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.

Chapter 5

Peter J White
Mark C Enright

Mathematical models in infectious disease epidemiology

The first mathematical model of infectious disease transmission was constructed by Bernoulli in 1760[1] to determine the effectiveness of variolation, a crude form of smallpox vaccination. Specifically he sought to calculate life tables for actuarial purposes assuming an eradication of smallpox in the population (for a review of Bernoulli's life and work, see Dietz & Heesterbeek[2]). Mathematical systems have been used extensively not only in the study of pandemic viral diseases such as AIDS, severe acute respiratory syndrome (SARS), influenza, smallpox and polio, but have also been successful in modeling transmission of bacterial pathogens including sexually transmitted infections (STIs) and methicillin-resistant *Staphylococcus aureus*, vector-borne diseases including malaria and diseases caused by macroparasites such as helminths.

Modern infectious disease epidemiology relies heavily on mathematical modeling to characterize the complex interactions between the biology and environment of human or animal hosts, the pathogens responsible for disease and, where present, vector species. Transmission models are vital for understanding the dynamics present in the spread of infectious agents in populations and for assessing the likely impact of public health interventions such as isolation, vaccination and chemotherapy. The defining characteristic of infectious diseases – that they are transmissible – means that an individual's risk of acquiring infection changes dynamically as levels of infection rise and fall in the population. Therefore vaccinating individuals or treating infectious individuals benefits not only the individual patient directly but also benefits others in the population indirectly by reducing their risk of acquiring infection through the reduction in levels of infection in the population. This was demonstrated by the effect of vaccination against pneumococcus: the incidence of infection with 'vaccine' serotypes fell in those who were older than the target group for vaccination who did not receive the vaccine, as well as those who were vaccinated.[3]

However, 'indirect' effects are not always beneficial and may even be harmful.[4] For example, whilst reducing levels of infection in the population through vaccination protects those who are not vaccinated as well as those who are vaccinated by reducing the overall rate of infection, those who do still get infected are older on average when they get infected and can suffer more severe disease outcomes, so vaccination can increase the overall rate of disease for a period, until infection is eliminated from the population.[4] Mathematical models can be constructed that can help predict these effects and aid the design of strategies to mitigate them.

Mathematical models of infectious disease transmission are increasingly being used to guide public health policy in the areas of naturally disseminated pathogens such as viruses, bacteria and fungi and in developing response measures to mitigate the human and environmental costs caused by the deliberate release of infectious agents, i.e. bioterrorism.[5] Examples include the control of an epidemic of foot-and-mouth disease in the UK in 2001,[6,7] the SARS outbreak of 2003,[8,9] planning control strategies for TB, HIV and STIs,[10–13] and planning for pandemic influenza,[14,15] as well as examining general principles of disease control.[16]

Importantly for infectious diseases, there is typically a complex nonlinear relationship between the size of an intervention and the benefits. This is due to the 'indirect' effects interventions have on those who are not treated by changing levels of infection in the population, which affects the risk of acquiring infection.

As the scale of an intervention (e.g. vaccination coverage or provision of treatment) increases from a low level, the benefits – reductions in levels of disease – increase faster than the costs, until disease has been reduced to a low level or even eliminated. This means that when health economic analyses are performed to determine the cost-effectiveness of interventions, it is essential that the models take account of the transmissible nature of infectious diseases and the 'indirect' effects of interventions.[17] One example is vaccination, where vaccinating only a small proportion of the population mostly benefits only those who receive the vaccine because it does little to interrupt transmission, i.e. the indirect effect is small. Vaccinating a large proportion of the population can prevent epidemics, providing a large indirect benefit to those who were not vaccinated in addition to the direct benefit. Another example is in the control of curable infections (e.g. STIs) through treatment: if treatment capacity is inadequate then there is a 'vicious circle' where failing to control transmission in the present results in more infections in the future, maintaining the inadequacy of treatment capacity.[13] Conversely, making a concerted effort to increase capacity can break this vicious circle and change it to a virtuous circle, where promptly treating a large enough proportion of infections reduces transmission rates, allowing a more intense focus on remaining disease and reducing future need for treatment, leading to significant cost savings.[13]

Infectious disease epidemiology is inherently multidisciplinary because the transmission of infection within a population is affected not just by the biologic characteristics of the infectious agent and its human host, but also by the patterns of contact between individuals, their use of health services, their response to public health interventions, etc. Mathematical models enable information from diverse sources, including social sciences, to be integrated. Infectious disease modeling is not a purely technical, 'mathematical' exercise – many modelers come from biologic or clinical backgrounds.

EPIDEMIOLOGIC DATA

The fundamental measures of the epidemiology of a pathogen in a population are the incidence and prevalence, which are sometimes confused in the nonspecialist literature. Incidence is the number of new infections arising per unit time, and is usually expressed at x% per year or x cases per 1000 per year or x cases per 100 000 per year.

Prevalence is the proportion of the population (usually expressed as a percentage) that is infected at a point in time and is measured by cross-sectional surveys. Incidence can be measured directly in longitudinal cohort studies, where a group of subjects is followed through time, or can be calculated from a series of cross-sectional prevalence surveys.[18] Case notifications are often used as a proxy measure for incidence; long-term datasets are available for a large number of infectious agents due to mandatory (notifiable) disease surveillance schemes in developed countries. Historical datasets containing valuable information on causes of death and age, in some cases going back for hundreds of years, are also available. These datasets are rather subjective but have been useful in examining epidemics such as the Black Death (plague) and smallpox.

Table 5.1 lists 30 disease agents that are currently notifiable by clinicians to the Health Protection Agency (HPA) in the UK and similar arrangements are in place with the Centers for Disease Control and Prevention (CDC) in the USA. In developed countries demographic and clinical information is collected by clinicians when examining patients and raw incidence data can therefore usually be stratified by sex, age, ethnicity, spatial location and other factors such as tobacco and alcohol consumption. This is important because incidence can vary markedly across different groups of people. Such stratified longitudinal studies, where populations are placed into discrete classes, are extremely useful in examining trends of infection rates. In many of these datasets the impact of vaccination on childhood infectious diseases is striking.

Models are tools used throughout science and medicine – they are used to derive diagnoses from observed signs and symptoms and test results. Formulating models mathematically facilitates rigorous analysis and allows quantitative predictions to be made of trends in disease burden and the impact of interventions. The benefit of quantitative analysis in research is that one can determine if a putative cause for an observed effect would have been strong enough to cause the effect – for example, a mathematical modeling analysis of the HIV epidemic in Uganda found that several modes of behavior change (delaying sexual debut, reducing numbers of sexual partners, increasing condom use) must have occurred to explain the observed decline in prevalence. This is because no single change was sufficient to account for the reduced disease burden observed.[19]

Crucially, models allow evaluation of the impact of interventions that have been implemented by allowing comparison with what would have happened in the absence of the intervention. Epidemics have 'natural dynamics', with incidence typically rising to a peak then declining in the absence of any intervention. Therefore, simply observing a decline in incidence following an intervention is not sufficient evidence to demonstrate its effectiveness.[19] Indeed, in some circumstances it is even possible to observe incidence continue to rise despite an effective intervention, due to an increase in the prevalence of infection.[20]

Another important use of models is in setting priorities for empiric research by determining the importance of different 'gaps' in knowledge. This is done by testing the 'sensitivity' of a model's behavior to changes in the value of parameters that are poorly estimated by current data – for example, there is uncertainty in the amount of protection that bacille Calmette–Guérin (BCG) vaccination offers against TB acquisition and against progression to disease in those who are infected. By testing how much varying these parameters affects a model's behavior we can determine how important it is to obtain more precise estimates of each parameter.

Table 5.1 Infectious diseases notifiable to the United Kingdom Health Protection Agency

- Acute encephalitis
- Acute poliomyelitis
- Anthrax
- Cholera
- Diphtheria
- Dysentery
- Food poisoning
- Leprosy
- Leptospirosis
- Malaria
- Measles
- Meningitis
- Meningococcal septicemia (without meningitis)
- Mumps
- Ophthalmia neonatorum
- Paratyphoid fever
- Plague
- Rabies
- Relapsing fever
- Rubella
- Scarlet fever
- Smallpox
- Tetanus
- Tuberculosis
- Typhoid fever
- Typhus fever
- Viral hemorrhagic fever
- Viral hepatitis
- Whooping cough
- Yellow fever

DYNAMICS OF INFECTIOUS DISEASE TRANSMISSION

The transmissibility of infectious diseases means that there is dynamic feedback between the prevalence of infection (or, more precisely, of infectious individuals) and the incidence of infection. This is why dynamic models are required for infectious diseases.

In a typical epidemic, the prevalence of infection rises initially as infection spreads. This causes an increase in the incidence of infection, which in turn causes prevalence to increase even faster – so the epidemic accelerates. Consequently, the supply of susceptible individuals becomes depleted (by their becoming infected) and the incidence falls, even though prevalence may continue to rise for a time. Eventually, the fall in incidence leads to a fall in prevalence because infections are 'lost' from the population (due to recovery, death or emigration) faster than they are replaced by the spreading of infection. In the longer term, the infectious agent may be able to persist in the population (i.e. become endemic) if there is a high enough rate of resupply of susceptible individuals due to birth, immigration, recovery from infection (if there is no lasting immunity) or waning of immunity (if applicable); otherwise the infectious agent will go extinct locally.

The key measure of an infectious agent's ability to spread in a population is the reproduction number (sometimes called the net reproduction number or effective reproduction number), $R(t)$, which is the mean number of new infections caused by a single infectious individual in the population of interest. (Note that '(t)' indicates that the value can change with time – see below.) A related quantity is the basic reproduction number R_0, which is defined as the mean number of new infections caused by a single infectious individual in a population of wholly susceptible individuals, i.e. the basic reproduction number is what the value of the reproduction number would be if the population were totally susceptible. It is important to understand that the reproduction number is specific to the particular infectious agent in the particular population at the particular time, and can be changed by interventions. The value of the reproduction number depends upon the average rate of transmission from an infectious individual and the average duration of infectiousness. An epidemic requires that $R(t)$ be greater than 1, so that the prevalence of infection increases because more than one new infection arises from the

average infected person before that person is 'lost' from the infected population. In the typical epidemic described above, depletion of the 'supply' of susceptible individuals causes $R(t)$ to fall, even though R_0 was not changing. In fact, $R(t)$ falls even as incidence rises; the increase in incidence is driven by the increase in prevalence, which initially 'overcomes' the effect of $R(t)$ falling.

Public health interventions aim to reduce and maintain $R(t)$ below 1, which may be achieved by reducing the average infectious period (e.g. through treatment or isolation) or the transmission rate (e.g. by closing schools and workplaces to combat SARS or influenza, or promoting condom use and reductions in numbers of sexual partners to combat STIs) or vaccinating people to remove them from the susceptible population. The higher the value of R_0, the harder an infection will be to control. In a homogeneous population (one where everyone has the same average risk of acquiring and transmitting infection) the relationship between R_0 and $R(t)$ is $R(t) = R_0 \times s$ where s is the proportion of the population that is susceptible. To prevent an epidemic by vaccination requires that s be reduced so that $R(t) < 1$ (i.e. that s be reduced below $1/R_0$), thus the greater the value of R_0 the smaller s must be. The critical vaccination threshold is the proportion of the population that must be successfully immunized to prevent an epidemic; for childhood infections such as measles which have high typical R_0 values this is typically >90% or even >95%.

It is important to realize that R_0 alone does not provide complete information on the transmission dynamics of an infectious agent. A highly infectious agent that spreads rapidly but which has a short infectious period could have the same R_0 value as another infectious agent that is much less infectious but which has a longer infectious period – the latter agent would spread more slowly but for longer.

COMPARTMENTAL MODELS OF INFECTION

The most common approach used in mathematical modeling of disease transmission is to divide or compartmentalize the study population with regard to their infection status (Fig. 5.1). Note that the structure of the model depends upon the natural history of the infection and so differs amongst infections. Important characteristics of the natural history of an infection are the incubation period (the time from the point of infection until the appearance of symptoms) and the latent period (the time from infection to becoming infectious). These periods vary greatly (from days to years, depending upon the infection) and either can be longer than the other. For HIV and influenza the latent period is shorter than the incubation period, with people becoming infectious before they become unwell, but for pulmonary TB they can be the same, with people becoming infectious at the time they become unwell.

In modeling there is a trade-off between complexity/realism and the ability to understand the model's behavior. Since even simple models can have complex dynamics it is important to make the model as simple as possible, whilst still capturing the essential features of the infection. For example, for gonorrhea the incubation period is often omitted from models[13] because it is short relative to the infectious period – and so has little effect on the dynamics of infection – while HIV's incubation period is long compared with the symptomatic late-stage period and so it is usually incorporated into models.[21] In the case of TB, most people with infection never develop infectious disease (they remain latently infected) and so models distinguish between these states.[22]

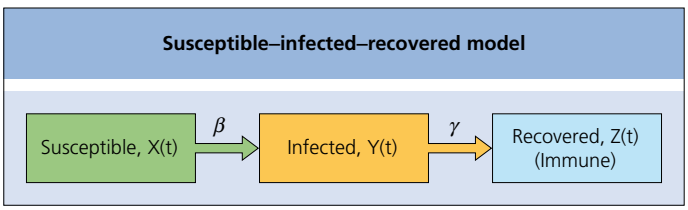

Fig. 5.1 The population is divided into three compartments according to whether they are Susceptible to infection, Infected (and infectious) or have Recovered from infection and are immune. Individuals who become infected move from the Susceptible compartment to the Infected compartment; the process of recovery subsequently moves them from the Infected compartment to the Recovered compartment. The parameters β and γ affect the rate of transmission of infection and the rate of recovery, respectively.

For a directly transmitted pathogen such as influenza, where acquired immunity (to a particular strain) is lifelong, the host population can be represented by three compartments containing the number of susceptible, infected (and infectious) and recovered (immune, non-infectious) individuals. In this example, the latent period is ignored, so individuals become infectious as soon as they become infected. This so-called called 'Susceptible–Infected–Recovered' (or 'SIR') model approach was first developed by Kermack & McKendrick in 1927,[23] elaborated upon more recently by Anderson & May,[4] and now forms the basis for many modern-day models of epidemics.

We present a simple example of an SIR-type model (see Fig. 5.1), which we apply to data from an outbreak of influenza in a boarding school in England.[24] Since the time period of the outbreak is short, we effectively have a 'closed' population: no one enters or leaves the population and there was no mortality due to infection, which simplifies our analysis. (Usually, one has to consider people entering the population through immigration and birth and leaving through emigration and death – and if the infection being modeled causes mortality then infected individuals have an additional disease-induced mortality rate which must be considered.)

Each compartment has a state variable that 'keeps track' of the number of individuals in that compartment, and how that number changes through time. In this case, the state variables are $X(t)$ for the Susceptible individuals, $Y(t)$ for the Infected individuals and $Z(t)$ for the Recovered individuals, where '(t)' indicates that the values can change with time. The total population size is $N(t)$, where $N(t) = X(t) + Y(t) + Z(t)$. (In this particular example, $N(t)$ does not change because the population is closed and there is no mortality.) The model consists of a set of differential equations which describe the rates that individuals flow between different compartments as they become infected, recover, die (if applicable), etc. The net rate of change in $X(t)$ is described by the differential equation $dX(t)/dt$, the net rate of change in $Y(t)$ is described by the differential equation $dY(t)/dt$, etc. In this example, there are two processes: infection and recovery.

The rate of infection (the number of people becoming infected per day) in the population depends upon the force of infection, the risk per susceptible individual of acquiring infection per unit time; and the number of susceptible individuals available to become infected, $X(t)$. The force of infection depends upon the prevalence of infection, $Y(t)/N(t)$, and the transmission parameter, β, which is a combination of the rate of contact between people in the population and the probability of transmission upon contact between an infected person and a susceptible person. Therefore, the force of infection is $\beta Y(t)/N(t)$ and the transmission rate is $X(t) \times \beta Y(t)/N(t)$, which is conventionally written as $\beta X(t)Y(t)/N(t)$. Since infection transfers people from the susceptible compartment ($X(t)$) to the infected compartment ($Y(t)$), the term $\beta X(t)Y(t)/N(t)$ appears negatively in $dX(t)/dt$ and positively in $dY(t)/dt$. (Note that the transmission parameter, β, does not change with time; changes in the infection rate are due to changes in $Y(t)/N(t)$ and $X(t)$.)

The rate of recovery (the number of people recovering per day) depends upon the per capita rate of recovery, γ, and the number of people who are infected, $Y(t)$, and is $\gamma Y(t)$. Since recovery transfers people from the Infected compartment ($Y(t)$) to the Recovered compartment ($Z(t)$), the term $\gamma Y(t)$ appears negatively in $dY(t)/dt$ and positively in $dZ(t)/dt$. (Note that the per capita rate of recovery, γ, does not change with time; changes in the recovery rate are due to changes in $Y(t)$.)

The equations of the model are:

$$\frac{dX(t)}{dt} = \frac{-\beta X(t)Y(t)}{N(t)}$$

$$\frac{dY(t)}{dt} = \frac{\beta X(t)Y(t)}{N(t)} - \gamma Y(t)$$

$$\frac{dZ(t)}{dt} = \gamma Y(t)$$

$$N(t) = X(t) + Y(t) + Z(t)$$

Note that this model is deterministic, i.e. random (stochastic) events are not considered. This is a common simplification that makes it much easier to gain insight into the fundamental dynamics of transmission because the effects of random chance, which cause fluctuations in the graph, are omitted. This model was fitted to data from an outbreak of influenza in a boarding school in England[24] (Fig. 5.2) by estimating values of β and γ.

The algebraic expression for R_0 depends upon the particular model. For this model, it can be derived simply. R_0 is the mathematical product of the transmission rate from a single infected individual in a wholly susceptible population and the average infectious period. The rate of transmission from a single infected individual when the population is wholly susceptible (i.e. when $Y(t) = 1$ and $X(t) = N(t)$; we ignore the fact that really $X(t) = N(t) - 1$ because one person is infected, because we assume that $N(t)$ is large) is:

$$\frac{\beta X(t)Y(t)}{N(t)} = \frac{\beta N(t).1}{N(t)} = \beta$$

The average infectious period is the reciprocal of the average recovery rate (the faster people recover, the shorter their infectious period), which is $1/\gamma$. Therefore $R_0 = \beta/\gamma$. The estimated values of β and γ obtained by fitting to the data were $\beta = 1.97$ per day, $\gamma = 0.47$ per day (corresponding to a mean infectious period of 2.12 days). Therefore the estimated R_0 value was 1.97 day^{-1}/0.47 day^{-1} = 4.18.

Note that there are other types of model used to represent infectious disease transmission, including stochastic compartmental models, individual-based network simulation models and spatial metapopulation models.[25]

EXAMPLES

Below we discuss SARS and influenza. Recent reviews of modeling of STIs and HIV[10,26,27] and TB[22] can be found elsewhere.

Example output of a 'SIR' (Susceptible–Infected–Recovered) model

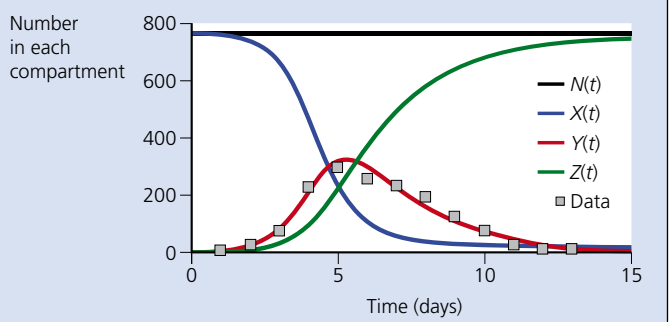

Fig. 5.2 Example output of a Susceptible–Infected–Recovered (SIR) model applied to data from an outbreak of influenza. Model parameters were adjusted to fit the number of Infected individuals, $Y(t)$, to the observed data.

Severe acute respiratory syndrome

The huge rise in the volume of international travel and the huge growth in population densities in many cities, especially in Asia, offer new challenges in controlling the spread of new epidemics. The first such epidemic, 'the first severe and readily transmitted disease of the 21st century' was SARS.[28] On November 16, 2002 the first human case of SARS was identified in Guangdong province in China. From here the disease is known to have spread quickly to other parts of Asia, then to Europe, the Americas and elsewhere, infecting >8000 individuals in 29 countries and killing at least 774 people.[8,28] SARS is caused by a coronavirus (SARS-CoV) normally found in wild animals such as the palm civet cat and Chinese ferret badger.[29] Early cases are thought to have involved zoonotic infection from animal reservoirs, a jumping of the species barrier, but genetic changes in the coronavirus enabled human-to-human transmission which accounted for the vast majority of cases in the global pandemic of 2002/3.

The causative agent of SARS was identified at an early stage of the outbreak and its genome was sequenced in a timely effort involving a multinational collaborative effort.[30] The sequence, published on May 30, 2003, showed it to be a single-stranded RNA virus containing 11 presumptive genes (open reading frames). Other RNA viruses include influenza and HIV and these are thought to be particularly prone to mutation as they are not proofread by DNA polymerases before transcription. The WHO global alert of March 2003 alerted countries to the spread of SARS and accurate case definitions were communicated to identify symptomatic carriers worldwide which led to the rapid control of the disease. This effective collaboration between health-care professionals in different countries and the sharing of patient and demographic data informed public health policy which limited the scale of the epidemic so that more than half of the countries affected reported fewer than 10 cases.

Sophisticated mathematical models of the SARS epidemic of 2002/3 have been developed, providing estimates of the key epidemiologic quantities and showing how disease transmission was controlled by effective intervention. The mean incubation period of SARS was reported to be between 4 and 6 days in most patients with a generation time of 8–12 days. The number of reported cases in the epidemic with time were used to estimate the initial growth rate (r, the rate of exponential increase in new cases at the start of an epidemic) which is related to the basic reproduction number (R_0) and the generation time (Tg, the mean period of time from a host becoming infected to infecting another individual) by the equation $R_0 = rTg + 1$. Using data from the initial period (before control measures were introduced) of the SARS epidemic in Hong Kong the epidemic growth rate was shown to be approximately 0.15 cases per day[31] which, when used in the above equation with $Tg = 10$, gives an R_0 value of 2.5, an estimate close to that made by analyses of data from Singapore by Lipsitch *et al.*[32] of between 2.2 and 3.6 days. Wallinga & Teunis[33] analyzed incidence data from four countries before and after the WHO global alert and subsequent control measures of March 2003. Before and after the alert, average $Z(t)$ values for Hong Kong, Vietnam, Singapore and Canada were 3.6 before the alert (0.7 after), 2.4 (0.3), 3.1 (0.7) and 2.7 (1), respectively. The reduction in the reproductive number in each country reflects the effectiveness of control measures such as quarantine and travel restrictions in curbing the epidemic.

SARS transmission was linked to close contact with another case and most of these were hospital-acquired infections of health-care workers or patients.[8,34] The mortality due to SARS calculated from WHO figures of 8098 probable SARS cases and 774 deaths gives a crude case fatality rate (CFR) of approximately 10.5%; however, the actual CFR is strongly positively correlated with age, with mortality in those over 65 years old exceeding 50% in a number of studies reviewed in Donnelly *et al.*[8] Estimating the case fatality rate of newly emerged pathogens is difficult as defining true cases can be problematic. The CFR may be overestimated if many subclinical infections go

uncounted. Alternatively, in epidemics where patients are hospitalized for lengthy periods, the CFR may be underestimated as patients can be recorded as cases before their outcome is known. In the 2003 SARS epidemic estimates of the CFR became more accurate as the fate of more patients became known but at the time this increasing CFR was wrongly taken as evidence that the virus was increasing in virulence.[8]

Influenza

In common with SARS, influenza is a disease that has been extensively modeled using data from past pandemics to predict the effectiveness of particular interventions in different disease scenarios. Influenza A is divided into subtypes based on differences in hemagglutinin (H) and neuraminidase (N) proteins. The 'Spanish flu' of the 1918 pandemic was an H1N1 lineage that was estimated to have caused up to 80 million deaths – many more than were killed in the First and Second World Wars combined. Small alterations to the influenza A genome occur by a process known as antigenic drift, where mutations increase in prevalence in the population driven by the selection of mutations in the viral genome in proteins exposed to the host immune system. These small changes lead to the differing severity of seasonal flu epidemics.

From a public health perspective the rarer phenomenon of antigenic shift is much more worrying. This results from the recombination of influenza genes to give novel combinations of virulence genes such as strains of avian flu of the H5N1 subtype that is widely feared to be the cause of the next pandemic wave of influenza. Past pandemic strains of influenza A are thought to have emerged from animal reservoirs following antigenic shifts that enabled them to become extremely pathogenic to humans, with surface proteins to which the host had little acquired immunity. Pandemic influenza differs from seasonal epidemic strains not only in the severity of infection but in other ways also. It is not restricted to the 'flu season' of the winter months and it tends to be most lethal in young children who presumably lack the immune memory of older patients.

In 2008 an epidemic of H5N1 influenza amongst Asian and African wild bird and poultry populations led to 385 cases of human disease resulting in 243 deaths – a case fatality rate of 63%.[35] These cases were overwhelmingly in individuals from Indonesia and Vietnam who had been in contact with poultry later shown to be infected with H5N1 strains. Control measures to destroy birds infected with H5N1 have been effective in limiting the number of human cases thus far but experts fear that a strain of highly pathogenic H5N1 will emerge that will acquire the ability to transmit between humans at high frequency which would lead to a global pandemic.

The 1918 influenza pandemic claimed between 50 and 100 million lives worldwide,[36] even although effective infection control procedures were in place in some areas (e.g. in some US cities). Bootsma & Ferguson[37] modeled the impact of infection control measures (such as the banning of mass gatherings, isolation, and improved hygiene and disinfection procedures) on transmission of the H1N1 epidemic in 16 US cities using historical datasets. The found that R_0 was reduced in cities with the most effective control measures (introduced at an early stage), which increased the length of the epidemic but reduced the overall and peak mortality. These data indicate that control measures may have a significant impact on a future H5N1 pandemic if introduced early in the course of the pandemic – in this case reducing mortality by 30–40%. The authors of this study caution against extrapolating data from the study too precisely on modern cities as family units and workplaces nowadays contain fewer people who are generally healthier. However, an H5N1 virus causing death in 14–33% of individuals infected would be expected to kill more individuals globally. This mortality could be mitigated by the rapid deployment of an effective H5N1 vaccine and the use of antiviral drugs such as neuraminidase and influenza A protein M2 inhibitors.[38]

Modeling pandemic influenza

A case fatality rate of 63% for avian influenza (from the number of cases and fatalities reported by WHO above) will be higher than the CFR during a pandemic as the number of cases reported so far will be an underestimate due to the non-inclusion of many nonfatal, mild and asymptomatic cases which will not be entered on the WHO reporting system. Additionally, some authors maintain that changes in the viral genome resulting in high rates of human-to-human transmission may cause a reduction in virulence in humans. From historical data the CFR for pandemic influenza A was calculated at between 0.1% (1957 and 1968 pandemics) and 2.5% (1918).[39] However, a recent article by Li *et al.*[39] suggests that a CFR of between 14% and 33% may be more realistic for a human transmissible H5N1 strain derived from an avian reservoir.

The examination of the SARS epidemic using mathematical models demonstrates some key qualities that enabled it to be contained effectively; however, when we compare these to features of past influenza epidemics and pandemics it appears that containing a future H5N1 pandemic will be much more difficult using similar containment/control measures. The generation time for influenza (4–6 days)[40] is much shorter than for SARS (8–12 days)[9,32] which means that influenza will spread much quicker than SARS given an overwhelmingly naive population (e.g. an H5N1 genotype epidemic). This will make a human transmissible H5N1 epidemic much harder to control than SARS. The R_0 of SARS of ~2.5 is similar if not higher than that estimated from reanalyses of pandemic influenza of 1.4–3.0,[40,41] but with a much higher expected CFR. In younger age groups in particular the impact of H5N1 would be expected to be much more costly not only in terms of disease but also from an economic viewpoint as a larger proportion and number of working-age individuals will be removed from the workforce by influenza.

One caveat about H5N1 that should be mentioned is that the factor that has primarily restricted transmission from birds to humans is that the sialic acid linkage favoured for binding to respiratory epithelium by highly pathogenic H5N1 is found primarily in the lower respiratory tracts of humans. This is one of the reasons that respiratory distress is such a common cause of death. The relative lack of these receptors in the upper respiratory epithelium causes much lower titers of virus to grow in nasal mucosa; hence, the virus is more pathogenic but much less transmissible by infected humans. In order to maintain its pathogenicity, the virus would need to maintain its tropism for the lower respiratory epithelial sialic acid receptor linkage and to acquire the ability to simultaneously bind to sialic acid linkages present in upper respiratory epithelium. It is quite possible (indeed likely) that the virus would have to compromise and be less pathogenic if it becomes more transmissible. Thus, a more highly transmissible H5N1 might be less transmissible than SARS and currently circulating strains of influenza A and less pathogenic than the strains that have to date been acquired directly from birds.

FUTURE RESEARCH

There is increasing integration between infectious disease modeling and empiric research in the field and laboratory. As noted above, models can be used to help set research priorities by determining which gaps in knowledge are most important epidemiologically. Increases in computing power make it possible to develop increasingly sophisticated simulation models and to use them in real-time to analyze outbreaks to determine whether interventions are

working and to guide policymakers in their response. DNA fingerprinting techniques are being used to identify clusters of transmission between individuals[42,43] and there is currently a lot of interest in synthesizing analysis of evolution and transmission dynamics, termed 'phylodynamics'.[44] Another area of research is in characterizing contact patterns between individuals in more detail[45] since this has important consequences for patterns of transmission. Finally, the use of modeling in planning and evaluating clinical trials has been advocated.[46]

REFERENCES

References for this chapter can be found online at http://www.expertconsult.com