Chapter 1
Introduction

Infectious diseases ranging from respiratory (influenza, common cold, tuberculosis, the respiratory syncytial virus), vector-borne (plague, malaria, dengue, chikungunya, and Zika) to sexually transmitted (the human immunodeficiency virus, syphilis) have historically affected the human population in profound ways. For example, the Great Plague, well known as the Black Death, was caused by the bacterium *Yersinia pestis* and killed up to 200 million people in Eurasia and about 30–60% of Europe's population during a 5-year span in the fourteenth century. At the time, the plague infection was thought to be due to some "bad air", but it was not discovered that bites of infected fleas were behind the pandemic until late 1890s. If the human civilization had known about the transmission mechanisms behind the plague infections, the epidemic's impact on morbidity and mortality could have been mitigated through basic public health interventions. This is to say that knowledge of the transmission processes and the natural history of infectious diseases in different environments represents invaluable actionable information for thwarting the spread of infectious diseases.

Fortunately, over the years human civilization has made great strides in increasing our understanding of the transmission dynamics of emerging and re-emerging infectious diseases. For instance, John Snow, known as the father of modern epidemiology, mapped the location of cholera cases during the 1854 epidemic in Soho, London, and made the link between the spatial distribution of cholera cases and a pump that he hypothesized as the source of the disease (Fig. 1.1). Following his observations, the pump was removed to avoid further exposures, and the number of cases subsided.

One remarkable and definite shift to the germ theory occurred during the "golden bacteriology" era during the second half of the nineteenth century. In fact, the 1889–1990 influenza pandemic is arguably the first influenza pandemic that occurred in a new and progressive state of knowledge about infectious disease transmission. This pandemic is better known as the "Russian Flu" because the rapid global spread of the pandemic virus can be traced back to Saint Petersburg, Russia in October

P. Yan, G. Chowell, *Quantitative Methods for Investigating Infectious Disease Outbreaks*, Texts in Applied Mathematics 70,
https://doi.org/10.1007/978-3-030-21923-9_1

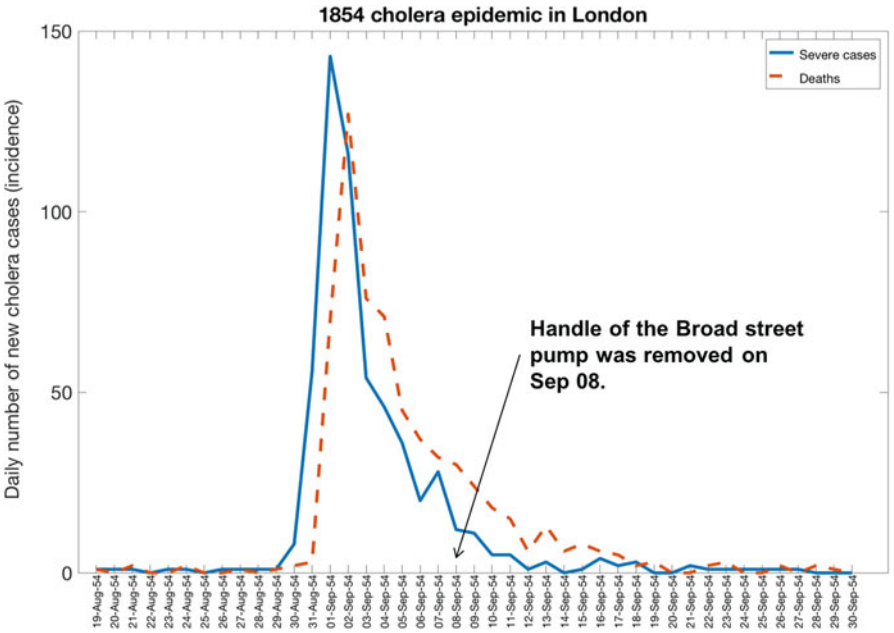

Fig. 1.1 The number of new cholera cases during the 1854 epidemic in Soho, London

1889 (Valleron et al. 2010). Moreover, it was the first pandemic to unfold in a world connected by rail and maritime transportation; it spread across Europe in approximately 6 weeks, with an estimated mean speed at 394 km/week (Valleron et al. 2010) and circulated around the world in just 4 months (Valleron et al. 2010).

Following the 1889–1990 influenza pandemic, in 1918 a novel influenza virus struck the world and killed 20–100 million people, a figure that easily exceeds the death toll associated with World War I (Johnson and Mueller 2002; Dahal et al. 2017; Mills et al. 2004). In the USA alone, about 675,000 people succumbed to the 1918 pandemic virus (Fig. 1.2). However, it was not discovered until years later that an influenza virus was responsible for this pandemic. One hundred years after the 1918 pandemic, we not only remember this devastating historic health disaster, it also serves as a stark reminder of the public health impact that the influenza virus continues to exert on the global population. The 1918 "Spanish Flu" pandemic represents one of the most important case studies for pandemic preparedness available today. However, locating death records to reconstruct the mortality impact of this pandemic requires the arduous task of searching for these documents in old cemeteries, public archives, parishes, and church records (Alonso et al. 2018; Simonsen et al. 2018).

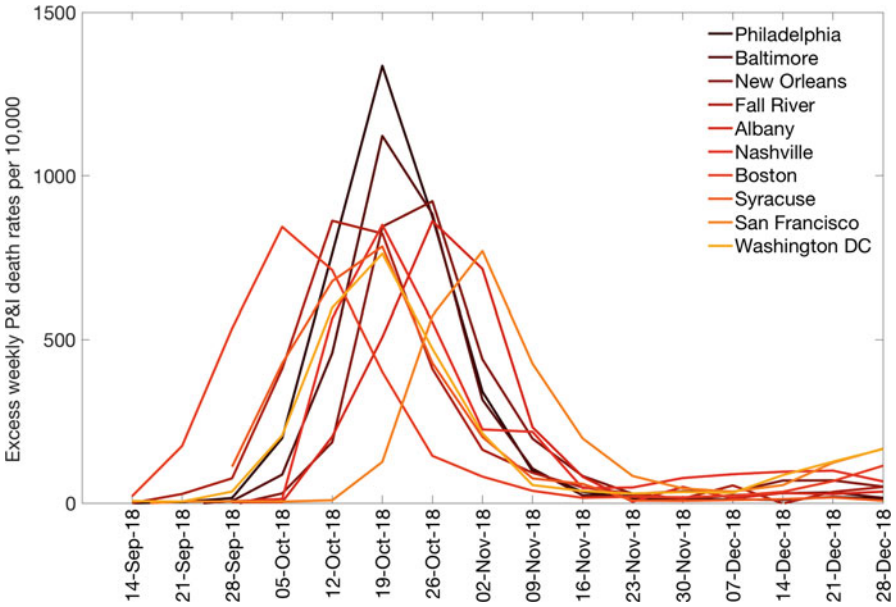

Fig. 1.2 Excess death rate associated with the 1918 influenza pandemic in US cities that exhibited the highest peak excess death rates

The application of mathematical and statistical tools to investigate and forecast evolving epidemics and pandemics has increased significantly during the last couple of decades from $\sim$50 to $>$800 publications per year (Fig. 1.3). The worldwide epidemic of acquired immunodeficiency syndrome (AIDS), caused by the human immunodeficiency virus (HIV), started in the early 1980s and accelerated the applications and developments of mathematical and statistical models. This contributed to the understanding of factors that promote transmission of HIV and of strategies for preventing transmission. While the number of studies that apply mathematical modeling to study infectious disease dynamics has rapidly increased over the last two decades (Fig. 1.3), the great majority of those studies are still associated with HIV/AIDS, although this trend has declined somewhat during the last decade, followed by tuberculosis. In addition, the number of studies associated with emerging infectious diseases such as Ebola, dengue, chikungunya, and Zika has been increasing during the last 5–10 years as a result of recent regional and global epidemics (Fig. 1.3).

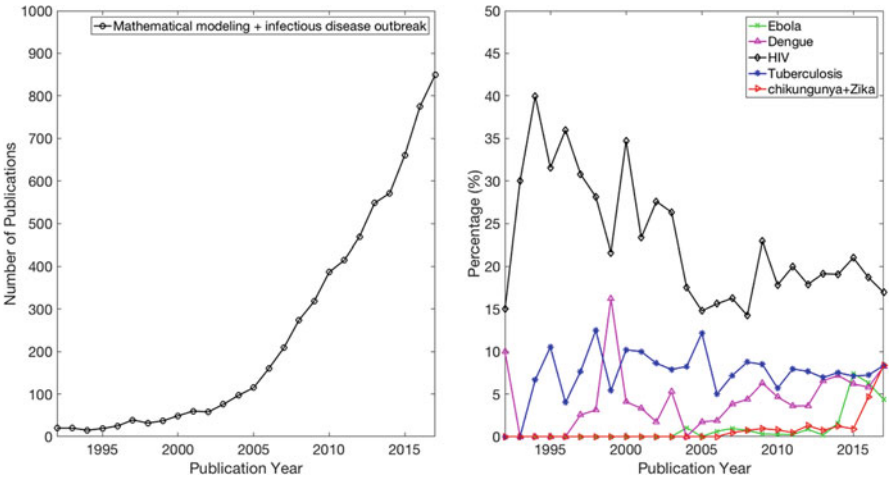

Fig. 1.3 Number of publications on mathematical modeling and infectious disease (left panel) and the fraction of those publications related to different infectious diseases (right panel) by publication year

1.1 The Motivation

Mathematical modeling plays an important role in ordering our thoughts and sharpening vague intuitive notions. Initial models are verbal descriptions that tend to become insufficient as soon as the scenarios become complicated. Mathematics provides a powerful language that forces us to be logically consistent and explicit about assumptions.

Over the years, we have encountered very interesting, inspiring, and challenging discussions at the end of workshops on infectious disease modeling with the following recurrent themes:

1. While most disease transmission models predict an expected exponential growth at the beginning of the epidemic, empirical data often exhibit sub-exponential growth patterns (Viboud et al. 2016). How do we best characterize these non-unique sub-exponential growth functions in the context of infectious disease modeling?
2. Are there many, even infinitely many, mechanisms that lead to the same or very similar sub-exponential growth functions?
3. Does a slower than expected initial growth at the beginning of the epidemic imply a smaller value of the basic reproduction number R_0, a key quantity in the field of infectious disease epidemiology (Anderson and May 1991; Diekmann and Heesterbeek 2000; Brauer 2006), as suggested by many transmission models?
4. What exactly does it mean when we say "deterministic models approximate their stochastic counterparts by the law-of-large numbers"? Are we referring to a population that is infinitely large or something else?

5. Which features of the population-based models, in which the exponential distribution is assumed at the individual levels, can be generalized with non-exponential distributions?
6. Regarding effectiveness of control measures against the spread of diseases, even if imperfect implementation in terms of coverage or compliance has been explicitly taken into account in the models, empirical observations often leave us with impressions that the control measure that "looks good" in theory "do not work at all" in practice. Are there more theories that could capture this phenomenon?
7. How do we reconcile the quantities as predicted by disease transmission models with observed data from outbreak investigations and public health surveillance?
8. The need for precise definitions of verbal descriptions in quantitative analyses. For instance,
   - What do we mean by "a case" when data from outbreak investigations and surveillance are presented as time-series of "number of cases"?
   - Are "generation intervals" consistently defined across literature in epidemiology and infectious disease models?
   - How do we characterize and compare "variability" among random variables, such as the infectious periods or the numbers of secondary infections transmitted by a primary infector?
9. What do we mean by "non-identifiability" when fitting models to data?

Of models formulated in mathematical languages, there are different types that are designed for different purposes.

Broadly speaking, there are mathematical models aimed at facilitating our understanding of the medical, biological, ecological, and social interactions that manifest the outbreaks and epidemics of infectious diseases in order to gain insight into specific questions or to generate theories about what must or might happen; and there are statistical models aimed at capturing the data generation process, for detecting general patterns, predicting epidemic trajectories, managing control strategies, or simply describing epidemic trends. Within both mathematical and statistical models, there are models designed at the population level in a phenomenological way versus models that are individual-based with which researchers aim to capture relevant mechanistic processes.

Individual-based models start from descriptions or assumptions about the evolution of the infectiousness and the natural history of the disease progression within an infected host. These include models for the latent periods, the infectious periods, the incubation periods, recovery, mortality, and so on. Some of the individual-based models also combine social contacts with the evolution of the infectiousness in terms of infectious contacts (Dietz 1995).

Phenomenological models can be deterministic or stochastic and include transmission dynamics models formulated using differential equations or stochastic processes as well as empirical growth functions, such as the generalized logistic growth models. Transmission dynamics models depend on tacit assumptions at the individual level.

The developments of many new statistical models and methods in the study of infectious diseases were driven by the HIV epidemic (Brookmeyer and Gail 1994). Data arising from infectious disease investigations pose unique challenges in classic statistical theory and practice because disease outbreak data do not arise from designed experiments. Each outbreak cannot be repeated naturally under identical conditions, whereas the large amount and multiple sources of clinical data, outbreak investigation data from non-conventional surveys, public health surveillance, and observational data from prevalence and incidence cohorts are collected addressing the same outbreak event. Before statistical methods are used to understand and control the epidemic, statistical models are needed to address the data generation processes, which not only include the epidemiologic and biologic processes that give rise to the disease outbreaks, but also the processes that dictate how data are observed and how a "case" is documented and reported.

When talking about "fitting the model to data," we tend to think of one type of model designed for a specific purpose. However, fitting a dynamic mathematical model to observed outbreak data (e.g., for the purpose of estimating important transmission parameters) involves all three levels of models: the population phenomenological model which depends on tacit assumptions of the individual-based model nested within it, and the statistical model that links the disease transmission process to the data generation process. Very often in practice, these different types of models are considered simultaneously even without the investigators' consciousness.

Driven by the HIV epidemic that started in the late 1970s, the outbreaks of the severe acute respiratory syndrome (SARS) in 2003, pandemic influenza preparedness, and preparedness for other emerging and re-emerging epidemics, the literature on infectious disease modeling has flourished during the past 40 years. However, most articles are confined within subdisciplines according to model characteristics and research focus. While the field of mathematical epidemiology has a long history (e.g., Ross 1911, 1928; Anderson and May 1991; Diekmann and Heesterbeek 2000; Keeling and Rohani 2008; Sattenspiel 2009; Allen 2010; Vynnycky and White 2010; Becker 2015; Andersson and Britton 2012; Manfredi and D'Onofrio 2013; Kermack and McKendrick 1927; Brauer 2006; Brauer and Castillo-Chávez 2001), formal efforts at connecting mathematical models with epidemiological data with the goal of calibrating models for predictive/forecasting purposes have only started to take hold during the last decade (Chretien et al. 2015; Biggerstaff et al. 2016; Chowell 2017; Viboud et al. 2018).

1.2 Structure of the Book with Brief Summary

Chapter 2 provides a review of basic concepts of probability and statistical models for the distributions of continuous lifetime data, closely related to individual-based models that describe the evolution of infectiousness and the natural history of the disease progression. We re-tell the story from a different angle with emphases on the shapes of hazard functions and tail properties of the lifetime distributions

instead of repeating the subject commonly found in a typical textbook on survival analysis. These characteristics have profound impacts on outcomes of the transmission dynamic models at the population level. We will discuss and compare two lifetime random variables, both in terms of magnitude and variability, together with the Laplace transform of lifetime distributions. These concepts will provide the foundations for most of the remaining chapters.

Chapter 3 addresses the distributions of random counts and counting processes, which are closely related to population-based phenomenological models. Section 3.2 provides a framework that links the continuous lifetime distributions at the individual level to the distributions of random counts at the population level. It also provides a historical account. Contemporary discussions on "super-spreading events" as seen in outbreak investigation data in SARS-like diseases are typically associated with transmissions along highly heterogeneous networks characterized by long tailed degree distributions (Lloyd-Smith et al. 2005). Similarly, in the context of incurring accidents, publications in actuarial science journals can be traced back to debates on proneness, contagion, or spells in the first half of the twentieth century that gave rise to important models such as the mixed-Poisson process and the Yule process. Section 3.3 lays the foundation for measuring the evolution of random counts over time, which are key measurements in all population-based models.

Chapter 4 focuses on behaviors of a disease outbreak during the initial phase, immediately after a single (or very few) infected individual are "seeded" into a very large susceptible population. The first part discusses extinction versus growth and relationships among three key parameters: the basic reproduction number R_0, the initial (exponential) growth rate r, and the probability of extinction δ are made and established. With the notion of the "prevalence cohort" (Fig. 4.8), we re-write the classic Lotka equation (4.36) as (4.40) under the assumptions about homogeneous mixing. It reveals that:

1. R_0 only depends on the average value of the infectious periods regardless of the variance or the exact distribution. In models without natural births and deaths in the population, the value of R_0 is not affected by the presence or absence of latent periods.
2. The probability of extinction δ depends on the specific distribution of the infectious periods but is not affected by the presence or absence of latent periods.
3. If the infectious disease does not become extinct during the first few generations, the initial (exponential) growth rate r depends on specific distributions for both the latent periods and the infectious periods.
4. Each of the mathematical relationships between R_0 and δ, and between R_0 and r, as found in the literature, is under a set of strict assumptions on the social contact process and the progression of infectiousness within infected individuals.

Therefore,

1. Given the fixed value $R_0 > 1$ and the infectious periods distribution, the model with latent periods has a smaller initial growth rate r than the one without.

2. Given the fixed value $R_0 > 1$ and the latent periods distribution, the more variable the infectious periods, the smaller the value of r.
3. Without specifying the distributions of the latent periods and the infectious periods, there is no order between the values of r and of R_0.
4. If $R_0 > 1$, without specifying the distribution of the number of secondary infections generated by the primary infectious individual (through the distribution of the infectious periods), there is no order between the values of δ and of R_0.
5. There is a direct relationship between r and δ, rarely mentioned in the literature, that $r = \beta(1 - \delta)$, provided that there is no latent period and that the number of infections produced by a typical infectious individual during a time interval of length x is Poisson distributed with mean value βx. This relationship does not depend on the distribution of the infectious period.

The second part of Chap. 4 emphasizes that the three parameters R_0, δ, and r are intrinsic in the sense that they represent the state of the system at (disease-free) equilibrium when the initially infected individuals are seeded. Section 4.5 presents growth patterns that are most likely to happen when the system moves away from the equilibrium condition. Many discussions are on empirically observed slower growth patterns that largely deviate from the exponential growth assumption (Chowell et al. 2015; Chowell 2017). We attempt to precisely define the sub-exponential growth functions in the context of infectious disease transmission and enlist several assumptions about the transmission dynamics that all lead to such early growth pattern, from the depletion of the susceptible population to scaling of epidemic growth shaped by various factors and their combination including the level of contact clustering and reactive behavior changes (Chowell et al. 2016) and to unobservable individual-level heterogeneity. A special sub-exponential growth function of the form, $(1 + rvt)^{1/v}$, $r, t > 0, 0 < v \leq 1$, is introduced in Chap. 4 which frequently appears in later chapters (6, 8 and 9) in examples and discussions.

Chapters 5 and 6 discuss compartment models when the outbreak moves beyond the initial phase. Much of Chap. 5 is the synthesis of previously published literature on both stochastic and deterministic transmission dynamic models, with our added perspectives. Our interest is to generalize some of the features of these models beyond the assumptions based on the exponential distribution on durations of various stages, and beyond the simple generalizations such as the Erlang distribution (which is a subset of the gamma distribution characterized by smaller variances compared to the exponential distribution with equal mean values). These discussions start in Sect. 5.5.2 and continue in Sect. 6.2.1. In these discussions, Laplace transforms of probability distributions are extensively used as tools to calculate transition probabilities among compartments and average durations within compartments. They are valid for arbitrary distributions without specific assumptions of these distributions. When these distributions are exponential, general results in Sects. 5.5.2 and 6.2.1 return to those published in the literature, such as the expression of the reproduction number as the non-negative eigenvalue of the next generation matrix (van den Driessche and Watmough 2008) as well as in examples in these sections.

We also point out a transcendental relationship among (4.43), (5.66), and (6.24). In these expressions, the Laplace transforms are tools to compare distributions ranked by variability which lead to Propositions 27 and 28 along with discussions in subsequent paragraphs.

Other distinct topics in Chap. 5 are empirical models to describe population-based phenomena without "mechanically" modeling the transmission dynamics at the level of individuals and interactions among individuals. These models are useful for curve fitting, as used in examples later in Chap. 8.

Models in Chap. 6 are more complex and involve intervention measures during the epidemic. Section 6.3 demonstrates a potential application of these models in the context of preparedness for an influenza-like acute respiratory infectious disease with numerical illustrations in hypothetical race-to-treat scenarios and with limited treatment supply. Section 6.5 discusses the impact of unobservable heterogeneity in treatment rates on effectiveness. This section addresses Question 6 in Sect. 1.1. We also draw the attention of the expression $(1 + \phi x v)^{-1/v}$ in (6.31) which echoes the sub-exponential growth function $(1 + rvt)^{1/v}$ in Chap. 4. This is because in both cases, a frailty model from survival analysis is used to model the unobservable heterogeneity among individuals.

Chapter 7 addresses Question 7, 8, and 9 in Sect. 1.1 and serves as a transition between the theoretical topics in previous chapters and Chaps. 8 and 9. The focus is on the data generating processes and statistical issues of fitting models to data. As repeatedly emphasized in Chaps. 4–6, population-based models involve tacit assumptions at the level of individuals, such as the exponential, gamma, or other distributions of the infectious periods. These are conceptual models to address general issues and general patterns, such as the prediction of "incidence" according to time at infection (which is usually unobservable). On the other hand, statistical models address the data generating processes, which include the epidemiology aspects but also the observational schemes, including "case definition," surveillance and reporting, and adjustments for observational biases. In each model, choices are made with respect to which aspects of "the real world" should be included in the description of the model and which should be ignored. These choices not only depend on the perceived importance of various factors, but also on the purpose of each of these models. Frequently, fitting a mathematical model, such as a transmission model, to data collected from surveillance and outbreak investigations involves three types of models (assumptions) that take place at the same time. This requires "nesting" one type of model within another. For example, the statistical model that describes data may involve assumptions of the mean and variance, and in some instances, the assumptions of specific distributions such as Poisson or negative binomial. In addition, the model also handles observational biases such as adjustment for reporting delays (Sect. 7.3). The mean of the statistical model may be a function of time with unknown parameters. This function may involve convolution structures, such as back-calculation (Sect. 7.4), to connect predictions from a conceptual model to expected values of observable outcomes. The conceptual model is thus embedded inside a statistical model. However, this will inevitably result in statistical issues such as non-identifiability (Sect. 7.2). This section mainly

discusses concepts, with a few examples as well as some simple methods where applicable. This is an important field that needs more research and development.

Chapters 8 and 9 focus more heavily on applications, although some models not covered in Chaps. 5 and 6 are presented such as metapopulation spatial models and individual-based network models (Chap. 9). Examples presented are based on a case study for the 2016 Zika epidemic in Antioquia, Colombia (Sect. 8.3), a case study of the 2016 epidemic of yellow fever in two areas of Angola: Luanda (the capital) and Huambo (Sect. 8.4), and a case study of the 2014 Ebola outbreak in Mali (Sect. 9.4).