Real-time epidemiology
Understanding the spread of SARS

Fitting transmission models to outbreak data allows key epidemiological parameters to be estimated and the impact of control measures to be evaluated. Christl Donnelly and Azra Ghani describe the use of such models in the 2003 SARS epidemic in Hong Kong.

Are infectious diseases still important?
Infectious diseases have been a major cause of mortality and thus an important force of natural selection in both animal and plant communities. Even in human infectious diseases remain a major public health problem in developing countries. Although mass immunisation and the use of anti-microbial drugs substantially reduced their impact in the developed world in the 20th century, continued progress is being put at risk by emerging pathogens (e.g. HIV/AIDS, new variant Creutzfeldt–Jakob disease, novel influenza strains) and drug-resistant strains of previously controlled diseases (e.g. tuberculosis). Similarly, novel and re-emerging zoonoses—infectious diseases transmitted from wild and domestic animals to humans, e.g. bovine spongiform encephalopathy (BSE)—pose substantial risks to animal and human health in many societies.

The volume and detail of epidemiological data on infectious agents have been increasing rapidly owing to the growth of computerised databases recording patterns of infection-related morbidity and mortality; the increased level of socio-spatial and temporal data stratification available for disease notification data and the increasing availability of molecular epidemiological data in the form of genome sequences and epidemiologically linked phylogenetic trees. New statistical and mathematical tools are required for analysis and interpretation of patterns to gain insights into the underlying biological and epidemiological processes.

What are transmission models and why use them?
To understand the factors determining the spread of infectious diseases it is necessary to model the interaction between the pathogen and its host, usually taking into account the biology of the disease and the social behaviour of its hosts. Such non-linear models, often referred to as mathematical models of infectious disease, are mechanistic in that they describe the underlying mechanism of infection to some degree. Such models are, in fact, no more or less mathematical than traditional epidemiological and statistical models. However, historically they have been developed within the mathematical biology literature rather than within the statistics literature.
Mechanistic transmission models share certain common features: they compartmentalise the population by disease state (into susceptible, infectious and recovered individuals, for instance) and represent the epidemic process as a feedback loop in which the rate at which susceptible individuals become infected is proportional to the number of infectious individuals in the population (the so-called “mass-action principle”). This central density-dependent non-linearity of epidemic transmission models generates the complex dynamics exhibited by many disease systems. The dramatic differences in the behaviour of different host–pathogen systems arise from the details of how the structure of the host and/or pathogen populations affects the infection contact process.

Like other statistical modelling efforts, the development of such mechanistic transmission models is an iterative process of model specification, parameter estimation and validation. In early applications of transmission models, parameter estimation methods often lacked robustness. Deterministic modelling approaches are increasingly augmented with stochastic frameworks that take account of variations in space and time that are often important determinants of disease persistence and extinction in host communities. Such models, informed by statistical analyses of data, enable quantitative design and evaluation of targetted disease prevention and control programmes.

The most important parameter describing the transmission dynamics of an infectious agent is the basic reproduction number, typically denoted \( R_0 \), defined to be the average number of secondary cases caused by a primary case in a fully susceptible population. Thus, if \( R_0 \) is greater than 1, then, in a deterministic framework, the number of infections will increase, though possibly quite gradually at first. Similarly, in stochastic models, infection incidence will most probably increase if \( R_0 \) is greater than 1, but there is a non-zero probability that the disease will go extinct. Thus, \( R_0 \) determines the proportion of infections that must be prevented to force an outbreak into decline.

Two parameters that indicate how such a reduction in infection incidence might be achieved are the disease generation time and the proportion of infections that occur prior to the appearance of symptoms, \( \Theta \). The generation time sets the pace of the epidemic, thus determining how quickly control measures must be enforced, whereas \( \Theta \) determines the proportion of infections that can be prevented solely by the isolation of symptomatic patients.

### Statistical challenges posed by infectious disease data

With the increasing realism and power of mechanistic transmission models come increasing complexity and consequent difficulties in model parameterisation and validation. Rigorous parameter estimation is particularly important in the application of non-linear models, because their qualitative behaviour can often change dramatically with even small changes in parameter values. For deterministic models with few parameters, it was often feasible to characterise the nature of different dynamical regimes completely and so to gain at least a qualitative understanding of the degree to which a model was consistent with epidemiological data. However, even such simple procedures become computationally burdensome as models become more detailed. Advances in statistical techniques and computing resources have greatly extended the range of situations where it is now feasible to apply increasingly complex transmission models; complete with parameter estimates, confidence intervals and goodness-of-fit assessment.

Many of the recent developments have been in the area of model specification and stochastic process theory. Analytically well-defined approximations to non-linear stochastic systems are giving greater insight into the aggregation and dispersion processes that generate between-host variability (e.g. in macro-parasite burden or in spatial disease clusters). Such methods are considerably less computationally demanding than full micro-simulation approaches but are still an order of magnitude more complex than traditional deterministic models.

The problem of parameter estimation can typically be reduced to that of maximising a non-linear multivariate function—the likelihood or posterior probability density. Even where the likelihood is easily evaluated, the problems inherent in finding the global maximum...
within highly complex multi-dimensional spaces are substantial. Recent progress in the development of numerical techniques to solve such problems has been considerable, with Markov chain Monte Carlo (MCMC) and novel deterministic algorithms being particularly promising. Where explicit evaluation of the likelihood is impractical—such as for many stochastic non-linear systems with incomplete data—stochastic integration techniques have also come to the fore. The accurate characterisation of uncertainty in parameter estimates and more general sensitivity analysis of model structure are particularly important for epidemiological modelling, where wide confidence intervals on some parameters may have serious implications for predictability, for instance of the effect of proposed control strategies.

“The accurate characterisation of uncertainty in parameter estimates and more general sensitivity analysis of model structure are particularly important for epidemiological modelling”

SARS analysis and modelling

The coronavirus which causes severe acute respiratory syndrome (SARS) spread rapidly, with the first identified human case in Guangdong Province, China, in November 2002 and 27 countries reporting SARS cases in 2003. The last known case of the initial epidemic experienced the onset of symptoms in June 2003 in Taiwan. (However, owing to differences in case definitions, the USA has reported probable cases of SARS with onsets of illness after July 5th, 2003.) Cases have subsequently arisen in Singapore, Taiwan and China due to laboratory-related infections and onward transmission. Worldwide surveillance identified 8966 clinically affected SARS cases, of whom 774 died as a result (http://www.who.int/csr/sars/).

The Department of Infectious Disease Epidemiology at Imperial College in London was approached in March 2003 by the Department of Community Medicine in the University of Hong Kong to collaborate with academic
colleagues and public health authorities, in order to inform policy makers through analysis and modelling of the SARS epidemic in Hong Kong. An outbreak response team was formed (as happened in 2001 during the British foot-and-mouth disease epidemic) with key members Roy Anderson, Steven Riley, Christophe Fraser, Neil Ferguson and ourselves. Assistance was also sought from Sir David Cox who provided virtually instantaneous answers to all our queries. Hong Kong officials and researchers were seeking any insights into transmission control that could be gained from statistical analysis and mechanistic transmission modelling.

In addition to developing statistical and mathematical approaches for epidemiological modelling and analysis, we advised our Hong Kong colleagues about the importance of systematic collection of clinical and demographic data on SARS cases and their contacts. The coordinated approach to data collection and disease control taken by the Hong Kong authorities, including the Hospital Authority, the Department of Health and the Police Force, recently received the Stockholm Challenge Award for the innovative use of information technology.

Such modelling efforts required key biological parameters to be estimated, particularly the distributions of times from infection to symptom onset, from symptom onset to hospitalisation, and from hospitalisation to death or recovery, as well as the case–fatality ratio, that is, the probability that a SARS-related case would die of SARS-related causes. These were estimable from the detailed case data collected on the SARS cases in Hong Kong, although survival analysis techniques were needed to account for the censoring in the data on the time from hospitalisation to death and recovery, that is, the uncertainty in the eventual outcome of cases still hospitalised at the time of analysis.

Ooward transmission varied substantially between individual cases. Two large clusters of cases were particularly prominent in Hong Kong and were termed “super-spread events” (SSEs)—rare events where, in a particular setting, individuals generated far more than the average number of secondary cases. In one case, at least 125 people were infected in early March 2003 in the Prince of Wales Hospital by the index patient for the Hong Kong epidemic. In the other, hundreds of people were infected in the Amoy Gardens housing estate, probably from an environmental source. Debate continues on whether these, and the SSEs in other SARS-affected countries, reflect a long tail of a single distribution of the number of secondary cases or a separate class of transmission events.

Our team developed a stochastic compartmental model utilising the parameter estimates obtained for the time delays between SARS infection, symptom onset, hospitalisation and death or recovery. A metapopulation approach allowed us to incorporate the substantial between-district variation in the incidence of SARS within Hong Kong. We developed stochastic models to predict both average trends and variability in case incidence, because chance fluctuations in case numbers can be important in the early stages of an epidemic. Such an approach gave us insight into whether observed changes in case incidence could have arisen by chance or genuinely reflected the impact of interventions.

Excluding the inherently unpredictable SSEs, we found that in Hong Kong the reproduction number dropped around March 19th from roughly 2.7 (95% confidence interval: 2.2–3.7) to near the threshold level 1, but with sufficient uncertainty (95% CI: 0.7–1.2) to warrant further efforts to control the spread of infections. It dropped again around April 10th to well below the threshold for disease persistence (95% CI: 0.1–0.4). As in other countries, the epidemic died out, with the last probable SARS case in Hong Kong experiencing disease onset in late May.

Real-time epidemiology and advice to policy makers

The SARS epidemic demonstrated how rapidly advances in the understanding of a new disease could be achieved through effective international collaborations, in this case coordinated by the World Health Organization. Clinical management benefited from the identification of the aetiological agent of the disease and the subsequent development of serological tests for infection. The identification of the zoonotic origin of the virus allowed officials to target the sources of new human infections to limit reintroductions which could spark off new outbreaks. Estimates of the incubation period informed the scope of contact tracing, and estimates of the case-fatality ratio provided both officials and the public with information on the severity of the disease. The fast-tracked peer review and publication of analyses of the on-going SARS epidemics in Hong Kong and other affected countries provided timely results to the widest possible audience.

“The coordinated approach to data collection and disease control taken by the Hong Kong authorities recently received the Stockholm Challenge Award for the innovative use of information technology”

The rational assessment of the level of response needed to control an outbreak requires accurate data to underpin analyses and careful assessment of both parameter estimates and the attached uncertainty. The proportionality of a particular policy will often be difficult to judge in the face of a novel infectious disease or the re-emergence of an old disease in a novel setting. However, well-informed planning can provide frameworks for decision making in the face of just such uncertainties. Contingency plans informed by the experience of the SARS epidemics have rightly placed a high priority on both surveillance and contact tracing, and transmission modelling has clarified the types of diseases that can be controlled with the straightforward public measures of isolation and contact tracing.

Statisticians can contribute substantially to both outbreak response and contingency planning efforts, from the initial stages of facilitating data collection and ensuring data integrity to informing policy makers on the meaning of CI values and the importance of sensitivity analyses. The threats posed by emerging diseases are serious, and the potential contributions of statisticians should not be underestimated.

Dwight Donnelly is Professor of Statistical Epidemiology and Aria Ghani is a Royal Society Dorothy Hodgkin Research Fellow in the Department of Infectious Disease Epidemiology, Imperial College London, Faculty of Medicine. They work at the interface of statistics, mathematical biology and epidemiology.