2 SIR: the workhorse model

The SIR model forms the backbone of a large part of infectious disease modeling. The letters stand for Susceptible, Infected & Infectious, and Recovered. The model describes the number of individuals in each of these three states over time. It is well suited to most directly and some indirectly transmitted infections. We'll reexamine its assumptions later in the notes.

The SIR model is part of a larger class of models called **compartmental models**. A compartmental model divides a population into different "compartments" based on certain characteristics or states. In the SIR model, the population is divided into compartments by disease state: susceptible, infected, and recovered individuals. The key idea behind a compartmental model is to track the flow of individuals between the compartments over time. Critically, when we say that the population is divided between the compartments, this also implies that individuals must be in one and only one compartment at any time. Consequently, if we total up the number of people in the compartments, that must be equal to the population size N.

$$S(t) + I(t) + R(t) = N(t) \; . \tag{1}$$

Compartmental models also treat every person in the same compartment as the same, seeing no difference between them. Although we'll focus on compartmental models in epidemiology, they are also used in other fields such as population dynamics, ecology, and transportation systems. Note that compartmental models are simplifications of real-world dynamics and make assumptions such as homogeneous mixing and linearity, which may not always hold.[1]

2.1 Deriving the SIR equations

We start with considering an individual in S who is, by definition, susceptible to infection. Here are a few simple assumptions about such a person:

- If the susceptible person comes into contact only with people who are susceptible or people who are recovered, then they should remain susceptible. In other words, to become infected in the SIR model a susceptible individual must come into contact with an infected person—a necessary condition.
- Just because someone comes into contact with an infected person does not guarantee they will become infected. In other words, contact with an infected individual is not sufficient to cause the susceptible person to become infected.

These two points tell us that to calculate whether someone becomes infected, we'll need to write down the rate at which they come into contact with infected people, and the probability that each contact results in an infection.

[1] Stay tuned, as we'll dig into some of these important topics later in the semester!

2.1.1 Contact rates

A core assumption of the SIR model is that the population is well mixed, meaning that when two people do come into contact, the probability that each one is in S, I, or R is proportional to their proportions in the population, and the first person's compartment is independent of the other one's compartment. This means that the probability that a randomly chosen contact is susceptible is S/N, is infected is I/N, and is recovered is R/N.

We'll assume that the total rate of contact is c contacts per person per unit time. Let's now imagine a single person in the population, named Attie, and ask at what rate he will come into contact with different types of people, such that his total contact rate adds up to c. Because the probability that each contact is susceptible is S/N, Attie will contact susceptibles at a rate of cS/N susceptibles per unit time. Similarly, he will contact cI/N infecteds and cR/N recovereds per unit time.[2] Our calculations above are for a single person named Attie, but of course there are S total people in the susceptible compartment, each of whom contacts people at these rates. Thus we can reason that the total number of contacts between S and I will be cI/N multiplied by S, i.e., $S(cI/N)$.

2.1.2 Probability of transmission

At what rate will new infections occur? Suppose that if a susceptible person comes into contact with an infected person, let p be the probability that the transmission takes place. We showed above that there are cSI/N total contacts between S and I per unit time. Each of those contacts results in a new infection with probability p, meaning that the rate at which new infections occur is $cpSI/N$ new infections per unit time. The number of individuals who become infected per unit time is called the **incidence**. This means that

$$\begin{aligned} \dot{S}(t) &= -\text{incidence} \,, \\ &= -\left(\frac{cpI}{N}\right) S \,, \end{aligned} \tag{2}$$

where we have used $\dot{S}$ to mean dS/dt, the rate of change of S over time.

You may have noticed that we kept parentheses around $\frac{cpI}{N}$ in the equation above. This is because the quantity $\frac{cpI}{N}$ is the per-capita rate at which susceptibles become infected. This quantity, the per-capita rate at which susceptibles become infected, has a special name: the **force of infection**. Typically, the force of infection, which may vary as our modeled dynamics unfold, is written as $\lambda(t)$, so we can write

$$\dot{S}(t) = -\lambda(t) S \,. \tag{3}$$

Another thing you may have considered is that, while p and c have clearly different meanings, their role in the force of infection is similar in the sense that they are both multiplied constants. Put differently, doubling

[2] Can you show that adding these three contact rates yields a total of c?

the contact rate $c \to 2c$ would have the same effect on the force of infection as doubling the per-contact transmission rate $p \to 2p$. As a result, of this multiplicative redundancy, we often write $\beta = pc$, where β is called the **transmission rate constant**, marrying contact and per-contact transmission into a single number.[3] As a result, you'll often see

$$\dot{S}(t) = -\frac{\beta SI}{N} \, . \tag{4}$$

2.1.3 Infection

The **natural history of infection** refers to the progression of an infectious disease in an individual, from the point of initial infection to the point of resolution or outcome. It includes the different stages and symptoms of the infection, as well as the duration of each stage. We'll come back to this when we discuss other models, including those for measles[4] and HIV.[5]

For the SIR model, we assume that the natural history of infection is that an individual who becomes infected is immediately infectious. Note that because we are using a compartmental model, we treat everyone in I as identical, meaning that no one person is more or less infectious than any other. People in the I compartment recover at a per-capita rate γ, meaning that the total rate of recovery of infected people is γI people per unit time. This allows us to write an equation for $\dot{I}$,

$$\begin{aligned} \dot{I}(t) &= \text{incidence} - \text{recovery} \\ &= \frac{\beta SI}{N} - \gamma I \, . \end{aligned} \tag{5}$$

Note that because of our compartmental assumption—that everyone must be in one and only one compartment—any loss from I due to recovery will be a gain in R, and therefore

$$\dot{R}(t) = \gamma I \, . \tag{6}$$

Writing down all three equations at once, we get the SIR model, which is a system of ordinary differential

[3] You will find that these three variable letters (S, I, R) and two parameter letters (β, γ), as well as the force of infection letter (λ), are almost universally used in the infectious disease modeling literature, so we use those letters here.

[4] The natural history of an infection with the measles virus begins with an incubation period of 7-14 days during which the infected person is not yet contagious but may already have the virus replicating in their body. After that, a prodrome stage of fever, cough, runny nose, and red eyes occur and then the typical rash appears. After a week or two, the rash disappears and the person is no longer contagious.

[5] In some cases, the natural history of infection can be divided into several distinct stages, such as an acute phase and a chronic phase. For example, HIV infection has a primary acute phase, followed by a chronic asymptomatic phase, and then a symptomatic phase.

equations (ODEs),

$$\begin{aligned}\dot{S} &= -\frac{\beta SI}{N} \\ \dot{I} &= \frac{\beta SI}{N} - \gamma I \\ \dot{R} &= \gamma I\end{aligned} \tag{7}$$

And that's it. That's the basic SIR model on which so many other models are built. It assumes that the population undergoes homogeneous mixing (is well mixed), and that there is no birth or death.[6]

2.2 Individuals or population proportions? A common normalization

It turns out to often be convenient to model an epidemic in terms of the *proportions* of individuals in the S, I, and R compartments. Let $s = \frac{S}{N}$ be the proportion of susceptibles, $i = \frac{I}{N}$ be the proportion of infecteds, and $r = \frac{R}{N}$ be the proportion of recovereds. Note that we can take the derivative with respect to time of both sides of these equations to get $\dot{s} = \frac{1}{N}\dot{S}$, $\dot{i} = \frac{1}{N}\dot{I}$, $\dot{r} = \frac{1}{N}\dot{R}$. This means

$$\begin{aligned}\dot{s} &= \frac{1}{N}\dot{S} & \dot{i} &= \frac{1}{N}\dot{I} & \dot{r} &= \frac{1}{N}\dot{R} \\ &= \frac{1}{N}\frac{-\beta SI}{N} & &= \frac{1}{N}\left(\frac{\beta SI}{N} - \gamma I\right) & &= \frac{1}{N}\gamma I \\ &= -\beta si & &= \beta si - \gamma i & &= \gamma i\end{aligned} \tag{8}$$

These equations represent the proportion of the total population in each of the compartments. Notice that these equations look similar to Eqs. (7) in their structure except that they simply lack a division by N in two of the equations.

WARNING: From time to time, you may see the SIR equations in the wild, and they may look like Eq. (8) (without the N) but may state that they model counts of susceptible, infected, and recovered individuals. How can this be so? Typically, this is because the authors have absorbed the factor $1/N$ into the definition of the contact rate c and thus into β. In general, we should just emphasizes that it is critical to be careful of the units being used for analysis. For this class, we will attempt to be consistent: variables S, I, and R measure people, while variables s, i, and r measure proportions.

2.3 The SIR flow diagram

It turns out that a convenient way to express compartmental models is using a **flow diagram**. A flow diagram uses a circle or square to represent each compartment, and arrows to show the possible flows

[6] If there's no birth or death, the population should be constant. Can you use Eq. (7) along with Eq. (1) to show that the population doesn't grow?

between compartments. Flows are typically labeled with the rates of flow between compartments.[7] A flow diagram for the SIR model is shown in Fig. 1.

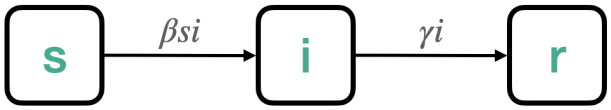

Figure 1: A flow diagram for the SIR model.

Flow diagrams are nice because they show us the structure of a model from a high level view. For example, when we look at Fig. 1, one clearly visible thing we can note is that there is no way out of the R compartment. Another thing we can note is that there is no way into the S compartment. These observations reveal to us that an assumption of our model is that after recovery, immunity is perfect—one cannot be infected again. For this reason, you may see some textbooks or models refer to R as "recovered/removed" because folks in R simply don't contribute to the dynamics after they arrive there.

2.4 SIR Model Ethics

Infectious disease epidemiology is about modeling infections among people, which means that we simply cannot pretend that our work in this course will be separated from ethical considerations if it is to ever be used in practice. While there isn't *that* much to talk about with the current model, it behooves us to get in the habit of thinking about ethics as we go, and that's why this section is here, even now.

One ethical consideration of the SIR model is that it treats people in each compartment as identical, and doesn't consider any differences between them. In reality, of course, there are differences between people. For example, contact rates vary by occupation, which means that exposure to certain infectious diseases may be higher for some than others. Consider the category of "essential workers" who staffed our clinics, grocery stores, fire stations, and power plants even as many others were staying safe at home during March, April, and May of 2020 during the COVID-19 pandemic. When we model a real population using the SIR model, we wash out those differences and treat everyone the same—a process that can blind us to differential impacts, and thus disparities, experienced by some groups vs others.

[7] These are pretty easy to make using Keynote or Powerpoint, by the way, especially if you use the "magnetic" arrows that stick to objects.