Epidemiology Primer*

Glenn Ledder

May 2022

Contents

[0]Adapted from [9].

Epidemiology is a rich area for mathematical modeling, thanks to the variety of diseases and the ongoing possibility that novel diseases, such as COVID-19, will arise. While the reader is undoubtedly most interested in COVID-19, a fully-adequate model for that disease is too complicated to serve as a starting point. Instead, we begin with a well-known model called the SEIR epidemic model.

Classification of Epidemiological Models

Several features are needed to classify epidemiology models. We present these in rough order of importance.

1. Transmission Type

   Infectious diseases can be divided into two subgroups, based on the transmission mechanism.

   - **Person-to-person** transmission involves either direct transmission through physical contact or indirect transmission through the environment, such as droplets that enter the air through sneezing or coughing.
   - **Vector-borne** transmission is required for diseases in which the pathogen has a complicated life cycle that requires multiple host species. An example is malaria, which is caused by a protozoan that lives part of its life in humans and part in mosquitoes.

   In this primer, we consider only person-to-person transmission.

2. Time Frame

   - **Epidemic** models have no mechanism for replenishment of susceptible people, so the epidemic burns out when there are not enough susceptible people left to keep the fire going. Of course such models are only valid for short-term scenarios.
   - **Endemic** models are designed for the long term. They always include at least one mechanism for replenishment of susceptibles, typically birth of susceptible individuals.

   In this primer, we consider epidemic models in Sections 1, 2 and 6 and endemic models in sections 3, 4, and 5.

3. Population Constancy

   The total population can be fixed by omitting demographic processes or making sure birth and death rates are equal. This simplifies the model as compared to the more common case of variable size populations and is typically done for all epidemic models as well as some endemic models.

4. Classes

   Although the choice of classes is not more fundamental than the choice of time frame, it is traditional to name models according to the list of classes used in them. In the SEIR model, for example,

   - S is *Susceptible*, for individuals who are at risk of catching the disease.
   - E is *Exposed*, for individuals who have been infected but are not yet infectious. This is a poor choice of term. In everyday language, we would say that a person has been 'exposed' when they have contacted an infectious person, regardless of whether they have caught the infection, but in epidemiology all members of the 'Exposed' class have been infected. As a compromise between accuracy and consistency, we will retain the class symbol E but use the term *Latent* as the class name.

- I is *Infectious*, for individuals who can transmit the disease.
- R is *Removed*, for individuals who are not currently infectious and are immune from further infection.

Note that the classes are epidemiological rather than clinical. Infectious people may or may not have symptoms. Removed people might still be sick, such as people with long COVID. If we are desirous of using a fixed population model, we can even count deceased people as part of the removed class.

5. Processes

   Epidemiological models need a list of processes that cause individuals to move from one class to another. The standard SEIR epidemic model has three processes:

   (a) a transmission process that moves susceptible individuals into the latent class;

   (b) an incubation process that moves latent individuals into the infectious class; and

   (c) a removal process that moves infectious individuals into the removed class.

6. Dynamical System Type

   Discrete-time models are based on algebra, while continuous-time models are based on calculus. Discrete-time models seem more intuitive, but continuous dynamical systems have much better mathematical properties and are more appropriate for settings, such as diseases, where events can occur at any time. We will follow the more common practice of using continuous models in epidemiology.

1 The SEIR Epidemic Model

Dynamical system models in epidemiology are constructed using a method called compartment analysis. The idea of this method is the same as that of accounting. If we are keeping a budget and want to know how much money is in each account, we could keep track of changes and make regular updates to the total rather than counting the money every day. In continuous models, the accounting is done via rates.

1.1 Model Construction

Figure 1.1 is a compartment diagram for the SEIR epidemic model, showing the four classes as compartments and the three processes as arrows that connect compartments.

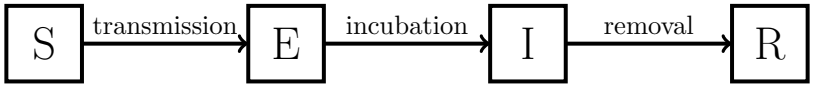

Figure 1.1: The SEIR epidemic model in words.

Notice what we might have included, but did not:

- We have no birth or death processes. Obviously births and natural deaths will occur during a disease outbreak. However, on an epidemic time scale of weeks or months, births and natural deaths only change the class counts to a limited extent. If there are disease-related deaths, we are counting the deceased as part of the removed class so that the population remains constant. Models cannot exactly match reality, and this should not be the goal. Too much detail makes models harder to understand without adding any predictive value.

- We have not included any modifications for public health measures, such as vaccination, isolation of infectious individuals, or quarantine of individuals through contact tracing. These can be added to the base model.

With the compartment diagram showing the process structure, it remains to quantify each of the processes. The standard SEIR model uses mass action incidence and spontaneous transitions, which mean that the risk of infection is proportional to the population of the infectious class and that transitions are modeled like radioactive decay.[1]

It is customary to use lower-case Greek letters for the rate constants of each process. The specific symbol for a particular process varies from one author to another. Most authors use β for the transmission rate constant, but the transition rate constant symbols vary widely. We'll use η for the incubation rate constant and γ for the removal rate constant. Filling in the details in the compartment diagram leads us to Figure 1.2.[2]

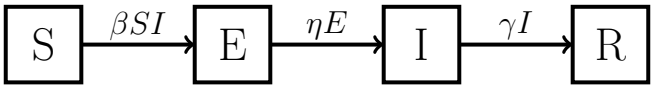

Figure 1.2: The SEIR epidemic model in symbols.

Working directly from the diagram gives us the differential equations that describe the rates of change in terms of the state of the system; for example, the term ηE contributes a rate of decrease to the differential equation for E and a rate of increase to the differential equation for I. Full model specification also requires initial conditions. We use lower-case letters for these to avoid confusion between the symbol for the initial value of R and the symbol $\mathcal{R}_0$, defined in Section 2. We also assume that the class sizes are given as fractions of the total constant population. The final model is then

$$\frac{dS}{dt} = -\beta SI\,, \qquad S(0) = s_0 > 0\,; \tag{1.1}$$

$$\frac{dE}{dt} = \beta SI - \eta E\,, \qquad E(0) = e_0 \geq 0\,; \tag{1.2}$$

$$\frac{dI}{dt} = \eta E - \gamma I\,, \qquad I(0) = i_0 \geq 0\,; \tag{1.3}$$

$$\frac{dR}{dt} = \gamma I\,, \qquad R(0) = r_0 \geq 0\,; \tag{1.4}$$

where

$$s_0 + e_0 + i_0 + r_0 = 1, \qquad e_0 + i_0 > 0. \tag{1.5}$$

These last two requirements (1.5) ensure that the initial population total is 1 and that there are some infected people to get the outbreak started.

1.2 Model Behavior

We defer a careful analysis of the model to Section 2. For now, we simply examine some numerical simulations of the model, shown in Figure 1.3. Note that the parameters η and γ are the reciprocals of the mean times for incubation and recovery; thus, these have been taken as 5 days and 10 days, respectively, roughly matching the data for COVID-19.[3] The specific scenario has everyone susceptible at the beginning except for a very small number of latent individuals who somehow contracted the disease prior to the simulation start.

Several features common to epidemic models are visible in the graphs.

[1] See [9], Sections 3.1 and 3.2.

[2] It is more common to label the arrows with the rates per unit—that is, βI, η, and γ—rather than the rates themselves. I prefer using the full rates so that translation to the differential equations is purely mechanical.

[3] We should emphasize that the SEIR model is not a particularly good choice for COVID-19, as it makes no distinction between symptomatic and asymptomatic cases.

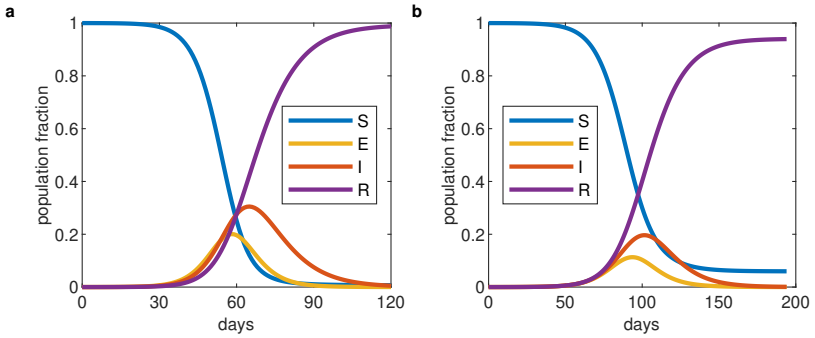

Figure 1.3: Simulation results for the SEIR epidemic model with $\eta = 0.2$, $\gamma = 0.1$, $e_0 = 0.0001$, $i_0 = r_0 = 0$; **a:** $\beta = 0.5$**; b:** $\beta = 0.3$**.**

1. The epidemic gets off to a slow start whenever the initial number of infected individuals (latent plus infectious) is small, in this case just one out of 10,000. Eventually the epidemic takes off in what appears to be exponential growth.

2. The latent and infectious classes eventually peak, with the latent peak preceding the infectious peak. The latent peak is smaller in this example because the latent period is shorter than the infectious period.

3. The epidemic ends with some fraction of the population still susceptible. The size of that fraction depends strongly on the parameter values.

These features will be prominent in the analysis of the next section.

Problems

1.1. Explain the SEIR model. Specifically,

(a) Why is the change in S proportional to both S and I?

(b) Why do the terms βSI, ηE, and γI appear in two different equations, and why are they positive in one instance and negative in the other?

(c) The model

$$\frac{dR}{dt} = kR$$

represents exponential growth of R. Why does

$$\frac{dR}{dt} = \gamma I$$

not represent exponential growth?

(d) Does the model clearly limit how large R can be?

1.2. Suppose a fraction p of infectious individuals self-isolate to reduce their contact rates to a factor of f times that for unisolated individuals. How does this change the SEIR model?

For each of Problems 1.3–1.5, sketch a compartment diagram and write down the appropriate differential equations. Each adds features to the base SEIR model.

1.3. Add a transition process where immunity is lost.
(This problem is continued in Problem 2.9.)

1.4. Add an asymptomatic stage (A) that occurs between the latent and symptomatic (I) stages. Assume that asymptomatic individuals are contagious.

1.5. Add an asymptomatic class (A) that is distinct from the symptomatic class (I). In other words, assume that exposed individuals become either asymptomatic (with probability p) or symptomatic, and that individuals in these classes recover rather than transitioning between A and I. Also assume that the infectiousness of asymptomatic individuals is only a fraction $f < 1$ of the infectiousness of symptomatic individuals.

2 Analysis of the SEIR Epidemic Model

Before conducting a thorough analysis of the model (1.1)–(1.5), we first need to introduce the principal parameter that identifies the ease with which a disease spreads through a susceptible population. To begin, we assume that the mean durations t_L and t_I of the latent and infectious periods are known. Then the transition rate parameters are given as[4]

$$\eta = \frac{1}{t_L}\,, \qquad \gamma = \frac{1}{t_I}\,. \tag{2.1}$$

2.1 The Basic Reproductive Number

The *basic reproductive number*, which is given the symbol $\mathcal{R}_0$ and read as 'R-nought', is the fundamental measure of the infectiousness of a disease.

Definition 2.1 *The **basic reproductive number** is the average number of secondary infections brought about by one infectious person in a population that is wholly susceptible.*

This definition of $\mathcal{R}_0$ sounds complicated, but it is much simpler if we break it down into parts. The transmission rate formula βSI tells us the average number of secondary infections *per day* in a population of any composition. If that population is wholly susceptible, then we get an average of βNI secondary infections per day. That is the number produced by the whole infectious class; with $I = 1$, we see that "the average number of secondary infections *per day* brought about by one infectious person in a population that is wholly susceptible" is βN. To get the basic reproductive number, we just need to take into account that one infectious person has, on the average, t_I days in which to produce secondary infections. Total is rate times time, so the basic reproductive number is

$$\mathcal{R}_0 = \beta N t_I. \tag{2.2}$$

While we have calculated the basic reproductive number in terms of β, the actual use of this formula is often to determine β from $\mathcal{R}_0$. Of course this means that we need to be able to determine $\mathcal{R}_0$ by some other means. You can look up values for well known diseases. The issue of how to estimate $\mathcal{R}_0$ for a novel disease is critically important to accurate modeling results; this will be addressed later in the section.

Note that the value of $\mathcal{R}_0$ is independent of the units used for population class sizes. If we change N, it is the value of β that makes a corresponding change. Assuming we have chosen $N = 1$ by design, we can rearrange (2.2) and replace t_I with $1/\gamma$ to obtain

$$\beta = \gamma \mathcal{R}_0. \tag{2.3}$$

We can therefore specify a particular disease using t_L, t_I, and $\mathcal{R}_0$ as the three fundamental disease parameters and use (2.1) and (2.3) to calculate the parameters that appear in the model.

[4] See Section 3.1 of [9].

Example 2.1.1 *Suppose the first generation consists of 10 infectious individuals. If $\mathcal{R}_0 = 3$, then each infectious person will generate an average of three new infections, for a total of 30 in the second generation. The epidemic will grow explosively as long as the population remains largely susceptible. Only when most of the population has been infected will we stop seeing more infections in the next generation than the previous one. In the end, nearly everyone will have got the disease.*

Now suppose instead that $\mathcal{R}_0$ is close to 1. If $\mathcal{R}_0 = 1.1$, there will be an average of 11 in the second generation. This is enough to keep the outbreak growing for a little while, but at a much slower rate than if $\mathcal{R}_0 = 3$, and a much smaller decrease in the susceptible population will be enough to stop the disease. Continuing with the comparison, if $\mathcal{R}_0 = 0.9$ then the second generation will average 9 individuals. Clearly the disease is unable to get a foothold. Thus, $\mathcal{R}_0 = 1$ is a critical value—a disease can only cause an epidemic outbreak if $\mathcal{R}_0 > 1$.

We can now appreciate why COVID-19 is such a serious problem. The most common infectious diseases just prior to December 2019 were the common cold and influenza, with $\mathcal{R}_0$ values on the order of 1.5 to 3. It is not uncommon for a person to have known exposures to someone with the flu and not get the disease. In contrast, the standard 20th century childhood diseases of measles, chicken pox, and mumps have $\mathcal{R}_0$ values of 10 or more. Before the development of the vaccines for these diseases, virtually everyone who was exposed caught them. The best estimate for the original strain of COVID-19 is $\mathcal{R}_0 = 5.7$ [14], which is a very large value compared to influenza. The delta and omicron strains are progressively more infectious. Initial estimates suggest that omicron has a basic reproductive number of at least 10, similar to that of mumps. Simulations show that in a society that completely ignores the threat, almost the entire population will get the disease in less than 2 months. (This is approximately the situation in Figure 2.1.)

2.2 Goals of the Analysis

Analytical methods (calculus and algebra) and numerical methods (approximation, nowadays with computers) are complementary in many ways. One of these is that the behavior illustrated by numerical simulations can yield conjectures that can subsequently be confirmed by analysis.

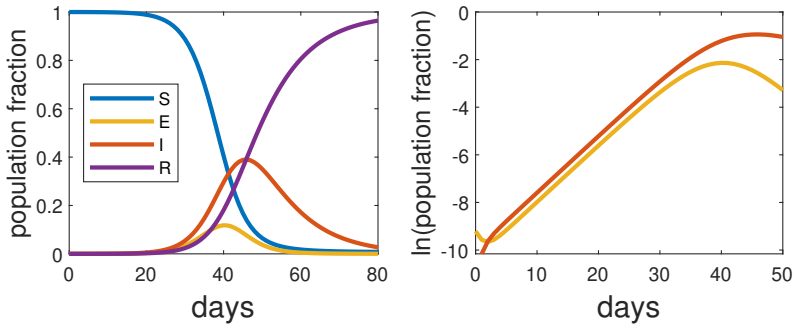

Figure 2.1: Simulation results for the SEIR epidemic model with $R_0 = 5$, $t_L = 2$, $t_I = 10$, $e_0 = 0.0001$, $i_0 = r_0 = 0$.

Figure 2.1 shows the results of a simulation using an initial condition of no infectious or removed individuals and only one latent individual per 10K population. The plot on the left shows the typical pattern of an epidemic outbreak. It takes a while to get started, but then the infection grows rapidly. Both the latent and infectious classes reach a peak and then drop off to 0, with the latent peak occurring earlier in time than the infectious peak. The plot on the right shows some very important detail that can only be seen on a logarithm plot. There is a very fast initial phase during which the latent population is decreasing. This is because we started without any infectious individuals, so new transmissions had to wait until the first batch of latent individuals became infectious. Then there is a significant period of time during which the graphs of $\ln(E)$ and $\ln(I)$ are linear *and parallel*. Only as the latent population comes close to its peak do the graphs

of $\ln(E)$ and $\ln(I)$ begin to curve downward. These same features appear with any realistic choices for the disease parameters, as long as the initial fractions of the exposed and infectious classes are small and $\mathcal{R}_0 > 1$. From this graph, we can make the following conjecture:

- After a short initial adjustment phase, there is a period in which the logarithms of the infected classes are linear with a common slope λ.

The model has six input parameters, the three that define the disease properties and three that define the starting point of the scenario. The goal of analysis is to study their impact on the model behavior. There are a number of possible outcomes we could be interested in. We focus on five of these:

1. The early-phase logarithmic slope parameter, λ, which tells us how rapidly the epidemic grows at the beginning;

2. The maximum infectious class size, I_{max}, which tells us how much impact the epidemic will have at the worst point in time;

3. The time at which the maximum infectious class size occurs, t_{max}, which tells us how much time there is to prepare for the peak;

4. The ending susceptible population, s_∞, which tells us how much of the population did not contract the illness and will be at risk in a subsequent scenario;

5. The total population infected during the scenario, $\Delta S = s_0 - s_\infty$. If we know what fraction of infected people die, we can use ΔS to estimate the total number of deaths in a given initial population.

Figure 2.2 frames the model analysis as a schematic diagram, with model outcomes as functions of model parameters.[5] Our goals are to determine how these model outcomes depend on the parameters and also to use this information to develop a method for determining $\mathcal{R}_0$ for a novel disease.

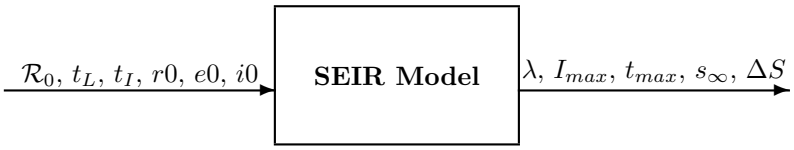

Figure 2.2: Schematic diagram of the SEIR model as a function.

The outcomes will need to be determined by a variety of methods. We can use analytical methods to compute λ and to derive an algebraic equation for s_∞. That equation cannot be solved using analytical methods, but it can be solved with numerical methods. Most epidemic model outcomes, such as I_{max} and t_{max} in the SEIR model, can only be determined by a fully numerical method.

2.3 Early-Phase Exponential Growth

Mathematical exploration of the insight from Figure 2.1 leads to the following result:[6]

Theorem 2.1 *Suppose the initial populations of the infected classes are small compared to that of the susceptible class. Then the SEIR model shows an extended exponential growth phase with*

$$I \approx I_0 e^{\lambda t}, \quad E \approx \rho I_0 e^{\lambda t}, \quad S \approx s_0 \approx 1 - r_0, \tag{2.4}$$

[5] See Section 1.4 of [9].
[6] Problem 2.6.

where λ is the positive solution of the equation

$$(\lambda + \eta)(\lambda + \gamma) = \eta\gamma s_0 \mathcal{R}_0 \,, \tag{2.5}$$

$$\rho = \frac{\lambda + \gamma}{\eta} \,, \tag{2.6}$$

and I_0 is a constant that represents the y-intercept of the straight line for the $\ln I$ plot.

Theorem 2.1 has two very important consequences. First, it gives us a way to estimate $\mathcal{R}_0$ from early data on the infectious class population. From data for $\ln(I)$, we can estimate the value of λ and then use (2.5) to estimate $\mathcal{R}_0$. This is the best way to estimate $\mathcal{R}_0$ for a novel disease, like COVID-19, provided it is done with data that precedes any public health measures.

The other important consequence of Theorem 2.1 is that it allows us to prescribe scenarios using only two initial conditions rather than three. For any scenario that starts with a small infected population, we can choose i_0 and then use $e_0 = \rho i_0$, where ρ is given by (2.6).

2.4 The End State

Dynamical systems often progress toward a fixed end state. These must be states for which all of the rates of change are 0. For endemic models, there are usually only one or two such states and there are straightforward methods to determine which is the end state.[7] The situation is much more complicated for epidemic models. To begin, note that a fixed end state must have no further changes in R. Since $dR/dt = \gamma I$, we can only have a fixed value of R if $I = 0$. We can similarly conclude that a fixed value $I = 0$ also requires $E = 0$. The chain stops here, however. There is no particular reason why a fixed end state should have $S = 0$ or any particular value of R. Based on the differential equation model (1.1–1.4), any state with values

$$S = s_\infty, \quad E = 0, \quad I = 0, \quad R = r_\infty = 1 - s_\infty, \tag{2.7}$$

with $0 \le s_\infty \le s_0$, could serve as the end state.

In the case of the SEIR epidemic model and a few other simple ones, the end state can be found using calculus. The idea is that the relationship between the variables S and R is determined by the differential equations for those two variables. If we assume that R is a function of S, rather than being an independent function of t, we can use the chain rule to obtain an expression for dR/dS. If that expression depends only on the variable S, as is the case here, then we can use integration techniques to find the one-parameter family of anti-derivatives for dR/dS. Only one of these anti-derivatives also satisfies the initial conditions for R and S. The result is as follows:[8]

Theorem 2.2 *The initial and final values of the susceptible population are related by the equation*

$$\ln s_0 - \ln s_\infty = \mathcal{R}_0(1 - r_0 - s_\infty). \tag{2.8}$$

Equation (2.8) is an analytical result that connects the initial values r_0 and s_0, the basic reproductive number $\mathcal{R}_0$, and the final value s_∞. It cannot be used to immediately calculate s_∞ for any set of input parameters because there is no way to solve the algebraic equation for s_∞. However, we can obtain several useful conclusions from the theorem:

1. The final state s_∞ depends only on the basic reproductive number and the initial conditions; it is unaffected by the values used for the rate constants η and γ. If these constants are relatively small for one scenario, then it will just take longer to reach the same final state.

[7] Analysis of long-term behavior is the subject of Section 4.
[8] Problem 2.8.

2. The final state cannot have $s_\infty = 0$ because there are no values one can choose for the other parameters to satisfy (2.8).

3. If we want to see how s_∞ depends on the infectiousness of the disease for a given initial scenario, we can use (2.8) to calculate $\mathcal{R}_0$ values from given values of s_∞ and then plot a graph of s_∞ vs $\mathcal{R}_0$, without ever solving the equation for s_∞.

If we want to be able to calculate s_∞ rather than approximating it from a graph, we need a numerical method, since (2.8) can't be solved for it using algebra. The equation can be manipulated in various ways to put it in the form $F(s_\infty) = 0$. There are a variety of well-documented numerical methods for solving such equations.[9] Scientific computing software, such as Matlab, includes built-in functions for this task. Some care is required, as the methods are harder to use in practice than in theory because of the requirement of having a good initial guess for some functions as compared to others.

Problems

2.1. Our best guesses for the original strain of COVID-19 are a basic reproductive number of 5.7 [14], an incubation period average of 5 days, and an infectious period average of 10 days. Assume an initial population that is entirely susceptible except for ten latent individuals per 100K. Run *SEIR_onesim.m* and describe what would have happened in a community that made no behavioral or public health adjustments.

2.2. [Smallpox]
The Incan Empire had a population of over one million when it was conquered by 168 Spanish Conquistadores in 1525. The Spanish had gunpowder weapons and horses, but these advantages would not have been sufficient to defeat the huge Incan army. (It took about 2 minutes to reload a single-shot arquebus, during which time the number of conquistadores would have been significantly reduced.) They also benefited by joining forces with peoples subjugated by the Incas, but that would not have happened without those peoples assessing the Spanish as having the advantage. Both historian William H. McNeill and natural scientist Jared Diamond have argued that the key factor in the Incan defeat was the European diseases the Spanish brought with them [13, 5]. To test this theory, set the basic reproductive number at 5, the incubation period at 12 days, and the infectious duration at 20 days, values that roughly match smallpox. Use a simulation to study the effect introduction of smallpox into Incan civilization would have had, even without considering the death toll of the disease. Discuss the implications of your findings.

2.3. For a more complete look at the effect of the basic reproductive number on epidemic progression, run *SEIR_comparison.m* using *R0* values 5, 3, 2, 1.5, and 1.25, with incubation period 5 days, infectious duration 10 days, and initial infectious fraction 0.001.

(a) Discuss the graphs, explaining why the effects of *R0* are what you see.

(b) Suppose a disease with *R0*=5 is combated with social distancing and measures to decrease transmission probability for each contact. What effect do you expect these social policies to have and why?

2.4. Use *SEIR_paramstudy.m* to study the effect of the basic reproductive number on epidemic outcomes. Use *R0* values from 0 to 6 and the original default values for the other parameters. Describe and explain the results, paying particular attention to the behavior near *R0*=1.

2.5. Use *SEIR_paramstudy.m* to do a more thorough study of the effect of the disease duration on epidemic outcomes. Use *R0*=2.5 and *tI* values from 4 to 12. Describe and explain the results. Pay particular attention to the axis limits.

[9] These include Newton's method, the secant method, the bisection method, and other less well-known methods. Details of these methods can be found in any introductory numerical analysis book.

2.6. Assume that $S \approx s_0$, $R \approx r_0$, $\ln I \approx \ln I_0 + \lambda t$, and $\ln E = \ln I + \ln \rho$ for some unknown values I_0 and ρ. Derive the results of Theorem 2.1 by substituting these assumptions into the differential equations (1.2)–(1.3).[10]

2.7. Assume that $\mathcal{R}_0$, η, γ, s_0, and i_0 are given, with i_0 small and e_0 unknown. Use the assumptions $S \approx s_0$ and $\ln E = \ln I + \ln \rho$ to derive a quadratic equation whose solution determines ρ. Find the positive solution(s) of this equation to obtain a formula that can be used to select an appropriate e_0 for a scenario that starts shortly after the beginning of an outbreak.[11]

2.8. Assume R is a function of $S(t)$. Use this assumption and the model (1.1)–(1.4) to derive a formula for dR/dS. Then use calculus and algebra to derive the result of Theorem 2.2.

2.9. (Continued from Problem 1.3).
Modify the SEIR program suite to incorporate a transition process where immunity is lost. Assume that the mean time over which immunity is lost is 100 days. Use the values from Figure 2.1 for comparison with that simulation. You will need to run the simulation for 200 days or more.

3 SIR Endemic Models

A variety of assumptions can be used to add demographics to an SIR model, turning it into an endemic disease model. These have varying degrees of realism, so it is always important to think about how realistic the assumptions are in a given context.

3.1 A Generic SIR Model with Demographics

The different versions of SIR endemic models are easier to compare if we embed them in a common framework by defining some generic functions, as shown in Figure 3.1. In any particular model, some of these will reduce to the simple version that appears in the SIR epidemic model.

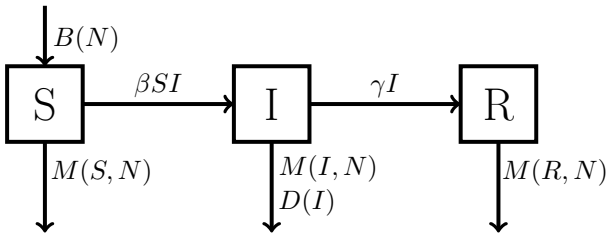

Figure 3.1: The generic SIR model with population demographics.

1. $B(N)$ is the birth rate, which could depend on the current total population N. By making the birth rate a function of the total population only, we are assuming that epidemiological status does not affect one's ability to have children. This is a reasonable assumption, as we will see that the infectious population is almost always small enough to have little effect on the birth rate.

2. $M(X, N)$ is the natural death rate for any population class X. Usually we will assume the natural death rate to be independent of population density, so we'll have μS as the death rate for susceptibles, μI as the death rate for infectives, and so on. If the death rate does depend on population density, then N will appear in the formula as well as X.

[10] Figure 2.1**b** justifies the assumptions when e_0 and i_0 are small; however, it does not justify the additional assumption that I_0 is the same as i_0 because of the small transient at the left edge of the graph. This is why we take I_0 as an unknown to be determined.

[11] This means that we are defining "time 0" to be *after* the initial transient in Figure 2.1**b**.

3. $D(I)$ is the disease-related death rate for class I. Where present, this will almost always be taken as a spontaneous transition process; that is, $D(I) = \alpha I$.

Our generic model assumes mass action incidence for the transmission process. One could instead assume standard incidence.[12] This may be appropriate for some disease scenarios, but we omit this option here because it has no bearing on the general question of how to add demographics to an epidemic model.

From the compartment diagram, the generic model is

$$S' = B(N) - \beta SI - M(S, N)\,, \tag{3.1}$$

$$I' = \beta SI - \gamma I - D(I) - M(I, N)\,, \tag{3.2}$$

$$R' = \gamma I - M(R, N)\,. \tag{3.3}$$

Adding these yields a total population equation, which will be used in place of one of the equation for R:

$$N' = B(N) - M(N, N) - D(I)\,. \tag{3.4}$$

Constant Population

The simplest way to add demographics to the SIR epidemic model is to assume a natural death process without density dependence, no disease-induced deaths, and a birth rate process that exactly balances the death rate; thus, $M(X, N) = \mu X$, $D(I) = 0$, and $B(N) = \mu N$. The population is then constant, as seen from (3.4). The equation for R is not needed, so we have a 2-component model:

$$S' = \mu N - \beta SI - \mu S\,, \tag{3.5}$$

$$I' = \beta SI - \gamma I - \mu I\,. \tag{3.6}$$

3.2 A Variable Population Model with Fixed Birth Rate

If we want to include disease-related deaths in an endemic disease model, we have to have a mechanism that allows either a higher birth rate or a lower natural death rate to compensate for the extra deaths, or else we have to have a net positive immigration rate. Since the rate of additional deaths depends on I rather than N, we can expect the population to be variable rather than fixed. This usually means that a 3-component model is required. While such a model could be written using R in addition to the usual S and I, it is often more convenient to use the total population N as the third component.

Fixed Birth Rate

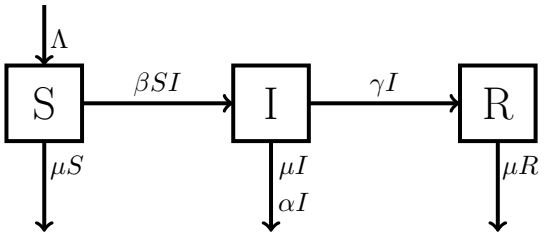

Figure 3.2: The SIR model with fixed birth rate, natural death, and disease-induced death.

The simplest way to incorporate disease-related death is to use a fixed birth rate Λ in place of a birth rate that balances the natural death rate (see Figure 3.2), resulting in the model

$$N' = \Lambda - \mu N - \alpha I\,, \tag{3.7}$$

[12] See Section 3.2 of [9].

$$S' = \Lambda - \beta SI - \mu S\,, \tag{3.8}$$

$$I' = \beta SI - \gamma I - \alpha I - \mu I\,. \tag{3.9}$$

A population greatly decreased by disease-induced death is going to have a lower birth rate than an unaffected population, with a risk of extinction if the death rate is too high. The fixed birth rate model is unrealistic in such a scenario. The death rate μN declines with population but the birth rate Λ does not; hence, the population cannot go extinct. This limits the use of the model to scenarios where disease-induced deaths are not a threat to overall population survival.[13]

This model decouples into a two-dimensional SI system along with an additional equation for N, allowing for two-dimensional methods of analysis. We'll see in Section 4 that the difficulty of model analysis increases significantly as the number of components in the system increases. It would therefore be mathematically beneficial to use the relatively simple but unrealistic fixed birth rate model whenever its results are comparable to those of models that are more realistic but less simple. We can do this for diseases with low mortality.

3.3 Scaling

If we want to simulate a specific disease scenario, we can use the dimensional versions of the models presented here. However, if we want to do analysis, it is much better to scale the models, partly to cut down the number of parameters and partly to facilitate the use of asymptotic methods to simplify the analysis. Here we focus on the model with fixed birth rate: (3.7)–(3.9). The issues involved in scaling any of the other models are similar.[14]

The first step in scaling an epidemiological model is to choose a reference population and a reference time. The reference population is based on the equilibrium value of N in the absence of disease-induced death, and we choose the lifespan of the population as the reference time.[15]

$$K = \frac{\Lambda}{\mu}\,, \qquad T_\mu = \frac{1}{\mu}\,, \tag{3.10}$$

We also define dimensionless parameters

$$\mathcal{R}_0 = \frac{\beta K}{\gamma + \alpha + \mu}\,, \qquad \epsilon = \frac{\mu}{\gamma + \alpha + \mu} \ll 1\,, \qquad d = \frac{\alpha}{\gamma + \alpha + \mu}\,, \tag{3.11}$$

which represent the basic reproductive number, the ratio of the natural death rate to the overall rate of removal from class I, and the fraction of infectious individuals who die from the disease, respectively.

The notation $\ll$ is used in asymptotic analysis to indicate a parameter that is assumed to be arbitrarily small at such time as this information is useful. This assumption is well justified for the parameter ϵ. Given a life expectancy of about 70 years, a disease duration of 3.5 weeks would yield a value $\epsilon = 0.001$. Given that most diseases have shorter durations, ϵ will often be even smaller than this. In many cases, the presence of a small parameter helps simplify the computations for linearized stability analysis.[16] The substitutions

$$X = Kx\,, \qquad \frac{d}{dT} = \mu \frac{d}{dt}\,,$$

where T is being used for dimensional time to reserve t for dimensionless time, yield the dimensionless model

$$n' = (1 - n) - \epsilon^{-1} di. \tag{3.12}$$

$$s' = (1 - s) - \epsilon^{-1} \mathcal{R}_0 si, \tag{3.13}$$

$$i' = \epsilon^{-1}(\mathcal{R}_0 si - i). \tag{3.14}$$

Terms with factors of ϵ^{-1} are processes that occur on a short time scale based on disease duration, while the terms without those factors are processes that occur on a long time scale based on demographic changes.

[13] This is a good time to remind the reader of the importance of matching models to scenarios rather than automatically using a model created for a different scenario.

[14] See [10] for a detailed presentation of scaling for models in epidemiology.

[15] It is equally good to use the mean time in the infectious class, $T_i = 1/(\gamma + \alpha + \mu)$. The choice is a matter of taste.

[16] Section 4.

3.4 Simulations

When setting up a simulation, one can either use the original dimensional model or a dimensionless version such as (3.12)–(3.14). The dimensional version will give graphs with easily interpreted time coordinates, but will also require more parameters to be specified.

For convenience in interpreting dimensionless time, we can define several fixed times and time coordinates. The critical fixed times are

$$T_i = \frac{1}{\gamma + \alpha + \mu}\,, \qquad T_\mu = \frac{1}{\mu}\,, \tag{3.15}$$

which are the expected duration of infectiousness and the expected lifespan in the absence of disease. These are the two parameters that are easiest to estimate for most disease scenarios.[17] For convenience, we assume that these are given in days and years, respectively. Once we have completed a simulation, we will want the graphs to show dimensional time in either days (t_d) or years (t_y). We need formulas to calculate ϵ in terms of T_i and T_μ and both dimensional times in terms of the dimensionless time t. These are

$$\epsilon = \frac{T_i}{365 T_\mu}\,, \qquad t_y = T_\mu t\,, \qquad t_d = 365 t_y. \tag{3.16}$$

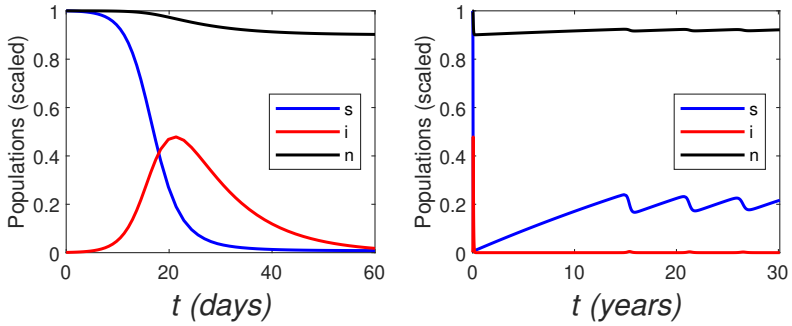

Figure 3.3: The development of a classic childhood disease starting with the initial epidemic, from Example 3.4.1.

Example 3.4.1 *Figure 3.3 shows the early and long-term course of an endemic disease with a disease duration of $T_i = 10$ days, a mortality probability of 10%, a basic reproductive number of 5, a time scale ratio $\epsilon = 0.0005$, corresponding to a life expectancy of roughly 55 years, and an initial state with 0.1% infectious. The plots show the combination of fast and slow processes. There is a massive first wave in which roughly half of the population is simultaneously infectious within a few weeks. By two months, the epidemic has run its course, with virtually nobody remaining in the susceptible class. Slow demographic processes allow the susceptible class to recover gradually while the disease continues to impact a tiny fraction of the population. After about 14 years, the susceptible population reaches roughly $s = 1/\mathcal{R}_0$, which is the level needed for herd immunity. Further births of susceptibles ignites a second epidemic. With such a small fraction of the population in the susceptible class, this second epidemic is small, with barely a blip in the graph of class i. Thereafter, new epidemics occur roughly every 7 years, each with slightly less intensity than the previous one. Over a period of several centuries, the system will approach a steady state at a population that is approximately $1 - d$ times the 'normal' population and with a susceptible fraction of roughly $1/\mathcal{R}_0$, all children. This scenario roughly describes the process by which measles and mumps became childhood diseases.*[a]

[a] It would have been more dramatic for these diseases than for the example, as $\mathcal{R}_0 \approx 10$ for mumps and even higher

[17] It is much easier to estimate T_i from data than to measure its component rates γ and α.

for measles.

3.5 Rescaling

The system (3.12)–(3.14) is conveniently scaled for simulations, but a rescaling makes the problem better suited for analysis. The issue is that infectious class sizes tend over time toward levels on the order of ϵ. Thus, long-time analysis occurs in the regime where i is small. We can make that smallness explicit by defining a new infectious variable y that will be $O(1)$ when $i = O(\epsilon)$. With the substitution

$$i = \epsilon y\,, \tag{3.17}$$

the system becomes

$$n' = 1 - n - dy, \tag{3.18}$$

$$s' = 1 - s - \mathcal{R}_0 sy, \tag{3.19}$$

$$y' = \epsilon^{-1}(\mathcal{R}_0 sy - y), \tag{3.20}$$

This rescaled version has the advantage that the parameter ϵ is now merely a time-scale parameter; that is, it does not affect the equilibrium solutions.

Problems

3.1. [SIR disease with fixed population]
Scale the base SIR model with fixed population (3.5)–(3.6). Compare the result with the fixed birth rate version in the text. What results will be the same for both models, and what results will be different? In particular, address the issue of how the disease impacts total population.

(This problem is continued in Problem 4.3.)

3.2. [SIS disease with fixed birth rate]

(a) Modify the model (3.7)–(3.9) to change from an SIR disease, where recovery confers immunity, to an SIS disease, where recovered patients are once again susceptible.

(b) Scale the model, as in the text. Explain why you do not need the s equation.

(c) Modify *ODEsim.m* to run a simulation, using the same parameters as in Example 3.4.1.

(d) Compare the results with those of Figure 3.3.

(This problem is continued in Problems 3.3 and 4.2.)

3.3. [SIS disease with fixed birth rate] (Continued from Problem 3.2.)
Consider the scaled SIS model with fixed birth rate, from Problem 3.2(b).

(a) Replace d by a parameter $\delta = \mu/\alpha = \epsilon/d$. Then set both derivatives equal to 0 to get algebraic equations for n and i. When $\mathcal{R}_0 > 1$, these will be the long-term stable (equilibrium) values of n and i for the population. The change of parameters means that these algebraic equations will depend only on two parameters instead of three.

(b) Solve the algebraic equations to determine the equilibrium values for n and i as functions of $\mathcal{R}_0$ and δ. We anticipate δ to be small, so write your answers using δ rather than δ^{-1}.

(c) Compare your results with the graphs from Problem 3.2. They should be consistent.

(d) Define x to be the ratio of i at equilibrium to n at equilibrium and determine it as a function of $\mathcal{R}_0$ and δ. Explain why this result has more biological significance than i separately.

(e) Explain the impact of $\mathcal{R}_0$ and δ on n and x and the biological meanings of the results. (It will help to get a simple approximation by keeping in mind that $\delta = \mu/\alpha$ is likely to be small.) In particular, address the question of whether a disease with only modest mortality can severely decrease a population. Think particularly in terms of animal populations in nature.

3.4. [SIR disease with fixed birth rate and loss of immunity]

(a) Modify the model (3.7)–(3.9) by adding a loss of immunity term θR to the S equation.

(b) Scale the model, as in the text, with $h = \theta/\mu$ as an additional parameter.

(c) Modify *ODEsim.m* to run two simulations, using the same parameters as in Example 3.4.1, along with $h = 0.5$ and $h = 5$.

(d) Compare the results with those of Figure 3.3.

(This problem is continued in Problem 4.5.)

4 Analysis of Endemic Models

Endemic disease models generally have multiple equilibria, including a disease-free equilibrium that is always present and an endemic disease equilibrium that is present when $\mathcal{R}_0 > 1$. Stability of these equilibria is determined by linearized stability analysis, which can use the eigenvalues of the Jacobian matrix or the Routh-Hurwitz conditions. These methods are described in this section.

4.1 Eigenvalues and Stability

We begin by summarizing the main stability result for linear systems of differential equations.

Theorem 4.1 (Stability of Equilibria for Autonomous Linear Systems) *The equilibrium solution* $\mathbf{x} = \mathbf{x}^*$ *for the equation* $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$, *where* $\det(\mathbf{A}) \neq 0$, *is asymptotically stable if and only if all real eigenvalues are negative and all complex eigenvalues have a negative real part.*

Most meaningful systems in biology are nonlinear, so a method that works only for linear systems is of limited value. Fortunately, nonlinear systems can be linearized around an equilibrium point, with the results of the approximate linear system usually being valid for the original nonlinear system. The key to doing so is to find the appropriate matrix, which is called the *Jacobian*.

Definition 4.1 *The* ***Jacobian*** *of an n-dimensional system of differential equations is the* $n \times n$ *matrix for which the entry in row* i *and column* j *is the partial derivative of the* i^{th} *function with respect to the* j^{th} *variable.*

Example 4.1.1 *The rescaled dimensionless SIR model from Section 3 is*

$$n' = 1 - n - dy.$$

$$s' = 1 - s - \mathcal{R}_0 sy,$$
$$y' = \epsilon^{-1}(\mathcal{R}_0 sy - y).$$

The Jacobian matrix for this system is

$$\mathbf{J}(n,s,y) = \begin{pmatrix} -1 & 0 & -d \\ 0 & -(1+\mathcal{R}_0 y) & -\mathcal{R}_0 s \\ 0 & \epsilon^{-1}\mathcal{R}_0 y & \epsilon^{-1}(\mathcal{R}_0 s - 1) \end{pmatrix}.$$

Suppose (x^*, y^*) is an equilibrium point for a 2-dimensional nonlinear system. If we zoom in on the equilibrium point in the phase plane, we approach a linear system with matrix given by the Jacobian at that point. We can determine the stability of the linearized system using Theorem 4.1. Because the behavior of solutions near an equilibrium point in the phase space doesn't change as we zoom in, it seems reasonable to expect that the nonlinear system and its linear approximation should have the same stability properties. This is generally true, although there are exceptions. The important result, which holds for systems of any dimension, is summarized in the following theorem:

Theorem 4.2 (Stability of Equilibria for Autonomous Nonlinear Systems) *Let $\mathbf{x}^*$ be an equilibrium point for a nonlinear autonomous system, where $\mathbf{x}$ is a vector containing the state variables of the system. If Theorem 4.1 applies to the linearized system having matrix $\mathbf{J}(\mathbf{x}^*)$, then the conclusion from that theorem applies to $\mathbf{x}^*$ in the corresponding nonlinear system.*

Example 4.1.2 *The system in Example 4.1.1 has a disease-free equilibrium with $n = 1$, $s = 1$, $y = 0$. At this point, the Jacobian is*

$$\mathbf{J}(1,1,0) = \begin{pmatrix} -1 & 0 & -d \\ 0 & -1 & -\mathcal{R}_0 \\ 0 & 0 & \epsilon^{-1}(\mathcal{R}_0 - 1) \end{pmatrix}. \tag{4.1}$$

This matrix is triangular, so the entries on the main diagonal are the eigenvalues. The first and second of these are clearly negative. The third is negative if $\mathcal{R}_0 < 1$; hence, this is the sufficient condition for stability of the disease-free equilibrium.

There is also an endemic-disease equilibrium with $s^* = \mathcal{R}_0^{-1}$, $y^* = 1 - \mathcal{R}_0^{-1}$, $n^* = 1 - dy$, provided $\mathcal{R}_0 > 1$. At this point, the Jacobian is

$$\mathbf{J}(n^*, s^*, y^*) = \begin{pmatrix} -1 & 0 & -d \\ 0 & -\mathcal{R}_0 & -\mathcal{R}_0 \\ 0 & \epsilon^{-1}(\mathcal{R}_0 - 1) & 0 \end{pmatrix}. \tag{4.2}$$

We could determine the stability of this equilibrium by finding eigenvalues; however, we will see an easier method for cases where the eigenvalues are not obvious.

Note that Theorem 4.2 does not apply in certain cases, such as when $\det \mathbf{A} > 0$, $\operatorname{tr} \mathbf{A} = 0$. This is not very common, but it is important to keep in mind that linearization does not always work.

Example 4.1.3 *The system*

$$\frac{dw}{dt} = ap, \qquad \frac{dp}{dt} = p(1-p-w), \qquad k > 0,$$

models the dynamics of a population (p) that is gradually poisoned by its own waste (w). The first

equilibrium equation requires $p = 0$, which automatically satisfies the second equation as well. Thus, all points of the form $(w, 0)$ are equilibria. The Jacobian is

$$\mathbf{J}(w,p) = \begin{pmatrix} 0 & a \\ -p & 1-2p-w \end{pmatrix},$$

so

$$tr\,\mathbf{J}(w,0) = 1-w\,, \qquad \det \mathbf{J}(w,0) = 0\,.$$

If $w < 1$, then the trace of the Jacobian is positive. In this case, Theorem 4.1 indicates that the equilibrium is unstable, and this result carries over to the nonlinear system. However, the theorem does not apply if $w > 1$; hence, we cannot draw any conclusion about stability.

4.2 The Routh-Hurwitz Conditions

So far, the general plan for analyzing a system of differential equations involves two basic computational steps:

1. Rewrite the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ as a polynomial equation for λ, which is called the *characteristic equation* of the matrix $\mathbf{A}$.

2. Check to see if all eigenvalues have a negative real part.

Unless the eigenvalues are obvious from the structure of the matrix, finding them involves a lot more work than actually needs to be done. This work is eliminated by employing some general mathematical results about the relationship between the coefficients of the polynomial in step 1 of the plan and the sign of the eigenvalues in step 2. The following theorem supplies the key fact for a 2-component system:

Theorem 4.3 (Roots of Quadratic Polynomials) *Both roots of the polynomial equation*

$$x^2 + bx + c = 0$$

have negative real parts if and only if

$$b, c > 0\,.$$

This theorem allows us to determine stability for a 2-dimensional system without having to solve the characteristic equation of step 2. We can also eliminate step 1 by connecting the coefficients of the polynomial that determines the eigenvalues directly to the entries in the matrix. Using this connection results in the Routh–Hurwitz[18] conditions, which provide stability criteria in terms of the entries in the matrix, thereby eliminating the need to actually compute $\det(\mathbf{A} - \lambda\mathbf{I})$.

Theorem 4.4 (Routh–Hurwitz Conditions for a System of Two Components) *The equilibrium solution (x^*, y^*) for a nonlinear system with Jacobian $\mathbf{A} = \mathbf{J}(x^*, y^*)$ is asymptotically stable if*

$$\operatorname{tr}\mathbf{A} < 0\,, \qquad \det\mathbf{A} > 0\,,$$

*where $\operatorname{tr}\mathbf{A}$ (called the **trace** of the matrix) is the sum of the elements on the main diagonal. The equilibrium solution is unstable if $\det\mathbf{A} < 0$ or $\operatorname{tr}\mathbf{A} > 0$.*

[18] "Routh" rhymes with "mouth."

The derivation of the result in Theorem 4.4 is straightforward. Calculating the quantity $\det(\mathbf{A} - \lambda\mathbf{I})$ for the general matrix identifies the coefficients b and c as $-\operatorname{tr}\mathbf{A}$ and $\det\mathbf{A}$, respectively.

Example 4.2.1 *The rescaled dimensionless SIR model from Section 3 decouples, so that the sy system can be studied independently of the additional equation for n:*

$$s' = 1 - s - \mathcal{R}_0 sy,$$

$$y' = \epsilon^{-1}(\mathcal{R}_0 sy - y).$$

The Jacobian matrix for this system is

$$\mathbf{J}(s,y) = \begin{pmatrix} -(1+\mathcal{R}_0 y) & -\mathcal{R}_0 s \\ \epsilon^{-1}\mathcal{R}_0 y & \epsilon^{-1}(\mathcal{R}_0 s - 1) \end{pmatrix}.$$

There is an endemic-disease equilibrium for this system with $s^* = \mathcal{R}_0^{-1}$ and $y^* = 1 - \mathcal{R}_0^{-1}$, provided $\mathcal{R}_0 > 1$. The Jacobian at this point is

$$\mathbf{J}(s^*, y^*) = \begin{pmatrix} -\mathcal{R}_0 & -1 \\ \epsilon^{-1}(\mathcal{R}_0 - 1) & 0 \end{pmatrix}.$$

The matrix has

$$tr\, J = -\mathcal{R}_0 < 0\,, \qquad \det A = \epsilon^{-1}(\mathcal{R}_0 - 1)\,;$$

therefore, the endemic-disease equilibrium is asymptotically stable whenever it exists.

The eigenvalue method is best when the eigenvalues of a matrix are obvious, usually because the matrix is triangular. Otherwise, the Routh–Hurwitz conditions require far less calculation than is required to find the eigenvalues for a two-component system.

The difficulty of stability calculations increases rapidly as system size increases. Generally we must use the eigenvalue method in such cases, but we can only do specific parameter values using computer software rather than general cases. However, for three-component systems it is often possible to obtain general results by using the corresponding Routh–Hurwitz conditions.

Theorem 4.5 (Routh–Hurwitz Conditions for a System of Three Components) *Let $\mathbf{A}$ be a 3×3 matrix. Let $\mathbf{A_k}$ be the 2×2 matrix obtained from $\mathbf{A}$ by deleting row k and column k. Define c_1, c_2, and c_3 by*

$$c_1 = -\operatorname{tr}\mathbf{A}\,, \quad c_2 = \sum_{k=1}^{3} \det\mathbf{A_k}\,, \quad c_3 = -\det\mathbf{A}\,,$$

where $\mathbf{A}$ is the Jacobian evaluated at the equilibrium point and $\operatorname{tr}\mathbf{A}$ is the sum of the diagonal elements of $\mathbf{A}$. Then the equilibrium solution of the nonlinear system is asymptotically stable if all three coefficients are positive and $c_1 c_2 > c_3$. The equilibrium solution is unstable if any of the coefficients is negative or $c_1 c_2 < c_3$.

Example 4.2.2 *In Example 4.1.2, we computed the Jacobian for the full 3-dimensional SIR model at the endemic-disease equilibrium:*

$$\mathbf{J}(n^*, s^*, y^*) = \begin{pmatrix} -1 & 0 & -d \\ 0 & -\mathcal{R}_0 & -\mathcal{R}_0 \\ 0 & \epsilon^{-1}(\mathcal{R}_0 - 1) & 0 \end{pmatrix}.$$

The coefficients of the characteristic polynomial are

$$c_1 = 1 + \mathcal{R}_0 > 0, \qquad c_3 = \epsilon^{-1}\mathcal{R}_0(\mathcal{R}_0 - 1) > 0,$$

$$c_2 = \mathcal{R}_0 + \epsilon^{-1}\mathcal{R}_0(\mathcal{R}_0 - 1) > \epsilon^{-1}\mathcal{R}_0(\mathcal{R}_0 - 1) > 0,$$

where we have used the existence requirement $\mathcal{R}_0 > 1$ to show that all three coefficients are positive. It is also clear that $c_1 c_2 > c_3$; hence, the endemic-disease equilibrium is stable whenever it exists.

In many problems, the formulas for the values of the variables at equilibria are complicated. In such cases, it is almost always better to use the Routh-Hurwitz conditions with algebraic simplification rather than inserting the complicated formulas. It is important to do the right algebraic steps at the right time, just like untying a knot requires the right manipulations in the right order. See Problem 4.1 for an example, and read Appendix G of [9] for a more general description of best practices in algebra.

Problems

4.1. [SIS disease with logistic growth and standard incidence]

(a) Determine the stability of all equilibria for the SIS model with logistic growth,

$$n' = \delta n(1 - n - wx)\,,$$
$$x' = \mathcal{R}_0 x(1 - \mathcal{R}_0^{-1} - x)\,,$$

where n and x are scaled variables that represent the total population and the infectious population fraction i/n, noting any restrictions on R_0 and w for their existence. Two algebraic strategies make this problem easy.

1. Don't multiply out products; instead take product rule derivatives. The first entry in the matrix (the partial derivative of the n' function with respect to n) is then

$$\delta(1 - n - wx) - \delta n$$

rather than

$$\delta(1 - 2n - wx)\,,$$

for example.

2. For the endemic disease equilibrium, use the equilibrium relations

$$1 - n^* - wx^* = 0\,, \qquad 1 - \mathcal{R}_0^{-1} - x^* = 0\,,$$

but do not substitute in the formulas for n^* or x^* that come from solving these equations.

(b) Explain why the algebraic advice given in (a) makes the problem easier.

4.2 [SIS disease with fixed birth rate] (Continued from Problem 3.2.)

(a) Determine the stability of the disease-free equilibrium for the SIS model with constant birth rate,

$$n' = \epsilon(1 - n - y)\,,$$
$$y' = \mathcal{R}_0 y(n - \mathcal{R}_0^{-1} - \delta y)\,,$$

Note that the results could depend on the parameter values.

(b) Determine the stability for the endemic disease equilibrium with $\mathcal{R}_0 = 2$, $\delta = 0.1$, $\epsilon = 0.01$.

(c) Discuss the results with reference to Problem 3.2.

(This problem is continued in Problem 4.6.)

4.3 [SIR disease with fixed population] (Continued from Problem 3.1.)

(a) Determine the stability of the disease-free equilibrium for the SIR model with fixed birth rate

$$s' = \epsilon(1 - s - \mathcal{R}_0 sy)$$
$$y' = \mathcal{R}_0 sy - y\,,$$

Note that the results could depend on the parameter values.

(b) Determine the stability for the endemic disease equilibrium with $\mathcal{R}_0 = 4$, $\epsilon = 0.001$.

(c) Discuss the results with reference to Problem 3.1.

(This problem is continued in Problem 4.7.)

4.4 [SEIR disease with fixed population]
The SEIR model with fixed population is given in dimensionless form as

$$s' = \epsilon(1 - s - bsy)\,,$$
$$x' = bsy - \nu x\,,$$
$$y' = \nu x - y\,,$$

where x and y are the exposed and infectious populations, rescaled because they are both $O(\epsilon)$ at equilibrium.

(a) Find the Jacobian.

(b) Determine the stability of the disease free equilibrium using the parameters $b = 5$, $\nu = 2$, $\epsilon = 0.001$.

(c) Repeat (b) for the endemic disease equilibrium.

(This problem is continued in Problem 4.8.)

4.5 [SIR disease with fixed birth rate and loss of immunity]
(Continued from Problem 3.4.)
Determine the existence and stability of the disease free equilibrium for the SIR model with fixed birth rate and limited immunity, given in dimensionless form as

$$n' = 1 - n - dy\,,$$
$$s' = 1 - s + h(n - s) - bsy\,,$$
$$y' = \epsilon^{-1}(bsy - y)\,,$$

where n and s are the scaled total and susceptible populations and y is the rescaled infectious population. The parameter d is the fraction of people with the disease who die of it, b is the rescaled infection coefficient, and h is the ratio of the mean lifespan to the mean duration of immunity, which could be larger than 1.[19]

(This problem is continued in Problem 4.9.)

4.6 [SIS disease with fixed birth rate]
(Continued from Problem 4.2.)

[19] This model generalizes the SIS model, which has $h \to \infty$, as well as the SIR model, although deriving the SIS model from it is nontrivial because the scaling we used is based on the assumption that the duration of immunity is long compared to the duration of the disease.

(a) Compute the Jacobian for the endemic disease equilibrium of the SIS model with constant birth rate,

$$n' = \epsilon(1 - n - mi)\,,$$
$$i' = \mathcal{R}_0 i(n - i - \mathcal{R}_0^{-1})\,.$$

In so doing, you will find it most convenient to retain the variable i^* rather than to replace it with a solution formula.

(b) Determine the stability of the endemic disease equilibrium. Note that the results could depend on the parameter values.

(c) Discuss the results with reference to Problems 3.2 and 4.2.

4.7 [SIR disease with fixed population]
(Continued from Problem 4.3.)

(a) Determine the stability of the endemic disease equilibrium ($i^* > 0$) for the SIR model with fixed birth rate

$$s' = \epsilon(1 - s - \mathcal{R}_0 sy)$$
$$y' = \mathcal{R}_0 sy - y\,,$$

Keep in mind that both existence and stability can depend on the parameter values.

(b) Discuss the results with reference to Problems 3.1 and 4.3.

4.8 [SEIR disease with fixed population]
(Continued from Problem 4.4.)

(a) Determine the existence and stability of the equilibria for the SEIR model with fixed population, given in dimensionless form as

$$s' = \epsilon(1 - s - bsy)\,,$$
$$x' = bsy - \nu x\,,$$
$$y' = \nu x - y\,,$$

where x and y are the exposed and infectious populations, rescaled because they are both $O(\epsilon)$ at equilibrium.

(b) Discuss the results with reference to Problem 4.4.

4.9 [SIR disease with fixed birth rate and loss of immunity]
(Continued from Problem 4.5.)

(a) Determine the existence and stability of the equilibria for the SIR model with fixed birth rate and limited immunity, given in dimensionless form in Problem 4.5. As part of the existence requirements, you must be concerned about the possibility that the disease renders the population nonviable.

(b) Discuss the results with reference to Problem 4.5.

5 An Endemic Model with Vaccination and Loss of Immunity[20]

Vaccination is generally incorporated into dynamical system disease models as a single-phase spontaneous transition process that ultimately moves everyone out of the susceptible class who does not contract the disease first. These assumptions are questionable on two grounds: first, the significant level of resistance to vaccination against COVID-19 suggests that vaccine refusal needs to be incorporated into infectious disease models; second, the dynamics of the single-phase spontaneous transition does not match the dynamics of vaccination during the rollout of a new vaccine. In this section, we consider a model designed to study the impact of vaccine refusal for an endemic disease. We can ignore problems of distribution and supply that are issues only during a vaccine rollout, but we cannot ignore vaccine refusal.

5.1 Model Development

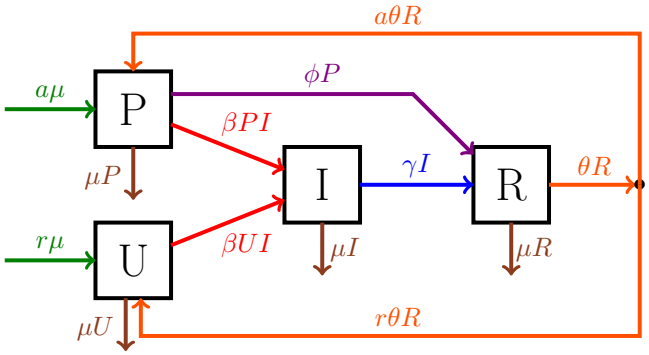

Figure 5.1: The PUIR endemic model.

The schematic in Figure 5.1 incorporates vaccination (violet), vaccine refusal, loss of immunity (orange), and demographic processes of birth (green) and death (brown) into a relatively simple disease model (red, blue). The class structure is PUIR, a modification of SIR in which the susceptible class is subdivided into (P)revaccinated and (U)nprotected subclasses. We assume a constant total population $N = 1$, achieved via a uniform death rate of μY for each class Y, with no disease-induced deaths, and compensated by a total birth rate μ. Prevaccinated and unprotected individuals become infected at rates βPI and βUI, respectively, and prevaccinated individuals move into the removed class by vaccination at rate ϕP. Infectious individuals are removed at rate γI and removed individuals lose immunity at rate θR. The influx of susceptibles due to both birth and loss of immunity is partitioned into fractions $a + r = 1$, where a is the fraction of individuals who accept vaccination and r the fraction that refuses vaccination. The special cases of no vaccine refusal and no loss of immunity are achieved by setting $r = 0$ and $\theta = 0$, respectively.

While the standard way of writing the model would be to use the mutually exclusive classes P, U, I, and R, it is more convenient to think of the model as having mutually exclusive classes S, I, and R, along with an additional state variable P.[21] Thus, we have

$$\frac{dP}{dt} = a\mu - \beta PI \qquad - \phi P + a\theta R - \mu P, \tag{5.1}$$

$$\frac{dS}{dt} = \mu - \beta SI \qquad - \phi P + \theta R - \mu S, \tag{5.2}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \qquad - \mu I, \tag{5.3}$$

$$\frac{dR}{dt} = \gamma I + \phi P - \theta R - \mu R, \tag{5.4}$$

[20] Adapted from [8].

[21] The total susceptible population S figures prominently in the equilibrium and stability analysis.

along with the algebraic equation $S + I + R = 1$.

To simplify the analysis, we scale the system using the mean lifespan $1/\mu$ as the time scale,[22] resulting in the model

$$\begin{aligned} I' &= \epsilon^{-1}(\mathcal{R}_0 S - 1)I, \\ P' &= a - (v+1)P + ahR - \epsilon^{-1}\mathcal{R}_0 PI, \\ S' &= 1 - vP - S + hR - \epsilon^{-1}\mathcal{R}_0 SI, \\ R' &= +vP - (h+1)R + (\epsilon^{-1} - 1)I, \\ 1 &= S + I + R, \end{aligned}$$

where the prime symbol refers to the derivative with respect to scaled time and the new dimensionless parameters are

$$\mathcal{R}_0 = \frac{\beta}{\gamma + \mu}, \qquad v = \frac{\phi}{\mu}, \qquad h = \frac{\theta}{\mu}, \qquad \epsilon = \frac{\mu}{\gamma + \mu} \ll 1. \tag{5.5}$$

Note that $\mathcal{R}_0$ is the basic reproductive number in the absence of vaccination; we will use $\mathcal{R}_v$ for the basic reproductive number with vaccination. The parameters h and v represent the expected number of times an individual loses immunity and receives a vaccination during their lifespan, respectively, and ϵ is the ratio of the mean time in class I to the mean lifespan. Since human infections last weeks while a normal lifespan is measured in years, the numerical value of ϵ is small enough that terms of $O(\epsilon)$ can be neglected when added to terms of $O(1)$. To prepare for asymptotic analysis, we rescale the model using

$$I = \epsilon Y,$$

thus obtaining the final version of the model as

$$Y' = \epsilon^{-1}(\mathcal{R}_0 S - 1)Y, \tag{5.6}$$
$$P' = a - (v+1)P + ahR - \mathcal{R}_0 PY, \tag{5.7}$$
$$S' = 1 - vP - S + hR - \mathcal{R}_0 SY, \tag{5.8}$$
$$R' = vP - (h+1)R + (1-\epsilon)Y, \tag{5.9}$$
$$1 = S + R + \epsilon Y. \tag{5.10}$$

5.2 The Vaccine-Reduced Basic Reproductive Number

In Section 2, we defined the basic reproductive number as the expected number of secondary infections produced by one infective in a wholly-susceptible population. This was a convenient simplification; the population should be whatever serves as the disease-free equilibrium. For a population with total population size 1 that is fully susceptible, this is the value defined as $\mathcal{R}_0$ in (5.5). A disease-free population in the PUIR model is not fully susceptible because it will have individuals in class R through vaccination. Instead, the disease-free susceptible population fraction will be some value $S_0 < 1$. The basic reproductive number in the presence of vaccination is then

$$\mathcal{R}_v = S_0 \mathcal{R}_0.$$

Solution of the algebraic system $P' = R' = 0$ for the disease-free equilibrium $I = 0$ yields the equilibrium removed fraction R_0. Conservation of population with $I = 0$ then yields $S_0 = 1 - R_0$, which we can rewrite using $a = 1 - r$ as

$$S_0 = \frac{(1+rv)(1+h)}{(1+rv)(1+h) + av}. \tag{5.11}$$

5.3 Effect of Vaccine Refusal on the Vaccine-Reduced Basic Reproductive Number

S_0 represents the infectiousness of the disease with vaccination, relative to the control case, and serves as one measure of the impact of vaccination and vaccine refusal. This value depends on the disease parameter

[22] The state variables are already scaled by the total population.

h, the public health parameter v, and the public refusal parameter r. For further analysis, we assume the public health policy

$$v = 2(1 + h), \tag{5.12}$$

which means that people will be vaccinated twice plus two times extra for each expected loss of immunity. A full accounting for the impact of vaccination would require a more complicated model that separates out young children from the older population; however, the question at hand is on the impact of vaccine refusal, which will depend less on vaccine administration protocols. Thus, it is reasonable to assume a slightly higher overall level of vaccination than would probably be achieved in practice, but administered without preference for small children. The issue of age-dependent vaccine administration should matter less for diseases with short-lived immunity, such as influenza and (as it becomes endemic) COVID-19; with high values of h the proportion of the prevaccinated class that consists of young children will be small. With (5.12), we obtain

$$S_0 = \frac{1 + 2r + 2rh}{3 + 2rh}. \tag{5.13}$$

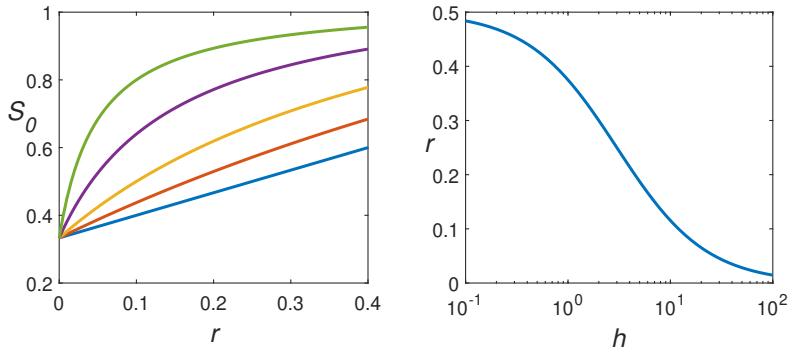

Figure 5.2: (a) The disease-free susceptible population fraction S_0 for the PSIR model, $h = 0, 1, 3, 10, 30$ (bottom to top); (b) The extent of vaccine refusal corresponding to a vaccine effectiveness reduction by 50%.

The results are illustrated in Figure 5.2. The panel on the left shows S_0 as a function of the vaccine refusal fraction for different values of the loss of immunity parameter h. At the prescribed vaccination level, the basic reproductive number is reduced to 1/3 of its base value when there is no vaccine refusal. The factor increases to 0.6 with vaccine refusal of 40% for a disease with no loss of immunity. As the rate of immunity loss increases, so does the impact of vaccine refusal. The right panel illustrates this result in a different way by showing the vaccine refusal fraction corresponding to a 50% decrease in vaccine impact ($S_0 = 2/3$). For diseases with lifetime immunity, 50% vaccine refusal reduces vaccine impact by 50%. The impact of refusal increases as the immune duration decreases, to the point where almost any level of vaccine refusal eliminates the public health vaccine benefit for diseases with very short immune duration.

While this model is not intended for COVID-19, it is reasonable to expect that its main results will hold for a more sophisticated model that does fit COVID-19. In light of ongoing vaccine refusal, we can expect that COVID-19 vaccination will have very little public health benefit aside from the individual benefits to those who remain fully vaccinated by receiving boosters when recommended by public health authorities.

5.4 Stability Analysis of the PUIR Model

As with most simple endemic disease models, the equilibrium and stability properties of the model are simple, albeit difficult to prove. We summarize the result in a theorem and defer the proof to the exercises.

Theorem 5.1 *The PUIR model (5.6–5.10) has the following properties:*

1. *There is a unique disease-free equilibrium, with susceptible population fraction given by (5.11). This equilibrium is asymptotically stable if and only if* $\mathcal{R}_v < 1$.
2. *There exists an endemic disease equilibrium if and only if* $\mathcal{R}_v > 1$. *This equilibrium is unique, with infectious population fraction*

$$Y^* = \frac{(q-v-2)+\sqrt{(q-v)^2+4rvq}}{2\mathcal{R}_0} + O(\epsilon), \quad q = \mathcal{R}_0(1+h) - h. \tag{5.14}$$

3. *The endemic disease equilibrium is asymptotically stable.*

5.5 Effect of Vaccine Refusal on Infection Prevalence

If the basic reproductive number is large enough for the endemic disease equilibrium to be stable, then vaccination and vaccine refusal have an effect on the average infectious population given by $I = \epsilon Y$ with Y from (5.14). To focus on the impact of vaccination, we consider the ratio of the equilibrium value Y^* to

$$y = \mathcal{R}_0^{-1}(q-1),$$

which is the equilibrium value in the absence of vaccination.

The results are shown in Figure 5.3. The bottom curve is the default for the case of no vaccine refusal. The loss of immunity parameter h has no effect because faster loss of immunity is being compensated by more frequent vaccine administration. When $\mathcal{R}_0 \le 3$, the vaccine is sufficient to eliminate the endemic equilibrium, but at very high $\mathcal{R}_0$ it is only sufficient to reduce the equilibrium infectious population by about one quarter. Increasing the vaccine refusal rate markedly decreases the benefit of the vaccine, particularly for diseases with relatively low $\mathcal{R}_0$. With vaccine refusal, vaccination is slightly less effective for diseases with faster waning of immunity, but this effect is still largely compensated for by the higher vaccination rate.

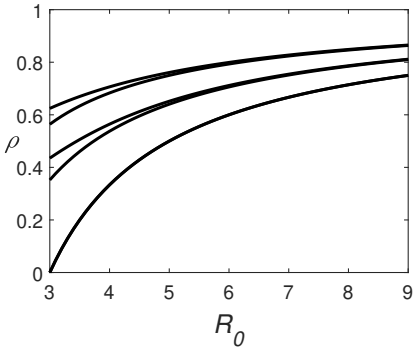

Figure 5.3: The ratio $\rho = Y^*/y$ of the equilibrium infected population relative to that in the absence of vaccination, with $(r,h) = (0,\text{any}), (0.2,0), (0.2,10), (0.4,0), (0.4,10)$ from bottom to top.

Problems

5.1. Derive (5.11).

5.2. Complete the proof of claim 1 of Theorem 5.1.

5.3. Prove claim 3 of Theorem 5.1 by using the Routh-Hurwitz conditions and the requirement $P \le S$ to prove that any endemic disease equilibrium must be asymptotically stable. Do this without attempting to solve the equations that define endemic disease equilibria.

5.4. Prove all of claim 2 of Theorem 5.1 except for the requirement $\mathcal{R}_v > 1$ and the formula for Y. This is done by deriving a single quadratic equation for the equilibrium value of P and proving that it has a unique solution in the acceptable range.

(a) Simplify the right-hand sides of the endemic disease equilibrium equations by using the parameter

$$q = \mathcal{R}_0(1 + h) - h$$

in place of h. (These right-hand side formulas will still include $\mathcal{R}_0^{-1}$ and a.) Then combine the two equations into a single quadratic equation of the form $F(P) = 0$.

(b) Evaluate the function F at 0, $\mathcal{R}_0^{-1}$ and ∞. Use the results to show that there is a unique solution in the interval $0 < P \leq \mathcal{R}_0^{-1}$.

5.5. Prove the remaining statements in claim 2 of Theorem 5.1.

(a) Use the results of Problem 5.4 to obtain the unique solution for the equilibrium value P^* and derive (5.14).

(b) Show that the condition $Y^* \leq 0$ is equivalent to the condition $\mathcal{R}_v \leq 1$. Explain why that proves the existence requirement $\mathcal{R}_v > 1$.

6 An Epidemic Model for a COVID-19 Scenario

As we have noted earlier, models need to be designed with a specific purpose in mind. Thus, there is no one "COVID-19" model.

For obvious reasons, the COVID-19 models that have generated the most interest are those with a goal of forecasting short-term trends such as ICU capacity and deaths. Models intended for this purpose require a large amount of detail, such as the age distribution of the population and the age-dependent frequencies of hospitalization and death. Such models require data not readily available and their complexity puts them outside the scope of this book.

Other models are intended to explore the possible impact of public health policies. These models do not require the same level of detail; indeed, extra detail may make it harder for the model to make meaningful predictions.

Because of their relative simplicity, models intended to predict the effect of policies need substantial revision when major events occur, such as the development of a vaccine or the change from one dominant variant to another. In this section, we consider a model the author created to explore the impact of public health policy at the beginning of testing and social distancing in March 2020.

COVID-19 was first identified in December 2019. By January 2020, the broad community of epidemiologists had identified the new disease as a danger. On March 11, the World Health Organization declared COVID-19 to be a pandemic. Within the next few weeks, most institutions and governments around the world had begun to impose public health policies such as social distancing and masking. The model presented here was originally created in late March 2020, at the beginning of interventions and public discourse about them. We consider a revised version created in January 2021 to better reflect what had been learned about the initial scenario.

6.1 Building the Model

The choice of model components is based on the natural history of the disease and the questions to be addressed. The logical starting point for COVID-19 is the SEIR model of Section 1. Several augmentations are required to model the March 2020 COVID-19 scenario.

1. We need an additional class for asymptomatic infectious patients (A). Compared to symptomatic patients, asymptomatic patients are less infectious, recover more quickly, and are less likely to be tested.

2. To assess the public health impact of the pandemic, we need to track either the number of hospitalized patients or the number of patients in ICUs, or both. For simplicity, our model includes an additional infectious class for hospitalized patients (H) and does not separately track ICU patients. To better account for the progression of the disease, we split the symptomatic infectious class into two groups: those who will not need hospitalization (I_1) and those who will (I_2).[23] We also add a separate class (D) for deceased individuals rather than incorporating these into the R class, as is done in the SEIR model. Thus, the model structure could be described as SEAIHRD.

3. To assess the impact of mitigation strategies on the pandemic, we need to build in testing, isolation of the sick, (possibly) quarantining of those known to be exposed, and the combined effect of social distancing and mask use. These factors do not change the class structure of the model, but they do change the formula for transmission.

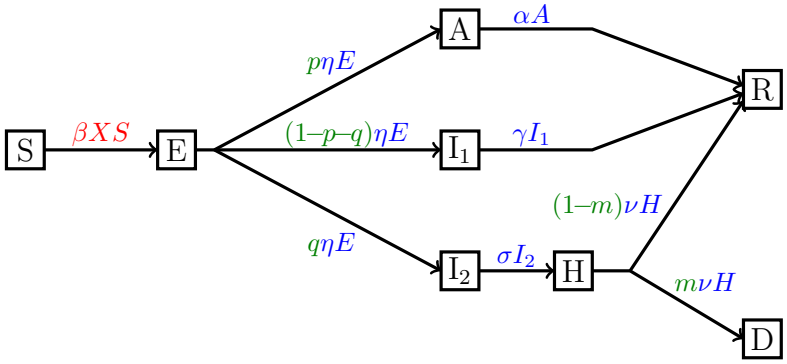

Figure 6.1: Schematic diagram of the SEAIHRD model. Red indicates transmission processes, blue indicates transition processes, and orange is for probabilities.

Figure 6.1 displays the compartment diagram that represents our COVID-19 model. The diagram is based on the following assumptions about the processes that move individuals among the classes.

1. Susceptible individuals become infected at a rate proportional to the susceptible population count S and an 'effective infectivity' count X. While the infectious class I is the only class capable of transmitting the disease in an SEIR model, the COVID-19 model has different categories of infectives with different levels of infectivity, which contribute to the effective infectivity in different ways (see below).

2. Latent individuals (class E) become infectious at rate ηE; a fraction p of these become asymptomatic, a fraction q become prehospitalized symptomatic, and the remainder become standard symptomatic.

3. Nonhospitalized infectives recover at rate γI_1 and asymptomatic individuals recover at rate αA. Prehospitalized infectives become hospitalized at rate σI_2.

4. Hospitalized individuals progress out of the hospital at rate νH; a fraction m of these die, while the rest recover.

5. Recovered individuals are immune for long enough that we can ignore possible loss of immunity.[24]

6. Deaths from unrelated causes and births are sufficiently small over the course of the epidemic that they can be ignored.

[23] Of course no infectious individual knows whether (s)he is going to be hospitalized, but models only track class counts rather than individual results.

[24] Duration of immunity was at least several months for the original strain. Models for the omicron variant cannot neglect loss of immunity.

These assumptions lead to the differential equations

$$\frac{dS}{dt} = -\beta XS\,, \tag{6.1}$$

$$\frac{dE}{dt} = \beta XS - \eta E\,, \tag{6.2}$$

$$\frac{dA}{dt} = p\eta E - \alpha A\,, \tag{6.3}$$

$$\frac{dI_1}{dt} = (1-p-q)\eta E - \gamma I_1\,, \tag{6.4}$$

$$\frac{dI_2}{dt} = q\eta E - \sigma I_2\,, \tag{6.5}$$

$$\frac{dH}{dt} = \sigma I_2 - \nu H\,, \tag{6.6}$$

$$\frac{dR}{dt} = \alpha A + \gamma I_1 + (1-m)\nu H\,, \tag{6.7}$$

$$\frac{dD}{dt} = m\nu H. \tag{6.8}$$

To complete the model, we need to define the quantity X that represents the effective infectious population, which is the number of individuals of class I needed to match the total infectivity of the actual population distribution. It is here that the complexity of COVID-19 dynamics is seen. A significant number of additional assumptions are needed. (We use $I = I_1 + I_2$ for simplicity.)

1. A fraction c of class I are identified by a positive test.[25] These confirmed cases have decreased infectivity because they are put into isolation.[26]

2. The infectivity of each unconfirmed symptomatic infective is 1 (without loss of generality because there is the additional rate constant β in the transmission rate formula).

3. Asymptomatics, confirmed infectives, and hospitalized infectives have infectivities of f_a, f_c, and f_h (all less than 1) relative to that of unconfirmed symptomatic infectives. The overall contribution of hospitalized infectives to the pandemic is small enough that we take $f_h = 0$.[27]

4. There is a 'contact factor' $\delta \leq 1$ that represents the level of risk from the average person's sum total of encounters, relative to normal. This parameter can be used to represent both physical distancing, which decreases the rate of encounters, and wearing of masks, which decreases the risk of each encounter. It is applied to unconfirmed infectives, both symptomatic and asymptomatic, but not to confirmed infectives (who are already in isolation).

With these assumptions, the effective number of infectives is

$$X = f_c cI + \delta[(1-c)I + f_a A]. \tag{6.9}$$

6.2 Parameterizing the Model

The model requires a large number of parameters, which we can classify by type:

1. Transition rate constants: η, α, γ, σ, ν. These are calculated as reciprocals of the mean times for each transition process, for which we had some useful data by late spring 2020.

[25] In spring of 2020, testing in most countries was restricted to people with symptoms. For a scenario in which anyone can choose to be tested, we should use c_i and c_a for the fractions of confirmed cases in I and A, and the formula for X (6.9) would need to be modified.

[26] The word 'quarantine' is incorrect here, as it refers to the isolation of individuals who have *not* tested positive, generally done because of known exposure–these individuals could be in any of classes S, E, A, I, or even R.

[27] Removing this process from earlier versions never made a visible change in any graphs. While it is undoubtedly true that some health care workers caught the disease from patients, a greater number of them probably caught the disease in the hospital cafeteria.

parameter	meaning	value	reference
f_a	relative infectivity of A	0.75	[3]
f_c	relative infectivity with isolation	0.1	estimate
m	deaths per H	0.25	[4]
p	asymptomatic fraction	0.4	[3]
p_c	fraction of confirmed cases	0.09	[3]
q_h	hospitalizations per confirmed case	0.12	[4]
t_2	mean early doubling time of H (days)	3 – 5	[**?**, 14]
$t_a = 1/\alpha$	mean infectious period for A (days)	8	[2]
$t_e = 1/\eta$	mean incubation time (days)	5	[7]
$t_h = 1/\nu$	mean hospitalization duration (days)	8	[6, 11]
$t_{i1} = 1/\gamma$	mean infectious period for I_1 (days)	10	[**?**, 12]
$t_{i2} = 1/\sigma$	mean transition time to H for I_2 (days)	6	[6]
c	confirmed cases per I	0.1 – 0.8	
δ	contact factor	0.1 – 1	

Table 6.1: Primary Parameter Values

2. Probabilities: p, q, m. Of these, p and m can be determined from data. The parameter q is difficult to determine directly from data. Instead, we use estimates for the fraction of confirmed cases (as of spring 2020) p_c and the fraction of confirmed cases that required hospitalization q_h and then calculate q as
$$q = \frac{q_h p_c}{1 - p}. \tag{6.10}$$
The value of p_c is at best a crude estimate, obtained by a study that tested everyone in a target population, including people with no known symptoms. The total number of deaths in a scenario will serve as a rough check on this parameter.

3. Infectivities: f_a, f_c. The first can be determined from data, while we can only estimate the second based on isolation behavior.

4. Transmission rate constant: β. This parameter is not easy to measure. As in Section 2.3, we will use a simplified early-phase model along with data on hospitalization counts (see below).

5. Public health parameters: c, δ: These will be specified as part of the scenario.

6. Initial conditions are the starting values for the state variables, which will be specified as part of the scenario. For some scenarios, the early-phase model and the initial hospitalization count determine the initial conditions for E, A, I_1, and I_2.

Table 6.1 contains the primary parameters taken directly from data or estimated.

6.3 Early-Phase Exponential Growth

For the first few weeks of the pandemic, the susceptible population fraction S can be approximated as a constant $S = 1$. This yields a linear model, which allows us to assume that all the infectious class counts are proportional to that of the latent class, with all growing exponentially with rate λ. The rate constant can be estimated from data using the doubling time equation
$$\lambda t_2 = \ln 2, \tag{6.11}$$
where t_2 is the doubling time of whichever class is easiest to measure. Given the difficulties of diagnosing COVID in the asymptomatic and mildly symptomatic, the best choice is the hospitalized class H. The model resolves into a series of formulas to determine β. We summarize the result here, along with the formula for the basic reproductive number, leaving the derivations as an exercise.

Theorem 6.1 *Suppose the initial population is almost entirely in the susceptible class. Then the SEAIHRD model shows an extended exponential growth phase with*

$$E \propto e^{\lambda t}, \quad A = aE, \quad I_1 = iE, \quad I_2 = jE, \quad H = hE, \quad S \approx 1, \tag{6.12}$$

where λ is determined from the hospitalization doubling time t_2 by

$$\lambda = \frac{\ln 2}{t_2}$$

and

$$a = \frac{p\eta}{\lambda + \alpha}, \quad i = \frac{(1-p-q)\eta}{\lambda + \gamma}, \quad j = \frac{q\eta}{\lambda + \sigma}, \quad h = \frac{\sigma j}{\lambda + \nu}.$$

Furthermore, the parameter β is given by

$$\beta = \frac{\lambda + \eta}{i + j + f_a a}$$

and the basic reproductive number is

$$R_0 = \beta \left[\frac{f_a p}{\eta} + \frac{(1-p-q)}{\gamma} + \frac{q}{\sigma} \right]. \tag{6.13}$$

The values for p_c and q_h yield $q = 0.018$. Doubling times from 3 days to 5 days ultimately yield $\mathcal{R}_0$ values from 6.8 to 3.9. In late March of 2020, the 'accepted' value of $\mathcal{R}_0$ for COVID-19, obtained by statistical analysis of known transmissions, was 2.6. This low value corresponds to a doubling time of 8 days, which is clearly too large in comparison with known data. The problem with the statistical analysis result is that the method misses most asymptomatic cases. The best estimate we have for the original strain, $\mathcal{R}_0 \approx 5.7$, was obtained using a method similar to ours and published in July 2020 [14]. Our method gives this value using a doubling time of 3.5 days, which is consistent with the known data.[28] While we do not have adequate data to reliably use the same method to estimate $\mathcal{R}_0$ for subsequent strains, we do have a rough estimate of a 2-day doubling time for omicron, with which our model (along with updates for the transition rate parameters) suggests a basic reproductive number of about 10.

6.4 Model Investigation

Figure 6.2 shows the default scenario for March 2020: a basic reproductive number of 5.7, no testing, masking, or social distancing, and a starting hospitalization level of one patient per 100K total population. Panel **a** shows the usual plot of susceptible, latent, total infectious, and removed, and we can see that the pattern is similar to an SEIR model with the same basic reproductive number. Note that the pandemic runs its course in about three months, with nearly everyone infected by the two-month mark. Panel **b** shows a peak of nearly 6000 new cases per day per 100K people occurring about one month from the beginning of the scenario, but of course this consists almost entirely of unconfirmed cases. Panels **c** and **d** show the projected death toll for the United States and the projected hospitalization count per 100K. The death toll in this scenario is 1.5 million, which seems consistent with what actually happened, given that there was some mitigation. Similarly, the actual hospitalization count was high enough to exceed the capacity for some regions (the dashed line in the plot is the average capacity in the US), and the dire prediction in the plot seems consistent with what would have happened without any mitigation.

The purpose of the model is to test different mitigation strategies. There are a lot of experiments we could do; here we show just two of them. Figure 6.3 shows the effect of testing by comparing simulation

[28] It is possible to match data reasonably well with an unrealistically low value for $\mathcal{R}_0$, but doing so requires that the value of mitigation strategies such as masking and testing be underestimated. If the goal is to make predictions rather than merely to match data, it is important that each of the parameter values be realistic; therefore, modelers should use $\mathcal{R}_0$ values derived from epidemiological outcomes rather than statistical analysis of case data.

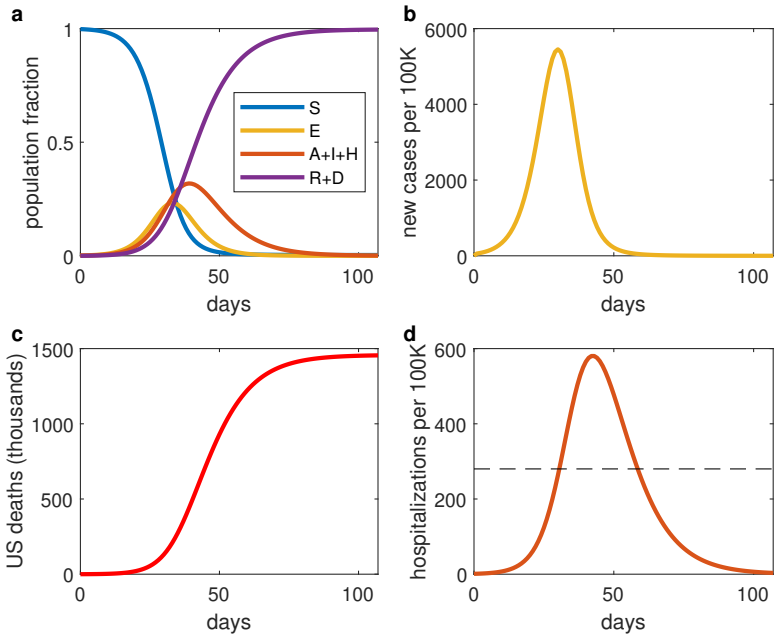

Figure 6.2: The SEAIHRD COVID-19 model default scenario – no testing or distancing.

plots similar to those of Figure 6.2 but with different values of the confirmed case fraction c, keeping all other parameters the same. The blue curves are the default scenario. Modest amounts of testing (50% or less) have only a small impact on the course of the pandemic. With $c = 0.75$, there is a noticeable decrease in the severity—for example, maximum hospitalization is reduced from about 6 per thousand to just under 4 per thousand—but it is still high enough to overwhelm the health care system. The total number of deaths is not reduced much by this change in testing. Only when testing becomes nearly universal do we see a large benefit to the health care system, but the total death count is still above one million for the United States. Of course the exact numbers are affected by our estimates for the parameters. The best conclusion to draw is that universal testing reduces the death toll by about 20%. Testing clearly has some value, but not enough without other mitigation strategies.

Figure 6.4 shows the results of an experiment that tests the importance of the contact factor. The results depend strongly on how δ compares to a critical value of roughly 0.2. This value reduces the effective basic reproductive number to 1. It is slightly more than $1/\mathcal{R}_0$ because lowering contact rates only affects the unisolated. Below the critical value of δ, we have the pandemic under control, but this is only an expedient while waiting for a treatment or a vaccine. The scenario ends quickly, but with nearly the entire population still susceptible. If we were to end our vigilance at that point, the pandemic would come roaring back like a wildfire that is only partially contained when the fire crews go home. As δ decreases from 1, we don't see much improvement in the total death count, but we immediately see improvement in the maximum hospitalization count. With the given parameter estimates, we need δ to be about 0.25 in order to decrease the U.S. death toll to the roughly 800,000 that we had when the vaccine rollout occurred in January 2021. This seems to be a reasonable figure. Many people decreased their transmission risk by much more than a factor of 4, while other people largely ignored public health recommendations.

Problems

6.1. Use *SEAIHRD_onesim.m* to reproduce Figure 6.2.

6.2. Use *SEAIHRD_onesim.m* to redo Figure 6.2, but with a contact factor of 0.5 instead of 1.0. Discuss the significance of a 50% reduction in contact rates.

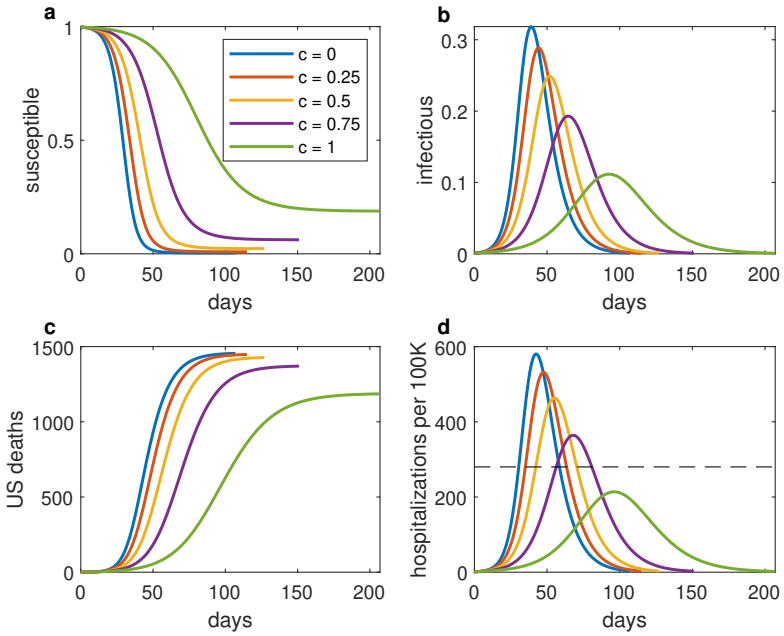

Figure 6.3: The SEAIHRD COVID-19 model with different levels of testing; other parameters are the default values.

6.3. Use *SEAIHRD_comparison.m* to reproduce Figure 6.3.

6.4. Explore the impact of the contact factor by using *SEAIHRD_comparison.m* to prepare a graph similar Figure 6.3, using contact factors of 1, 0.8, 0.6, and 0.4. Assume $c = 0.25$, which is a reasonable guess for the early stages of testing. Discuss the impact of the contact factor in this range.

6.5. Repeat Problem 6.4 using contact factors of 0.24, 0.22, 0.2, 0.18, and 0.16. Discuss the impact of the contact factor in this range.

6.6. Use *SEAIHRD_paramstudy.m* to reproduce Figure 6.4.

6.7. Repeat Problem 6.6, but with testing fractions $c = 0.5$ and $c = 0.75$. Discuss how different levels of testing affect the impact of the contact factor.

6.8. Some U.S. officials claimed in the early stages of the pandemic that the main effect of more testing is a higher reported case count, while public health experts argued that testing was an essential component of public health policy. These contrasting views can be explored by varying the testing rate in a scenario that otherwise matches the early stages of the pandemic. Modify *SEAIHRD_paramstudy.m* to prepare a graph similar to Figure 6.4, but with c as the independent variable. Assume $\delta = 0.3$, which is a reasonable guess for the early stages of public health intervention. Discuss the results.

6.9. On April 1st, 2020, Dr. Anthony Fauci, head of the U.S. National Institute for Allergy and Infectious Diseases, warned that even with aggressive measures, the total number of deaths from COVID-19 in the United States during the time required to control the initial outbreak could be 100,000 to 200,000. (This figure was considered outrageously high by many at the time. In the event, the 200,000 deaths had occurred by late September.) He also suggested that the most aggressive measures could reduce these numbers noticeably. Assess Dr. Fauci's claims, assuming a testing rate of $c = 0.3$.

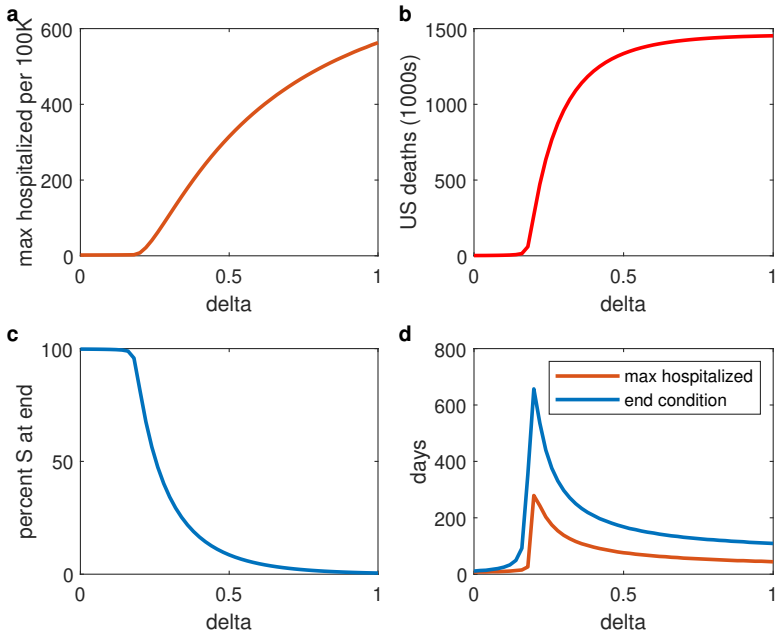

Figure 6.4: The effect of social distancing and masking on the SEAIHRD COVID-19 model outcomes, assuming a low level of testing ($c = 0.1$).

(a) Use trial and error with *SEAIHRD_comparison.m* to identify the range of values for δ that yield deaths between 100K and 200K. A few experiments should suffice to get an estimate to the nearest 0.001 for each level. Remember that you can use multiple values with *SEAIHRD_comparison.m*—but not too many.

(b) For a more in-depth look at the effect of contact factor on death counts, run *SEAIHRD_paramstudy.m* with a range of contact factors and examine the graph of total deaths. Try the ranges $0 \leq \delta \leq 1$ to get a picture for the full range and $0.1 \leq \delta \leq 0.2$ to get a picture that focuses on the critical range.

(c) Given that masks at that time reduced transmission by about a factor of 4 and physical distancing by about a factor of 2, does the model support the claim that more aggressive measures would have significantly impacted the death rate? Also comment on how the epidemic duration and final susceptible percentage changes and what these results mean for public health.

6.10. In an interview on March 18, 2020, a U.S. government official said "If we can get all America to pitch in for the next 15 days, we can flatten the curve." This suggests that a 15-day lockdown would have had a permanent benefit. In mid-April, other government officials recommended 45 days. To address the implied claim of a permanent benefit from a limited lockdown, we need to modify the base function *seaihrd* so that `delta` can change in the middle of a scenario from a low value (the lockdown) to 1 (post-lockdown) and modify *SEAIHRD_comparison.m* to utilize the new function.

The changes to *SEAIHRD_comparison.m* just add a new parameter called `lockdays`:

1. Change line 81 so that `xvals` is assigned to the name `lockdays`.
2. Change the function call in line 83 by making the function name `seaihrd2` and adding an additional item `lockdays` at the beginning of the argument list.

The changes to *seaihrd.m* are more substantial. Start by changing the name of the function to `seaihrd2` and saving the file as *seaihrd.m*. Then add `lockdays` at the beginning of the argument list in the `function` statement. The COMPUTATION section needs some significant changes:

1. Some code needs to be inserted that sets `delta` to 1, either because there is no lockdown or because the lockdown has elapsed. These are two different triggers, so they must be implemented separately.

   (a) At the beginning of the COMPUTATION section, add an `if` block that sets `delta = 1` whenever `lockdays==0`.[29]

   (b) After the statement that saves the results at time t+1, add an `if` block that sets `delta = 1` and `summ = sum(Y(2:6))` whenever `t==lockdays`.

2. Code needs to be inserted so that the check for the end condition (the `if-else` construction at the end of the `for` loop) is only checked after the lockdown. To do this, put the `if-else` block at then end of the loop inside an `if` statement that triggers when `t>lockdays`.

In addition to the program changes, set the default scenario data to $\delta = 0.1$ and $c = 0.1$, along with the usual default values for the other parameters. Note that the post-lockdown value of $\delta = 1$ is built in to the new function *seaihrd2*.

(a) Run the modified program *SEAIHRD2_comparison.m* using lockdown times of 0 days, 15 days, and 45 days as `xvals`.[30] Describe the effect these limited lockdowns have on the course of the epidemic. Also explain why the results come out this way. How accurate were the claims that a lockdown of limited duration would have had a permanent benefit?

(b) We set the total infectivity for unconfirmed symptomatics (δ) to match our assumed infectivity of confirmed symptomatics ($f_c = 0.1$). This is a reasonable guess for how much we can reduce transmission risk with maximum effort. However, we can also test the most optimistic assumption possible by rerunning the experiment with $f_c = 0$ and $\delta = 0.05$, reflecting a 20-fold reduction in transmission risk. This would eliminate all transmission during the limited lockdown. Does this help? Why or why not?

References

[1] F. Brauer, C. Castillo-Chavez, Z. Feng, *Mathematical Models in Epidemiology*. Springer-Verlag, New York, 2019.

[2] Byrne A.W. et al. 2020. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases, *BMJ Open*, **10**, DOI: 10.1136/bmjopen-2020-039856.

[3] The Centers for Disease Control. *COVID-19 Pandemic Planning Scenarios*, Sept 10 2020, Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#five-scenarios, October 2020.

[4] The COVID Tracking Project. *Our Data*. Retrieved from https://covidtracking.com/data, October 2020.

[5] Diamond J. *Guns, Germs, and Steel*. W.W. Norton, New York (1997)

[6] Faes C., Abrams S., Van Beckhoven D., Meyfroidt G., Vlieghe E., Hens N. 2020. Time between symptom onset, hospitalization and recovery or death: statistical analysis of Belgian COVID-19 patients. *Int J Environ Res and Public Health*, doi: 10.3390/ijerph17207560.

[7] Lauer S.A., Grantz K.H., Bi Q., Jones F.K., Zheng Q., Meredith H.R., Azman A.S., Reich N.G., and Lessler J. 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, doi: 10.7326/M20-0504.

[29] Don't forget the semicolon at the end of the assignment statement.

[30] Examine the plots carefully as a way of checking that your program is doing what it is supposed to do. The plots for a lockdown time of 0 should be identical to what you get in *SEAIHRD_onesim* with $\delta = 1$ and the right value of c. The other plots should start out looking like a simulation with a smaller value of δ.

[8] Ledder G. Incorporating vaccination into compartment models for infectious diseases. Submitted to *Math Bios. Eng.* (2022)

[9] Ledder G. *Mathematical Modeling for Epidemiology and Ecology*, Springer, in press (2022)

[10] Ledder G. Scaling for dynamical systems in biology. *Bull Math Bio*, **79**: 2747–2772 (2017)

[11] Lewnard J.A. et al. 2020. Incidence, clinical outcomes, and transmission dynamics of severe coronavirus disease 2019 in California and Washington: prospective cohort study, *BMJ*, **369**, DOI: 10.1136/bmj.m1923.

[12] Liu Y, Yan L-M, Wan L, et al. 2020. Viral dynamics in mild and severe cases of COVID-19. *Lancet Infect Dis* **20**: 656–657. doi: 10.1016/S1473-3099(20)30232-2.

[13] McNeill WH. *Plagues and Peoples*. Anchor Press, Garden City, NY (1976)

[14] Sanche S, YT Lin, C Xu, E Romero-Severson, N Hengartner, and R Ke. High contagiousness and rapid spread of severe respiratory syndrome coronavirus 2. *Emerging Infectious Diseases*, **26**: 1470–1477 (2020) doi: 10.3201/eid2607.200282.

[15] P. van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.*, **180** (2002), 29–48, doi: 10.1016/S0025-5564(02)00108-6.