Chapter 4

Mathematical Epidemiology

4.1 The Biology of Infectious Disease

A major question in public health is whether we can understand the details of how an infectious disease propagates through a population. Accurate models of infectious disease dynamics may allow for consideration of many important problems:

- How effective does a vaccine need to be in order to prevent the occurrence of **epidemics**, which are abnormally large numbers of infections occurring over a relatively short period of time? Even without eradicating the disease completely, it may be possible to keep the incidence to a low, relatively constant, rate.

- How effective does a vaccine need to be — and how complete must the vaccination program be — to successfully eradicate a disease completely? The only successful example worldwide is small pox; there has been a strong effort over the past decade to finally eradicate polio, but pockets of incidence in some areas of Africa and the Indian sub-continent have proven difficult.

- How effective are quarantine procedures at reducing the severity of an epidemic?

- What responses to an outbreak are reasonable and effective, and what are ineffective and promote unnecessary fear and doubt?

We will be concerned with populations of individuals, some of whom are susceptible to the disease, some of whom are infected, and (perhaps) some who are immune. Again, we will be concerned with the variation in these population numbers over time, and thus there are many similarities to the population dynamics models we have been discussing so far. However there are also important differences to be aware of.

4.1.1 Causes of infectious disease

Before we begin our discussion of models, it is worthwhile to outline some of the key aspects of infectious diseases and their transmission. The key factor that distinguishes an infectious disease is the existence of some infectious agent that is the underlying cause of the disease;

transmission of this agent can lead to new infections. Conversely, non-infectious diseases (heart disease, diabetes, most cancers, etc) are typically a result of pathologies that are restricted to an individual; the disease may have a causative agent (for example a potent carcinogen leading to cancer, or asbestos leading to asbestosis), but this agent can not be passed from an infected individual.

The infectious agent may be one of several different classes of biological agent:

Virus. This class includes many of the most common infectious diseases: influenza, measles, small pox, polio, HIV, herpes, chicken pox, the common cold, rabies, and many many more.

Microparasite. These are single-cell organisms (microbes) that live (at least for a portion of their life) within a larger organism; the parasite may live within the host's cells or simply within the non-cellular matrix of a host's tissue. Diseases of this class include: malaria, tuberculosis, cholera, gonorrhea, diptheria, salmonella, botulism and anthrax. In some of these (cholera, diptheria, botulism, and salmonella, for example), disease symptoms are caused by a toxin released by the bacteria. Exposure to the toxin (but not to the parasite) may thus cause symptoms, but not an infection. This is often the case in food poisoning by botulism and salmonella.

Macroparasites. Again, these are organisms that live a portion of their life in the body of a host, but these are macroscopic, multi-cellular organisms. These include internal parasites (flat worms, round worms, and some larval insects) as well as external parasites (lice, fleas, ticks, etc.) Often macroscopic parasitic infections can vary a great deal in degree of infection (that is, in the number of parasites present), which may be important to consider.

Prion. Quite recently, a novel infectious agent has been described, underlying a number of related neurological diseases: Creutzfeld-Jacob Disease, bovine spongiform encephalopathy (Mad Cow Disease), scrapie, and kuru, for example. Unlike all the above, which all include genetic material (DNA), prions consist of a single protein constituent. Typically, exposure to a toxic protein would not result in a transmissible disease. However, prion proteins are able to convert naturally occurring proteins into the prion form; this provides a mechanism for transmission.

The key characteristic of an infectious disease is that there is a transmissible, causative agent which can replicate within the host. Thus, the total amount of the infectious agent can grow over time. This is in contrast to a toxin, which can not increase in quantity over time without an external source.

4.1.2 Transmission of infectious disease

In addition to variations in the infectious agent, there are many different mechanisms by which infectious diseases may be transmitted. In some cases, transmission is from one infected individual to another:

- through direct (close-range) contact, as is the case for both viral chicken pox and most external macroparasites (such as lice).

- through indirect contact, such as through virus-containing water droplets expelled by sneezing or coughing. This is the case for many viral diseases (colds, the flu, small pox, etc), and some microparasitic diseases (such as tuberculosis).
- through sexual contact, which includes viral (HIV, hepatitis C), microparasitic (gonorrhea) and macroparasitic (pubic lice).
- through direct blood–blood contact, from open sores (often coupled with sexual activity), shared hypodermic needles, or blood transfusions. Many sexually transmitted diseases can also be transmitted through this mechanism.

In other cases, transmission is from direct contact with the infectious agent present in the environment, again in different forms:

- Contaminated water supplies are a major source of transmission of viral (e.g. hepatitis A) and microparasitic (e.g. cholera) disease in the developing world. A lack of adequate waste treatment systems allows fecal matter from infected individuals to enter the water supply, where the infectious agent can remain for an extended period of time.
- Food sources may be contaminated through similar mechanisms, but also through other means. For example, salmonella can be present in chicken and eggs from an inherited infection from the mother. Prion diseases are largely passed through food, in particular by the eating of meat from an infected individual.
- Some disease agents can be transmitted by skin contact, without ingestion. Hookworm, for example, is most often passed by walking barefoot on regions containing infected fecal matter, which can be very prevalent in rural areas, since the parasite infects most mammals, including livestock.

The last point brings up an important third mode of transmission — direct or indirect contact with a **vector**, which are other organisms that host the infectious agent for some period. Thus, livestock can serve as a vector for the transmission of hookworm, and most mammals can serve as vectors for rabies (through bites). These are examples of diseases where the vectors are closely related species to humans, and the disease can pass freely between species. In other cases, the infectious agent spends distinct phases of its life in different species. Malaria conforms to this model, being caused by an infectious microbe that spends part of its life cycle inside Anopheles mosquitoes. Here transmission contact is direct, through bites. Other vectors pass the disease through indirect contact, as is the case of bilharzia. This disease is caused by a parasitic worm that spends part of its life in aquatic snails, and it passed to humans by skin contact with water containing infected snails (and thus also free-swimming worms). Transmission back to worms is through eggs passed into the water from feces and urine.

4.1.3 Terminology of infectious disease

In discussing (and modeling) infectious diseases, it is important to understand the key terms used to discuss various characteristics of a disease. First, the length of infection can vary widely:

Acute infections describe a (relatively) short term illness, in which the patient recovers — and no longer carries the infectious agent — after the course of infection. This is the case for tuberculosis, malaria, influenza, and many others

Chronic infections , on the other hand, describe persistent infections over extended periods of time; the infectious agent is continuously carried by the infected individual indefinitely. Examples of chronic infections are the human immunodeficiency virus (HIV) and the herpes virus.

It is important to note that chronic infection does not necessarily imply chronic display of symptoms nor continuous infectivity. For example, roughly 68% of the American population carries the herpes simplex virus (HSV-1) which causes oral cold sores. Some infected individuals never display symptoms, and those that do only do so sporadically; individuals are infectious only for a short period of time before and during an outbreak of symptoms.

We also have specific terms to discuss the display of symptoms from infection.

Latency describes a period of time during which an individual is infected but does not display symptoms. Essentially all diseases have some period of latency following initial infection, but this length of time can vary from days to years. In chronic infections, periods of latency can also occur between recurrences of symptoms.

Severity is used to discuss whether the symptoms are relatively minor, an inconvenience discomfort, or debilitating and/or life-threatening.

Mortality has a very specific meaning of the fraction of infected individual who die of the disease, rather than recovering.

Among directly transmissible diseases, the **infectivity** of an individual describes how likely transmission to another is; depending on the disease, an individual may or may not be *contagious* through out the period of infection. It is important to not that the level of contagiousness does not necessarily match symptomatic severity. For example, individuals infected with the chicken pox virus are most contagious during early stages of infection, at which time the symptoms are indistinguishable from the flu or the common cold.

Additionally, we must consider issues of **recovery**, by which we typically mean elimination of the infectious agent from the body. However, this may or may not correspond to a complete physical recovery of the infected individual. For example, polio often results in long term (permanent) physical disability for infected individuals. Also, there is a very important consideration of **acquired immunity** — do recovered individuals become resistant to future infection, and is this immunity complete or only partial?

The above terminology has focused on fundamental properties; we also sometimes wish to discuss the behavior of a (possibly) infected population.

An epidemic is a period of time with an incidence of occurrence that is significantly above the "expected" or "average".

Endemic is the term used to describe a disease that has a persistent level of infection in a population over an extended period.

Pandemic is used to describe an epidemic whose scope goes beyond an isolated population (eg a control, state, etc), and rather is present in multiple countries and/or regions.

Note that epidemic can arise in an endemic population, by some (perhaps unknown) factor causing increased infectivity, or in a population without persistent infection by the introduction of one or a few new cases.

4.2 Developing Models of Infectious Disease

Among the simplest diseases to build models of are microparasitic infections passed by contact (direct or indirect) between individuals. In this case, a reasonable model can be made without consideration of "extra" factors, such as the water supply, animal vectors, and so on. We begin by defining classes of individuals, according to their disease status:

1. **Susceptible** individuals are those who could become infected; the number of susceptible individuals will be denoted S.

2. **Latent** or **exposed** individuals are those who harbor the infectious agent, but are not themselves contagious; their numbers will be denoted E.

3. **Infective** or **contagious** individuals are infected and can pass the disease to others; their numbers will be denoted I.

4. **Removed** individuals are not susceptible, latent nor infective — and thus they can not play a direct role in disease progression; their numbers will be denoted R. This class can have multiple biological interpretations, including: individuals with immunity, either due to vaccination or due to a prior infection from which they have recovered; isolated or quarantined individuals, who are infected, but can not pass on the disease; individuals who have died from the disease. In some cases, these classes may be directed models as individual groups.

First we will consider the case of disease progression over the short term — that is, where the time scale of the disease dynamics is short with respect to the host lifespan. In this case, it is reasonable to ignore natural population growth/decline. We generally refer to models based on the types of individuals included in the model, and how these individuals pass between groups.

4.2.1 The simple epidemic — An SI model

The simplest case we can consider is a disease in which individuals may become infected, but never recover, and always remain contagious; this is known as a simple epidemic. The only populations involved are susceptibles (S) and infectives (I), and we can schematically represent the disease progression by:

$$S \longrightarrow I \tag{4.1}$$

Because of this, we refer to this as an "SI" model. If $f(S, I)$ is a function of S and I representing the rate of infection or *incidence* of disease, then:

$$\frac{dS}{dt} = -f(S, I) \qquad \frac{dI}{dt} = f(S, I) \tag{4.2}$$

One of the simplest forms of $f(S, I)$ involves a linear response to both S and I:

$$f(S, I) = \beta SI \tag{4.3}$$

where β is an infectivity constant, describing the likelihood that a single infective individual would pass the disease to an arbitrary susceptible. For simplicity, we will ofter denote this as:

$$S \xrightarrow{\beta SI} I \tag{4.4}$$

Now, in a closed population (*i.e.* without birth and death, and without immigration and emigration), the total populations is a constant, N, giving:

$$S + I = N \qquad \rightarrow \qquad S = N - I \tag{4.5}$$

Thus, we have:

$$\frac{-dS}{dt} = \frac{dI}{dt} = \beta I(N - I) = \beta NI\left(1 - \frac{I}{N}\right) \tag{4.6}$$

This is simply the form of the logistic equation, $\frac{dN}{dt} = R_o N\left(1 - \frac{N}{K}\right)$, which was discussed as a model for single species population growth with a limited environmental carrying capacity. In this case, the key variable is the infected population size, the intrinsic growth rate is βN, and the environmental carrying capacity is N. Thus, a simple epidemic that occurs faster than the birth and death rate will result in universal infection.

Do any diseases fit this profile? Consider the Herpes simplex virus, Type-I (HSV-1), which results in oral cold sores. While individuals with HSV-1 are not continuously infective, the probability that an arbitrary infected individual is contagious at any given time may be taken as a constant, and thus the model is reasonable. The prevalence of HSV-1 is 68% in the US population, 80% in Australia, and 99% in Morocco, with essentially all countries falling in this range. Thus, we do, in fact find very high prevalence. Deviations may be partially due to a lack of consideration for births, which introduce new susceptible individuals into the population. In support of this, in Scandanavia, the overall prevalence is between 75 and 80%, while the prevalence in children is only about 40%.

Note that the intrinsic growth rate depends on N, the total population size. Since we are generally dealing with the population in a given region, it is more appropriate to consider this to be a population density — the number of people in a given area. The bilinear growth model is based on this assumption, describing the observation that the frequency of contact between individuals will increase with density. This dependence on total population density will arise frequently in our disease models.

4.2.2 Disease with recovery — An SIS model

When individuals recover from an infection, they may or may not acquire immunity to future infections. In general, viral infections often lead to immunity, while bacterial infections are

less likely to do so, but this should not be taken as any sort of rule. Additionally, acquired immunity may only be partial. First, consider the case of no immunity; after recovery, infected individuals return to the susceptible class. If again we presume that all infected individuals are infective, we obtain an SIS model that we represent as:

$$S \longrightarrow I \longrightarrow S \tag{4.7}$$

Again, we will have an incidence function, $f(S, I)$, but also a recovery function, $g(I)$, which should only depend on the infected population. If we assume a constant *per capita* recovery rate, we obtain a linear recovery model: $g(I) = \gamma I$ Combining this with a bilinear infectivity model gives:

$$S \xrightarrow{\beta SI} I \xrightarrow{\gamma I} S \tag{4.8}$$

The differential equations describing each population are thus:

$$\frac{dS}{dt} = -\beta IS + \gamma I = -(\beta S - \gamma)I \qquad \frac{dI}{dt} = -\frac{dS}{dt} = (\beta S - \gamma)I \tag{4.9}$$

It often makes more sense to discuss the fraction of the population that is infected or susceptible, rather than dealing with total numbers, and thus we define:

$$\hat{S} = \frac{S}{N} \qquad \hat{I} = \frac{I}{N} \tag{4.10}$$

Reconsidering the differential equations:

$$\frac{-d\hat{S}}{dt} = \frac{d\hat{I}}{dt} = \frac{1}{N}\frac{dI}{dt} = \frac{1}{N}(\beta S - \gamma)I = (\beta S - \gamma)\hat{I} = (\beta N\hat{S} - \gamma)\hat{I} \tag{4.11}$$

Again, in a closed population, $S + I = N$, so $\hat{S} + \hat{I} = 1$.

Now, what can we learn from this model? First, we note that if $\beta N\hat{S} - \gamma < 0$ (and thus $\beta N\hat{S} < \gamma$) the number of infectives will decrease over time. Also, we know that $\hat{S} \leq 1$, and thus $\beta N\hat{S} < \beta N$. Now, if $\beta N < \gamma$, then $\beta N\hat{S} - \gamma < 0$ for any choice of $\hat{S}$, and thus the infection will *always* die out. This makes logical sense, since we can think of βN as the rate at which one infected individual infects others, and γ is the rate of recovery — if, on average, individuals recover faster than than infect others, the infection can not be sustained. Also, note that this criterion depends on the total population size (density), N.

Now, if $\beta N\hat{S} - \gamma > 0$ ($\hat{S} > \frac{\gamma}{\beta N}$), then the infective population will grow, but $\hat{S}$ will fall. Thus, at some point, we may expect that we will reach $\hat{S} = \frac{\gamma}{\beta N}$, and there will be no further growth. Formally,

$$\frac{d\hat{I}}{dt} = (\beta N\hat{S} - \gamma)\hat{I} = \left(\beta N(1 - \hat{I}) - \gamma\right)\hat{I} = \left((\beta N - \gamma) - \beta N\hat{I}\right)\hat{I} \tag{4.12}$$

where we have used $\hat{S} = 1 - \hat{I}$. Note that because $-\frac{d\hat{S}}{dt} = \frac{d\hat{I}}{dt}$, the $\hat{S}$ and $\hat{I}$ null clines are identical, and thus the stationary points are simply defined as solutions to $\frac{d\hat{I}}{dt} = 0$, of which there are two. The first:

$$\hat{I} = 0 \qquad \hat{S} = 1 \tag{4.13}$$

is the trivial solution — if no individuals are infected, then the entire population remains in the susceptible state. For a small perturbation away from this state (*i.e.* $\hat{I}$ is very close to, but not equal, zero) , $\hat{S} = 1 - \hat{I} \approx 1$. Thus, we find that $\frac{d\hat{I}}{dt}$ is negative so long as $\beta N < \gamma$, and is positive otherwise:

$$\frac{d\hat{I}}{dt} < 0 \quad \rightarrow \quad \left(\beta N\hat{S} - \gamma\right)\Delta\hat{I} < 0 \quad \rightarrow \quad \beta N\hat{S} - \gamma < 0 \quad \rightarrow \quad \beta N - \gamma < 0 \tag{4.14}$$

As a negative $\frac{d\hat{I}}{dt}$ is a necessary and sufficient condition for stability, we find that this criterion determines stability of the *non-endemic* stationary point. The second solution is given by:

$$\beta N(1 - \hat{I}) - \gamma = 0 \rightarrow 1 - \hat{I} = \frac{\gamma}{\beta N} \rightarrow \hat{I} = 1 - \frac{\gamma}{\beta N} \tag{4.15}$$

$$\hat{S} = 1 - \hat{I} = \frac{\gamma}{\beta N} \tag{4.16}$$

Thus, we see that an *endemic* infection can arise, with a prevalence of $1 - \frac{\gamma}{\beta N}$. This stationary point will be biologically relevant only when $\frac{\gamma}{\beta N} < 1$ ($\beta N > \gamma$); it should also be clear that the stationary point will be stable under this condition. Thus we see that this ratio is the key determinant of whether the disease dies out or becomes endemic. It is important to note that while β and γ are characteristics of the disease itself (infectivity and recovery, respectively), the population size (or density, as discussed above) also plays a role. In high population densities, the likelihood of a given infection dying out is very small, and the prevalence of an endemic disease will be larger. This result has important ramifications for understanding the differences in disease propagation in urban environments (high population density) and rural environments (low density).

Why does population density matter? As the population density grows, the overall *transmission rate* will increase with square of the density, as expected for random interactions between N individuals. On the other hand, the overall *recovery rate* increases linearly, because recovery is a process that depends only on an individual.

4.2.3 Disease with acquired immunity — An SIR model

As discussed earlier, many diseases (as especially viral infections) result in immunity after an initial infection. This is the case for chicken pox, measles, and many more, and forms the basis of effectiveness of vaccination programs. In this case, we must explicitly consider a distinct population of individuals who have had the disease but recovered, as they will be immune. For simplicity, we will presume that all individuals become perfectly immune. Taking the same expressions for transmission and recovery as from the SIS model, we obtain the schematic representation:

$$S \xrightarrow{\beta SI} I \xrightarrow{\gamma I} R \tag{4.17}$$

Now in this case we have three variables to consider, and the differential equations governing the dynamics are:

$$\frac{dS}{dt} = -\beta IS \qquad \frac{dI}{dt} = \beta SI - \gamma I \qquad \frac{dR}{dt} = \gamma I \tag{4.18}$$

and with a constant total population size:

$$S + I + R = N \tag{4.19}$$

Again shifting to fractional populations, $\hat{S} = \frac{S}{N}$, $\hat{I} = \frac{I}{N}$, and $\hat{R} = \frac{R}{N}$, we obtain:

$$\frac{d\hat{S}}{dt} = \frac{1}{N}\frac{dS}{dt} = -\frac{1}{N}\beta IS = -\beta N\hat{I}\hat{S} \tag{4.20}$$

$$\frac{d\hat{I}}{dt} = \frac{1}{N}\frac{dI}{dt} = \frac{1}{N}(\beta SI - \gamma I) = \beta N\hat{I}\hat{S} - \gamma\hat{I} = (\beta N\hat{S} - \gamma)\hat{I} \tag{4.21}$$

$$\frac{d\hat{R}}{dt} = \frac{1}{N}\frac{dR}{dt} = \frac{1}{N}\gamma I = \gamma\hat{I} \tag{4.22}$$

with $\hat{S}+\hat{I}+\hat{R} = 1$. Note that the expression for the infected population is identical to the SIS model, but that the rate of change for the susceptible population is strictly negative. Also, note that the recovered population does not directly contribute to any of these expressions, but plays an indirect role in maintaining a constant total population size.

Now, consider the stationary points of the system. Taking $\frac{d\hat{R}}{dt} = 0$:

$$\frac{d\hat{R}}{dt} = \gamma\hat{I} = 0 \rightarrow \hat{I} = 0 \tag{4.23}$$

If $\hat{I} = 0$, then:

$$\frac{d\hat{S}}{dt} = -\beta N\hat{I}\hat{S} = 0 \quad \text{and} \quad \frac{d\hat{I}}{dt} = (\beta N\hat{S} - \gamma)\hat{I} = 0 \tag{4.24}$$

Thus, *any* state with no infections is a stationary point, and there are no others. This makes perfect sense, as with infections presents, the susceptible population can only decrease, and must continue to do so until there are no more infectives; of course when no infectives are present, the disease can not spread.

What happens if some (small) number of infectives are introduced into a population? Since the expression for the rate of change of $\hat{I}$ is the same as for the SIS model, much of the logic carries over. That is, if $\beta N\hat{S} < \gamma$, then $\frac{d\hat{I}}{dt} < 0$, and the infected population will steadily decrease, and if $\beta N < \gamma$, this will be true for any value of $\hat{S}$. On the other hand, if $\beta N\hat{S} > \gamma$, $\frac{d\hat{I}}{dt} > 0$ and the infected population will grow; at the same time, the susceptible population will decrease, and eventually the state will reach a point where $\beta N\hat{S} = \gamma$. Now, $\beta N\hat{S} - \gamma = 0$ is a *null cline* for $\hat{I}$, and thus at this point the infected population will stop growing. However, although $\frac{d\hat{I}}{dt} = 0$ at this point, if $\hat{S} \neq 0$ and $\hat{I} \neq 0$, then $\frac{d\hat{S}}{dt} \neq 0$, and thus the susceptible population will continue to decrease. This will result in crossing the the $\hat{I}$ null cline into the regime where $\beta N\hat{S} < \gamma$, and the infected population will begin to decline. These dynamics, an increase followed by a decrease, are precisely what we mean by an epidemic.

These dynamics are explained well by a plot of $\hat{I}$ versus $\hat{S}$. Given an initial recovered population, $\hat{R}_o$, we now that at any time, $\hat{R} \geq \hat{R}_o$, since the recovered population can only grow. Thus:

$$\hat{I} + \hat{S} + \hat{R} = 1 \rightarrow \hat{I} = 1 - \hat{R} - \hat{S} \rightarrow \hat{I} \leq 1 - \hat{R}_o - \hat{S} \tag{4.25}$$

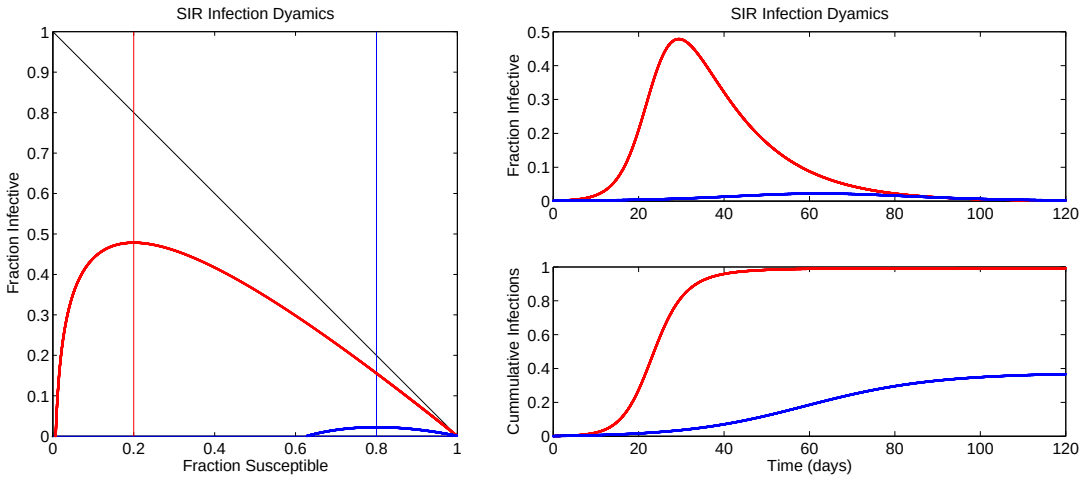

Figure 4.1: Dynamics of a simple SIR model, given an initial 0.1% infective population in an otherwise naïve (non-resistant) population. The blue curve represents a model with a recovery rate four-fold faster than the red.

In other words, the dynamics are restricted to a region where $\hat{I}$ is below a diagonal line of slope -1 and intercept $1-\hat{R}_o$. In a wholly susceptible (naïve) population, the intercept is 1. Of course, we are also restricted to the first quadrant, $\hat{S} \geq 0$ and $\hat{I} \geq 0$. Now, the null clines of $\hat{I}$ are given by, $\hat{I} = 0$, which is simple the x-axis, and by $\beta N \hat{S} - \gamma = 0$. This second cline is a vertical line of $\hat{S} = \frac{\gamma}{\beta N}$. To the left of this line, all motion will be towards reduced $\hat{I}$ and $\hat{S}$ (down and to the left), while to the right of the line, motion will be towards increased $\hat{I}$ and decreased $\hat{S}$ (up and to the left). When the x-axis is reached ($\hat{I} = 0$), motion will stop.

This explains several of our observations, including the rise and then fall of the infected population under certain conditions, and the strict decay of the infected population under others. Any time the initial susceptible population is below $\frac{\gamma}{\beta N}$, there will be monotonic decay in the infected population, and anytime the initial susceptible population is to the right of the line, the infected population will first grow, then decline. One key result of this is that regardless of the initial populations, the maximum possible susceptible population after an initial infection is $\frac{\gamma}{\beta N}$.

How would vaccination affect these results? If individuals have been vaccinated, we can think of the initial population of recovered/immune individuals as being non-zero. Thus, the bounding line of allowed values for $\hat{I}$ and $\hat{S}$ will be have a lower intercept, and the space of possible motion will be pushed down and to the left. For a given number of initially infected individuals, we may think of our starting position being shifted to the left. This will both increase the likelihood of being in the decay regime initially, and decrease the time taking to reach that point otherwise. Thus, as expected, vaccination should reduce to total number of infected individuals.

Total infections in an SIR model

We have seen above that the maximal number of susceptibles after an epidemic is $\frac{\gamma}{\beta N}$, and thus the minimal number of infected individuals is $1 - \frac{\gamma}{\beta N}$ (if there are no individuals initially

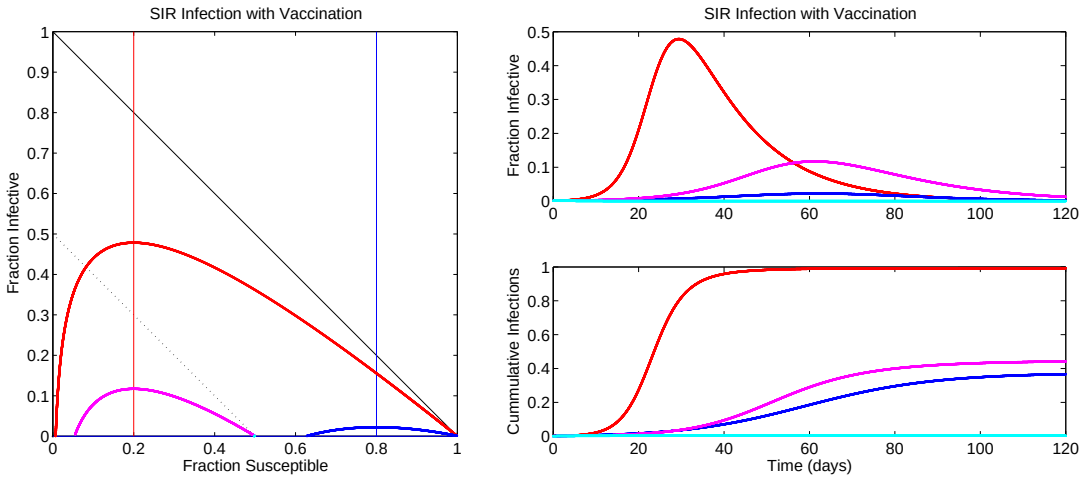

Figure 4.2: Effects of vaccination on the dynamics of a simple SIR model. Red and blue curves correspond to an initial 0.1% infective population in an otherwise naïve (non-resistant) population, while magenta and cyan curves begin with a 50% resistant population. Note that the infective population never grows in the cyan case.

immune). But what will the actual number of individuals who get the disease be?

One approach to this would be to integrate the expressions with respect to time, and take the limit of $t \to \infty$, but this typically requires a numerical solution implemented computationally. An alternative approach is to note that $\hat{S}$ decreases monotonically over time, and thus we may think of $\hat{S}$ as a replacement variable for time. What we then need are the expressions $\frac{d\hat{I}}{d\hat{S}}$ and $\frac{d\hat{R}}{d\hat{S}}$, which are given quite simply by:

$$\frac{d\hat{I}}{d\hat{S}} = \frac{d\hat{I}/dt}{d\hat{S}/dt} = \frac{\beta N \hat{I}\hat{S} - \gamma \hat{I}}{-\beta N \hat{I}\hat{S}} = -1 + \frac{\gamma}{\beta N}\frac{1}{\hat{S}} \tag{4.26}$$

$$\frac{d\hat{R}}{d\hat{S}} = \frac{d\hat{R}/dt}{d\hat{S}/dt} = \frac{\gamma \hat{I}}{-\beta N \hat{I}\hat{S}} = -\frac{\gamma}{\beta N}\frac{1}{\hat{S}} \tag{4.27}$$

Note that again we see the same key ratio of infectivity and recovery.

Now, if we take the reciprocal of $\frac{d\hat{R}}{d\hat{S}}$, we get:

$$\frac{d\hat{S}}{d\hat{R}} = -\frac{\beta N}{\gamma}\hat{S} \tag{4.28}$$

which is simply a linear function of $\hat{S}$, which we know has an exponential solution:

$$\hat{S}(\hat{R}) = \hat{S}_o e^{\frac{-\beta N}{\gamma}\hat{R}} \tag{4.29}$$

Where $\hat{S}_o$ is the fraction susceptible when $\hat{R} = 0$. Let X represent the total fraction of individuals who have been infected (and subsequently recovered) at any time; X does not include currently infected, but not recovered, individuals. Then, if we begin with no immune

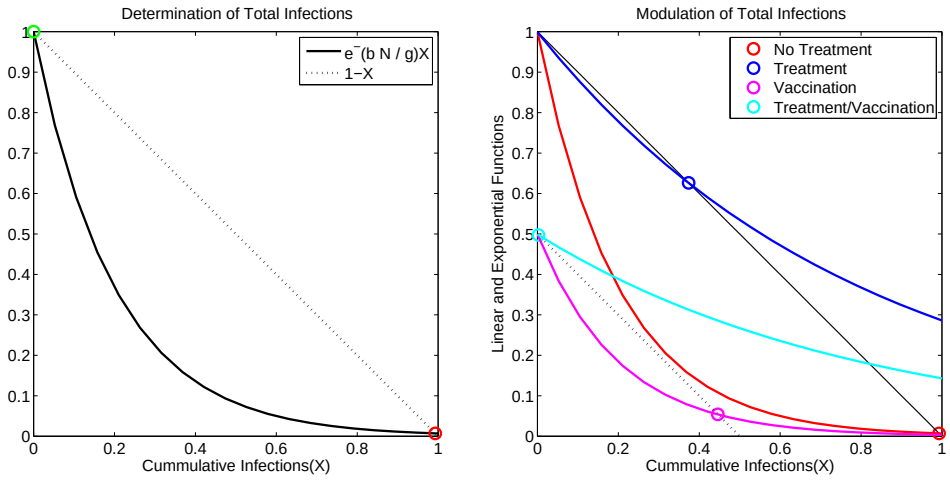

Figure 4.3: Determining the total infected population of an SIR model. The left hand curve shows the intersection of a linear $(1 - X)$ and exponential $(e^{\frac{-\beta N}{\gamma}X})$ curve. The left panel shows the results for systems with vaccination (50% initial resistance) and treatment (a four-fold increase in recovery rate). Treatment alone shifts changes the curvature of the exponential, but not the y-intercept or the line; vaccination shifts both curves, giving a similar cummulative infection total; combining both leads to no crossing of the curves in the biologically relevant domain (all infections immediately die out).

individuals, at the end of the epidemic the total fraction of individuals that are susceptible or recovered are:

$$\hat{S}_f = 1 - X \qquad \hat{R}_f = X \tag{4.30}$$

since $\hat{I}_f = 0$. Also, note that in this case, $\hat{S}_o = 1 - \delta\hat{I}$, where $\delta\hat{I}$ is the fraction of infected individuals introduced into the population at the outset; this value will typically be very small, and thus $\hat{S}_o \approx 1$. Substituting these expressions into our result for $\hat{S}(\hat{R})$:

$$1 - X = \hat{S}_o e^{\frac{-\beta N}{\gamma}X} \tag{4.31}$$

The total fraction of infected individuals is given by the intersection of the straight line $1-X$ and the decaying exponential $\hat{S}_o e^{\frac{-\beta N}{\gamma}X}$. Again, what about vaccination? In this case, we have:

$$\hat{S}_f = 1 - \hat{R}_o - X \qquad \hat{R}_f = R_o + X \tag{4.32}$$

and thus we are interested in the intersection of the straight line $1 - \hat{R}_o - X$ and the exponential $e^{\frac{-\beta N}{\gamma}\left(\hat{R}_o + X\right)}$. The straight line is shifted downward — with no change in slope — by $\hat{R}_o$. The exponential we can re-write:

$$\hat{S}_o e^{\frac{-\beta N}{\gamma}\left(\hat{R}_o + X\right)} = \hat{S}_o e^{\frac{-\beta N}{\gamma}\hat{R}_o} e^{\frac{-\beta N}{\gamma}X} \tag{4.33}$$

and thus we see that the exponential term is scaled by $e^{\frac{-\beta N}{\gamma}\hat{R}_o}$, which, since all the constants are positive, will always be less than unity. Note that the term $\hat{S}_o e^{\frac{-\beta N}{\gamma}\hat{R}_o}$ must correspond to the total susceptible fraction when $X = 0$, which is given by $1 - \hat{R}_o - \delta\hat{I}$, where again

$\delta\hat{I}$ is the initial fraction of the population infected. Scaling the exponential in this way can be thought of as a shift of the curve to the left, resulting in a change in y-intercept. The combination of scaling the exponential and lowering the line gives a net reduction in the value X where they intersect. Thus, we see that vaccination reduces the net impact of an epidemic by a quantifiable amount. We can similarly use this approach to quantify the effective of a public health intervention (lowering β and/or increasing γ, or the combination of the two.

4.2.4 Long term disease models with birth and death

For diseases with long term dynamics, it is essential to consider births and deaths within the population. To do so, we make a few additions to our initial models. First consider births; the birth rate should be a function of the population size. It is possible that infected, susceptible and recovered individuals may have different per capita birth rates, but we will assume that these are all the same. Thus, the birth rate is a function only of total population size. In the simplest model, this will be a linear function, αN. We additionally presume that all births become part of the susceptible population. In other words, there is no *vertical transmission* of the disease from mother to fetus.

Death we will consider in two classes. First, there will be a natural per capita death rate, which (as with the birth rate) is the same for all populations; again, the simplest form is the linear model, δN. Additionally, we may consider the possibility of disease related deaths. We will assume that this applies only to the infected population, and that recovered individuals have no increased risk of death. We may again treat this with a linear model, but only in terms of the infected population, χI.

Schematically, the SIR model discussed above, then becomes:

$$\begin{array}{ccccccc} & & & & X & & \\ & & & & \uparrow{\scriptstyle \chi I} & & \\ \xrightarrow{\alpha N} & S & \xrightarrow{\beta IS} & I & \xrightarrow{\gamma I} & R & \\ & \downarrow{\scriptstyle \delta S} & & \downarrow{\scriptstyle \delta I} & & \downarrow{\scriptstyle \delta R} & \\ & D & & D & & D & \end{array} \tag{4.34}$$

where D and X are place-holder variable to describe the population that has died of natural or disease-related causes, respectively; they will not explicitly be involved in any of the expressions in our model. Note that we are treating the natural death of each group separately, but that:

$$\delta N = \delta(S + I + R) = \delta S + \delta I + \delta R \tag{4.35}$$

since $N = S + I + R$. The differential equations for the model are thus:

$$\frac{dS}{dt} = \alpha N - \beta IS - \delta S \tag{4.36}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \delta I - \chi I \tag{4.37}$$

$$\frac{dR}{dt} = \gamma I - \delta R \tag{4.38}$$

Birth and death with constant population size.

Although we are considering birth and deaths, we may still choose to consider the total population size to be a constant. Most animal populations reach a steady state, and thus if we wish to consider an endemic disease that remains in the population for generations, this may be a reasonable assumption. Human populations have been an exception to this, with consistent growth over the past several thousand years, at least. However, in many developed countries, population growth is leveling out, so this trend may not continue. We will consider the case of constant population size in detail, but it is important to keep in mind that there are cases where this may not be an ideal assumption. For example, populations are rapidly growing in many developing countries, and populations can shrink in the case of a severe epidemic with a large rate of death by disease.

A stable population size requires that the birth rate perfectly balances the death rate. If there are no disease related deaths,, this gives:

$$\alpha N = \delta S + \delta I + \delta R = \delta N \qquad \therefore \qquad \alpha = \delta \tag{4.39}$$

That is, the per capita birth rate and the per capita natural death rate must be the same. When there are disease related deaths, we have:

$$\alpha N = \delta N + \chi I, \qquad \chi > 0 \qquad \therefore \qquad \alpha > \delta \tag{4.40}$$

Let us consider the first case, and begin by using fractional populations, $\hat{S} = \frac{S}{N}$, $\hat{I} = \frac{I}{N}$, and $\hat{R} = \frac{R}{N}$. The differential equations become:

$$\frac{d\hat{S}}{dt} = \alpha - \beta N \hat{I}\hat{S} - \alpha\hat{S} = \alpha\left(1 - \hat{S}\right) - \beta N \hat{I}\hat{S} \tag{4.41}$$

$$\frac{dI}{dt} = \beta N \hat{I}\hat{S} - \gamma\hat{I} - \alpha\hat{I} - \chi\hat{I} = \left(\beta N\hat{S} - (\gamma + \alpha)\right)\hat{I} \tag{4.42}$$

$$\frac{d\hat{R}}{dt} = \gamma\hat{I} - \alpha\hat{R} \tag{4.43}$$

Analysis of dynamics and stationary points

Now, we have a more complicated model, so we should begin our consideration of dynamics by characterizing the null clines. First, $\frac{d\hat{S}}{dt} = 0$ gives:

$$\alpha\left(1 - \hat{S}\right) - \beta N\hat{I}\hat{S} = 0 \rightarrow \beta N\hat{I}\hat{S} = \alpha\left(1 - \hat{S}\right) \rightarrow \hat{I} = \frac{\alpha}{\beta N}\left(\frac{1}{\hat{S}} - 1\right) \tag{4.44}$$

There is a single null cline, with $\hat{I}$ varying with the reciprocal of $\hat{S}$. Now, for $\frac{d\hat{I}}{dt} = 0$ we have:

$$\left(\beta N\hat{S} - (\gamma + \alpha)\right)\hat{I} = 0 \tag{4.45}$$

This gives *either* $\hat{I} = 0$, *or*:

$$\beta N\hat{S} - (\gamma + \alpha) = 0 \rightarrow \beta N\hat{S} = (\gamma + \alpha) \rightarrow \hat{S} = \frac{(\gamma + \alpha)}{\beta N} \tag{4.46}$$

There are two null clines, one a constant (zero) of $\hat{I}$, and the other a non-zero constant $\hat{S}$. Finally, $\frac{d\hat{R}}{dt} = 0$ gives:

$$\gamma\hat{I} - \alpha\hat{R} = 0 \rightarrow \alpha\hat{R} = \gamma\hat{I} \rightarrow \hat{R} = \frac{\gamma}{\alpha}\hat{I} \tag{4.47}$$

There is a single, linear null cline.

Stationary points are intersections of these. First consider $\hat{I} = 0$. Then:

$$0 = \frac{\alpha}{\beta N}\left(\frac{1}{\hat{S}} - 1\right) \rightarrow \frac{1}{\hat{S}} - 1 = 0 \rightarrow \hat{S} = 1 \tag{4.48}$$

and:

$$\hat{R} = \frac{\gamma}{\alpha}0 = 0 \tag{4.49}$$

This gives $(\hat{S}, \hat{I}, \hat{R}) = (1, 0, 0)$, which is the unaffected (disease absent) steady state. Is this point stable? Consider $\frac{d\hat{I}}{dt}$, whose sign will be determined by the expression $\beta N\hat{S} - (\gamma + \alpha)$, since $\hat{I}$ is strictly positive. Now, we are concerned with points very near the stationary point, and thus $\hat{S} \approx 1$. Thus:

$$\beta N\hat{S} - (\gamma + \alpha) \approx \beta N - (\gamma + \alpha) \tag{4.50}$$

If $\frac{\beta N}{\gamma+\alpha} < 1$, then this will be negative, and the small number of infections will die out — the point is stable. If, on the other hand, $\frac{\beta N}{\gamma+\alpha} > 1$, this expression will be positive, and a small initial number of infections will grow — the stationary point will not be stable. Again we see that the key ratio defining an infection that will die out is the determinant of stability of this point.

Now, we have a second stationary point, which we obtain beginning with the second null cline for $\hat{I}$, $\hat{S} = \frac{(\gamma+\alpha)}{\beta N}$. This gives:

$$\hat{I} = \frac{\alpha}{\beta N}\left(\frac{1}{\hat{S}} - 1\right) = \frac{\alpha}{\beta N}\left(\frac{\beta N}{\gamma + \alpha} - 1\right) = \left(\frac{\alpha}{\gamma + \alpha} - \frac{\alpha}{\beta N}\right) \tag{4.51}$$

and in the expression for $\hat{R}$:

$$\hat{R} = \frac{\gamma}{\alpha}\hat{I} = \frac{\gamma}{\alpha}\left(\frac{\alpha}{\gamma + \alpha} - \frac{\alpha}{\beta N}\right) = \left(\frac{\gamma}{\gamma + \alpha} - \frac{\gamma}{\beta N}\right) \tag{4.52}$$

We see that there is a non-zero value of $\hat{I}$, and thus this steady state represents an endemic state, with constant infected population fraction of $\hat{I} = \left(\frac{\alpha}{\gamma+\alpha} - \frac{\alpha}{\beta N}\right)$. Again note that there is a dependence on the total population density. Also note that while there is an endemic steady state, the system does *not* approach it monotonically. Rather, an initial infection results in an *epidemic*, followed by an endemic phase at much lower levels of prevalence.

This behavior becomes very significant when we consider the structure of populations on the global level. Populations consist of many *relatively* isolated populations, with frequent contact within a group, and limited (but non-zero) contact between them. This structure exists at multiple levels: economic regions, countries, cities, and even workplaces and schools; the time frame over which contact can be considered limited depends on the level of structure under consideration.

Now consider the following scenario. First, a new disease occurs in some city (perhaps from a rare transfer from an animal). An epidemic spreads through the population, but is soon contained. However, a very low level of endemic infection remains. Assume that because of a strong public health response, no infected individuals leave the city, and thus the epidemic seems to be a one time concern. After some time — when there are no longer concerns about the disease, because of very low prevalence — one of the rare (at a given time) infected individuals travels to a new city. This can initiate a new *epidemic*, because immunity in the new population is near zero and the susceptible population is large.

The infection stopped spreading rapidly in the first city due to acquired immunity — although the level of infection at a given point in time (prevalence) is low, many individuals are infected at some time or another. As a result, a large portion of the population is in the recovered (immune) state. Of course, this is precisely the motivation for the use of vaccination to prevent future epidemics. If a portion of the population is artificially moved into the immune (recovered) population by inoculation with an attenuated version of the infectious agent, then there is never a large susceptible population to propagate an epidemic.

4.2.5 Incorporating vaccination into an SIR model

How can our model be extended to account for this? One approach is to added a term that describes the "rate of vaccination"; this would moved individuals directly from the susceptible population into the recovered population, and the simplest form would be linear in susceptibles, μS. This leads to the system of differential equations:

$$\frac{dS}{dt} = \alpha N - \beta IS - \delta S - \mu S \tag{4.53}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \delta I - \chi I \tag{4.54}$$

$$\frac{dR}{dt} = \gamma I + \mu S - \delta R \tag{4.55}$$

While not unsolvable, the existence of all three variable in the last expression will complicate the analysis.

A second approach may be considered when vaccinations occur in early childhood. If we assume that vaccination occurs before any significant risk of infection, and that a fraction of new born children are vaccinated, we can simply divide new births into to categories, those that are vaccinated and those that are not. If p is the fraction vaccinated, and $q = 1 - p$ is the fraction unvaccinated, the schematic model becomes:

$$\begin{array}{ccccccc} \xrightarrow{q\alpha N} & S & \xrightarrow{\beta IS} & I & \xrightarrow{\gamma I} & R & \xleftarrow{p\alpha N} \\ & \downarrow{\scriptstyle \delta S} & & \downarrow{\scriptstyle \delta I} & & \downarrow{\scriptstyle \delta R} & \\ & D & & D & & D & \end{array} \tag{4.56}$$

Again assuming a constant population size with no disease-related death (and thus $\alpha = \delta$, the differential equations for the model are:

$$\frac{dS}{dt} = q\alpha N - \beta IS - \alpha S \tag{4.57}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I \tag{4.58}$$

$$\frac{dR}{dt} = p\alpha N + \gamma I - \alpha R \tag{4.59}$$

and in normalized units:

$$\frac{d\hat{S}}{dt} = q\alpha - \beta N\hat{I}\hat{S} - \alpha\hat{S} \tag{4.60}$$

$$\frac{dI}{dt} = \beta N\hat{I}\hat{S} - \gamma\hat{I} - \alpha\hat{I} \tag{4.61}$$

$$\frac{d\hat{R}}{dt} = p\alpha + \gamma\hat{I} - \alpha\hat{R} \tag{4.62}$$

Again, we begin by defining the null clines. First, $\frac{d\hat{S}}{dt} = 0$ gives:

$$\alpha\left(q - \hat{S}\right) - \beta N\hat{I}\hat{S} = 0 \rightarrow \beta N\hat{I}\hat{S} = \alpha\left(q - \hat{S}\right) \rightarrow \hat{I} = \frac{\alpha}{\beta N}\left(\frac{q}{\hat{S}} - 1\right) \tag{4.63}$$

Note that this is almost identical to the result above, with a simple substitution of $\frac{q}{\hat{S}}$ for $\frac{1}{\hat{S}}$. Now, for $\frac{d\hat{I}}{dt} = 0$ we have exactly the same expressions as previously, as vaccination does not directly affect the infectives; the null clines are *either*:

$$\hat{I} = 0 \quad \text{or} \quad \hat{S} = \frac{(\gamma + \alpha)}{\beta N} \tag{4.64}$$

Finally, $\frac{d\hat{R}}{dt} = 0$ gives:

$$p\alpha + \gamma\hat{I} - \alpha\hat{R} = 0 \rightarrow \alpha\hat{R} = p\alpha + \gamma\hat{I} \rightarrow \hat{R} = \frac{p\alpha + \gamma\hat{I}}{\alpha} = p + \frac{\gamma}{\alpha}\hat{I} \tag{4.65}$$

This is just the expression without vaccination with p added.

Now for the intersection of these. First we have $\hat{I} = 0$ which gives:

$$\frac{\alpha}{\beta N}\left(\frac{q}{\hat{S}} - 1\right) = 0 \rightarrow \hat{S} = q \tag{4.66}$$

and:

$$\hat{R} = p + \frac{\gamma}{\alpha}0 = p \tag{4.67}$$

This, of course, is the non-disease state, where the populations of $\hat{S}$ and $\hat{R}$ are determined solely by the vaccination fraction.

To consider the stability of this point, again we consider the sign of $\frac{d\hat{I}}{dt}$, which will be determined by the expression $\beta N\hat{S} - (\gamma + \alpha)$, since $\hat{I}$ is strictly positive. We are concerned with points very near the stationary point, and thus $\hat{S} \approx q = 1 - p$. Thus:

$$\beta N\hat{S} - (\gamma + \alpha) \approx (1 - p)\beta N - (\gamma + \alpha) \tag{4.68}$$

If $\frac{\beta N}{\gamma+\alpha}(1-p) < 1$, then this will be negative, and the small number of infections will die out, but if, on the other hand, $\frac{\beta N}{\gamma+\alpha}(1-p) > 1$, this expression will be positive, and a small initial number of infections will grow. We see that the key ratio defining an infection that will die out is modulated by $1-p$, where p is the fraction vaccinated — as the fraction of the population vaccinated is increased, the threshold for when an infection will result in an epidemic is increased, just as expected. Quantitatively, the shift in the threshold is proportional to the fraction of the population that is not vaccinated.

To understand the second stationary point, we begin with $\hat{S} = \frac{\gamma+\alpha}{\beta N}$ which gives:

$$\hat{I} = \frac{\alpha}{\beta N}\left(\frac{q}{\hat{S}} - 1\right) = \frac{\alpha}{\beta N}\left(\frac{q\beta N}{\gamma+\alpha} - 1\right) = \left(\frac{q\alpha}{\gamma+\alpha} - \frac{\alpha}{\beta N}\right) \tag{4.69}$$

and:

$$\hat{R} = p + \frac{\gamma}{\alpha}\left(\frac{q\alpha}{\gamma+\alpha} - \frac{\alpha}{\beta N}\right) = p + \frac{(1-p)\gamma}{\gamma+\alpha} - \frac{\gamma}{\beta N} = p\left(1 - \frac{\gamma}{\gamma+\alpha}\right) + \left(\frac{\gamma}{\gamma+\alpha} - \frac{\gamma}{\beta N}\right) \tag{4.70}$$

Now, this will only be physically meaningful when $\hat{I} > 0$, which requires:

$$\left(\frac{q\alpha}{\gamma+\alpha} - \frac{\alpha}{\beta N}\right) > 0 \rightarrow \frac{q\alpha}{\gamma+\alpha} > \frac{\alpha}{\beta N} \rightarrow q > \frac{\gamma+\alpha}{\beta N} \tag{4.71}$$

Note that this is the same key ratio that appears repeatedly. Converting to fraction vaccinated gives:

$$1 - p > \frac{\gamma+\alpha}{\beta N} \rightarrow p < 1 - \frac{\gamma+\alpha}{\beta N} \tag{4.72}$$

Thus, we see that as long as there is a vaccination fraction of *at least* $1 - \frac{\gamma+\alpha}{\beta N}$, the endemic state does not exist — the disease can be eliminated without 100% vaccination.

Now, let's consider two diseases that have (or had) systematic vaccination programs. Small pox has as value of $\frac{\beta N}{\gamma+\alpha}$ less than roughly 3–5 (for birth rates and population densities typical of the developed world). This gives a required vaccination fraction of between $1 - \frac{1}{3} = 67\%$ and $1 - \frac{1}{5} = 80\%$; thus at most an 80% effective vaccination rate is needed to eliminate the endemic disease. In fact, small pox is the only disease that has been globally eradicated. Measles, on the other hand, has a value of $\frac{\beta N}{\gamma+\alpha}$ of about 12, giving a required vaccination fraction of $1 - \frac{1}{12} = 92\%$; to eliminate measles would require 92% effective vaccination, and the disease does in fact remain endemic in the population although major epidemics no longer occur.

4.3 Models of vector-borne diseases

Vector-borne diseases can not be modeled by the simple models outlined above because they is not transmitted directly from individual to individual. As a vector is need for transmission, modeling of the disease epidemiology will be incomplete without consideration of that population as well. As examples, rabies incidence can be reduced by minimizing the number of feral animals present in a city; the spread of the bubonic plague in Medieval Europe was driven by poor control of rodent populations in the cities; and a major focus of controlling both malaria and dengue fever is the eradication of mosquito populations.

4.3.1 Modeling Malaria

A good example for an initial model is malaria. Humans (and other mammals) contract malaria by being bitten by infected *Anopheles* mosquitoes, and the mosquitoes similarly become infected when they bite an infected mammal. To a first approximation, neither humans nor mosquitoes become immune after an infection; this is not strictly true for humans, but it takes multiple exposures to build up any significant level of immunity.

To develop a model, we will need to consider both the human and mosquito populations. In each species, we will consider an SIS disease:

$$\begin{array}{ccccccc} & & D_i & & D_i & & \\ & & \uparrow_{\delta_i S_i} & & \uparrow_{\delta_i I_i} & & \\ \xrightarrow{\alpha_i N_i} & S_i & \xrightarrow{f_i(S_i,I_j)} & I_i & \xrightarrow{\gamma_i I_i} & S_i & \\ & & & \downarrow_{\chi_i I_i} & & & \\ & & & X_i & & & \end{array} \tag{4.73}$$

with linear terms for birth, death (both disease-related and otherwise) and recovery, just as in our previous models. However, the infection function ($f_{j\to i}(S_i, I_j)$) for species i must be a function of both the population of susceptible individuals (S_i) and the population of infected *vectors* (species j, I_j). We can schematically represent the whole system as follows:

$$\begin{array}{lccccccc} & & & \uparrow_{\delta_H S_H} & & \uparrow^{\delta_H I_H}_{+\chi_H I_H} & & \\ \text{Human :} & \xrightarrow{\alpha_H N_H} & S_H & & \xrightarrow{f_{M\to H}(S_H,I_M)} & I_H & \xrightarrow{\gamma_H I_H} & S_H \\ & & & & \Updownarrow & & & \\ \text{Mosquito :} & \xrightarrow{\alpha_M N_M} & S_M & & \xrightarrow{f_{H\to M}(S_M,I_H)} & I_M & \xrightarrow{\gamma_M I_M} & S_M \\ & & & \downarrow_{\delta_M S_M} & & \downarrow^{\delta_M I_M}_{+\chi_M I_M} & & \end{array} \tag{4.74}$$

where for simplicity, we have neglected the populations of deaths, and combine the arrow for disease and non-disease-related deaths. The double arrow represents the cross-species interaction.

Mosquito to human incidence function, $f_{M\to H}$: This will depend on the number of human susceptibles and mosquito infectives. First, presume that the human population is large, and thus, mosquitoes never have trouble finding a target to bite. We can think of this in terms of the predator response models that were discussed earlier, and we expect (for example, based on the Holling's disk equation model) that the *per capita* mosquito biting rate will be a constant (β). Also presume that mosquitoes bite randomly — they can not choose to select an infected or susceptible individual based on some preference. Thus, the *per capita* bite rate of susceptible human individuals will be the overall bite rate, scaled by the fraction of the human population that is susceptible, $\beta\frac{S_H}{N_H}$. Now consider that each bite by an infected mosquito (of a susceptible human) will lead to infection with some constant probability, $P_{M\to H}$. The final incidence function will then be the product of this probability, the bite rate of susceptible humans, and the total number of infective mosquitoes:

$$f_{M\to H} = P_{M\to H}\beta\frac{S_H}{N_H}I_M \tag{4.75}$$

Human to mosquito incidence function, $f_{H\to M}$: We will derive this expression using similar logic. In this case, we will be concerned with the *per capita* biting rate of infected humans, which will be the base biting rate, scaled by the fraction of infecteds, $\beta\frac{I_H}{N_H}$. Again, we will consider that each bite of an infected human by a susceptible mosquito leads to infection with some probability, $P_{H\to M}$, which need *not* be the same as the mosquito to human infection probability. Taking the product of these two terms with the total number of susceptible mosquitoes gives:

$$f_{H\to M} = P_{H\to M}\beta\frac{I_H}{N_H}S_M \tag{4.76}$$

4.3.2 Malaria dynamics with constant populations

To simplify our analysis we will begin with a model with fixed population sizes. For the human population, we presume that we are interested in relative short time periods, and thus that we can ignore births and deaths (*i.e.* $\alpha_H = \delta_H = 0$). We will also ignore disease-related death ($\chi_H = 0$). This is not a perfect assumption — malaria can be a fatal disease, with 881,000 deaths (worldwide) from the disease in 2006. However, in the same time period, there were an estimated 247 million cases, and thus the mortality rate is below half a percent. Thus, the effect of disease-related deaths on total population size can reasonably be neglected, and that is the key point here.

As the life cycle of the mosquito is much faster than that of humans (typically a week to a month), ignoring mosquito births and deaths will not be reasonable on any time scale of interest to humans. We may still assume a constant population size, by equating the birth and death rates ($\alpha_M = \delta_M \neq 0$). Note that we are also presuming that there are no disease-related deaths for mosquitoes, and that infected mosquitoes reproduce at the same rate as susceptibles. Again, this may not be a perfect assumption, as there is likely some negative effect on the insect, but the effect is not highly significant.

With these assumptions, the overall model becomes:

$$\begin{array}{lccccc}
\text{Human :} & & S_H & \xrightarrow{P_{M\to H}\beta\frac{S_H}{N_H}I_M} & I_H & \xrightarrow{\gamma_H I_H} & S_H \\
 & & & \Updownarrow & & & \\
\text{Mosquito :} & \xrightarrow{\alpha_M N_M} & S_M & \xrightarrow{P_{H\to M}\beta\frac{I_H}{N_H}S_M} & I_M & \xrightarrow{\gamma_M I_M} & S_M \\
 & & \downarrow{\scriptstyle \delta_M S_M} & & \downarrow{\scriptstyle \delta_M I_M} & &
\end{array} \tag{4.77}$$

with differential equations:

$$\frac{dS_H}{dt} = -P_{M\to H}\beta\frac{S_H}{N_H}I_M + \gamma_H I_H \tag{4.78}$$

$$\frac{dI_H}{dt} = +P_{M\to H}\beta\frac{S_H}{N_H}I_M - \gamma_H I_H \tag{4.79}$$

$$\frac{dS_M}{dt} = \alpha N_M - P_{H\to M}\beta\frac{I_H}{N_H}S_M + \gamma_M I_M - \alpha S_M \tag{4.80}$$

$$\frac{dI_M}{dt} = +P_{H\to M}\beta\frac{I_H}{N_H}S_M - \gamma_M I_M - \alpha I_M \tag{4.81}$$

As we have done in the other models with a constant population size, we rewrite this in normalized units. Note that since both populations are of constant size, $\hat{S}_H = 1 - \hat{I}_H$ and $\hat{S}_M = 1 - \hat{I}_M$, which leads to, $\frac{d\hat{S}_H}{dt} = \frac{-d\hat{I}_H}{dt}$and $\frac{d\hat{S}_M}{dt} = \frac{-d\hat{I}_M}{dt}$. Thus, we need only be concerned with the infected populations:

$$\frac{-d\hat{S}_H}{dt} = \frac{d\hat{I}_H}{dt} = \frac{1}{N_H}\frac{dI_H}{dt} = \beta P_{M\to H}\frac{N_M}{N_H}\hat{S}_H\hat{I}_M - \gamma_H\hat{I}_H \tag{4.82}$$

$$\frac{-d\hat{S}_M}{dt} = \frac{d\hat{I}_M}{dt} = \frac{1}{N_M}\frac{dI_M}{dt} = \beta P_{H\to M}\hat{I}_H\hat{S}_M - \gamma_M\hat{I}_M - \alpha\hat{I}_M \tag{4.83}$$

One thing immediately stands out — the rate of new human infections (first term) depends on the relative populations (overall) of mosquitoes to humans ($\frac{N_M}{N_H}$), while the rate of new mosquito infections is independent of this number. This is essentially a result of being in the saturated regime of the (mosquito) predator response curve — if there are more mosquitoes, humans will be bitten more often, as each mosquito bites at a constant rate, but larger human populations do not lead to more bites by a given mosquito. As is the number of bites per human that influences human infection, and the number of bites per insect that influences mosquito infection, we get this non-symmetry in the model.

Now, what are the null clines in this model? For $\frac{d\hat{I}_H}{dt} = 0$, we have:

$$\beta P_{M\to H}\frac{N_M}{N_H}\hat{S}_H\hat{I}_M - \gamma_H\hat{I}_H = 0 \to \beta P_{M\to H}\frac{N_M}{N_H}\hat{S}_H\hat{I}_M = \gamma_H\hat{I}_H \tag{4.84}$$

Using $\hat{S}_H = 1 - \hat{I}_H$ gives:

$$\beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_M - \beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_H\hat{I}_M = \gamma_H\hat{I}_H \to \beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_M = \left(\gamma_H + \beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_M\right)\hat{I}_H \tag{4.85}$$

which can be rearranged to give:

$$\hat{I}_H = \frac{\beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_M}{\gamma_H + \beta P_{M\to H}\frac{N_M}{N_H}\hat{I}_M} = \frac{\beta P_{M\to H}N_M\hat{I}_M}{\gamma_H N_H + \beta P_{M\to H}N_M\hat{I}_M} \tag{4.86}$$

Now, this may look complicated, but in fact, it is an expression of the form:

$$\hat{I}_H = \frac{A\hat{I}_M}{B + A\hat{I}_M} \tag{4.87}$$

with $A = \beta P_{M\to H}N_M$ and $B = \gamma_H N_H$. This is the same form as, for example, the Holling's disc equation, with linear behavior ($\hat{I}_H \approx \frac{A}{B}\hat{I}_M$) at low values of $\hat{I}_M$, and saturating behavior ($\hat{I}_H \approx 1$) at high values — we may thus sketch a plot of $\hat{I}_H$ versus $\hat{I}_M$. However, recall that $\hat{I}_M$ is limited to values between 0 and 1, and thus the saturating portion of the curve will only be reached if $A >> B$, or $\beta P_{M\to H}N_M >> \gamma_H N_H$. Conversely, if $\beta P_{M\to H}N_M << \gamma_H N_H$, the linear regime will always hold (and since this ratio defines the slope, the slope will be very small).

Now, similarly consider $\frac{d\hat{I}_M}{dt} = 0$, where we have:

$$\beta P_{H\to M}\hat{I}_H\hat{S}_M - \gamma_M\hat{I}_M - \alpha\hat{I}_M = 0 \to \beta P_{H\to M}\hat{I}_H\hat{S}_M = \gamma_M\hat{I}_M + \alpha\hat{I}_M = (\gamma_M + \alpha)\,\hat{I}_M \quad (4.88)$$

Again, using $\hat{S}_M = 1 - \hat{I}_M$ gives:

$$\beta P_{H\to M}\hat{I}_H\left(1 - \hat{I}_M\right) = (\gamma_M + \alpha)\,\hat{I}_M \to \hat{I}_H = \frac{(\gamma_M + \alpha)}{\beta P_{H\to M}}\frac{\hat{I}_M}{(1 - \hat{I}_M)} \quad (4.89)$$

Again we have $\hat{I}_H$ as a function of $\hat{I}_M$, but this cline has qualitatively very different behavior. At low values of $\hat{I}_M$, this too, behaves linearly ($\hat{I}_H = \frac{(\gamma_M+\alpha)}{\beta P_{H\to M}}\hat{I}_M$), but as $\hat{I}_M$ approaches 1, the denominator approaches zero, and thus the function approaches infinity. As $\hat{I}_H$ can not exceed 1, this null cline will always reach the boundaries of allowed space before $\hat{I}_M$ reaches 1.

Thus, in summary, we see that:

- The $\hat{I}_M$ null cline begins with a slope of $\frac{(\gamma_M+\alpha)}{\beta P_{H\to M}}$, and the slope then monotonically increases.
- The $\hat{I}_H$ null cline begins with a slope of $\frac{\beta P_{M\to H}N_M}{\gamma_H N_H}$, and the slope then monotonically decreases.

There are two possible, qualitatively distinct, possibilities. First, we may have:

$$\frac{(\gamma_M + \alpha)}{\beta P_{H\to M}} > \frac{\beta P_{M\to H}N_M}{\gamma_H N_H} \to (\gamma_M + \alpha)\gamma_H N_H > \beta^2 P_{H\to M}P_{M\to H}N_M \quad (4.90)$$

In this case, the cline with the increasing slope begins with a higher slope than the other, and thus the two null clines intersect only at $(0, 0)$. Below both null clines, the fraction of mosquito infectives will decrease, while that of human infectives will increase; similarly, above both null clines, the fraction of human infectives will decrease while mosquito infections will rise. In between the two null clines, both populations will decrease. This behavior indicates that the $(\hat{I}_M, \hat{I}_H) = (0, 0)$ stationary point is attracting (stable) in this parameter regime. Biologically, this means that the infection will die out in both populations over time. Note that each side of the inequality describing this threshold behavior can be thought of as a product of two terms, one for each population. On the left hand side, we have recovery terms, and on the right hand side we have infection terms; the geometric average of recovery terms for each population must out-weigh the geometric average of infection terms for the infection to die out.

If this inequality is not satisfied — that is, if:

$$\frac{(\gamma_M + \alpha)}{\beta P_{H\to M}} < \frac{\beta P_{M\to H}N_M}{\gamma_H N_H} \leftrightarrow (\gamma_M + \alpha)\gamma_H N_H < \beta^2 P_{H\to M}P_{M\to H}N_M \quad (4.91)$$

then the cline with the increasing slope will begin with a lower slope, and thus there will be a non-zero crossing point. Again, below both null clines, the fraction of mosquito infectives will decrease, while that of human infectives will increase, and above both null clines, the fraction

of human infectives will decrease while mosquito infections will rise. In between the two null clines, and to the right of (and above) the intersection, both populations will decrease, while between the clines but below (and to the left of) the intersection, both infected populations will increase. Thus, we see that this stationary point is stable, and the $(0,0)$ point becomes unstable.

To find this point of intersection, we simply equate the two null clines:

$$\frac{\beta P_{M\to H} N_M \hat{I}_M}{\gamma_H N_H + \beta P_{M\to H} N_M \hat{I}_M} = \frac{(\gamma_M+\alpha)}{\beta P_{H\to M}} \frac{\hat{I}_M}{(1-\hat{I}_M)} \tag{4.92}$$

which simplifies to:

$$\begin{aligned}
\beta^2 P_{H\to M} P_{M\to H} N_M (1-\hat{I}_M) &= (\gamma_M+\alpha)(\gamma_H N_H + \beta P_{M\to H} N_M \hat{I}_M) && (4.93)\\
\beta P_{M\to H} N_M \left(\beta P_{H\to M} + \gamma_M + \alpha\right) \hat{I}_M &= \beta^2 P_{H\to M} P_{M\to H} N_M - (\gamma_M+\alpha)\gamma_H N_H && (4.94)\\
\hat{I}_M &= \frac{\beta^2 P_{H\to M} P_{M\to H} N_M - (\gamma_M+\alpha)\gamma_H N_H}{\beta P_{M\to H} N_M \left(\beta P_{H\to M} + \gamma_M + \alpha\right)} && (4.95)\\
\hat{I}^*_M &= \frac{\beta^2 P_{H\to M} P_{M\to H} N_M - (\gamma_M+\alpha)\gamma_H N_H}{\beta^2 P_{H\to M} P_{M\to H} N_M + \beta P_{M\to H} N_M (\gamma_M+\alpha)} && (4.96)
\end{aligned}$$

Given, $\hat{I_M}$, we can easily find $\hat{I}_H$:

$$\begin{aligned}
\hat{I}^*_H &= \frac{(\gamma_M+\alpha)}{\beta P_{H\to M}} \frac{\hat{I}^*_M}{(1-\hat{I}^*_M)} && (4.97)\\
&= \frac{\beta^2 P_{H\to M} P_{M\to H} N_M - (\gamma_M+\alpha)\gamma_H N_H}{\beta^2 P_{H\to M} P_{M\to H} N_M + \beta P_{H\to M} N_H \gamma_H} && (4.98)
\end{aligned}$$

where the algebra is left as an exercise. While fairly complicated, we find that we can determine the steady-state infected fractions of both the human and mosquito populations.s

Note that the key comparison (of $\beta^2 P_{H\to M} P_{M\to H} N_M$ and $(\gamma_M+\alpha)\gamma_H N_H$) involves both total human population density (N_H and mosquito population density (N_M). An increase in the mosquito population will increase the infectivity term, making endemic infection more likely, while increases in the human population term increase the recovery term, and make endemic infection less likely. This is in direct contrast to models with direct transmission, in which increased population lead to increases in disease prevalence.

4.3.3 Parallels with sexually transmitted diseases

While this model was developed in the context of a vector-borne disease, there are many similarities with diseases that are transmitted directly from individual to individual, but primarily through heterosexual contact. In this case, we have a male and a female population, and we can write a model as:

$$\begin{array}{lccccc}
\text{Female :} & S_F & \xrightarrow{\beta_{M\to F} S_F I_M} & I_F & \xrightarrow{\gamma_F I_F} & S_F \\
 & & \Updownarrow & & & \\
\text{Male :} & S_M & \xrightarrow{\beta_{F\to M} S_M I_M} & I_M & \xrightarrow{\gamma_M I_M} & S_M
\end{array} \tag{4.99}$$

where we have assumed a simple bi-linear transmission function, and no birth or death.