



Mathematical models of disease transmission

MANISH KUMAR¹, Dr. ARUN KUMAR²

Dept .Of Mathematics, OPJS UNIVERSITY, CHURU RAJASTHAN

Abstract |

Mathematical analysis and modelling is central to infectious disease epidemiology. Here, we provide an intuitive introduction to the process of disease transmission, how this stochastic process can be represented mathematically and how this mathematical representation can be used to analyse the emergent dynamics of observed epidemics. Progress in mathematical analysis and modelling is of fundamental importance to our growing understanding of pathogen evolution and ecology. The fit of mathematical models to surveillance data has informed both scientific research and health policy. This Review is illustrated throughout by such applications and ends with suggestions of open challenges in mathematical epidemiology.

INTRODUCTION

Infectiousness

A characteristic of the infected individual that determines the rate of infection of susceptible members of the population and can be broken down into biological, behavioural and environmental components

Mathematical representation and analysis of infectious diseases has been central to infectious disease epidemiology since its inception as a discipline more than a century ago^{1,2}. In recent years, detailed electronic surveillance of infectious diseases has become widespread owing to the advent of improved computing, electronic data management, the ability to share and deposit data over the internet, and rapid diagnostic tests and genetic sequence analysis. These ongoing developments have increased the application of mathematical models to both the generation and testing of basic scientific hypotheses and the design of practical strategies for disease control. Mathematical analyses and models have successfully explained previously puzzling observations and played a central part in public health strategies in many countries^{3,4}. Fundamental to the growing importance of mathematical epidemiology has been the integration of mathematical models with rigorous statistical methods to estimate key parameters of these models and test hypotheses using available data. In the absence of reliable data, mathematics can be used to help formulate hypotheses, inform data-collection strategies and determine sample sizes, which can permit discrimination of competing hypotheses. In this way, mathematics is “no more, but no less, than a way of thinking clearly about the problem in hand” (Ref. 5). The extent and quality of available data can be variable. Ideally, data should be analysed using models that adequately describe the observed dynamics and patterns of interest, and the mechanisms that generate these observations. Models should be as simple as possible, but not so simple that the conclusions drawn are altered by the consideration of additional realistic complexity. Unnecessary complexity can obscure fundamental results and is almost as undesirable as over-simplification. Indeed, model choice — the process of deciding which model complexities are necessary — is a central part of mathematical modelling of infectious diseases. The aim of this Review is to provide an introduction to mathematical modelling of infectious disease transmission and demonstrate how a pathogen’s natural history and ecology determine the outcome of epidemics.

Disease transmission

Epidemics of infectious diseases among humans and other animals result from the transmission of a pathogen either directly between hosts or indirectly through the environment or intermediate hosts. The efficiency of transmission depends on the infectiousness of the infected host (or hosts) and the susceptibility of uninfected individuals who are exposed to infection. Infectiousness comprises three major components: biological, behavioural and environmental. Biological infectiousness depends on the excretion of infectious pathogens and can relate simply to viral or bacterial load in specific anatomical sites or in a more complex way to the pathogen's life cycle (for example, the periodic release of *Plasmodium* spp. gametocytes during human malaria) (FIG. 1). The dynamics of the pathogen in the body depend, in turn, on the features of an individual host's immune system, including innate and acquired immunity; features of the pathogen, such as the dynamics of pathogen replication and spread within the host, the initial infectious dose, virulence factors or sensitivity to drugs; and the interaction between the genetic determinants of disease progression in both the pathogen and the host. Behavioural infectiousness depends on the contact patterns of an infected individual and, if relevant, the contact patterns of intermediate hosts or vectors. The nature of these contacts varies depending on the disease and route (or routes) of transmission. For the intimate contacts that are required for the transmission of sexually transmitted infections, the frequencies of contact are highly dispersed, such that a few individuals have multiple contacts whereas most have few⁷. For the more casual conversational contacts or handshakes that can result in the transmission of respiratory infections, the frequencies of contact are typically less dispersed and the number of such contacts that are reported by individuals cluster around the average⁸. Environmental infectiousness depends on the location and environment of the infected individual. The environment is important for the survival of the pathogen outside the host and for the survival of intermediate hosts and vectors, which can affect the efficiency of transmission. Climatic variation in temperature or rainfall drives seasonal patterns of disease incidence for many infections (for example, cholera, influenza and polio)^{9,10}. In some cases, the environmental component of infectiousness can result in an unusually large number of secondary infections. For example, the 'superspreading' event that led to various severe acute respiratory syndrome (SARS) infections among residents of the Amoy Gardens estate in Hong Kong in 2003 was the result of unusual environmental conditions and airflow dynamics that spread the virus through aerosolized faeces¹¹. The susceptibility of an individual also has biological, behavioural and environmental components: for example, susceptibility might depend on immune memory, patterns of contact and location. After contact between an infectious pathogen and a susceptible individual, transmission occurs with a probability that is a function of infectiousness and susceptibility. For example, the probability of transmission of HIV-1 during unprotected heterosexual intercourse depends on the viral load of the infectious individual and whether either individual has an ulcerative sexually transmitted disease¹².

Mathematical models of disease transmission Mathematical models can be used to link the biological process of transmission and the emergent dynamics of infection at the population level. At its simplest, an epidemic can be described by enumerating who infected who and when. At the root of this transmission tree there is at least one index case. Two key statistical properties of an epidemic that can be linked to the underlying infection process are the offspring distribution and the generation-time distribution (FIG. 2). Offspring distribution If Y is the number of secondary infections that are generated by a single infected individual, then the probability function that describes the distribution of Y is referred to as the offspring distribution, denoted $p(y)$ (that is, the probability that the number of secondary infections Y that are caused by a single infectious individual is equal to y). The number of secondary infections depends on the infectiousness of the index case over time τ since they themselves became infected. We represent this quantity by a function that is denoted $\beta(\tau)$. In some cases, it might be useful to explicitly decompose $\beta(\tau)$ as a

product of biological, behavioural and environmental infectiousness. For example, $\beta(\tau)$ might be the product of biological infectiousness (FIG. 1) and contact rates. It would then be of interest to predict the impact of an intervention that modifies contact rates. Infectiousness can be used as a basis for modelling the transmission process by specifying that the probability that an infected person will infect another person during a small time interval between τ and $\tau + \delta\tau$ as $\beta(\tau)s\delta\tau$, in which $\delta\tau$ is an arbitrary small measure of time and s is the probability that the other person is susceptible to infection. Summing up or integrating infectiousness over the time since infection provides the expected number of offspring (Equation 1).

$$E(Y) = \int_0^{\infty} \beta(\tau) s \tau d\tau \quad (1)$$

In a population in which the pattern of contact between infectious and susceptible individuals is random, the expected number of offspring $E(Y)$ is exactly equivalent to the reproduction number for an infection $R^{3,13,14}$. In a naive, entirely susceptible population ($s = 1$), this is called the basic reproduction number (R_0). In a population with non-random mixing, the reproduction number is a more complex function of the offspring distribution and patterns of mixing. The reproduction number is a fundamental quantity in infectious disease epidemiology that relates to the individual but is also a fundamental determinant of the dynamics of infection at the population level. Crucially, an epidemic can only occur if the reproduction number R is greater than one. This threshold property means that estimation of R provides important information about the potential for disease transmission and the impact of disease control^{15–19}. If a group of individuals all have the same infectiousness, then the number of secondary infections that are caused by each infectious individual is a random number drawn from the Poisson distribution with mean R . In reality, however, the infectiousness and susceptibility of an individual are influenced by many different factors. Individuals can be categorized by any factors that are considered to be important for infectious disease transmission and can be labelled $i = 1, 2, \dots, n$ (n denotes the number of different categories of individuals; Diekmann and Heesterbeek¹³ refer to these as h -states). In a randomly mixing population, the offspring distribution is derived from the mixture of the distributions for each individual, weighted according to their frequency. For example, the susceptible–infected–recovered (SIR) model assumes a constant infectivity β while an individual remains infected and a constant rate of recovery from infection α such that the time that is spent infectious is exponentially distributed²⁰. Infectious individuals can therefore be categorized by their infectious period, and summing over the distribution of infectious periods provides an offspring distribution that follows a geometric distribution with mean $R_0 = \beta/\alpha$ (a mixture of Poisson distributions with exponentially distributed means). This distribution is highly skew and over-dispersed compared with the Poisson distribution (the variance is more than the mean). Therefore, the stochastic SIR model shows considerable variability between individuals in their contribution to transmission. The offspring distribution has been estimated from surveillance data for outbreaks of several infectious diseases, including SARS, foot-and-mouth disease, measles, smallpox and plague^{18,21}. Typically, these estimates indicate variation in the number of secondary infections that is equivalent to, or even greater than, that of the geometric distribution. This is broadly consistent with the more general observation that a disproportionate amount of disease transmission results from a small fraction of all infected individuals²². Knowledge of the offspring distribution is crucial to understanding disease dynamics if the number of infected individuals is small^{23–26}. For example, outbreaks can occur even when R is less than 1, owing to the random nature of the transmission process. A description of the size distribution of these outbreaks follows from knowledge of the offspring distribution^{27–29}.

. Generation-time distribution

If τ is the time that has elapsed between one person being infected and that person infecting someone else, then the probability distribution of these times is the generation-time distribution, denoted $w(\tau)$ ³⁰. This

distribution is related to infectiousness through a simple rescaling of infectiousness by the reproduction number

(Equation 2).

$$w(\tau) = \beta(\tau)/R \quad (2)$$

In randomly mixing populations, in which individuals vary in their infectiousness, the generation-time distribution is a weighted sum across these individuals. However, if mixing patterns are not random with respect to infectiousness, the relationship between infectiousness and the generation-time distribution is more complex, and work remains to be done in this area^{31,32}. Estimation of the generation-time distribution is rarely easy. A first approach is to estimate it directly from $\beta(\tau)$ by defining a pathogen-shedding or excretion score and measuring it longitudinally after infection (for example, FIG. 1). The problem with this approach is that the relationship between shedding and infectiousness can be complicated, and behavioural effects are not considered. A second approach is to track infectious individuals who have known times of infection and observe subsequent transmission events within defined risk groups³⁰. Such an approach has been followed, for example, for HIV-1 (Ref. 33) and SARS15. Care needs to be taken to account for immunity of contacts to infection, to adjust for the fact that transmission events are not usually observed, but rather some surrogate of infection is recorded (such as symptoms or microbiological diagnosis) and for censored observations (when individuals cannot be followed up for a particular reason or have not completed their infectious period). From individuals to epidemics The dynamics of an infectious disease at the population level represent an average of all infected individuals. Even if the contribution of an infected person to the epidemic is unpredictable, the emergent population dynamics often conform to mathematical expectations. The random effects among individuals tend to cancel each other out as the number of infected individuals increases — the law of large numbers. Therefore, even if the underlying distribution of the number of secondary cases is highly skew, an epidemic will progress smoothly as long as the expected incidence at each observation is reasonably large. If the incidence of infection is small, however, more complex and resurgent epidemic dynamics might be expected (FIG. 3). Even if the probability distribution for the number of secondary infections varies between individuals in a systematic way, reflecting, for example, underlying behavioural tendencies or biological differences, the change in these characteristics over the course of an epidemic is often gradual and predictable. For example, the behavioural and demographic characteristics of individuals who are infected during a heterosexual HIV-1 epidemic typically change in a gradual and predictable way as the infection moves from individuals who are at high risk, such as sex workers, to individuals at lower risk, such as those in stable partnerships^{34,35}. The expected evolution of an epidemic in which transmission follows the mathematical representation described in Equation 1 was first examined in detail by Kermack and McKendrick in 1927 (Ref. 36). If $I(u)$ is a function that describes the number of individuals who are newly infected at time u (incidence) and at some later time $t = u + \tau$ the infectiousness of these individuals is $\beta(\tau)$, then the number of people they will infect over the next incremental time step of duration $\delta\tau$ will be Poisson distributed with mean $I(u) \beta(\tau) \delta\tau = I(t - \tau) \beta(\tau) \delta\tau$. The total expected number of people infected at time t , which must take account of possible secondary infections for all previously infected individuals, is obtained by summing this quantity over all possible values of τ . This statement can be expressed as a sum or integral that is known as the renewal equation (Equation 3).

$$\int_0^t I(t - \tau) \beta(\tau) d\tau \quad (3)$$

The infectiousness function can be seen from Equation 2 to be the product of the reproduction number and the generation-time distribution, and substituting this relationship into the renewal equation immediately provides an estimator for the reproduction number³² (Equation 4).

$$\int_0^{\infty} R(t) = I(t) / I(t-\tau) w(\tau) d\tau \quad (4)$$

This estimator assumes that the generation-time distribution does not change over time. However, this is not always the case — for example, patient isolation shortens generation times — and in these cases the estimator needs to be adjusted for this effect³². Equation 4 was used recently to estimate the extent of seasonal variation in polio transmission in India and to analyse the effect of the vaccine programme in reducing transmission¹⁷. Related estimators have been used to explore the transmission dynamics of SARS³⁷, of the 2001 United Kingdom foot-and-mouth disease epidemic in cattle and sheep flocks¹⁹ and of HIV-1 epidemics in Europe^{38,39}. Of particular interest has been the estimation of the basic reproduction number R_0 of influenza pandemics. Although historical data have been too limited for detailed analysis of incidence trends, it has been possible to estimate the exponential growth rate r of the early epidemic from weekly incidence data. Substituting an exponential function for growth in the number of infections —

$$I(t) = I(t_0) \exp(r(t-t_0))$$

into the renewal equation (Equation 3) yields the so-called Lotka–Euler estimating equation^{32,40} (Equation 5).

$$\int_0^{\infty} R_0 = 1 / w(r) e^{-rt} \tau d\tau \quad (5)$$

This equation easily yields analytical forms for the relationship between R_0 and r for a given generation-time distribution. For the growth rate that was observed in London during the 1918 pandemic of $r = 0.20$ per day, reported estimates of the reproduction number have ranged from 1.6 to 2.5 depending on the assumed generation-time distribution, which is poorly quantified from historical data (FIG. 4). Assessing the potential impact of interventions In a susceptible population, each individual initially infects R_0 new individuals on average, and so any intervention must prevent at least $R_0 - 1$ out of every R_0 infections to result in a reproduction number $R \leq 1$ and control of the infection. In other words, the critical efficacy of interventions in reducing transmission must be Equation 6.

$$1 - \frac{1}{R_0} = 1 - \frac{1}{R_0} \quad (6)$$

This does not vary linearly with R_0 ; for example, the effort required to control an infection with $R_0 = 2$ is 50% greater than the effort required to control an infection with $R_0 = 1.5$, despite infectiousness being only 33% higher. Detailed simulations of influenza pandemic control and mitigation policies are consistent with this simple analytical insight, demonstrating the limited impact of control for values of R_0 that are greater than ~ 2 (Refs 41–43). This highlights the crucial importance of estimates of this parameter such as those that are presented in FIG. 4. Estimates from other historical data vary on either the higher or lower side of this number depending on the data and assumptions that are made^{44,45}. The renewal equation (Equation 3) can be used to predict the impact of interventions during an epidemic that modify infectiousness. For example, Fraser et al.⁴⁶ used this model to explore the impact of isolation or treatment and contact-tracing measures for SARS, smallpox, influenza and HIV-1 epidemics,

Mixing patterns

We have so far focused on populations in which infectious contacts between individuals are assumed to occur at random. This assumption of random mixing is frequently made and rarely accurate, although the resulting model often captures the essence of the infectious disease dynamics and parameter estimates may be relatively unbiased. Consequently, the assumption of random mixing should always be tested, and we now discuss situations in which this assumption must be relaxed to ensure robust conclusions from mathematical analyses. Restricted mixing might occur as a result of geographical, behavioural or social factors. The implication of such population structures for the spread of disease depends on the natural history of the

infection and routes of transmission. For example, infections that are transmitted by insect vectors depend less on the distribution of individuals among households than infections that require direct contact. Similarly, for directly transmitted infections, the patterns of contact that are important will vary, from the intimacy that is required for sexually transmitted infections to the less intimate handshake or conversation that can lead to the transmission of respiratory infections. In the simplest case of non-random mixing, individuals may be similar with respect to their infectiousness and susceptibility to infection, but the population may be divided into distinct communities, with infrequent contacts between communities. This could be considered to approximate the situation for childhood infections, such as measles, among different towns or villages.

If infection is introduced to each community simultaneously, the resulting epidemic will follow the expected pattern of spread for an unstructured population²⁰. If communities are distributed across a landscape and infection is introduced to just one community, however, travelling waves of infection can occur^{20,47}, as shown for the spread of measles before the introduction of vaccination in the United Kingdom⁴⁸. Waves of infection during the epidemic season spread out from large cities to smaller towns. Non-random mixing has important implications for the impact of vaccination. If individuals who are missed by a vaccination programme tend to be from the same community, coverage must be greater than the critical threshold p_c (Equation 6), as transmission can be sustained within the unvaccinated community⁴⁹. In The Netherlands, for example, overall vaccine coverage against measles is high, but low coverage in specific communities owing to parental refusal to vaccinate children has led to localized disease outbreaks, with a reproduction number that is greater than one⁵⁰. Populations typically consist of individuals that vary in their infectiousness and susceptibility to infection. If patterns of mixing result in a correlation between the infectiousness of individuals and the infectiousness of the individual who infected them (infectiousness correlates with susceptibility), then it becomes important to capture this heterogeneity and patterns of mixing in the mathematical representation of transmission, as they affect both R_0 and the long-term dynamics of infection⁵¹. Such correlations might arise because of age-related or social determinants of behaviour, or from environmental factors, such as the slums that facilitate faecal–oral transmission of diarrhoeal disease. If infectiousness and susceptibility correlate, then the basic R_0 can be calculated from the next-generation matrix^{3,13}. In such cases, R_0 is typically greater than the weighted average of the individual reproduction numbers (as for random mixing) because early in the epidemic the most infectious individuals dominate transmission. If a control programme is applied uniformly across the population, the critical reduction in transmission that will bring R below the threshold in a structured epidemic model with non-random mixing is still determined by $p_c = 1 - 1/R_0$, and, as R_0 is typically greater in a heterogeneous population, p_c is generally greater if there is heterogeneity. If, however, it is feasible to target the most highly infectious individuals, then this critical threshold declines. The challenge is to identify the more infectious individuals before they transmit infection and to intervene appropriately. An example of a strategy that uses information on heterogeneity is the targeting of larger households for vaccination against smallpox, which was shown to be significantly more effective than random vaccination⁵²

Saturation Infection with a pathogen can, in some cases, kill the host, induce host protective immune responses or activate immune memory, which reduces the chance of subsequent re-infection. The resulting decline in the number of individuals that are susceptible to infection is known as saturation. If the immune response is ineffective and the infection is long-lived, such as in some chronic infections, repeat infection with the same species of pathogen can result in ‘super-infection’. For infections such as intestinal nematodes, flukes, filarial nematodes and tapeworms, usually referred to as macroparasites, this process of super-infection results in an aggregated distribution of the parasites among the host population and enables sexual reproduction of the parasite. For these macroparasitic infections, mathematical models that focus on the number of parasites and their distribution among hosts tend to yield more insight compared with models that take the number of infected hosts as their fundamental unit³. Saturation leads to non-linear disease

dynamics, in which the incidence of infection is a function of both the current number of infected individuals and the number of susceptible individuals. What this function should be is still debated^{3,53}. Two of the most common approaches are to assume density dependence, in which the number of contacts that are made by an individual increases with population numbers, such that incidence depends on the number of infected individuals (incidence = βxy , in which β is infectiousness, x is the number of susceptible individuals and y is the number of infected individuals), and to assume frequency dependence, in which the number of contacts is independent of population size and incidence depends on the fraction of contacts that are infected (incidence = $\beta xy/n$). Tests of the fit of these functions for the incidence of childhood infections and sexually transmitted diseases have typically found support for largely frequency-dependent transmission, with limited dependence on population numbers^{3,54}. Saturation in an epidemic can be global if a significant fraction of the overall population is no longer susceptible, or local if non-random mixing results in a cluster of infected or immune individuals. Local saturation is most clearly illustrated by the important phenomenon of clustering of infections within households. The distribution of individuals among households is a crucial determinant for the spread of many directly transmitted infections, such as influenza, trachoma, measles and smallpox. Within a household, all susceptible individuals could become infected following the initial infection event and the potential for further transmission of disease is therefore restricted, despite the ongoing infectiousness of household members. This can complicate estimation of the generation-time distribution for an infectious disease. The spread of infection in this case might be more easily described by the time between the initial infection of households (the household generation-time distribution) and the average number of households that are infected by a single infected individual in a household (the household reproduction number)^{32,55,56}. Simple models that assume random mixing do not readily capture the effects of local saturation in limiting the spread of infection. Much recent work has focused on measuring and describing contact patterns and examining the impact of local contact-network structure on infectious disease dynamics^{7,57–65}. Usually, local saturation occurs rapidly, and subsequent dynamics are driven by global contacts, which means that an epidemic in a highly structured population can show rapid initial growth, suggesting a high R_0 , even though subsequent dynamics can be consistent with a much lower reproduction number⁵⁸. The effect of global saturation is to slow and eventually reverse the growth in the number of infections. If the susceptible population is not replenished, global saturation gives rise to the classic bell-shaped epidemic curve and the infection will become extinct²⁰. If individuals recover to an immune class, an epidemic in a closed population will not necessarily infect all susceptible individuals. The relationship between R_0 and the final fraction of the population that is infected (z) is provided by the transcendental equation³⁶ (Equation 8).

$$z = 1 - \exp(-R_0 z) \quad (8)$$

This final-size equation is surprisingly robust to the consideration of additional infectious stages, latency, arbitrarily distributed transmission rates for infected individuals and certain types of spatial contact processes, and allows the potential scale of an epidemic to be estimated as soon as an estimate for R_0 is available⁶⁶. However, for more complex patterns of contact its application is limited, and it no longer applies if susceptible individuals are introduced to the population through births, uptake of risk behaviour, recovery from infection (for example, bacterial infections such as trachoma or gonorrhoea; susceptible–infective–susceptible (SIS)-type infection), during the waning of protective immunity (for example, polio; susceptible–infected–recovered–susceptible (SIRS)-type infection) or after

Model choice The criteria that define an appropriate mathematical model with which to address a scientific question should be based on the principle of parsimony — choose the simplest model that explains the data — and the ability of the model to answer the question of interest. Inevitably, there will be an element of subjectivity in the choice of model, and in some cases the model might need to be revised in the light of new

observations, which would lead to an iterative process of model development. The ability of a model to explain data can be formally assessed using methods for statistical inference. These can be categorized into three broad approaches: frequentist, Bayesian and information theoretic (reviewed in Refs 73–75). All three approaches are based on different uses of the concept of likelihood. The likelihood of a model based on an observed set of data is proportional to the probability of the observed data if that model is true, the constant of proportionality being arbitrary. Likelihood can be used to estimate parameters and compare models. In particular, likelihood is a key measure in deciding whether additional complexity in model structure is warranted in terms of improving our ability to describe data. The exact comparison will depend on the inference framework (for measures such as the likelihood ratio, the Akaike and Bayesian information criteria^{73–75}), but all these methods balance the gain in likelihood that is achieved by extending a model with the increase in the number of parameters that are needed to describe that model. To calculate a likelihood, the transmission model must be stochastic and capable of predicting the probability of different outcomes (as the likelihood is the probability that the data was generated by that model). A formal starting point for defining a likelihood for a time series of observations of an infectious disease system could therefore be a stochastic model that predicts the distribution of the number of people in different states at different times. The sources of stochasticity could arise from the transmission process itself (the offspring distribution) and from the observation or sampling process. Infection events depend on prevalent infectious cases, which, in turn, depend on earlier infection events, as clearly shown by the renewal equation (Equation 3). Therefore, all observations in an epidemic time series are correlated and should be viewed as part of a single observation of the epidemic process. For example, random effects that occur early in an epidemic, when the number of infections is small, significantly affect subsequent epidemic spread and can lead to different epidemic dynamics for the same underlying parameters (FIG. 3d). For this reason, repeated and possibly incomplete observations of independent epidemics are often easier to assess than detailed, but temporally correlated data, from a single epidemic. The difficulty of dealing with temporally correlated data is that, even for small epidemics, the likelihood can be difficult to write down and thus calculate for a given model and data. Several exact and approximate approaches have been reviewed in Refs ^{76,77}. In particular, Bayesian methods are beginning to result in progress in this area, by treating unknown events, such as individual infection times, as random variables to be estimated. A particular simplification arises for ‘memoryless’ (‘Markovian’) transmission models. These are equivalent to SIR and related models, in which individuals occupy ‘compartments’ (for example, infectious or recovered) and the rate of leaving the compartment depends only on the current state of the epidemic and not its history. In such models, if the state variables, such as the number of susceptible individuals, are known at a point in time, then incidence over the next time step is an independent random variable^{76,78}. The problem then requires inference of the unknown state variables, rather than disentangling the likelihood for the whole epidemic. For example, this approach has been applied to data on Ebola⁷⁹, hospital-acquired infections⁸⁰ and measles⁸¹. Studies have often simplified and approximated the problem of estimating a true likelihood by treating observations as being independent despite the bias that this can introduce. For example, a deterministic model can be made to fit observed data by assuming that the model prevalence or incidence corresponds to the expectation of an underlying distribution. Observations are then assumed to be drawn from this distribution, which is typically chosen to reflect some sort of random sampling process. This approach has been used by the United Nations to estimate HIV-1 prevalence and AIDS mortality over time from national seroprevalence data⁸². To our knowledge, little work has been done to assess the biases that are caused by using such approximate likelihoods or least-squares to fit epidemic models.

Open challenges We expect our understanding of the epidemiology of infectious diseases to improve in the near future, but progress will depend on multidisciplinary and multilevel systematic approaches across

immunology, genetics, ecology and epidemiology. Mathematical analysis and modelling will play a central part, and, although difficult to predict in detail, we end by outlining some outstanding challenges.

Pathogen evolution. Mathematical analyses of the evolution of pathogens in response to the selective pressure of the host immune system will play an important part in the design and interpretation of interventions with drugs or vaccines that alter pathogen genetic diversity. For example, the impact of a vaccine that only targets specific pathogen strains (for example, pneumococcal vaccines) on subsequent disease dynamics and pathogen diversity could be predicted. Progress in this area might also lead to an improved ability to match vaccine manufacture to circulating pathogen antigenic types for rapidly evolving infections, such as influenza⁸³. As sensitive genetic typing becomes routine in disease surveillance, this will not only improve our understanding of evolutionary processes, but also directly shed light on the transmission network. The development of coupled evolutionary–epidemic models and joint inference methods will be crucial in making the most of these data.

Statistical methods. The number of new infections at a particular time and place depends on unknown variables, such as population immunity, which can be strongly correlated over time and space. Furthermore, infections can often go unreported or undiagnosed, leading to many ‘missing’ observations. Statistical methods to estimate the parameters of models that describe a disease’s natural history and transmission for incomplete, highly correlated surveillance data have yet to be developed. Progress in this area is being made (reviewed, for example, in Refs 76,77), but much remains to be done, particularly in the context of spatial correlations. Contact patterns. Many ‘laws’ of infectious disease epidemiology, such as mass action or the relationship between the average age at infection and the basic reproduction number, rely to some extent on the assumption of random infectious contacts among individuals in a population. Non-random mixing leads to local saturation effects that can undermine these laws and lead to more complex epidemics that have a range of implications for control. Mathematical descriptions of nonrandom mixing need to go beyond simply dividing the population into discrete compartments to a more complete representation of the contact patterns that lead to infectious disease transmission. Computer simulations could provide insights, but the development of analytical approximations of the contact process would be far more powerful if they could provide simple rules (and predictions) for the emergent dynamics of an epidemic. This might require a different type of mathematical model and language, perhaps motivated by ongoing research into network or graph theory⁶¹. The collection and analysis of empirical data that describe both the contact process and the transmission tree will be central to this research. Empirical data that describe the contact process could come from questionnaires, such as those used to investigate conversational contacts that are important for the spread of respiratory infections^{7,84}, from routine surveys of travel and commuting patterns or even from mobile-telephone global-positioning data. Empirical data that describe the transmission tree will depend on the collection of robust information about the contacts of sampled individuals, the contacts of these contacts and so on. A major challenge for theoreticians will be the integration of information that has been gathered at all levels, from household contacts to the global travel patterns of individuals. **Pathogen ecology.** The dynamics of a pathogen and the impact of control might depend not only on the interaction of the pathogen with the host population, but also on the interaction of the host with other pathogens and hosts. It seems that cycles in the incidence of whooping cough in the pre-vaccine era were dependent on changes in population density that were caused by measles mortality⁸⁵. Declines in syphilis in the United States in the 1990 have been partly attributed to AIDS mortality among the sexually active core transmission group⁸⁶. Viral respiratory infections can increase shedding of other respiratory co-infections, thereby facilitating their transmission⁸⁷. Bacterial and viral sexually transmitted infections can increase the transmission of HIV-1 (Ref. 12). Identifying and modelling all the interactions that are important for a particular pathogen might be difficult in a particular setting owing to their number and complexity. For example, the dynamics of HIV-1 in a population can depend

on the dynamics of a number of bacterial and viral sexually transmitted infections that facilitate HIV-1 transmission, in addition to changes in risk behaviour and patterns of mixing. The dynamics of these other infections are, in turn, mutually interdependent, which further complicates the analysis. An understanding of these dynamics and their underlying determinants could be made easier by comparative analysis across different settings and time periods that differ only in a small number of variables. In some cases, it might be possible to gain further insight by experimental manipulation of the system⁸⁸. A more complete understanding of the ecology of a pathogen could lead to better policies for infectious disease control.

Conclusion The dynamics of infectious diseases and their hosts can be complex and the impact of control programmes can be difficult to predict. Underlying this complexity are simple stochastic processes, some of which are described in this Review. The challenge to the infectious disease epidemiologist is to disentangle the underlying determinants of the emergent behaviour of epidemics and identify rules that could enable this behaviour to be predicted and effective controls to be identified.

REFERENCE

1. Heesterbeek, H. in *Ecological Paradigms Lost: Routes of Theory Change* (eds Cuddington, K. & Beisner, B.) 81–105 (Elsevier, Burlington, Massachusetts, 2005).
2. Dietz, K. & Heesterbeek, J. A. P. Daniel Bernoulli's epidemiological model revisited. *Math. Biosci.* 180, 1–21 (2002).
3. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford Univ. Press, 1991).
4. Glasser, J., Meltzer, M. & Levin, B. Mathematical modeling and public policy: responding to health crises. *Emerg. Infect. Dis.* 10, 2050–2051 (2004).
5. May, R. M. Uses and abuses of mathematics in biology. *Science* 303, 790–793 (2004).
6. Johnson, A. M. et al. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. *Lancet* 358, 1835–1842 (2001).
7. Edmunds, W. J., O'Callaghan, C. J. & Nokes, D. J. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc. R. Soc. Lond. B* 264, 949–957 (1997). A first attempt to measure the contact patterns that result in the transmission of respiratory infections.
8. Loosli, C. G., Lemon, H. M., Robertson, O. H. & Appel, E. Experimental airborne influenza infection: I. Influence of humidity on survival of virus in air. *Proc. Soc. Exp. Biol. Med.* 53, 205–206 (1943).
9. Grassly, N. C. & Fraser, C. Seasonal infectious disease epidemiology. *Proc. R. Soc. Lond. B* 273, 2541–2550 (2006).
10. Altizer, S. et al. Seasonality and the dynamics of infectious diseases. *J. Anim. Ecol.* 9, 467–484 (2006).
11. Yu, I. T. S. et al. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N. Engl. J. Med.* 350, 1731–1739 (2004).
12. Gray, R. H. et al. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* 357, 1149–1153 (2001).
13. Diekmann, O. & Heesterbeek, J. A. P. *Mathematical Epidemiology of Infectious Diseases: Model building, Analysis and Interpretation* (ed. Levin, S.) 1–303 (Wiley, Chichester, 2000).

14. Heesterbeek, J. A. P. A brief history of R_0 and a recipe for its calculation. *Acta Biotheor.* 50, 189–204 (2002).
15. Lipsitch, M. et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 300, 1966–1970 (2003).
16. Riley, S. et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* 300, 1961–1966 (2003).
17. Grassly, N. C. et al. New strategies for the elimination of polio from India. *Science* 314, 1150–1153 (2006).
18. Haydon, D. T. et al. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc. R. Soc. Lond. B* 270, 121–127 (2003).
19. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* 413, 542–548 (2001).
20. Bailey, N. T. J. *The Mathematical Theory of Infectious Diseases and Its Applications*. 2nd edn 1–413 (Griffin, London, 1975).
21. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* 438, 355–359 (2005). Empirically driven, theoretical exploration of the implications of variation in individual infectiousness for emergent disease dynamics.
22. Woolhouse, M. E. et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* 94, 338–342 (1997).
23. Jansen, V. A. A. et al. Measles outbreaks in a population with declining vaccine uptake. *Science* 301, 804 (2003). An illustration of the application of branching process theory to surveillance data on measles outbreaks in England and Wales to estimate the underlying reproduction number and potential for more widespread transmission.
24. Gay, N. J., De Serres, G., Farrington, C. P., Redd, S. B. & Papania, M. J. Assessment of the status of measles elimination from reported outbreaks: United States, 1997–1999. *J. Infect. Dis.* 189 (Suppl. 1), 36–42 (2004).
25. Ferguson, N. M., Fraser, C., Donnelly, C. A., Ghani, A. C. & Anderson, R. M. Public health risk from the avian H5N1 influenza epidemic. *Science* 304, 968–969 (2004).
26. Matthews, L. & Woolhouse, M. E. J. New approaches to quantifying the spread of infection. *Nature Rev. Microbiol.* 3, 529–537 (2005).
27. Becker, N. On parametric estimation for mortal branching processes. *Biometrika* 61, 393–399 (1974).
28. Farrington, C. P. On vaccine efficacy and reproduction numbers. *Math. Biosci.* 185, 89–109 (2003).
29. Jagers, P. *Branching Processes With Biological Applications* 1–282 (Wiley, London, 1975).
30. Fine, P. E. M. The interval between successive cases of an infectious disease. *Am. J. Epidemiol.* 158, 1039–1047 (2003).
31. Svensson, A. A note on generation times in epidemic models. *Math. Biosci.* 208, 300–311 (2007).
32. Fraser, C. Methods for estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2, e758 (2007).
33. Wawer, M. J. et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J. Infect. Dis.* 191, 1403–1409 (2005).
34. Wasserheit, J. N. & Aral, S. O. The dynamic topology of sexually transmitted disease epidemics: implications for prevention strategies. *J. Infect. Dis.* 174 (Suppl. 2), 201–213 (1996).

35. Garnett, G. P. The geographical and temporal evolution of sexually transmitted disease epidemics. *Sex. Transm. Dis.* 78 (Suppl. 1), 14–19 (2002).
36. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* 115, 700–721 (1927). One of the earliest and fullest explorations of the mathematical representation of infectious disease transmission.
37. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* 160, 509–516 (2004).
38. Amundsen, E. J., Stigum, H., Rottingen, J. A. & Aalen, O. O. Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden. *Epidemiol. Infect.* 132, 1139–1149 (2004).
39. White, P. J., Ward, H. & Garnett, G. P. Is HIV out of control in the UK? An example of analysing patterns of HIV spreading using incidence-to-prevalence ratios. *AIDS* 20, 1898–1901 (2006).
40. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. Lond. B* 274, 599–604 (2007). A clear description of how the reproduction number can be estimated from early epidemic growth and the dependence of the estimate on the generation-time distribution.
41. Ferguson, N. M. et al. Strategies for mitigating an influenza pandemic. *Nature* 442, 448–452 (2006).
42. Germann, T. C., Kadau, K., Longini, I. M. & Macken, C. A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* 103, 5935–5940 (2006).
43. Longini, I. M. et al. Containing pandemic influenza at the source. *Science* 309, 1083–1087 (2005).
44. Mills, C. E., Robins, J. M. & Lipsitch, M. Transmissibility of 1918 pandemic influenza. *Nature* 432, 904–906 (2004).
45. Lessler, J., Cummings, D. A. T., Fishman, S., Vora, A. & Burke, D. S. Transmissibility of swine flu at Fort Dix, 1976. *J. R. Soc. Interface* 4, 755–762 (2007).
46. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* 101, 6146–6151 (2004). An illustration of the power of simple mathematical approaches to answer important policy questions. This analysis made it clear why different micro-simulations of smallpox transmission and control provide different answers.
47. Mollison, D. Spatial contact models for ecological and epidemic spread. *J. R. Stat. Soc. B* 39, 283–326 (1977).
48. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* 414, 716–723 (2001).
49. Glass, K., Kappey, J. & Grenfell, B. T. The effect of heterogeneity in measles vaccination on population immunity. *Epidemiol. Infect.* 132, 675–683 (2004).
50. van den Hof, S. et al. Measles outbreak in a community with very low vaccine coverage, the Netherlands. *Emerg. Infect. Dis.* 7, 593–597 (2001).
51. Woolhouse, M. E. J. et al. Epidemiological implications of the contact network structure for cattle farms and the 20–80 rule. *Biol. Lett.* 1, 350–352 (2005).
52. Ball, F. & Lyne, O. Optimal vaccination schemes for epidemics among a population of households, with application to Variola minor in Brazil. *Stat. Methods Med. Res.* 15, 481–497 (2006).
53. McCallum, H., Barlow, N. & Hone, J. How should pathogen transmission be modelled? *Trends Ecol. Evol.* 16, 295–300 (2001).

54. Bjornstad, O. N., Finkenstadt, B. F. & Grenfell, B. T. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* 72, 169–184 (2002).
55. Becker, N. G. & Dietz, K. The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* 127, 207–219 (1995).
56. Ball, F., Mollison, D. & Scalia-Tomba, G. Epidemics with two levels of mixing. *Ann. Appl. Probab.* 7, 46–89 (1997).
57. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* 316, 1298–1301 (2007).
58. Keeling, M. The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.* 67, 1–8 (2005).
59. Parham, P. E. & Ferguson, N. M. Space and contact networks: capturing the locality of disease transmission. *J. R. Soc. Interface* 3, 483–493 (2006).
60. Ferrari, M. J., Bansal, S., Meyers, L. A. & Bjornstad, O. N. Network frailty and the geometry of herd immunity. *Proc. R. Soc. Lond. B* 273, 2743–2748 (2006).
61. May, R. M. & Lloyd, A. L. Infection dynamics on scalefree networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 64, 066112 (2001).
62. Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 180–184 (2004).
63. Kretzschmar, M. & Morris, M. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* 133, 165–195 (1996).
64. Ghani, A. C. & Garnett, G. P. Risks of acquiring and transmitting sexually transmitted diseases in sexual partner networks. *Sex. Transm. Dis.* 27, 579–587 (2000).
65. Eames, K. T. D. & Keeling, M. J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl Acad. Sci. USA* 99, 13330–13335 (2002).
66. Ma, J. L. & Earn, D. J. D. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* 68, 679–702 (2006).
67. Koelle, K. & Pascual, M. Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* 163, 901–913 (2004).
68. Grassly, N. C., Fraser, C. & Garnett, G. P. Host immunity and synchronized epidemics of syphilis across the United States. *Nature* 433, 417–421 (2005).
69. Aron, J. L. & Schwartz, I. B. Seasonality and period doubling bifurcations in an epidemic model. *J. Theor. Biol.* 110, 665–679 (1984).
70. Nguyen, H. T. H. & Rohani, P. Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. *J. R. Soc. Interface* 5, 403–413 (2008).
71. Bauch, C. T. & Earn, D. J. D. Transients and attractors in epidemics. *Proc. R. Soc. Lond. B* 270, 1573–1578 (2003). A demonstration of how the interaction of random effects and non-linear dynamics can explain the observed dynamics of endemic childhood infections.
72. Hastings, A. Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* 19, 39–45 (2004).
73. Edwards, A. W. F. *Likelihood* 2nd edn 1–296 (Johns Hopkins Univ. Press, Baltimore, 1992).
74. Cox, D. R. *Principles of Statistical Inference* 1–236 (Cambridge Univ. Press, 2006).
75. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: a Practical Information– Theoretic Approach* 2nd edn 1–488 (Springer, New York, 1998). R E V I E W S 486 | june 2008 | volume 6 www