

Chapter 1

Why This Book? An Introduction

1.1 Terms for Germs

For the sake of simplicity, let us start by naming an infectious agent a “germ.” There are countless germs that can infect human, animal, and plant hosts. Germs can be transmitted directly between hosts via respiratory air droplets or bodily fluids (e.g., saliva, blood, or secretions from sexual organs). Germs can also be transmitted indirectly through an intermediary source, for instance via mosquitoes, ticks, rodents, environmental particles (e.g., contaminated water and food) or contaminated blood products. Germs evolve and transform while new germs emerge regularly, implying their supply can be considered infinite. A broad distinction is often made between microscopically small germs with relatively short life spans, which replicate within their hosts (often called microparasites such as viruses, bacteria, and fungi), and much larger germs with relatively longer life spans (often called macroparasites such as parasitic worms). Many germs live inside or on the surface of their hosts’ bodies without causing illness or even discomfort. In fact, hosts even depend for their survival on germs (e.g., bacteria in the human gut). However, when germs cause disease in their hosts they are often referred to as pathogens. So when we talk about infectious *diseases* we imply that these are caused by pathogens, which are transmissible between hosts, either directly or indirectly. Infectious diseases have been an important cause of sickness and death throughout the history of mankind. With the agricultural revolution, the world population grew and concentrated in clusters. The density of human hosts thus reached levels that allowed continued local (endemic) transmission of a number of lethal pathogens. It seems therefore that diseases like plague, smallpox, measles, and cholera slowed population growth after the initial high growth rates between 10,000 and 5,000 years ago. Most pathogens in humans emerged and spread out from a local community, as they travelled along with their hosts. Nowadays, they quickly establish anywhere environmental conditions allow them to.

1.2 Models of Infectious Diseases

Mathematical models can take many forms, but essentially they describe a system through mathematical equations. They allow studying how a system changes from one state to the next, as well as the relation between variables used in the equations that define the system. There is a difference with statistical models, which are used to study relations between different variables based on data and to make inferences based on these relations.

The field of infectious disease modeling has been the focus of ever increasing research activity over the last 30 odd years. Figure 1.1 illustrates this by showing the evolution over time of retrieved publications using the search string “model* AND (mathematic* OR statistic* OR simulat*) AND (infect* OR communicable OR epidemic* OR vaccin* OR immuni* OR virus OR viral)” in topics of the Scientific Citation Index (SCI expanded, in ISI Web of Science).

It can be seen that the quantity of publications jumped up in the early 1990s, an observation which holds for many fields of science, likely because in this period access to personal computers and the use of the internet became widely established.

The continuing rising interest in the specific field of infectious disease modeling since the 1990s was likely fueled by various evolutions and events, amongst which we single out the following:

1. Increasing research and computing capacity in multidisciplinary fields, such as mathematical epidemiology and biology, biostatistics, and health economics.

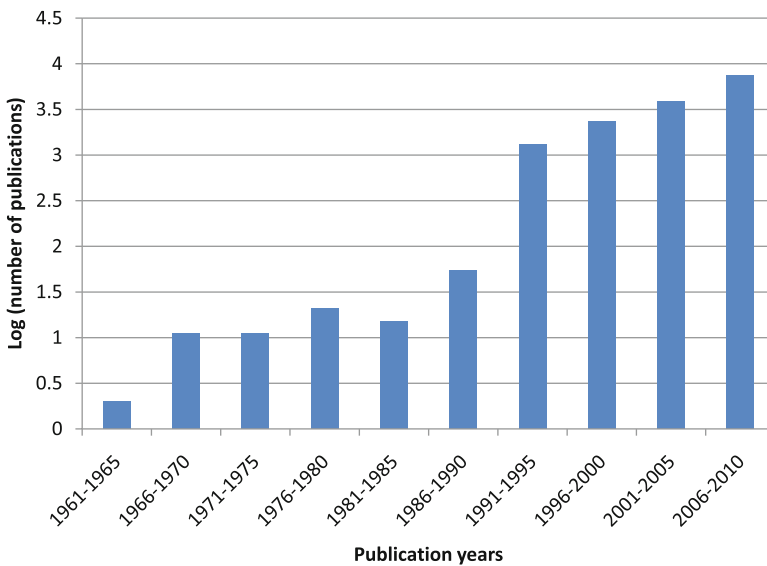


Fig. 1.1 Evolution of publications on models of infectious diseases (on a log-10 scale)

2. Expanded use of mathematical models as part of new standard procedures for evidence-based health policy, often as an implicit part of economic evaluation of pharmaceutical products.
3. New or expanding public health emergencies of international concern (e.g., HIV/AIDS since the 1980s, SARS in 2003, and pandemic influenza A (H1N1)v in 2009).
4. Increasing knowledge base for measuring and categorizing relevant information at both the pathogen and the host level (e.g., genomics, testing of cellular immunity, social contact patterns).

These observations are by no means to imply that infectious disease modeling hardly existed prior to the 1990s. In fact, the central concepts and the foundations for the main techniques used in this field were developed over a century ago and can be traced back to seminal work by d'Alembert (1761), Bernoulli (1766), Ross (1916), McKendrick (1926), Kermack and McKendrick (1927, 1932, 1933), and Muench (1934). Yet the application of these techniques remained much more rare and less internationally dispersed than we can observe today.

While many of the publications in this field are about improving model methodology, by far most of the publications apply a certain modeling approach to a specific problem.

In the application of models of infectious diseases, we distinguish two general aims of such models: forecasting and understanding. By forecasting we mean that projections are made of infections and their consequences under various scenarios of interest (e.g., the time evolution of the number and age distribution of people infected with measles using various scenarios for vaccination schedules and vaccine uptake rates). In relation to forecasting, models are increasingly applied in the slipstream of evolutions 1, 2, and 3 described above. By understanding we mean that models are used that mimic a particular process for infectious disease development or transmission with the aim to improve our knowledge of the process itself, rather than produce estimates of outcomes of this process. In such models the qualitative form and order of magnitude of the results are more important than the exact quantity they represent. Evolutions 1 and 4 above would have been strong drivers for more research on models in relation to understanding the underlying mechanisms of infectious disease transmission, evolution, and development.

1.3 Where Does This Book Fit in the Field?

There have already been many textbooks on infectious disease modeling. Probably the best known and most influential the book written by Anderson and May (1991), who, like Kermack and McKendrick (1927, 1932, 1933), approached the subject from an epidemiologist's angle, using intuitively appealing and elegant mathematical derivation. Their text book has had a major influence on the application of deterministic transmission models, to a wide range of infections. Other textbooks

that followed in the same tradition have extended the mathematics and the range of models considered. Examples here include Farrington (2008) and Keeling and Rohani (2008). The recent book by Vynnycky and White (2010) also fits within that tradition. Their book is elaborate in scope, covering many modeling techniques on a very diverse range of infections and populations. They also provided program code so that the avid reader could apply the techniques and models they discussed. Other types of textbooks in this category are provided by Isham and Medley (1996) and Krämer et al. (2010), who edited and bundled contributed papers covering a variety of specific topics in infectious disease modeling by many different authors.

Notwithstanding that many authors of the above textbooks are mathematicians or statisticians to begin with, other textbooks have approached the subject of modeling infectious diseases or epidemics more from a mathematical rather than an epidemiological viewpoint. The pioneering text book that we consider to fit this bill was produced in 1975 by Bailey (1975). More recent examples are provided by Becker (1989), Daley and Gani (1999), Diekmann and Heesterbeek (2000), and Capasso (2008).

Additionally, there have been textbooks focusing on much more specific subjects and their associated specific challenges. For instance, Andersson and Britton (2000) focused on stochastic models of epidemic data, whereas Halloran et al. (2010) covered the design and analysis of observational vaccine studies.

The book you are reading now aims at filling a gap in the latter tradition of textbooks by zooming in on a specific subject. Our book presents a range of modern statistical and mathematical techniques to estimate parameters that are of pivotal importance in infectious disease modeling. The applications we show in this book are on microparasitic pathogens (with a strong focus on viruses), causing infectious disease in humans. We provide model syntax, as well as R code, with which the statistical and modeling analyses can be applied. We also provide datasets to enable exploration of the techniques. Although it is generally recognized that parameters like the force of infection and the transmission rates between infected and susceptible persons are of very high influence, especially in applications of forecasting, relatively little attention has been given to estimating these parameters to the best of our ability. In view of intensified research in this specific area over the past decade, we considered it the right moment to summarize these developments in this book. We strive not only to explain what to do, given the nature of the data, but we also show how to do it. In that sense this text book is hands-on.

1.4 A Road Map for This Book

The book starts well and truly in Part II, by introducing pivotal epidemiological parameters, describing their properties as well as those of different prevailing mathematical models that mimic the spread of microparasitic pathogens between human hosts. Readers already familiar with these rather basic concepts can press on to Part III.

Part III describes the various datasets used throughout the book. This should give the reader clear insights into the kind of data that are used. The most commonly used data throughout the book are age-specific seroprevalence data and social contact frequencies (to which we also refer as social mixing patterns). Some techniques are also illustrated on incidence data. In Part IV, a wide range of statistical methods is explained and illustrated to derive the force of infection from seroprevalence and incidence data. We describe and illustrate parametric, semiparametric, and nonparametric techniques that use serological test results in their traditional sense (as categorical variables), as well as in their crude form (as antibody levels). Part V shows how to estimate a matrix of transmission probabilities from serology or social mixing pattern data alone or by using the combination of these data. Furthermore, methods to estimate the basic reproduction number are also shown. Here too, various methods are proposed and illustrated. The final part, Part IV, integrates the statistical estimation methods for the pivotal infectious disease parameters (as described in Parts IV and V) in the mathematical model framework described in Part II. Thus at the end, we've come full circle: the modern methods for statistical derivation and inference are integrated in and applied to the mathematical model frameworks we introduced at the very beginning of the book.

We endeavored to demonstrate the methods on a good mix of different applications. These include airborne close contact infections (such as mumps, parvovirus B19, rubella, tuberculosis, and varicella), feco-oral infections (such as hepatitis A), and sexually transmitted and/or blood borne infections such as hepatitis B, hepatitis C, and HIV/AIDS.

We included a short appendix in which we introduce the software package R and several statistical concepts such as maximum likelihood, bootstrap-methods, etc.

We hope you will find this book a useful source of information and inspiration and—perhaps less desirable—transpiration.