to assess factors that contribute to an increased risk of an adverse outcome. The definition of adverse outcome varies, but generally includes death, admission to an ICU requiring mechanical ventilation, and development of acute respiratory distress syndrome. Most studies identify older age as a factor that increases the probability of an adverse event.[45,64–66,68,80,82] The strong association between age and CFR is also clearly seen in national case reports, with CFR estimates in those aged over 60 years particularly high.[47] Furthermore, no deaths from SARS occurred in children, who had fewer complications and less severe symptoms.[83,84]

The presence of co-morbidities, including diabetes mellitus, hypertension, coronary artery disease, and chronic obstructive pulmonary disease, also significantly increased the risk of adverse outcomes and death from SARS,[46,64,65,80,82,85] and probably helps to explain the strong association between the CFR and age. Indeed, in some cohorts, deaths almost exclusively occurred in patients with other co-morbidities.[69,85] Other notable factors contributing to higher CFRs were sex (with men at higher risk than women),[45,86] high lactate dehydrogenase concentration at presentation,[45,66] and higher viral loads.[87]

Most patients with SARS received some treatment that was based on previous experience with respiratory infections and evolved over the course of the epidemic. For this reason, most reports of the success of different treatments are observational and may be subject to treatment allocation bias. Common treatments included administration of antibiotics, ribavirin, and corticosteroids,[46,64,69,82,88,89] with the HIV-1 antiviral drug lopinavir also tested later in the epidemic in Hong Kong.[90] In one study in Guangzhou, China, patients were randomly allocated to one of four treatment regimens, with the best response seen in the group receiving early high-dose steroids.[88] However, as was the case

Table 2. Correlates of mortality and mid-epidemic estimates of case-fatality rates

First author	Publication date	Location	Number of patients	Analysis method	Estimated mortality rate	Significant correlates	Non-significant factors
Lee[45]	April 7, 2003	Hong Kong	138	Logistic regression	3·6% died by day 21	Age (p=0·007) Sex (p=0·01)	··
WHO[77]	April 11, 2003	Worldwide	2781	Deaths divided by SARS cases	4%	Age: higher death rate in older patients in Canada	··
Booth[46]	May 6, 2003	Canada	144	Proportional hazards multivariate analysis	6·5% at 21 days	Diabetes: RR 3·1 (95% CI 1·4–7·2) Other comorbid disease:* RR 2·5 (95% CI 1·1–5·8)	Age ⩾60 years RR 1·4 (95% CI 0·95–2·1)
Donnelly[47]	May 7, 2003	Hong Kong	1425	Non-parametric and parametric estimation allowing for censoring	14·9% (non-parametric) 18·2% (parametric)	Age (non-parametric): <60 years 6·8%, ⩾60 years 55·0% Age (parametric): <60 years 13·2%, ⩾60 years 43·3%	··
WHO[48]	May 7, 2003	Worldwide	··	"More reliable methods" than used previously[77]	14–15% overall: 11–17% Hong Kong, 13–15% Singapore, 15–19% Canada, 5–13% China	Age: <25 years <1%, 25–44 years 6%, 45–64 years 15%, ⩾65 years >50%	··
Fowler[80]	July 16, 2003	Toronto	38 adults admitted to ICU	Fisher's exact test and logistic regression	34% at 28 days	Age, diabetes	Sex, occupation (healthcare worker *vs* non-healthcare worker), ischaemic cardiac disease, chronic pulmonary disease.
Lew[67]	July 16, 2003	Singapore	199	Logistic regression of early or intermediate recovery *vs* late recovery or death	10·1% at 28 days	Age: OR for 1 yr increase 1·04 (95% CI 1·01–1·09); APACHE II score: OR for 1 unit increase 1·2 (95% CI 1·05–1·4)	Sex, asthma, diabetes, hypertension, chronic renal failure.
Chan[65]	Aug, 2003	Hong Kong	115	Proportional hazards models	15·7% by May 31, 2003 (outcome known in 100 patients), 10% at 21 days	Age >60 years: HR 3·5 (95% CI 2·8–29·1); diabetes or cardiac disease: HR 9·1 (95% CI 2·8–29·1); other comorbid conditions:† HR 5·2 (95% CI 1·4–19·2)	··
Choi[66]	Nov 4, 2003	Hong Kong	267	Proportional hazards models	12% at 3 months	Age >60 years: HR 5·1 (95% CI 2·3–11·3)	··
Shen[81]	Feb, 2004	Beijing, China	77	Fisher's exact test (two-tailed)	··	Onward transmission: 75% super-spreaders,‡ 16% others.	··

*Defined as chronic obstructive disease, cancer, and cardiac disease. †Defined as hypertension, asthma, and chronic renal failure. ‡Shen and colleagues[81] arbitrarily defined super-spreaders to be those attributed as the source of SARS in at least eight other persons. HR=hazard ratio; ICU=intensive care unit; OR=odds ratio; RR=relative risk; ··=not reported.

for the observational studies, the study was not fully randomised since this fourth group consisted of patients diagnosed later in calendar time than the other three groups.

Transmission dynamics of SARS

Traditional epidemiological approaches cannot be used to assess the population-level risk posed by an emerging infectious disease. The expected number of cases on any given day is determined by the current size of the outbreak, the transmissibility of the disease, and the mixing behaviour of the population, with the infection process causing positive feedback, which results in highly non-linear trends in case incidence over time. Mathematical epidemic models[14] describe (with varying levels of realism) the underlying mechanisms and dynamics of disease progression in the infected individual and transmission in the population. They are therefore also known as mechanistic or dynamical models. It is now common practice for the continuing population-level risk from infectious disease to be assessed using such models.

The most important concept underlying the dynamics of infectious disease epidemics is that of the reproduction number R_t. This is defined to be the average number of new infections caused by one infectious case, over the whole course of that individual's infectious period. At the very start of an outbreak (t=0), the basic reproduction number, R_0, is defined to be the average number of secondary cases caused by the index case in an entirely susceptible population. If R_0 is greater than 1, then an infectious disease outbreak has the potential to establish itself, resulting in an epidemic that will infect a substantial proportion of the population if there is no significant change in either the behaviour (ie, reduced

Table 3. Mathematical transmission models fitted to data

First author	Publication date	Model*	Stochastic	Data†	Explicit SSEs	Mixing	Other key assumptions	Fitting methods	Results
Razum[91]	May 17, 2003	Exponential	No	HK 21/2–5/4	No	Homogeneous	..	LS to cumulative case numbers	Explains why models should not be fitted to cumulative case numbers
Riley[92]	June 20, 2003	SEIHR/D	Yes	HK 26/2–30/4	Yes	Metapopulation (homogeneous within districts)	Interventions reduced both community and hospital transmission; infectiousness reduced by 80% after hospital admission; used realistic incubation distributions.	ML to incidence; used waiting times estimated from individual case reports.	R_0 excluding SSEs=2·7 reduced to 0·14 by end of epidemic; SSE contribution of order 0·3.
Lipsitch[93]	June 20, 2003	SEIR	No	HK 15/2–28/4; World 16/11–20/5	No	Homogeneous epidemic was	Assumed the case, matched growing exponentially (ie, there were no reductions in transmission caused by interventions).	For a given first model to final cumulative case numbers; serial interval estimated from Singapore outbreak.	R_0=2·2–3·6
		Branching process	Yes	HK 15/2–19/4	No	Homogeneous	Assumed that there were no reductions in transmission caused by interventions.	Bayesian estimation with negative binomial distribution of secondary infections and Weibull distribution of serial intervals, both fitted to Singapore data.	R_0 posterior mode=2·2, 95% credible interval 1·5–7·7
Galvani[94]	Aug 8, 2003	Exponential	No	All WHO data 18/3–11/5	No	Homogeneous	..	LS to cumulative case numbers.	Find a negative correlation between doubling time and CFR.
Chowell[95]	Sept 7, 2003	SEIHR	No	World, HK, Canada, Ontario 31/3–14/4	No	Homogeneous	Assumed the epidemic was growing exponentially.	LS to cumulative case numbers; most parameters fixed to plausible values.	R_0=1·1–1·2
Ng[96]	Sept 10, 2003	SEIR	No	HK 17/3–12/5; Beijing, Inner Mongolia 21/4–12/5	No	Homogeneous	Assumed epidemic of unknown virus providing widespread protection to SARS resulted in decline in cases.	LS to cumulative case numbers.	Did not calculate R_0; found that the model had difficulty explaining rapid decline of case numbers.
Choi[53]	Oct 1, 2003	SIHR/D	No	Canada 25/2–26/5	No	Homogeneous	Assumed discrete generations, with a fixed infectious/ incubating period of 5 days and time to death or recovery of 14 days; assumed no hospital transmission.	Fitted by trial and error to cumulative case and death reports.	R_0=1·5, CFR=30%
Wang[97]	Nov 6, 2003	SEQIR	No	Beijing 27/4–2/6	No	Homogenous	Distinguish between suspected and probable cases.	Fit empirical time-dependent rates in simplified model to incidence.	R_0=1·1–3·3
Zhou[98]	Dec 12, 2003	Curve fit	No	Beijing 21/4–24/6; HK 17/3–23/6; Singapore 17/3–30/5	No	Homogenous	..	LS to cumulative case numbers; fit an empirical curve.	R_0=2·7 (Beijing), 2·1 (HK) , 3·8 (Singapore), using method based on initial growth rate.[93]

Continued on next page

Table 3. Mathematical transmission models fitted to data (continued)

First author	Publication date	Model*	Stochastic	Data†	Explicit SSEs	Mixing	Other key assumptions	Fitting methods	Results
Wallinga[99]	Sept 15, 2004	Branching process	Yes	HK, Vietnam, Singapore, Canada	No	No assumptions	Assume homogenous infectiousness	ML of who-infected-whom matrix and serial interval based on Singapore data	Detailed R_t curves, around 3 excluding SSEs, with large reduction to 0·7 after March 12.

*In their simplest form, such model structures divide individuals into three compartments: susceptible (S), infected (I), and recovered (R), with recovered individuals assumed to be immune to further infection; for this reason, such models are often called SIR models. Extensions of SIR models have included additional classes of individuals: exposed (E, also known as latent), hospitalised (H), quarantined (Q), and dead (D). †Region and dates from which data were obtained for analysis. CFR=case fatality rate; HK=Hong Kong; LS=least squares; ML=maximum likelihood; SSE=super-spreading event; ··=not applicable.

mixing) or the susceptibility of the population (ie, through vaccination). If disease spread is not controlled, the number of new cases each day will not start to decline until the pool of susceptible people has been substantially depleted.

The objective of disease control is therefore to reduce R_t to below 1 as quickly as possible, either by reducing contact rates in the population through public-health campaigns or improving hospital infection control. Reducing R_t to below 1 causes a rapid reduction in case incidence long before a significant proportion of the population has been affected. However, the relative ease with which an epidemic can be controlled is critically dependent on the magnitude of R_0.[14] To control an epidemic with R_0=10 through vaccination, 90% of the population need to be immunised, whereas if R_0=2, the level of coverage required drops to 50%.

Table 3 summarises the structure, fitting method, and results of modelling studies published during and after the SARS epidemic. R_0 values from between 1·05 and 7·7 have been estimated for SARS in different settings.[53,92–95,97–99] SSEs (see below) are included in the model presented by Riley and colleagues,[92] but excluded from their R_0 estimate of 2·1–3·7 for SARS in Hong Kong. Wallinga and Teunis[99] used an elegant method based on individual cases to infer the most likely network of contacts between cases from partial exposure data and estimates of the generation time distribution. This method allows direct non-parametric estimation of the mean and distribution of R_t through time. They calculated R_t to be 3 before the WHO global alert on March 12, 2003, and 0·7 afterwards. Lipsitch and colleagues[93] give a much broader range of 1·1–7·7 for R_0. These estimates were derived using two methods, one fitting exponential growth to three pre-selected timepoints for data from Hong Kong, Canada, and Singapore, and a second, more sophisticated approach, resembling that used by Wallinga and Teunis.[99] Overall, SARS can be classified as moderately transmissible, with smallpox being slightly more transmissible (R_0=4–10)[100] and measles being much more transmissible (R_0=7–45).[101]

All the regional SARS outbreaks were controlled relatively quickly, implying that R_t was rapidly brought below 1. However, the reason for this drop in R_t is the subject of some debate. Riley and colleagues[92] showed that very little of the drop in R_t can be attributed to changes in the speed with which people are admitted to hospital, and suggested that it was due to a general reduction in overall mixing in the community, coupled with improved infection control in hospitals. Lloyd-Smith and co-workers[102] use a relatively complex model of community and hospital transmission to show that it was improved infection control in hospital that contributed most to reductions in transmissibility.

The reproduction number is, of course, not the only factor determining how difficult it is to control an epidemic. The generation time, T_g (also called the serial interval), the average time between a person being infected and infecting others, also plays an important role in determining the feasibility of any control measures. Diseases with a very short generation time, such as influenza (T_g=3 days), spread so rapidly that reactive control measures such as contact tracing are unlikely to be practical. Conversely, smallpox may be comparably transmissible but has a much longer generation time (T_g >14 days), making contact tracing, pre-infectious diagnosis, and isolation or treatment much more feasible.

The third key factor determining the likely success of simple public-health interventions (such as isolation or contact tracing) against an emerging disease is when infectiousness occurs during disease progression.[103] Diseases for which a substantial proportion of transmission occurs before the onset of clear symptoms make reactive control measures such as case isolation less effective. However, if symptoms nearly always precede the onset of substantial infectiousness (as was the case with SARS), rapid and effective diagnosis, hospital admission, and isolation of clinical cases is predicted to be a highly effective means to control transmission. This issue is the topic of a recent analytical study,[103] which examined the formal relation between the outcome of public-health measures, R_0, and the proportion of pre-symptomatic or asymptomatic transmission.

The main benefit of mechanistic models, compared with purely descriptive models, is their ability to allow the exploration of hypothetical situations (table 4). This can take the form of examination of the impact of a range of potential control options on case incidence (such as the imposition of movement restrictions,[92] or improved quarantine and contact tracing[93]), or the investigation of disease spread in a novel setting (such as Japan, where there were no SARS cases[104]). Such investigations do, by definition, involve extrapolation beyond the observed data. However, when presented with careful sensitivity analyses that show the extent to which key results depend on model assumptions, these studies can provide valuable insights to scientists and public-health policy makers.

Table 4. Mathematical transmission models used to explore hypothetical situations

First author	Publication date	Model*	Stochastic	Explicit SSEs	Mixing	Other key assumptions	Parameter choice	Results
Riley[92]	June 20, 2003	SEIHR/D	Yes	Yes	Meta-population (homogeneous within districts)	..	From their best fit model (above)	Movement restrictions between districts would have been able to stop an otherwise uncontrolled Hong-Kong-like epidemic.
Lipsitch[93]	June 20, 2003	SQEIHR	No	No	Homogeneous	Assumed quarantining occurred instantaneously after contact with infective; assumed patients could be perfectly isolated in hospitals.	From their best fit model (above).	Quarantine and accelerated isolation could be expected to control SARS.
Lloyd-Smith[102]	July 30, 2003	SQEIHR	Yes	No	Separate core-group of health-care workers, otherwise homogeneous	Assumed quarantining occurred instantaneously after contact with infective; used realistic incubation distributions.	From earlier studies.	Control of nosocomial transmission was key to controlling SARS.
Nishiura[104]	March 1, 2004	SQEIHR	No	No	Homogeneous	Same model as Lipsitch.[93]	From Lipsitch.[93]	If SARS were to re-emerge in an environment where it could be controlled (such as Japan), the number of people infected would most strongly depend on the initial number of cases.
Masuda[105]	Mar 31, 2004	Individual-based simulation	Yes	Yes	Realistic "small-world" social network	..	From earlier studies and from Singapore contact tracing data.	SSEs did not arise from highly-connected individuals, but were a different transmission process; transmission patterns were not consistent with a scale-free social network.
Fraser[103]	April 7, 2004	Individual-based model, with isolation and quarantining	Yes	No	Homogenous	Model explores interplay between appearance of symptoms and changing infectiousness as a function of time since infection.	Based on collated studies of SARS, HIV, influenza, and smallpox.	Because infectiousness does not peak until long after symptoms, SARS can be contained by isolation alone, though quarantining helps counter logistical delays; smallpox, which is more infectious, can be contained using isolation and quarantining; HIV and pandemic influenza cannot.

*In their simplest form, such model structures divide individuals into three compartments: susceptible (S), infected (I), and recovered (R), with recovered individuals assumed to be immune to further infection; for this reason, such models are often called SIR models. Extensions of SIR models have included additional classes of individuals: exposed (E, also known as latent), hospitalised (H), quarantined (Q), and dead (D).

Heterogeneity in transmission: the role of SSEs

Heterogeneity in contact rates or infectiousness has been recognised as a key factor in determining patterns of infectious disease spread for many years.[14] However, for SARS the importance of such heterogeneity was particularly underscored by the occurrence of a few dramatic SSEs in which single individuals were responsible for infecting many times more individuals than the average (given by R_t). The examples of patient 1 in Hong Kong, who infected 10 people in the Metropole Hotel (known as Hotel M) and additional people after his admission to St Paul's Hospital, Hong Kong,[106] and the Amoy Gardens[107,108] cluster in Hong Kong are the best known, but patients who generated large numbers (>10) of secondary cases were also identified in Singapore (with at least five such patients[49]) and Canada.[52] However, super-spreading individuals are not unique to SARS. Their existence has been well documented for tuberculosis,[109] measles,[110,111] and smallpox,[112] and they are believed to have occurred in other diseases including Ebola[113] and the zoonotic transmission of monkeypox.[114] Furthermore, the importance of a small number of individuals with high rates of partner change is critical to the epidemiology of many sexually transmitted infections.[115,116]

There will, of course, be variability in the number of secondary cases from any primary case owing to random variation, even without any underlying variation due to characteristics of the primary case. However, if all cases have identical levels and durations of infectiousness with constant contact probabilities, such variation is expected to be Poisson.[117] Once variation in the duration of infectiousness period is allowed for, higher than Poisson variance is expected, with negative binomially distributed numbers of secondary cases expected for exponentially distributed infectiousness.[93,117–119]

However, key to the debate surrounding SSEs is whether such events are merely the extreme tail of a continuous distribution[93] or they represent a distinct separate class of cases.[92,105] Although some of the SSEs (particularly those in Hong Kong) seem too extreme to have arisen from an underlying continuous distribution, it should be noted that estimating the frequency of SSEs from case data in a single region is subject to severe selection biases. This is because in

the earliest stages of a local outbreak the occurrence of an SSE dramatically lowers the chances of that outbreak becoming extinct by chance.[93] Hence, SSEs are more likely to have occurred early in the outbreaks in those locations where large outbreaks were seen (eg, Toronto, Hong Kong, Singapore). Ideally, one would like to characterise the overall distribution of secondary case numbers for SARS before and after controls were introduced. To do this rigorously would require a global analysis, given the early importance of international spread. However, the detailed contact tracing data required for such an analysis does not exist in some areas, and are incomplete for others where large outbreaks were seen.

Irrespective of whether SSEs are a discrete class of transmission events or the tail of a distribution, it cannot be assumed that variation in secondary case numbers is primarily due to biological variation in the amount of virus shed by patients and hence their infectiousness. Variation in contact rates with other individuals in the population is also likely to have been important, and may indeed have been the dominant factor explaining SSEs. Such variation might be in the frequency of direct contacts (eg, large numbers of medical personnel saw the index patient in the Prince of Wales Hospital in Hong Kong[120]) or indirect contacts (eg, unusual modes of viral spread in the Metropole Hotel and Amoy Gardens[107,108]). Characterisation of heterogeneity in contact rates has been the topic of much research in infectious disease epidemiology,[14,121,122] and various modelling approaches (including stratified population models[123–126] and individual-based network models[127]) have been developed to incorporate such heterogeneity. Masuda and colleagues[105] use a network-based approach to model SARS, and concluded that SSEs are best explained by an increase in infectiousness in a few individuals, rather than extreme contact-rate heterogeneity. However, these conclusions are dependent on the investigators' simplifying assumptions about network structure. Overall, identifying the biological, social, or environmental causes for SSEs is important for the development of strategies for efficiently preventing or controlling such events, since the optimal choice of tactics to be employed will depend on the causative mechanisms.

Conclusions

Despite the substantial achievements already made in understanding the origin and determinants of spread of the SARS epidemics, important questions remain unanswered. These include clarification of how, if at all, seasonality contributed to the epidemic patterns observed; understanding precisely how transmission took place within particular settings (eg, hospital wards); gaining insight into the extent to which differences in social networks contributed to heterogeneity in SARS transmission; and determining the zoonotic origins of the virus. Answering these questions will depend mainly on the reliability and availability of the relevant data and will require multiple methodological approaches.

For example, investigating the impact, if any, of seasonality on transmission would require coordinated modelling of the large SARS epidemics (in China, Hong Kong, Taiwan, Singapore, and Canada) to separate the effects of temporal changes in humidity, temperature, and other environmental factors from the effects of temporal changes in epidemiological factors, such as contact tracing and reduced mixing. Greater understanding of within-ward transmission would be gained into risk factors associated with both infectiousness and susceptibility from detailed stochastic modelling of patient and healthcare worker contacts. This work could usefully build on published Markov chain Monte Carlo models of nosocomial transmission.[128]

As well as giving greater insights into processes underlying the SARS epidemic, the models and estimation methods developed will strengthen the set of analytical tools available for the analysis of future epidemics. In each case, surveillance and data quality are fundamental to providing sound foundations to underpin analyses and conclusions. Contingency plans developed on the basis of the experience of the SARS epidemics have rightly placed a high priority on both surveillance and contact tracing.[129,130]

Epidemiological modelling has clarified the types of diseases that can be controlled with the straightforward public measures of isolation and contact tracing.[103] More detailed modelling will be required to further clarify the potential impact of further measures, including restrictions on both short-range and long-range movements of people. A review of recent smallpox modelling[131] cautioned that modelling efforts should not set the nearly impossible goal of identifying the best public-health strategy in advance of an epidemic, but should identify sets of recommended actions with associated decision rules for adaptive management as an epidemic unfolds.

Search strategy and selection criteria

Data for this review were identified by searches of PubMed, ISI Web of Science, Medline, and references from relevant articles. Search terms were "severe acute respiratory syndrome", "SARS", "incubation period", "mortality", "attack rate", and "model". Only English language papers were reviewed. Publication dates (based on electronic publication for those many SARS publications published in this fast-track manner) are given in tables to indicate the timeline of data availability and analysis.

Conflicts of interest

We declare that we have no conflicts of interest.

References

1 WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 (based on data as of the 31 December 2003). http://www.who.int/csr/sars/country/table2004_04_21/en/ (accessed Sept 22, 2004).

2 Centers for Disease Control and Prevention. Revised US surveillance case definition for severe acute respiratory syndrome (SARS) and update on SARS cases: United States and worldwide, December 2003. *MMWR Morb Mortal Wkly Rep* 2003; **52:** 1202–06.

3 Centers for Disease Control and Prevention. Revised CSTE SARS surveillance case definition [appendix 1 of supplement B of the public health guidance for community-level preparedness response to severe acute respiratory syndrome (SARS)]. Atlanta, GA: Centers for Disease Control and Prevention, 2004.

4 Lim PL, Kurup A, Gopalakrishna G, et al. Laboratory-acquired severe acute respiratory syndrome. *N Engl J Med* 2004; **350:** 1740–45.

5 Orellana C. Laboratory-acquired SARS raises worries on biosafety. *Lancet Infect Dis* 2004; **4:** 64–64.

6 Paterson R. SARS returns to China. *Lancet Infect Dis* 2004; **4:** 64.

7 WHO. China's latest SARS outbreak has been contained, but biosafety concerns remain—update 7.

http://www.who.int/csr/don/2004_05_18a/en/ (accessed Sept 30, 2004).
8 Normile D. Mounting lab accidents raise SARS fears. *Science* 2004; **304:** 659–61.
9 Guan Y, Zheng BJ, He YQ, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. *Science* 2003; **302:** 276–78.
10 Zhong NS, Zheng BJ, Li YM, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* 2003; **362:** 1353–58.
11 United Nations, Population Division, Department of Economic and Social Affairs. The world at six billion, 1999. http://www.un.org/esa/population/publications/sixbillion/sixbillion.htm (accessed Oct 4, 2004).
12 United Nations Centre for Human Settlements, Habitat. The state of the world's cities 2001. http://www.unchs.org/istanbul+5/statereport.htm (accessed Oct 4, 2004).
13 Wacholder S, Hartge P. Case-control study. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: Wiley, 1998: 503–14.
14 Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford: Oxford Science Publications, 1991.
15 Bernoulli D. Essai d'une nouvelle analyse de la mortalite causee par la petite verole et des advantages de l'inoculation pour la prevenir. *Mem Math Phys Acad R Sci Paris* 1760: 1–45.
16 Ross R. Some a priori pathometric equations. *Br Med J* 1915; **1:** 546–47.
17 Ross R. An application of the theory of probabilities to the study of a priori pathometry, I. *Proc R Soc A* 1916; **92:** 204–30.
18 Ross R. An application of the theory of probabilities to the study of a priori pathometry, II. *Proc R Soc A* 1917; **93:** 212–25.
19 Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc A* 1927; **115:** 700–21.
20 Marra MA, Jones SJ, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. *Science* 2003; **300:** 1399–404.
21 Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003; **300:** 1394–99.
22 Snijder EJ, Bredenbeek PJ, Dobbe JC, et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 2003; **331:** 991–1004.
23 Gibbs AJ, Gibbs MJ, Armstrong JS. The phylogeny of SARS coronavirus. *Arch Virol* 2004; **149:** 621–24.
24 Stavrinides J, Guttman DS. Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J Virol* 2004; **78:** 76–82.
25 Lio P, Goldman N. Phylogenomics and bioinformatics of SARS-CoV. *Trends Microbiol* 2004; **12:** 106–11.
26 Qin E, Zhu Q, Yu M, et al. A complete sequence and comparative analysis of a SARS-associated virus (isolate BJ01). *Chin Sci Bull* 2003; **48:** 941–48.
27 Eickmann M, Becker S, Klenk HD, et al. Phylogeny of the SARS coronavirus. *Science* 2003; **302:** 1504–05.
28 Lee CW, Jackwood MW. Spike gene analysis of the DE072 strain of infectious bronchitis virus: origin and evolution. *Virus Genes* 2001; **22:** 85–91.
29 Lee CW, Jackwood MW. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch Virol* 2000; **145:** 2135–48.
30 Rest JS, Mindell DP. SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect Genet Evol* 2003; **3:** 219–25.
31 Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 2004; **303:** 1666–69.
32 Yu D, Li H, Xu R, et al. Prevalence of IgG Antibody to SARS-associated coronavirus in animal traders: Guangdong Province, China, 2003. *MMWR Morb Mortal Wkly Rep* 2003; **52:** 986–87.
33 Guan Y, Peiris JS, Zheng B, et al. Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *Lancet* 2004; **363:** 99–104.
34 Normile D, Enserink M. Tracking the roots of a killer. *Science* 2003; **301:** 297–99.
35 Enserink M, Normile D. Search for SARS origins stalls. *Science* 2003; **302:** 766–67.
36 WHO. Visit of Second WHO Expert Team on SARS to Guangdong Province: report and recommendations 5–12 May 2003. http://www.cbc.ca/disclosure/archives/031118_sars/docs/WHOExpertTeamReport.pdf (accessed Sept 22, 2004).
37 Ruan YJ, Wei CL, Ee AL, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* 2003; **361:** 1779–85.
38 Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000; **288:** 1789–96.
39 Zeng F, Chow KY, Leung FC. Estimated timing of the last common ancestor of the SARS coronavirus. *N Engl J Med* 2003; **349:** 2469–70.
40 Brookmeyer R. Incubation period of infectious diseases. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: Wiley, 1998: 2011–16.
41 Sun J. Interval censoring. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: Wiley, 1998: 2090–95.
42 Tsang K, Ho P, Ooi G, et al. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003; **348:** 1977–85.
43 WHO. Preliminary clinical description of severe acute respiratory syndrome. www.who.int/csr/sars/clinical/en/ (accessed Sept 22, 2004).
44 Poutanen SM, Low DE, Henry B, et al. Identification of severe acute respiratory syndrome in Canada. *N Engl J Med* 2003; **348:** 1995–2005.
45 Lee N, Hui D, Wu A, et al. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003; **348:** 1986–94.
46 Booth CM, Matukas LM, Tomlinson GA, et al. Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area. *JAMA* 2003; **289:** 2801–09.
47 Donnelly CA, Ghani AC, Leung GM, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 2003; **361:** 1761–66.
48 WHO. Update 49—SARS case fatality ratio, incubation period. http://www.who.int/csr/sars/archive/2003_05_07a/en/ (accessed Sept 22, 2004).
49 Leo YS, Chen M, Heng BH, et al. Severe acute respiratory syndrome—Singapore, 2003. *MMWR Morb Mortal Wkly Rep* 2003; **52:** 405–11.
50 Wu W, Wang JF, Liu PM, et al. A hospital outbreak of severe acute respiratory syndrome in Guangzhou, China. *Chin Med J* 2003; **116:** 811–18.
51 Avendano M, Derkach P, Swan S. Clinical course and management of SARS in health care workers in Toronto: a case series. *Can Med Assoc J* 2003; **168:** 1649–60.
52 Varia M, Wilson S, Sarwal S, et al. Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada. *Can Med Assoc J* 2003; **169:** 285–92.
53 Choi BCK, Pak AWP. A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths. *J Epidemiol Community Health* 2003; **57:** 831–35.
54 WHO Department of Communicable Disease Surveillance and Response. Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). WHO/CDS/CSR/GAR/2003.11. http://www.who.int/csr/sars/en/WHOconsensus.pdf (accessed Sept 22, 2004).
55 Olsen SJ, Chang H, Cheung TY, et al. Transmission of the severe acute respiratory syndrome on aircraft. *N Engl J Med* 2003; **349:** 2416–22.
56 Chow KY, Lee CE, Ling ML, Heng DMK, Yap SG. Outbreak of severe acute respiratory syndrome in a tertiary hospital in Singapore, linked to an index patient with atypical presentation: epidemiological study. *BMJ* 2004; **328:** 195–98.
57 Meltzer MI. Multiple contact dates and SARS incubation periods. *Emerg Infect Dis* 2004; **10:** 207–09.
58 Xu J, Yang M, Liu Z. Clinical analysis of patients with severe acute respiratory syndrome in Beijing area[in Chinese]. *Zhonghua Jie He He Hu Xi Za Zhi* 2003; **26:** 683–85.
59 Huo N, Lu H, Xu X, et al. The clinical characteristics and outcome of 45 early stage patients with SARS[in Chinese]. *Beijing Da Xue Xue Bao* 2003; **35** (suppl)**:** 19–22.
60 Lu H, Huo N, Xu X, et al. The epidemiologic characteristics of patients with severe acute respiratory syndrome (SARS) [in Chinese]. *Beijing Da Xue Xue Bao* 2003; **35** (suppl)**:** 8–11.
61 He X, Shen Z, Ning F, et al. Analysis on the epidemiological features and the transmission of an imported severe acute respiratory syndrome case in Beijing [in Chinese]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2003; **24:** 557–60.
62 He J, Xu R, Yu D, et al. Severe acute respiratory syndrome in Guangdong Province of China: epidemiology and control measures [in Chinese]. *Zhonghua Yu Fang Yi Xue Za Zhi* 2003; **37:** 227–32.
63 Li Q, Zeng G, Ou J, Guo G. Epidemiological study of the transmission chain of a severe acute respiratory syndrome outbreak [in Chinese]. *Zhonghua Yi Xue Za Zhi* 2003; **83:** 906–09.
64 Peiris JS, Chu CM, Cheng VC, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 2003; **361:** 1767–72.
65 Chan JW, Ng CK, Chan YH, et al. Short term outcome and risk factors for adverse clinical outcomes in adults with severe acute respiratory syndrome (SARS). *Thorax* 2003; **58:** 686–89.
66 Choi KW, Chau TN, Tsang O, et al. Outcomes and prognostic factors in 267 patients with severe acute respiratory syndrome in Hong Kong. *Ann Intern Med* 2003; **139:** 715–23.
67 Lew TW, Kwek TK, Tai D, et al. Acute respiratory distress syndrome in critically ill patients with severe acute respiratory syndrome. *JAMA* 2003; **290:** 374–80.
68 Chan MH, Wong VW, Wong CK, et al. Serum LD1 isoenzyme and blood lymphocyte subsets as prognostic indicators for severe acute respiratory syndrome. *J Intern Med* 2004; **255:** 512–18.
69 Sung JJ, Wu A, Joynt GM, et al. Severe acute respiratory syndrome: report of treatment and outcome after a major outbreak. *Thorax* 2004; **59:** 414–20.
70 Xiao Z, Li Y, Chen R, Li S, Zhong S, Zhong N. A retrospective study of 78 patients with severe acute respiratory syndrome. *Chin Med J (Engl)* 2003; **116:** 805–10.
71 Lau JT, Lau M, Kim JH, Tsui HY, Tsang T, Wong TW. Probable secondary infections in households of SARS patients in Hong Kong. *Emerg Infect Dis* 2004; **10:** 235–43.
72 Goh DL, Lee BW, Chia KS, et al. Secondary household transmission of SARS, Singapore. *Emerg Infect Dis* 2004; **10:** 232–34.
73 Pang X, Zhu Z, Xu F, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. *JAMA* 2003; **290:** 3215–21.
74 Leung GM, Chung P-H, Tsang T, et al. SARS-CoV antibody prevalence in all Hong Kong patient contacts. *Emerg Infect Dis* 2004; **10:** 1653–56.
75 Lau JT, Fung KS, Wong TW, et al. SARS transmission among hospital workers in Hong Kong. *Emerg Infect Dis* 2004; **10:** 280–86.
76 Vu HT, Leitmeyer KC, Le DH, et al. Clinical description of a completed outbreak of SARS in Vietnam, February-May 2003. *Emerg Infect Dis* 2004; **10:** 334–38.
77 WHO. SARS epidemiology to date. http://www.who.int/csr/sars/epi2003_04_11/en/ (accessed Sept 22, 2004).
78 Altman LK. Death rate from virus more than doubles, varying sharply by country. *New York Times*, April 22, 2003.
79 CNN. SAR becoming deadlier: officials (published April 25, 2003). http://www.cnn.com/2003/HEALTH/04/24/sars.death/ (accessed Sept 22, 2004).
80 Fowler RA, Lapinsky SE, Hallett D, et al. Critically ill patients with severe acute respiratory syndrome. *JAMA* 2003; **290:** 367–73.
81 Shen Z, Ning F, Zhou WG, et al. Superspreading SARS events, Beijing, 2003. *Emerg Infect Dis* 2004; **10:** 256–60.
82 Gomersall CD, Joynt GM, Lam P, et al. Short-term outcome of critically ill patients with severe acute respiratory syndrome. *Intensive Care Med* 2004; **30:** 381–87.
83 Babyn PS, Chu WC, Tsou IY, et al. Severe acute respiratory syndrome (SARS): chest radiographic features in children. *Pediatr Radiol* 2004; **34:** 47–58.
84 Wong GW, Li AM, Ng PC, Fok TF. Severe acute respiratory syndrome in children. *Pediatr Pulmonol* 2003; **36:** 261–66.
85 Wong WW, Chen TL, Yang SP, et al. Clinical characteristics of fatal patients with severe acute respiratory syndrome in a medical center in Taipei. *J Chin Med Assoc* 2003; **66:** 323–27.
86 Karlberg J, Chong DS, Lai WY. Do men have a higher case fatality rate of severe acute respiratory syndrome than women do? *Am J Epidemiol* 2004; **159:** 229–31.
87 Chong PY, Chui P, Ling AE, et al. Analysis of deaths

during the severe acute respiratory syndrome (SARS) epidemic in Singapore: challenges in determining a SARS diagnosis. *Arch Pathol Lab Med* 2004; **128:** 195–204.
88 Zhao Z, Zhang F, Xu M, et al. Description and clinical treatment of an early outbreak of severe acute respiratory syndrome (SARS) in Guangzhou, PR China. *J Med Microbiol* 2003; **52** (pt 8)**:** 715–20.
89 Tsang K, Seto WH. Severe acute respiratory syndrome: scientific and anecdotal evidence for drug treatment. *Curr Opin Investig Drugs* 2004; **5:** 179–85.
90 Chu CM, Cheng VC, Hung IF, et al. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* 2004; **59:** 252–56.
91 Razum O, Bacher H, Kapaum A, Junghanss T. SARS, lay epidemiology and fear. *Lancet* 2003; **361:** 1739–40.
92 Riley S, Fraser C, Donnelly CA, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science* 2003; **300:** 1961–66.
93 Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 2003; **300:** 1966–70.
94 Galvani AP, Lei X, Jewell NP. Severe acute respiratory syndrome: temporal stability and geographic variation in case-fatality rates and doubling times. *Emerg Infect Dis* 2003; **9:** 991–94.
95 Chowell G, Fenimore PW, Castillo-Garsow MA, Castillo-Chavez C. SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism. *J Theor Biol* 2003; **224:** 1–8.
96 Ng TW, Turinici G, Danchin A. A double epidemic model for the SARS propagation. *BMC Infect Dis* 2003; **3:** 19.
97 Wang WD, Ruan SG. Simulating the SARS outbreak in Beijing with limited data. *J Theor Biol* 2004; **227:** 369–79.
98 Zhou GF, Yan GY. Severe acute respiratory syndrome epidemic in Asia. *Emerg Infect Dis* 2003; **9:** 1608–10.
99 Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004; **160:** 509–16.
100 Eichner M, Dietz K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am J Epidemiol* 2003; **158:** 110–17.
101 Wallinga J, Levy-Bruhl D, Gay NJ, Wachmann CH. Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiol Infect* 2001; **127:** 281–95.
102 Lloyd-Smith JO, Galvani AP, Getz WM. Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. *Proc R Soc Lond B Biol Sci* 2003; **270:** 1979–89.
103 Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. *Proc Natl Acad Sci USA* 2004; **101:** 6146–51.
104 Nishiura H, Patanarapelert K, Sriprom M, Sarakorn W, Sriyab S, Tang IM. Modelling potential responses to severe acute respiratory syndrome in Japan: the role of initial attack size, precaution, and quarantine. *J Epidemiol Community Health* 2004; **58:** 186–91.
105 Masuda N, Konno N, Aihara K. Transmission of severe acute respiratory syndrome in dynamical small-world networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004; **69** (3 pt 1)**:** 031917 (published online March 31, 2004).
106 Tsang T, Lai-Yin T, Pak-Yin L, et al. Update: outbreak of severe acute respiratory syndrome—worldwide, 2003. *MMWR Morb Mortal Wkly Rep* 2003; **52:** 241–48.
107 Ng SKC. Possible role of an animal vector in the SARS outbreak at Amoy Gardens. *Lancet* 2003; **362:** 570–72.
108 Yu ITS, Li YG, Wong TW, et al. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N Engl J Med* 2004; **350:** 1731–39.
109 Valway SE, Sanchez MPC, Shinnick TF, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* 1998; **338:** 633–39.
110 Chen RT, Goldbaum GM, Wassilak SGF, Markowitz LE, Orenstein WA. An explosive point-source measles outbreak in a highly vaccinated population: modes of transmission and risk-factors for disease. *Am J Epidemiol* 1989; **129:** 173–82.
111 Paunio M, Peltola H, Valle M, Davidkin I, Virtanen M, Heinonen OP. Explosive school-based measles outbreak: intense exposure may have resulted in high risk, even among revaccinees. *Am J Epidemiol* 1998; **148:** 1103–10.
112 Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID. Smallpox and its eradication. Geneva: WHO, 1988.
113 Khan AS, Tshioko FK, Heymann DL, et al. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo, 1995. *J Infect Dis* 1999; **179** (suppl 1)**:** S76–86.
114 Altman LK. One prairie dog plays critical role in Wisconsin. *New York Times*, June 14, 2003: 10.
115 Yorke JA, Hethcote HW, Nold A. Dynamics and control of the transmission of gonorrhea. *Sex Transm Dis* 1978; **5:** 51–57.
116 Garnett GP. The geographical and temporal evolution of sexually transmitted disease epidemics. *Sex Transm Infect* 2002; **78** (suppl 1)**:** i14–19.
117 Farrington CP, Kanaan MN, Gay NJ. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* 2003; **4:** 279–95.
118 Cairns AJG. Epidemics in heterogeneous populations. 2. Nonexponential incubation periods and variable infectiousness. *IMA J Math Appl Med Biol* 1990; **7:** 219–30.
119 Lloyd AL. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* 2001; **60:** 59–71.
120 Tomlinson B, Cockram C. SARS: experience at Prince of Wales Hospital, Hong Kong. *Lancet* 2003; **361:** 1486–87.
121 Grenfell BT, Anderson RM. The estimation of age-related rates of infection from case notifications and serological data. *J Hyg Camb* 1985; **95:** 419–36.
122 Farrington CP, Kanaan MN, Gay NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *J R Stat Soc C Appl Stat* 2001; **50:** 251–83.
123 Hethcote HW, Yorke JA. Gonorrhea: transmission dynamics and control. New York: Springer, 1984.
124 Schenzle D. An age-structured model of pre- and post-vaccination measles transmission. *IMA J Math Appl Med Biol* 1984; **1:** 169–91.
125 Sattenspiel L, Dietz K. A structured epidemic model incorporating geographic mobility between regions. *Math Biosci* 1995; **128:** 71–91.
126 Keeling MJ, Rand DA, Morris AJ. Correlation models for childhood epidemics. *Proc R Soc Lond B Biol Sci* 1997; **264:** 1149–50.
127 Eubank S, Guclu H, Kumar VS, et al. Modelling disease outbreaks in realistic urban social networks. *Nature* 2004; **429:** 180–84.
128 Starr JM, Campbell A. Mathematical modeling of Clostridium difficile infection. *Clin Microbiol Infect* 2001; **7:** 432–37.
129 CDC Department of Health and Human Services. Public health guidance for community-level preparedness and response to severe acute respiratory syndrome (SARS), version 2. Atlanta, GA: CDC, 2004.
130 Health Protection Agency. Interim contingency plan for severe acute respiratory syndrome (SARS). http://www.hpa.org.uk/infections/topics_az/SARS/pdfs/SARSContingencyDec03.pdf (accessed Sept 22, 2004).
131 Ferguson NM, Keeling MJ, Edmunds WJ, et al. Planning for smallpox outbreaks. *Nature* 2003; **425:** 681–85.