

Chapter One

Introduction

In the fall and winter of 1918-19 a deadly epidemic of influenza, commonly known as the Spanish flu, erupted in Europe. Soldiers returning home at the end of World War I carried the epidemic to all parts of the world, eventually resulting in the death of at least 20-40 million people and perhaps significantly more (Crosby, 1989; Johnson and Mueller, 2002; Potter, 2001). The major epidemic was preceded by a short and less severe wave that occurred in the spring of 1918. This wave was similar in severity to other influenza epidemics and consequently was barely noticed by medical authorities (Crosby, 1989; Johnson and Mueller, 2002; Potter, 2001), although like the later, more serious wave, it was quickly carried throughout the world by soldiers and other travelers. It has also been suggested that outbreaks of an unusual influenza-like illness observed in England and France in 1916 and 1917 may actually have represented early outbreaks of the same flu strain as that which caused the major pandemic, and that these earlier outbreaks seeded the population of Europe and set the stage for the severe pandemic of the succeeding years (Oxford, 2001; Oxford et al., 2002). Furthermore, in at least some parts of the world the Spanish flu virus continued to circulate until at least 1920 (Johnson and Mueller, 2002).

In February 2003 the World Health Organization received reports of an outbreak of an unusual respiratory illness in China, with 305 cases reported and 5 deaths (Peiris et al., 2003). For the next 5 months, the world watched as this disease, given the name Severe Acute Respiratory Syndrome, or SARS, was carried throughout the world, eventually resulting in 774 reported deaths. From its start in China, the epidemic spread first to Hong Kong, then Vietnam, Singapore, Canada, and elsewhere, and although it eventually reached 26 countries on 5 continents, outbreaks with significant numbers of deaths were limited to only a few locations.

Both epidemics were caused by viruses that spread mainly through respiratory droplet transmission, both were carried rapidly across the globe, and yet the relative impact of the two epidemics differed markedly. What accounts for these differences? Were they due to dif-

ferences in the biology of the two viruses or in the biological response of the human host? Or were they due to differences in the patterns of social contact within and among populations? Was it a combination of both? Why did the Spanish flu kill tens of millions of people while the SARS epidemic killed less than a thousand? What effects did human responses have on the spread, morbidity, and mortality of these epidemics?

A variety of approaches are being used to find answers to these and other questions. Virologists have been called in to identify the viruses that cause the diseases and determine whether specific biological features of the viruses influence their transmission and severity of the disease process. It took a mere 6 weeks from the time the World Health Organization received reports of the outbreak in China until scientists successfully isolated the cause of SARS, a brand new coronavirus (Peiris et al., 2003). The strain of the influenza virus responsible for the 1918-19 epidemic has been isolated from preserved tissues of soldiers who died in the epidemic as well as from the tissues of at least one epidemic victim whose body was preserved in the permafrost layer in Alaska (Taubenberger et al., 2000). Extensive studies have been conducted to try to determine if the 1918 strain possessed any unique biological characteristics that would explain its unusual virulence, but, so far, a definitive answer continues to elude researchers (Taubenberger and Morens, 2006a).

Table 1.1 compares these and other essential features of the 1918-19 influenza epidemic and the 2003 SARS epidemic. A few biological or epidemiological factors are clearly different for the two diseases, especially the relative transmissibility (the ability to be transmitted from one person to another), the importance of asymptomatic cases in spreading the diseases, and the ages at highest risk for infection, but the importance of these differences for explaining the observed epidemic patterns has not yet been determined. Furthermore, although massive public health responses were mounted to try to stop the spread of both diseases, the nature of the measures attempted and the chances of their being successful differed significantly, given that one epidemic occurred at the end of a long World War in the early 20th century while the other occurred at the beginning of the 21st century at a time of relative world prosperity.

The 2003 SARS epidemic brought home the message to health authorities that many characteristics of modern society have increased the risk that infectious disease epidemics will spread quickly across time and space in the decades to come. Yet it is clear from looking at the history of the 1918-19 flu epidemic that this is not really a

new phenomenon. In fact, our renewed interest in the 1918-19 epidemic has dramatically increased fears of new world-wide pandemics, precisely because the levels of travel present during the early 20th century were so much lower than now, and yet they were sufficient to spread a disease that resulted in the death of tens of millions of people. Although the SARS epidemic was not the world-wide pandemic that scientists feared, it still managed to spread to nearly every continent on Earth. This clearly points out how crucial it is to understand how, when, and why epidemics spread across the landscape so that effective planning, preparation, and control measures can be in place before a disaster occurs.

Humans have long recognized that travelers carry diseases from place to place. Setting limits on movement to control the geographic spread of diseases has been a common strategy since at least the time of the Black Plague epidemics in 14th-century Europe. In fact, the word quarantine is derived from the Italian words *quarantins* and *quarnta giorni*, which refer to a forty-day period during which ships, their goods, crew, and passengers were isolated in the Port of Venice during the 14th and 15th centuries (Markel, 1997). Italian authorities believed that an isolation period of 40 days would be sufficient to dissipate the causes of infections (Dorolle, 1968; Matinovic, 1969; Markel, 1997; Miller, 1993; Musto, 1988; Spencer, 1967).

Recognition of the importance of an activity in disease transmission does not guarantee that it will be addressed by scientists, however, and the spatial aspects of disease spread have more often than not been omitted from mathematical models, which have tended to stress who becomes infected, when they become infected, and why they become infected, but not where the transmission occurs and where the disease is spreading. Who, when, and why are important questions with answers that are necessary in order to determine how resources can be targeted to treat cases and institute preventive measures, but where the disease is predicted to spread is equally important.

The diffusion of a disease across a landscape is aided by the presence in space of a susceptible population and prevented by barriers of nonsusceptible persons or by empty space (Meade and Earickson, 2000). Geographic models can help us understand where a disease is likely to go given the local structure of barriers and susceptible populations, and can also help to determine where barriers should be placed in order to prevent further spread. For example, although the limited extent of the SARS epidemic is not fully understood, it is almost certain that kinship links and travel to visit relatives and friends on different continents provide at least a partial explanation of why

Table 1.1 A comparison of the influenza and SARS viruses

Feature	1918-19 flu	SARS
<i>Observed characteristics:</i>		
Epidemic length	multiple waves for at least 1 yr	single epidemic of 8 months duration
Cases	not known	8000+
Deaths	20-40 million	774
<i>Geographic characteristics:</i>		
Geographic distribution	world-wide	26 countries on 5 continents
Pattern of spread	variable depending on wave	China to Hong Kong to Vietnam, Singapore, Canada, elsewhere
Mechanism of spread	mostly rail, ship; troop movements	air travel
<i>Mode of transmission, cause, and primary risk factors:</i>		
Mode of transmission	respiratory droplets	respiratory droplets, fecal-oral?
Cause	new strain of known human virus	new human virus
Ages most affected	under 60; young adults especially hard hit	over 60
Specific risk factors	soldier; pregnant female	health care setting
<i>Epidemiological parameters:</i>		
Infectiousness	moderate	moderate
Transmissibility	moderate	relatively low, but possibly superspreaders
Estimated R_0 *	1.5-2.5	2-4 early on; later 1.2-1.6
Incubation period	1-4 days	2-10 days; median 4-7 days, mean 6 days
Asymptomatic cases	significant	not important
Transmission before onset of symptoms	possible	not significant
Transmission after onset of symptoms	possible	possible
Global case fatality rate	variable; average 3%; up to 80% or more	about 10%

* R_0 is the average number of secondary cases caused by a single infectious person introduced into a population consisting only of susceptible persons

Canada was the only place outside Asia that experienced significant numbers of deaths from the disease. As noted in Gould (1989), “ignoring the spatial dimensions of [an] epidemic [is] like predicting the time of an eclipse, but being unable to tell people where they [can] see it.”

The earliest work on the geography of disease centered on mapping, a practice that began as early as the 18th century (Cliff, 1995). By the middle of the 19th century maps of disease distributions were in widespread use and began to be used for hypothesis testing. Mapping is still the foundation of much work on infectious disease spread, and a number of new and complex statistical techniques have been developed to aid in this research. Nonetheless, although progress has been made with mapping techniques, other methods have been developed that provide better explanations for how and, more importantly, *why* infectious diseases spread across the landscape. Many of these methods have been drawn from disciplines other than geography, most notably ecology, epidemiology, and mathematics.

In this book we focus on one major area of study, mathematical epidemiology. Much of the work dealing with spatial aspects of infectious diseases traces its roots to the spread of animal and plant diseases and questions about the spatial distribution of resources. In large part, however, we have chosen to concentrate our discussion on the transmission and geographic spread of human infectious diseases and so we, of necessity, omit many studies that contribute to a wider understanding of the spatial spread of infectious diseases. Furthermore, although there is much theoretical work on this topic that has been of great value to the field of mathematical epidemiology, because of space limitations and personal interests our emphasis in this book is on applications of geographic models. Consequently, we focus on describing and evaluating the methods and results of models that have had their predictions tested by data, although promising approaches that have not yet been validated with existing data are still considered.

1.1 MATHEMATICAL MODELS AND THE GEOGRAPHIC SPREAD OF EPIDEMICS

Most disease models have been used historically to understand the naturally occurring introduction of known diseases and their subsequent spread within and, in some cases, across populations. This still is an important goal of disease modeling activities, but two sit-

uations in the modern world have brought epidemic modeling much more into the limelight. These situations include the emergence of new pathogens, such as SARS or the 1918-19 influenza, not seen before in human populations and the deliberate release of pathogens into human populations, or bioterrorism. By their very nature, the biology and epidemiology of new pathogens are not well understood. And epidemics due to the deliberate release of pathogens are likely to be fundamentally different from natural epidemics, for the simple reason that the pathogen release is likely to involve more than one source and be placed to maximize the rate of spread through a population.

Traditional statistical and mathematical analyses of data from past epidemics may not be suitable to deal with either of these situations, primarily because relevant data and knowledge on which to base the analyses do not exist. As this book will illustrate, mathematical modeling techniques are an important addition to the arsenal of epidemiological tools, especially since they can take advantage of the very limitations in data that compromise other techniques. Computers and computational strategies have become sophisticated enough to allow the development and analysis of more complex and realistic mathematical models. These models are based on an understanding of the fundamental biology of a host-pathogen interaction, and as long as a new disease is relatively similar to known diseases, models can be developed that reflect the underlying biology. In addition, known or suspected differences between the biology of previously known diseases and the new disease can be built into the structure of a model. The model can then be used to predict the outcome of an epidemic, even though humans may have had little prior experience with the disease. Although the predictions are likely to be imperfect, they provide some information with which to respond to the disease.

To take a recent example, mathematical models were used in the very early days of the SARS epidemic to help determine not only how serious the epidemic might become, but also to explore the potential impact of different proposed control measures (e.g., Chowell et al., 2003, 2004; Lipsitch et al., 2003; Riley et al., 2003). Insights from these models were used to show that the virus, if unchecked, could cause a significant epidemic, but that basic epidemiological control measures such as patient isolation and contact tracing could have a substantial impact on the extent of the epidemic. These activities on the part of public health authorities proved to play a major role in limiting the spread of the 2003 epidemic. The structure of the models was overly simplified, especially with regards to heterogeneities in contact and transmission, which were shown to be significant during the epidemic

(Dye and Gay, 2003), but, nonetheless, the models provided important guidance to public health authorities at a critical time when little other information was available.

So besides helping to predict and control the spread of new pathogens like SARS, what else can mathematical models do? They have a number of important uses. For instance, they can be used more generally to help elucidate important patterns in epidemiological data from any epidemic and can help us to further our understanding of the forces that generate these patterns. Sometimes the resulting insights may appear to be common sense in retrospect, but their significance may not be recognized until illustrated with a well-posed model. Mathematical models can also be used to help figure out how important different types of data may be for understanding and predicting disease spread and they can point out fundamental uncertainties in the existing data.

One of the most important roles of epidemic models is that they provide a way to experiment on human populations without actually doing invasive research that would be ethically unacceptable or technically unfeasible. For example, mathematical models can be used to help determine what might happen if a person was infected with deadly disease and then released from the hospital while still infectious to resume normal activities in a community. Essentially, mathematical models are best used as a way to enhance our basic understanding of how a complex system works. Furthermore, a model that is well-structured and adequately tested can be used to answer “what-if” questions about the behavior of a system that may aid in both the prediction of future behavior and the development of ways to alter that behavior, if desired. This process allows authorities to choose the control strategy most likely to be successful given existing, but often limited, knowledge and put it into place before an emergency occurs.

In a practical sense, a common goal of epidemic models is to aid in testing the feasibility of different control strategies at the community or larger scale. In order to do this, however, mathematical models must address a number of questions about the patterns of epidemic spread. A good model must be able to capture the most important details of the mechanism of spread both within a population and across the space that links different communities. However, a mathematical model is always a simplified description of reality. An important task facing a modeler, therefore, is to decide the level of detail at which the model will attempt to describe the system of interest. Although all models incorporate simplifying assumptions, a model that is too

simple will not represent reality adequately, while a model that incorporates too much of the detail in the real world will result in less general results that may not be applicable to situations other than the one being modeled.

There are no hard and fast rules about how much detail to incorporate into a model, but there are a number of important considerations to keep in mind. The intended use of a model is probably the most important factor guiding model formulation. At the very beginning of the modeling process it is essential to know the questions one wants answered, because the structure of the model needs to be adequate to address those questions. For instance, if spatial dynamics are of interest then the model must contain some description of the spatial structure.

It is also essential to have a thorough understanding of the particular biological system being modeled so that the model structure is an adequate representation of that system. Lack of detailed knowledge of a particular process, however, does not prevent its inclusion within a model, but this will necessitate making additional assumptions regarding the process.

In a perfect world a modeler would have available any data needed to estimate the model parameters, but, unfortunately, data are almost always inadequate in real situations. Sometimes a modeler is in the position to collect the necessary data, but more often the availability of data dictates the modeling approach to be employed. In the absence of good-quality epidemiological data, the benefit of generating highly detailed models is questionable, as the lack of data prohibits parameter estimation and model validation. When data are available, the behavior of the model can be compared to that seen in reality; differences in behavior may indicate deficiencies of the model, which require modifications to be made.

Often these modifications include the incorporation of additional biological processes within the model, with a corresponding increase in the number of variables needed to specify the state of the system and the number of parameters needed to specify the model. In some situations, such parameters can be estimated independently of the data set at hand. For instance, in an epidemiological setting, demographic parameters can be estimated from population census data, without reference to data regarding the disease. If this is not the case, then the inclusion of additional parameters can be problematic because their values will have to be estimated using the disease data set.

Although statistical techniques can be used to provide these esti-

mates, simultaneous estimation of several parameters can involve a complex optimization process, and general statistical theory shows that the more parameters that are to be estimated from a given finite data set, the less precise their estimates become. Care must also be taken to avoid overfitting a model to a given data set: the flexibility of a model increases as its number of parameters increases, and so it should not be surprising that a model with many parameters can fit a given data set better than a model with few parameters. In general, epidemic modelers should employ the practice known throughout science as Occam's razor, or the strategy of developing the simplest model that is consistent with observed behavior.

The reliability of models is another important consideration. Policymakers must have confidence in the recommendations that come from an epidemic model. This confidence can only come from comparisons of model predictions with data from actual epidemics — animal, plant, and human. At the very least a model should be able to explain some of the observed epidemic patterns, although it is important to realize that detailed geographic patterns of an epidemic cannot be replicated exactly, partly because of the inherently stochastic nature of real epidemics and partly because of a lack of adequate spatial data. For example, models of the geographic spread of rabies (e.g., Ball, 1985; Jeltsch et al., 1997; Murray et al., 1986; Murray, 1987), can reproduce the year by year wave advances of the disease measured in kilometers, but cannot predict where the disease will be in a more localized area. Recent advances in modeling techniques using hybrid models that combine dynamic models and statistical analyses have begun to more effectively model the actual heterogeneities present in real landscapes (e.g., Smith et al., 2002), but data of sufficient resolution to estimate needed model parameters are not yet available (Grenfell, 2002).

Mathematical models for the geographic spread of human infectious diseases require knowledge about detailed short-term mobility patterns of humans in the course of their daily activities, since those activities are directly responsible for disease spread across space. Human movement patterns are complicated and difficult to study, however. Data sets that would aid in understanding these patterns are probably available but scattered in various proprietary locations. For example, Baroyan and colleagues used airline and other transportation data to estimate mobility in their studies of the geographic spread of flu epidemics (cf. Baroyan et al., 1969, 1971; Baroyan and Rvachev, 1978; Rvachev and Longini, 1985), a strategy that has been used by several other investigators, including Aguirre and Gonzalez (1992),

Bonabeau et al. (1998), Flahault et al. (1988, 1994), and Freeman (2002). Eubank and colleagues have developed epidemic models that have at their core detailed data on the use of bus transportation in a western U.S. city (Eubank et al., 2004). Other potential sources of data on population travel patterns might include traffic counters along major roads or records from credit card use or hotel stays. The difficulty of acquiring and condensing such complex data into usable packets has led some researchers (e.g., Kaplan, 1989; Kaplan and Lee, 1990) to ask the question of whether detailed information about daily activities or complex social structures is truly necessary to understand observed patterns of disease, and, if so, exactly how detailed the data and the model using it need to be in order to adequately reflect the real situation.

Useful models must also be able to be generalized to different infectious diseases and geographical conditions or must be readily adaptable to specific conditions. As conditions change in the progress of the disease, the model must be able to be updated quickly to the new conditions. Of course, the results from the model must be available quickly so that they can be used before the real-time epidemic runs its course.

A particular strength of modeling approaches is that they can allow not only for generalizations to different infectious diseases, but also for generalizations across many different systems — epidemic modelers can often draw upon theory from other population biology settings. In particular, epidemiological systems can often be viewed as predator-prey processes, and so there are many direct analogies between ecological and epidemiological theory. The large body of work on ecological invasions is particularly informative with regard to the study of disease invasion. Conversely, many ecologists have shown interest in epidemiological systems because of the importance of disease in regulating the sizes of natural populations and because better data are often available for epidemiological systems than for ecological ones.

These crossovers among different areas of population biology are often reflected in the types of questions that are asked by modelers of disease processes. For example, epidemiologists are often most interested in practical questions related to epidemic prediction and control strategies. Hence, a modeler taking an epidemiological approach to developing models that take advantage of the large data sets on childhood diseases in Western Europe and North America may ask questions about when the next epidemic will occur, where it is likely to spread, and how effective different control strategies may prove in

the face of an epidemic. Ecologists, on the other hand, are often more interested in understanding the underlying population dynamics and how they change over time, so an ecological modeler might use the same data set and a similarly structured mathematical model to answer how and why disease levels fluctuate within a community over long periods of time and whether such variation is due to random effects or nonlinear (“chaotic”) dynamics.

1.2 STRUCTURE OF THIS BOOK

The remainder of this book will illustrate how mathematical models have been applied to understanding the geographic spread of recent and potential epidemics occurring in modern human and domesticated animal populations. Our goal is to provide content that will be accessible to a wide audience that includes both students and professionals in the biological, epidemiological, and social sciences as well as in mathematics. In-depth discussions of the structure and results of models that have been used to study and understand the patterns of spread of particular epidemics occurring in the recent past will be interspersed with discussions of the essential mathematical concepts and techniques used in these applications.

In Chapter 2 we introduce many of the important concepts used to motivate, structure, and analyze epidemic models in general before we delve into the additional complexity of models concerned with geographic spread. In the remaining chapters we alternate chapters dealing with applications of mathematical and computer models to the geographic spread of infectious diseases with chapters centered on issues related to model structures and analysis. Readers who carefully work through these chapters will be introduced to a number of modeling approaches as well as specific and extensive examples of their use within the literature.