Notes On $\mathcal{R}_0$

James Holland Jones [*]
Department of Anthropological Sciences
Stanford University

April 25, 2020

1 The Basic Reproduction Number in a Nutshell

The basic reproduction number, $\mathcal{R}_0$, is defined as the expected number of secondary cases produced by a single (typical) infection in a completely susceptible population. It is important to note that $\mathcal{R}_0$ is a dimensionless number and not a rate, which would have units of time^{-1}. Some authors incorrectly call $\mathcal{R}_0$ the "basic reproductive rate."

We can use the fact that $\mathcal{R}_0$ is a dimensionless number to help us in calculating it.

$$\mathcal{R}_0 \propto \left(\frac{\text{infection}}{\text{contact}} \right) \cdot \left(\frac{\text{contact}}{\text{time}} \right) \cdot \left(\frac{\text{time}}{\text{infection}} \right)$$

More specifically:

$$\mathcal{R}_0 = \tau \cdot \bar{c} \cdot d \tag{1}$$

where τ is the transmissibility (i.e., probability of infection given contact between a susceptible and infected individual), $\bar{c}$ is the average rate of contact between susceptible and infected individuals, and d is the duration of infectiousness.

2 The SIR Epidemic Model

It is pretty clear how we calculate $\mathcal{R}_0$ given information on transmissibility, contact rates, and the expected duration of infection. But how do we know that this quantity defines the epidemic threshold of a particular infection? To understand this, we need to formulate an epidemic model. The model we use is called an SIR model, where SIR stands for "Susceptible-Infected-Removed."

For simplicity, we will deploy several assumptions:

1. Constant (closed) population size, N

---

[*]Present Address: Department of Earth System Science, Y2E2 Building, 473 Via Ortega, Room 140 Stanford, CA 94305-4216; phone: 650-721-2484, fax: 650-725-5099; email: `jhj1@stanford.edu`

2. Constant rates (e.g., transmission, removal rates)

3. No demography (i.e., births and deaths)

4. Well-mixed population

A well-mixed population is one where any infected individual has a probability of contacting any susceptible individual that is reasonably well approximated by the average. This is often the most problematic assumption, but is easily relaxed in more complex models.

In our closed population of N individuals, say that S are susceptible, I infected, and R are removed. Write $s = S/N$, $i = I/N$, $r = R/N$ to denote the fraction in each compartment.

The SIR model is then:

$$\frac{ds}{dt} = -\beta s i \tag{2}$$

$$\frac{di}{dt} = \beta s i - \nu i \tag{3}$$

$$\frac{dr}{dt} = \nu i \tag{4}$$

where $\beta = \tau \bar{c}$ and is known as the effective contact rate, ν is the removal rate. By assumption all rates are constant. This means that the expected duration of infection is simply the inverse of the removal rate: $d = \nu^{-1}$.

What are the conditions for an epidemic? An epidemic occurs if the number of infected individuals increases, i.e., $di/dt > 0$

$$\beta s i - \nu i > 0$$

$$\frac{\beta s i}{\nu} > i$$

At the outset of an epidemic, nearly everyone (except the index case) is susceptible. So we can say that $s \approx 1$. Substituting $s = 1$, we arrive at the following inequality

$$\frac{\beta}{\nu} = \mathcal{R}_0 > 1$$

Since $\beta = \tau \bar{c}$ and $d = \nu^{-1}$, we see that we have derived our expression for $\mathcal{R}_0$ given in equation 1. This little bit of mathematical trickery explains why we have that cumbersome phrase "in a completely susceptible population" tacked onto our definition for $\mathcal{R}_0$.

2.1 Heterogeneous Contact

Following the insights of Hethcote and Yorke (1984) on the role of partnership heterogeneity in permitting continued gonorrhea transmission, Anderson et al. (1986) derived an analytical relationship for the impact of such behavioral heterogeneity on the basic reproduction number for HIV infection.

Begin again with a population of N individuals that is divided into k subgroups, N_i, where i is the average number of sexual partners per unit time. $N_i = Np_i$, where p_i is the proportion of the population in the ith activity class. We use an SI model, since there is no recovery from HIV infection (at least there wasn't in 1986).

Assume (and this is a big assumption) that the risk of transmission increases *linearly* with the average number of partners. Let $\lambda = \tau I/N$ be the force of infection, where τ is the probability of transmission given contact.[1] Thus, the rate of formation of new infections in the ith activity class is $i\lambda$. Let $X_i(t)$ and $Y_i(t)$ be the number of susceptible and infectious individuals in activity class i respectively. Our SI model is thus written as:

$$\frac{dX_i}{dt} = -i\lambda X_i \tag{5}$$

$$\frac{dY_i}{dt} = i\lambda X_i - \nu Y_i \tag{6}$$

At time t, assuming proportionate mixing (another big assumption) the overall force of infection is:

$$\lambda(t) = \frac{\tau \sum_i i Y_i(t)}{\sum_i i N_i(t)} \tag{7}$$

Integrate equation 5 to get

$$X_i(t) = N_i \exp\left[-i\psi(t)\right], \tag{8}$$

where $\psi(t)$ is simply the summed force of infection up to time t:

$$\psi(t) = \int_0^t \lambda(g)dg.$$

Substitute 7 into 6 and sum to produce a differential equation for $\lambda(t)$:

$$\frac{d\lambda}{dt} = \lambda \left(\tau \sum_i i^2 e^{-i\psi} p_i / \sum_i i p_i - \nu \right). \tag{9}$$

At the outset of an epidemic, we can assume that approximately everyone in each activity class is susceptible (i.e., $X_i \approx N_i$) and the summed force of infection is zero, $\psi = 0$, since there's been no epidemic yet! This simplifies equation 9 considerably. Note that, after simplification, the term in the numerator $(\sum_i i^2 p_i)$ is the expectation of a squared variable, $\mathbb{E}(i^2)$. This is related to variance, which by definition is $\mathbb{V}(i) = \mathbb{E}(i - E(i))^2 = \mathbb{E}(i^2) - \mathbb{E}(i)^2$. So $\mathbb{E}(i^2) = \mathbb{E}(i) + \mathbb{V}(i)$. The denominator $(\sum_i i p_i)$ is the expected (or mean) number of contacts $\mathbb{E}(i)$. Call the variance in contacts σ^2 and the mean number of contacts μ. Noting that $\beta = \tau\mu$ and defining the coefficient of variation as $c = \sigma/\mu$, we get the relationship:

---

[1] Anderson et al. (1986) use β for this probability, but this is confusing given more current usage where β is more conventionally the product of transmissibility and the mean rate of contact between susceptibles and infectious individuals. We therefore use τ in the derivation rather than β.

$$\frac{d\lambda}{dt} = \lambda \left[\beta(1 + c^2) - \nu\right]. \tag{10}$$

Solving for $\dot{\lambda} > 0$, we recover the value of $\mathcal{R}_0$ for heterogeneous contacts.

$$\mathcal{R}_0 = \mathcal{R}_0^{(M)}(1 + c^2), \tag{11}$$

where $\mathcal{R}_0^{(M)}$ is the mean-field solution for $\mathcal{R}_0$.

3   Epidemic Thresholds in Structured Populations

3.1   Next Generation Matrix: Intuitive Approach

If $\mathcal{R}_0$ is the number of secondary infections produced by a single typical infection in a rarefied population, how do we define it when there are multiple types of infected individuals. For example, what is a *typical* infection in a vector-borne disease like malaria? What about a sexually transmitted infection where there are large asymmetries in transmissibility (like HIV)? Or what about a multi-host pathogen like influenza?

It turns out that there is a straightforward extension of the theory for structured epidemic models (Diekmann et al. 1990; van den Driessche and Watmough 2002). The mathematics behind this theory is not especially difficult, but it does involve scary German terms that are not familiar to the non-engineers in our midst. The key concept is that we now need to average the expected number of new infections over all possible infected types.

Assume that we have a system in which there are multiple discrete types of infected individuals (e.g., mosquitoes and humans; women and men; or humans, dogs, and chickens). We define the *next generation matrix* as the square matrix $\mathbf{G}$ in which the ijth element of $\mathbf{G}$, g_{ij}, is the expected number of secondary infections of type i caused by a single infected individual of type j, again assuming that the population of type i is entirely susceptible. That is, each element of the matrix $\mathbf{G}$ is a reproduction number, but one where who infects whom is accounted for.

Once we have $\mathbf{G}$, we are one step away from $\mathcal{R}_0$. The basic reproduction number is given by the *spectral radius* of $\mathbf{G}$. The spectral radius is the also known as the dominant eigenvalue of $\mathbf{G}$. The next generation matrix has a number of desirable properties from a mathematical standpoint. In particular, it is a non-negative matrix and, as such, it is guaranteed that there will be a single, unique eigenvalue which is positive, real, and strictly greater than all the others. This is $\mathcal{R}_0$.

For illustrative purposes, we will limit our discussion to the case where there are two classes of infected individual. The next generation matrix is thus 2×2. For the 2×2 matrix, there is a fairly simple formula for the eigenvalues that everyone who took algebra in high school should know. Define

$$\mathbf{G} = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$$

The eigenvalues of $\mathbf{G}$ are:

$$\lambda_\pm = \frac{T}{2} \pm \sqrt{(T/2)^2 - D}$$

where $T = a + d$ is the trace and $D = ad - bc$ is the determinant of matrix $\mathbf{G}$.

Say you have a sexually transmitted disease in a completely heterosexual population. Define f as the expected number of infected women and m as the expected number of infected men given contact with a single infected member of the opposite sex in a completely susceptible population. The next generation matrix is

$$\mathbf{G} = \left[\begin{array}{cc} 0 & f \\ m & 0 \end{array} \right]$$

$\mathcal{R}_0$ is thus $\sqrt{mf}$. It is worth noting that this is the geometric mean of the expected number of female and male secondary cases.

When you have more than two discrete types of infectious classes, the formula for calculating the eigenvalues is not as simple, but the same ideas apply. In practice, we typically calculate eigenvalues numerically using software like R, MATLAB, or Mathematica.

3.2   Next Generation Matrix: More Formal Approach

A review paper by Heffernan et al. (2005) provides a nice readable introduction for calculating $\mathcal{R}_0$ in structured population models. The notation I use here follows their usage.

Consider the next generation matrix $\mathbf{G}$. It is comprised of two parts: F and V^{-1}, where

$$F = \left[\frac{\partial F_i(x_0)}{\partial x_j} \right] \tag{12}$$

and

$$V = \left[\frac{\partial V_i(x_0)}{\partial x_j} \right] \tag{13}$$

These are square matrices of the partial derivatives of new infections (F_i) and transfers between different compartments (V_i). The rank of these matrices is the number of distinct classes of infections. For example, in a simple two-sex SIR model, the next generation matrix would be 2×2 since there are two classes of infection (i.e., female and male). x_0 is the disease-free equilibrium state. This matrix should be non-negative, irreducible, and primitive.

$\mathcal{R}_0$ is then the dominant eigenvalue of the matrix $\mathbf{G} = FV^{-1}$.

Example: SEIR Epidemic   Consider a Susceptible-Exposed-Infected-Removed (SEIR) Epidemic. This is an appropriate model for a disease where there is a considerable post-infection incubation period in which the exposed person is not yet infectious.

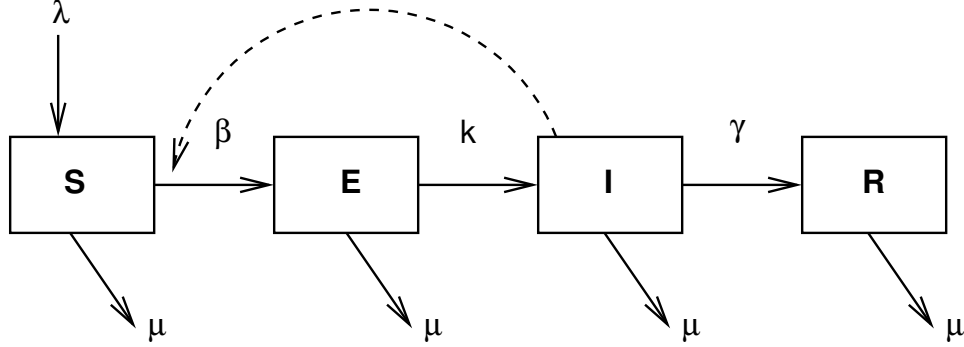The simple SEIR model consists of a set of four differential equations:

Figure 1: State diagram for the SEIR model. β is the effective contact rate, λ is the "birth" rate of susceptibles, μ is the mortality rate, k is the progression rate from exposed (latent) to infected, γ is the removal rate.

$$
\begin{aligned}
\dot{S} &= -\beta S I + \lambda - \mu S & (14)\\
\dot{E} &= \beta S I - (\mu + k)E & (15)\\
\dot{I} &= kE - (\gamma + \mu)I & (16)\\
\dot{R} &= \gamma I - \mu R & (17)
\end{aligned}
$$

where β is the effective contact rate, λ is the "birth" rate of susceptibles, μ is the mortality rate, k is the progression rate from exposed (latent) to infected, γ is the removal rate.

Note that in the absence of infection, the equilibrium number of susceptible is $S^* = \lambda/\mu$.

To calculate the next generation matrix for the SEIR model, we need to enumerate the number of ways that (1) new infections can arise and (2) the number of ways that individuals can move between compartments. There are two disease states but only one way to create new infections. New infections arise through the exposed class. There are only two infection classes E and I. Our F matrix collects the partial derivatives of the rate of new infections with respect to changes in the infection classes:

$$
F = \begin{pmatrix} 0 & \frac{\beta\lambda}{\mu} \\ 0 & 0 \end{pmatrix}, \tag{18}
$$

where we substitute the disease-free equilibrium for S.

The V matrix collects the partial derivatives of the various ways to move between the two infection states:

$$
V = \begin{pmatrix} k+\mu & 0 \\ -k & \gamma+\mu \end{pmatrix} \tag{19}
$$

$\mathcal{R}_0$ is the leading eigenvalue of the matrix FV^{-1}. This is reasonably straightforward to calculate since FV^{-1} is simply a 2×2 matrix.

$$\mathcal{R}_0 = \frac{k\beta\lambda}{\mu(k+\mu)(\gamma+\mu)} \tag{20}$$

It is interesting to note that $\mathcal{R}_0$ is also the product of the rate of production of (1) new exposures and (2) new infections, as it should be.

4   What is a Generation?

In demography, $\mathcal{R}_0$ represents the ratio of total population size from the start to the end of a generation, which is, roughly, the mean age of childbearing. $\mathcal{R}_0 = e^{rT}$, where r is the instantaneous rate of increase of the population. So what is a generation in an epidemic model? Generations in epidemic models are the waves of secondary infection that flow from each previous infection. So, the first generation of an epidemic is all the secondary infections that result from infectious contact with the index case, who is of generation zero. If $\mathcal{R}_i$ denotes the reproduction number of the ith generation, then $\mathcal{R}_0$ is simply the number of infections generated by the index case, i.e., generation zero. Now, these numbers are typically small and are therefore susceptible to random sampling error. Consequently, we talk about expected (i.e., averaged over many epidemics) numbers of secondary cases produced by generation zero. See Diekmann and Heesterbeek (2000) or Heesterbeek (2002) for a more complete discussion of this topic.

Figure 2 shows a schematic representation of an epidemic. The zeroth generation of the epidemic is the index case – patient zero – indicated in red. The number of secondary infections generated by the case in generation zero is $\mathcal{R}_0 = 3$. In the first generation (blue), $\mathcal{R}_1 = 6/3 = 2$. The second generation (cyan) has $\mathcal{R}_1 = 12/6 = 2$.

The generation time of an epidemic is typically called the *serial interval* which we will denote V. Anderson and May (1991) show how we can use the serial interval to estimate $\mathcal{R}_0$ at the outset of an outbreak. Assuming constant mortality, they show that

$$\mathcal{R}_0 = 1 + rV, \tag{21}$$

where r is the exponential rate of increase of infections. Assuming the outbreak started with a small number of infections (e.g., 1), we estimate this simply as $r = \log(I(t))/t$.

For example, on 26 January 2020, the WHO situation report indicated that there were approximately 2000 COVID-19 cases in China. On the same day, Imperial College epidemiologist Neil Ferguson told the Guardian newspaper that he believed there were already 100,000 infections in China. Based on previous estimates from the 2003 SARS outbreak, the serial interval can be assumed to be 5-7 days. Early epidemic reporting suggest that the spillover event occurred some time around 1 December 2019, meaning that 26 January was 56 days into the epidemic (allowing for a one-day reporting lag). So at a minimum, we can estimate $\mathcal{R}_0 \approx 1 + 5 \times \log(2000) = 1.68$. At the other end, we have $\mathcal{R}_0 \approx 1 + 7 \times \log(100000) = 2.44$. Note that the shorter serial interval, perhaps counter-intuitively, means that $\mathcal{R}_0$ is lower. This is because, for a given epidemic size, the number of infections has taken more generations to get to that size.

4.1 Doubling Time of an Epidemic

$\mathcal{R}_0$ also gives us information about the doubling time of an epidemic. This is a pretty intuitive statistic that can effectively communicate the potential severity of an outbreak. Denote the doubling time of the epidemic T_D, then

$$T_D = \frac{\log(2)V}{\mathcal{R}_0 - 1}. \tag{22}$$

An outbreak with $\mathcal{R}_0 = 2.5$ and a serial interval of $V = 5$ will double every 2.3 days.

We can also calculate the doubling time of an ongoing outbreak from case counts on two different days.

$$T_D = \frac{t_1 - t_0}{\log_2\left(\frac{N_1}{N_0}\right)} \tag{23}$$

CDC reports that there were 10 confirmed COVID-19 cases in the USA on 21 February 2020 and 41 nine days later on 1 March. This suggests a doubling time of 4.2 days for *confirmed cases*. Given the almost complete lack of testing at this point in the epidemic, this is probably quite a bit greater than the actual doubling time of actual infections.

5 Will An Epidemic Infect Everyone?

Will an epidemic, once it has taken off in a population, eventually infect everyone? In order to answer this question, we want to know how i changes with respect to the "fuel" for the epidemic, s.

We thus divide equation 4 by equation 3.

$$\frac{di}{ds} = -1 + \frac{\nu}{\beta s}$$

We solve this equation by first multiplying both sides by ds

$$di = (-1 + \frac{\nu}{\beta s})ds$$

We then integrate and do a little algebra, yielding

$$\log(s_\infty) = \mathcal{R}_0(s_\infty - 1) \tag{24}$$

This is an implicit equation for s_∞, the number of susceptibles at the end of the epidemic. When $\mathcal{R}_0 > 1$, this equation has exactly two roots, only one of which lies in the interval $(0,1)$. $\log(s_\infty) = 0$ (i.e., everyone remains susceptible since there is no epidemic) will always satisfy equation 24 (plug it in and see!). When $\mathcal{R}_0 > 1$, the other solution is the actual value of the final size. This is the one we really care about. If $\mathcal{R}_0 \leq 1$, the only value that satisfies equation 24 is $s_\infty = 1$. In words, at the end of the epidemic, everyone will still be susceptible (i.e., no one gets infected).

Figure 3 shows a geometric interpretation of equation 24 for various values of $\mathcal{R}_0 > 1$ in black. The point where the line $\mathcal{R}_0(s_\infty - 1)$ is tangent to the curve $\log(s_\infty)$ indicates the value of s_∞ for a given $\mathcal{R}_0$. As $\mathcal{R}_0$ gets larger, the final size of the epidemic gets larger as well (i.e., the fraction susceptible at the end of the epidemic gets smaller). Figure 3 also shows the solution when $\mathcal{R}_0 = 1$ in red. The only tangent to the curve $y = \log(s_\infty)$ occurs at $s_\infty = 1$, meaning that essentially none of the total population becomes infected when an infection is sub-critical.

Figure 4 plots the final size $(1 - s_\infty)$ for values of $\mathcal{R}_0$ ranging from 1-6. We can see that the curve is highly concave and approaches value of the whole population very slowly. We note the final size for $\mathcal{R}_0 = 2.5$, which is approximately the consensus estimate for the ongoing COVID-19 epidemic. The fraction remaining susceptible at the end of the epidemic is $s_\infty = 0.107$, so 89.3% of the population will ultimately become infected (assuming nothing changes).

The conclusion we can draw from all this analysis is that, in general, a fraction of the population will escape infection. That is, $s_\infty < 1$. This is one of the fundamental insights of mathematical theory of epidemics. It is very important to remember that final-size calculation assume that nothing changes in the epidemic. Public-health interventions (e.g., social distancing, quarantine, vaccination) and medical interventions (e.g., curing sick people) will reduce the effective reproduction number (i.e., the expected number of secondary cases produced by a case at some point later in the epidemic). This is indeed the goal of epidemic control!

6   Optimal Virulence: Pathogen Life History Evolution

But enough about you, let's talk about me for a while. It's instructive to think about epidemics from the pathogen's perspective. Pathogens bear biological information in their nucleic acids. This information varies from one copy of a pathogen to another, and the ability of a pathogen to persist and multiply can be a function of this variability. We therefore have fulfilled the necessary and sufficient conditions for natural selection. Pathogens evolve.

We will consider a model in which transmissibility and disease-induced mortality trade-off introduced by van Baalen and Sabelis (1995). This interaction is mediated by virulence, which we will take to be proportional to parasite burden or parasitemia. More copies of a virus (say) means that *conditional on contact* with an infected individual, the pathogen is more likely to be transmitted. However, more viral copies means the host is sicker – and potentially dead. Dead hosts do not transmit and very sick hosts are less likely to be up and interacting with susceptibles.

Consider a directly-transmitted infection from which there is no recovery (e.g., Herpes Simplex). The population experiences a baseline mortality rate, μ, and a disease-induced mortality δ.

$$ds/dt \;\; = \;\; -\beta si - \mu s \tag{25}$$

$$di/dt \;\; = \;\; \beta si - (\mu + \delta)i, \tag{26}$$

It is straightforward to show that the basic reproduction number is given by:

$$\mathcal{R}_0 = \frac{\beta}{\mu + \delta} > 1, \tag{27}$$

9

The parameter μ is independent of the epidemic, but the parameters, β and δ can conceivably be functions of virulence, which we denote by x.

An Evolutionary Stable Strategy (ESS) is a phenotype that can not be invaded by a rare mutant. Loosely speaking, it represents the optimal phenotype. The ESS virulence occurs where $d\mathcal{R}_0/dx = 0$.

Differentiate 27 with respect to x using the quotient rule:

$$\frac{d\mathcal{R}_0}{dx} = \frac{\beta'(\mu + \delta) - \delta'\beta}{(\mu + \delta)^2} = 0$$

Rearranging and evaluating $\beta(x)$ and $\delta(x)$ at the ESS values of x (denoted x^*), we have:

$$\frac{d\beta(x)}{d\delta(x)} = \frac{\beta(x^*)}{\mu + \delta(x^*)} \tag{28}$$

This result has a nice geometric interpretation. The ESS virulence occurs where a line rooted at the origin is tangent to the curve that relates β to δ. This result is known as the Marginal Value Theorem and has applications in economics and ecology as well as epidemiology. The MVT model for optimal virulence is plotted in figure 5. In the lower curve, the tangent line hits further out on the horizontal axis and mortality is higher.

Frank (1996) has a very thorough review of models of the evolution of virulence.

7 Malaria Models and Their Many Quantities

Human Blood Index. Entomological Inoculation Rate. Vectorial Capacity. Human Biting Rate. Malaria models are confusing. The complexity of this vector-borne disease means that mathematical models describing it have many parameters and such a wide variety of people have worked on malaria models over such a long time (Sir Ronald Ross received the Nobel Prize in Medicine in 1911 for the first malaria model) that the menagerie of parameters and derived quantities can be daunting to the student trying to understand malaria transmission biology and control.

Smith and McKenzie (2004) have written a nice paper in which they summarize this welter of information. More recently, Mandal et al. (2011) have presented a comprehensive review of different malaria models that have been employed over the years. In these notes, I will follow the notation of Smith and McKenzie (2004), which has become approximately standard.

C "Vectorial Capacity": Transmission potential of a mosquito population in the absence of *Plasmodium*. Expected number of humans infected per infected human per day, assuming perfect transmission efficiency (i.e., $b = c = 1$). $C = \frac{ma^2 e^{-gn}}{g} = \frac{ma^2 p^n}{-\log(p)}$.

EIR "Entomological Inoculation Rate." $EIR = maZ = \varepsilon SZ = \varepsilon\beta = mg\beta = \frac{ma^2 cXe^{-gn}}{g + acX}$.

HBI "Human Blood Index," proportion of mosquitoes that have ever fed on a human $HBI = a/(a + g)$.

HBR "Human Biting Rate." $HBR = ma = \varepsilon S$.

IC Individual Vectorial Capacity: the expected number of infectious bites from a single vector after feeding on an infectious host. $IC = cP_eS$.

P_e Probability of surviving the length of the *Plasmodium* incubation period. $P_e = e^{-gn}$.

Q Proportion of mosquito blood meals on humans.

S "Stability Index," the expected number of mosquito bites on humans over a mosquito's lifetime. $S = a/g$. Because g is constant and lifespan is therefore exponentially distributed, S is also the number of bites after a mosquito becomes infected.

V Slope of β with respect to X evaluated at $X = 0$ (i.e., zero prevalence in humans). This value describes the total contribution to vectorial capacity of a single mosquito over its lifetime.

X Proportion of humans who are infectious.

Y Proportion of infected mosquitoes. $Y = \frac{acX}{g+acX}$.

Z Proportion of mosquitoes infectious ("Sporozite Rate"). $Z = \frac{acX}{g+acX}e^{-gn} = YP_e$.

β Lifetime transmission potential. $\beta = \frac{a^2bcXe^{-gn}}{g(g+acX)}$.

ε Rate of mosquito emergence per human per day.

$\eta(A)$ Proportion of surviving mosquitoes age A that have ever bitten a human. $\eta(A) = 1 - e^{-aA}$.

$\lambda(A)$ Proportion of a mosquito cohort surviving to age A. $\lambda(A) = e^{-gA}$

$\mu(A)$ Proportion of mosquitoes age A that are infectious. $\mu(A) = 1 - e^{-acX(A-n)}$.

$\nu(A)$ Proportion of mosquitoes age A that ever become infected. $\nu(A) = 1 - e^{-acXA}$. Note that acX is the rate at which mosquitoes become infected.

a Expected number of bites on humans per mosquito. $a = Qf$.

b Transmission efficiency from infectious mosquito to susceptible human.

c Probability of infection of an uninfected mosquito by biting an infectious human.

f Mosquito feeding rate. $1/f$ is then the interval between blood meals.

g Force of mortality (i.e., instantaneous mortality rate) for mosquitoes. $g = -\log(p)$

m Equilibrium mosquito density per human $m = \varepsilon/g$.

n Length of incubation period.

p Probability of a mosquito surviving 1 day. $p = e^{-g}$.

7.1 Some Derived Quantities

It is first worthwhile reviewing expectation and weighted averages of continuous variables, since these are used in many of the derived measures in malaria models.

Say we have a random variable X which can only take on values greater than or equal to zero (as most continuously distributed quantities of biological interest are). The expected value of X is given by

$$\mathbb{E}(X) = \int_0^\infty x f(x) dx,$$

where $f(x)$ is the probability density function of x. In words, you sum (i.e., integrate) the product of all possible values of X and the probability of observing a particular value of X. By definition,

$$\int_0^\infty f(x) dx = 1.$$

Say that we want to weight an average by something that does not sum to unity. We can force our weighting function to behave like a probability by dividing by its sum:

$$\bar{x}_w = \frac{\int_0^\infty x w(x) dx}{\int_0^\infty w(x) dx},$$

where $\bar{x}_w$ stands for a weighted average of X and $w(x)$ is a weight function.

The Human Blood Index (HBI) is a weighted average of $\eta(A)$, the proportion of mosquitoes age A that have ever bitten a human. We weight the $\eta(A)$ values by the survivorship of mosquitoes to age A.

$$HBI = \frac{\int_0^\infty \eta(A)\lambda(A) dA}{\int_0^\infty \lambda(A) dA} = \frac{a}{a+g}. \tag{29}$$

The proportion of infected mosquitoes, Y, is the survival-weighted average of the age-specific fraction infected $\nu(A)$

$$Y = \frac{\int_0^\infty \nu(A)\lambda(A) dA}{\int_0^\infty \lambda(A) dA} = \frac{acX}{g+acX}. \tag{30}$$

The proportion of mosquitoes that are infectious Z is a survival-weighted average of the age-specific proportion infectious $\mu(A)$:

$$Z = \frac{\int_0^\infty \mu(A)\lambda(A) dA}{\int_0^\infty \lambda(A) dA} = YP_e = \frac{acX}{g+acX} e^{-gn}. \tag{31}$$

The lifetime transmission potential is simply survival-weighted sum of the product of transmission efficiency b, expected number of bites on humans per mosquito a, and the age-specific proportion of mosquitoes infected $\mu(A)$. Note that this is not a weighted average, but is simply a sum:

$$\beta = \int_0^\infty ba\mu(A)\lambda(A) dA = \frac{a^2 bcX e^{-gn}}{g(g+acX)} \tag{32}$$

7.2   So What is the Vectorial Capacity Anyway?

Vectorial capacity, C, is a measure which is essentially independent of the prevalence of *Plasmodium* infection. It represents the transmission potential of a mosquito population. Vectorial capacity is very similar to $\mathcal{R}_0$ in one crucial way. It represents the expected number of humans infected per infected human per day (assuming perfect transmission) in a completely susceptible human population, $X = 0$. Once we put in transmissibility (b and c) and the duration of infectiousness r, we have a measure directly analogous to $\mathcal{R}_0$.

Vectorial capacity is the product of three quantities: (1) the emergence rate of mosquitoes per human per day, ε, (2) the squared number of mosquito bites on humans after mosquitoes become infectious (the stability index, S), and (3) the probability of a mosquito surviving the *Plasmodium* incubation period, P_e. The stability index is squared because it takes two bites to transit malaria: the one that infects the mosquito and the one that infects the person.

$$
\begin{align}
C &= \varepsilon S^2 P_e \tag{33} \\
&= mg\left(\frac{a}{g}\right)^2 e^{-gn} \tag{34} \\
&= \frac{ma^2 e^{-gn}}{g} = \frac{ma^2 p^n}{-\log(p)} \tag{35}
\end{align}
$$

These latter two expressions are both encountered in the literature (remember that $p = e^{-g}$ and $g = -\log(p)$).

It is worth noting that this last form is often referred to as the *Ross Equation* or the *Ross-MacDonald Equation*, and is commonly used for vector-borne diseases. Smith et al. (2012) note that because it lacks a canonical form, it is probably more accurate to refer to the Ross-MacDonald *family* of equations.

7.3 $\mathcal{R}_0$ for a Simple Malaria Model

The simplest model for malaria is as follows:

$$
\begin{align}
\dot{X} &= mabY(1 - X) - rX \tag{36} \\
\dot{Y} &= acX(e^{-gn} - Y) - gY \tag{37}
\end{align}
$$

where a is the expected number of bites on humans per mosquito, m is equilibrium mosquito density per human, n is length of incubation period, g is the daily force of mosquito mortality, b is the transmission efficiency from infectious mosquito to susceptible human, c is the probability of infection of an uninfected mosquito by biting an infectious human, X is the proportion of infected humans, Y is the proportion of infectious mosquitoes.

We can calculate $\mathcal{R}_0$ from the next generation matrix:

$$
F = \begin{pmatrix} 0 & abm \\ ace^{-gn} & 0 \end{pmatrix}
$$

and

$$V = \begin{pmatrix} -r & 0 \\ 0 & -g \end{pmatrix}$$

making

$$FV^{-1} = \begin{pmatrix} 0 & -\frac{abm}{g} \\ -\frac{ace^{-gn}}{r} & 0 \end{pmatrix}$$

The dominant eigenvalue of FV^{-1} is

$$\mathcal{R}_0 = \sqrt{\frac{ma^2 e^{-gn} bc}{rg}}$$

It is important to note that Smith and McKenzie (2004), Smith et al. (2005), and Smith et al. (2007) use $\mathcal{R}_0^2$, though the call it $\mathcal{R}_0$. Either way, its threshold behavior $\mathcal{R}_0 > 1$ remains the same.

The derivation of $\mathcal{R}_0$ here is done for only the simplest model. Mandal et al. (2011) present a variety of different malaria models and their associated values of $\mathcal{R}_0$. Smith et al. (2007) present a series of more realistic formulae for $\mathcal{R}_0$ which allow for heterogeneous biting, immunity, and finite population effects, all features of real environments that can substantially change $\mathcal{R}_0$.

8 Epidemics in Multi-Host Communities

8.1 A Graphical Approach

Bob Holt and colleagues present a very important heuristic framework for thinking about the persistence of pathogens in multi-species communities (Holt et al. 2003).

All of the following figures depict zero-growth isoclines for the pathogen. plot the mix of species at which $\mathcal{R}_0 = 1$. The interior of the space inscribed by the isocline represents the space where the pathogen can not be maintained in the community. is a central technique in theoretical ecology. It is also widely used in microeconomics, where such curves are typically called "indifference curves."

The basic theory for this work is given in Tilman's classic volume (Tilman 1982). All but the last of the isoclines were anticipated by Tilman in his work on competition and the structure of ecological communities.

The basic model – the "noninteraction" case (fig. 6) – is one in which pathogen persistence is predicated on a critical threshold of either species 1 or species 2. The pathogen goes extinct only if *both* species are under their critical densities (N_1 and N_2).

In the "weakly interacting" case (fig. 7), a mix of species makes pathogen persistence more likely than a monoculture of either species. That is, a small number of S_1 can substitute, albeit inefficiently, for S_2 being below threshold. (and vice-versa)

The "substitutable" case (fig. 8) characterizes a community where S_1 and S_2 efficiently substitute for each other. A constant ratio of substitutability (the slope of the line) applies at

all host densities. In the special case of perfect substitutability applies when a single S_1 can be substituted for a singe S_2 with regard to maintaining pathogen persistence.

It between-host transmission is more efficient than within-host transmission, the isocline bends inward (fig. 9). Thus, in the "complementary" case, pathogen persistence is much more likely in a multi-host community than in single host populations.

For vector-borne pathogens with complex life cycles, passage through an intermediate host is obligate for the perpetuation of the transmission cycle. Frequently, passage through the ultimate host is also obligate. Elimination of either intermediate or ultimate hosts from the community will lead to pathogen extinction. Thus, in the "alternating" case (fig. 10), a critical threshold exists for one or both species. As long as both host species co-exist above their minimum critical densities, the presence of a mix of both hosts makes pathogen persistence more efficient – this is why the isocline bends inward.

The "inhibitory" case (fig. 11) is unusual, as the slope of the isocline is *positive*. For the given plot, a critical threshold of S_1 is required for pathogen persistence. The presence of *any* of S_2 means that there must be more S_1. Pathogen persistence will be much less likely in multi-host communities.

Holt et al. (2003) note that this particular isocline does not appear in Tilman's original typology for competitive communities. There are some very interesting applications in infectious disease ecology that we will discuss in class. The first is the case of intact mammalian communities in Eastern North American woodlands and their diluting effect on Lyme disease transmission (LoGiudice et al. 2003). The second is discussed by Cohen and Gürtler (2001), in which the presence of peri-domestic chickens acts as a sink for Chagas' disease transmission.

Multi-host communities change our expectations regarding the pathogen evolution. Specialist pathogens are expected to evolve toward intermediate virulence (Lenski and May 1994), as discussed in the van Baalen and Sabelis (1995) paper cited in section 6. For generalist pathogens, hosts which are not essential for pathogen fitness will not exert a sufficient selective force to push the pathogen toward reduced virulence. This is why some zoonoses are so pathogenic. For example, *Echinococcus multilocularis* is a cestatode worm enzootic in Central European foxes. When it spills over into human populations, the case fatality rate can exceed 98%!

References

Anderson, R., G. Medley, R. M. May, and A. Johnson (1986). A preliminary study of the transmission dynamics of the Human Immunodeficiency Virus (HIV), the causative agent of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology 3*, 229–263.

Anderson, R. M. and R. M. May (1991). *Infectious diseases of humans: Dynamics and control.* Oxford: Oxford University Press.

Cohen, J. and R. Gürtler (2001). Modeling Household Transmission of American Trypanosomiasis. *Science 293*, 694–698.

Diekmann, O. and H. Heesterbeek (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* New York: Wiley.

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology 28*(4), 365–382.

Frank, S. A. (1996). Models of parasite virulence. *Quarterly Review of Biology 71*(1), 37–78.

Heesterbeek, J. A. P. (2002). A brief history of R_0 and a recipe for its calculation. *Acta Biotheoretica 50*(3), 189–204.

Heffernan, J., R. Smith, and L. Wahl (2005). Perspectives on the basic reproduction ratio. *Journal of the Royal Society Interface 2*, 281–293.

Hethcote, H. W. and J. A. Yorke (1984). *Gonorrhea: Transmission dynamics and control*, Volume 56 of *Lecture Notes in Biomathematics*. Berlin: Springer-Verlag.

Holt, R. D., A. P. Dobson, M. Begon, R. G. Bowers, and E. M. Schauber (2003). Parasite establishment in host communities. *Ecology Letters 6*(9), 837–842.

Lenski, R. E. and R. M. May (1994). The Evolution of Virulence in Parasites and Pathogens: Reconciliation between 2 Competing Hypotheses. *Journal of Theoretical Biology 169*(3), 253–265.

LoGiudice, K., R. S. Ostfeld, K. A. Schmidt, and F. Keesing (2003). The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. *Proceedings of the National Academy of Sciences, USA 100*(2), 567–571.

Mandal, S., R. Sarkar, and S. Sinha (2011). Mathematical models of malaria: A review. *Malaria Journal 10*(1), 202.

Smith, D. L., K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie (2012). Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLOS Pathogens 8*(4), e1002588.

Smith, D. L., J. Dushoff, R. Smow, and S. I. Hay (2005). The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature 483*, 492–495.

Smith, D. L. and F. E. McKenzie (2004). Statics and dynamics of malaria infection in *Anopheles* mosquitoes. *Malaria Journal 3*(1), 13.

Smith, D. L., F. E. McKenzie, R. W. Snow, and S. I. Hay (2007). Revisiting the Basic Reproductive Number for Malaria and Its Implications for Malaria Control. *PloS Biology 5*(3), 0531–0542.

Tilman, D. (1982). *Resource competition and community structure*. Princeton: Princeton University Press.

van Baalen, M. and M. Sabelis (1995). The dynamics of multiple infection and the evolution of virulence. *American Naturalist 146*(6), 881–910.

van den Driessche, P. and J. Watmough (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences 180*(1), 29–48.
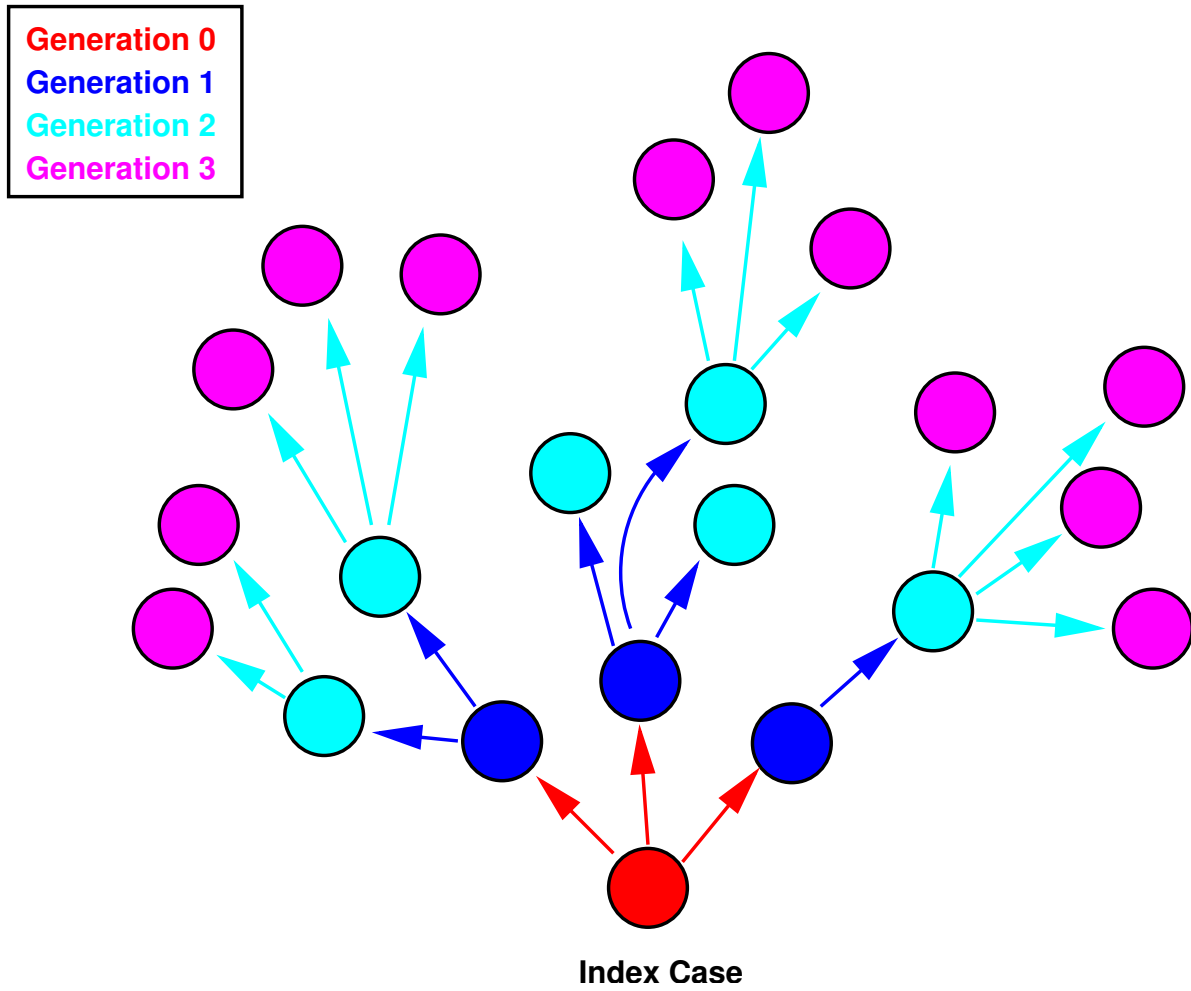
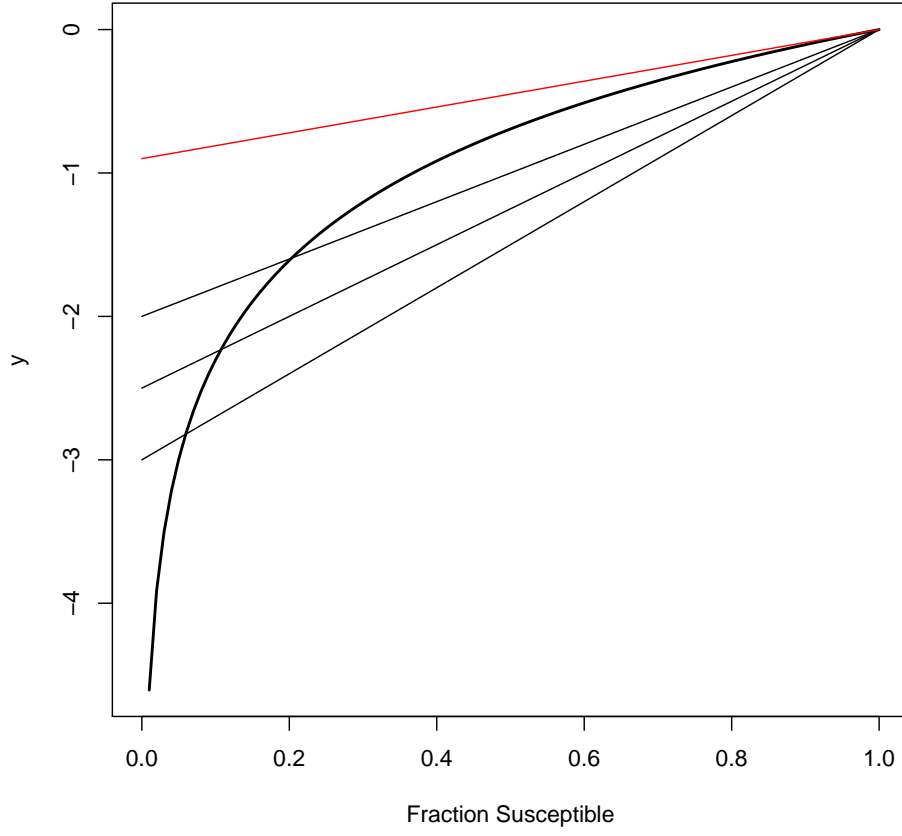Figure 2: Graphical depiction of "generations" in an epidemic.

Figure 3: Geometric interpretation of equation 24. The fraction remaining susceptible at the end of the epidemic occurs where the line $\mathcal{R}_0(s_\infty - 1)$ intersects the curve $\log(s_\infty)$. The four lines from the bottom are for $\mathcal{R}_0 = 2$, $\mathcal{R}_0 = 2.5$, $\mathcal{R}_0 = 2$ and $\mathcal{R}_0 = 1$ (in red). For $\mathcal{R}_0 > 1$, there is an interior solution. For $\mathcal{R}_0 \leq 1$ the only solution is at $s_\infty = 1$.
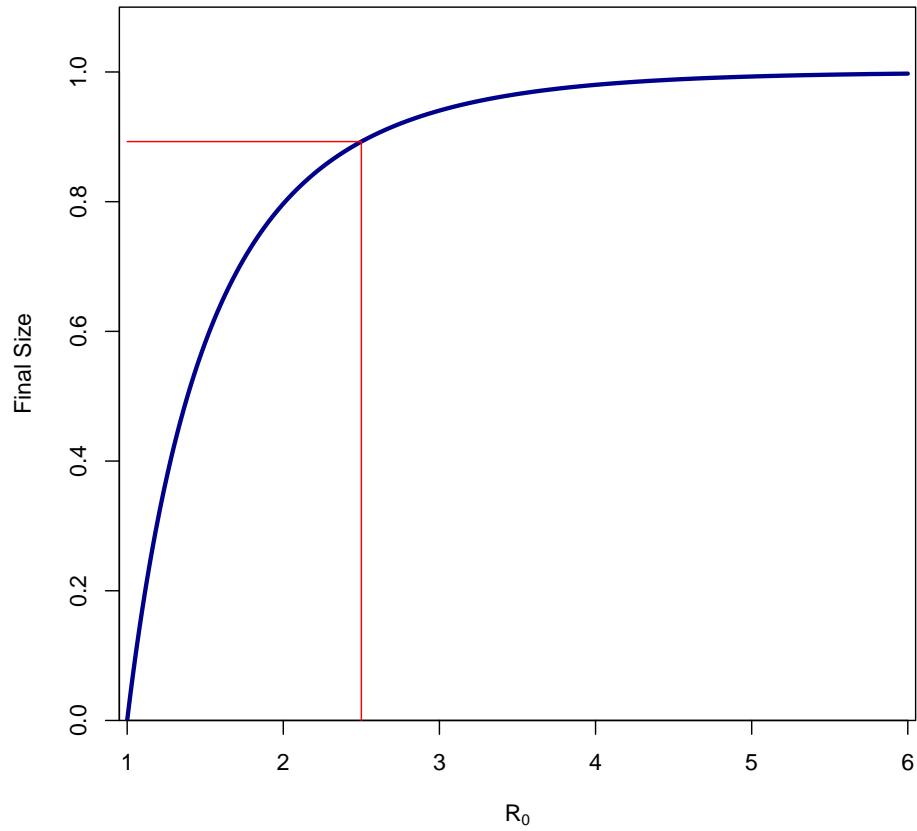
Figure 4: A plot of the final size of an epidemic with values of $\mathcal{R}_0$ ranging from $\mathcal{R}_0 = 1$ to $\mathcal{R}_0 = 6$ Thin red lines indicate the final size for an epidemic with $\mathcal{R}_0 = 2.5$, the more-or-less consensus estimate of the reproduction number for COVID-19. The corresponding final size is 89.3% of the population infected.
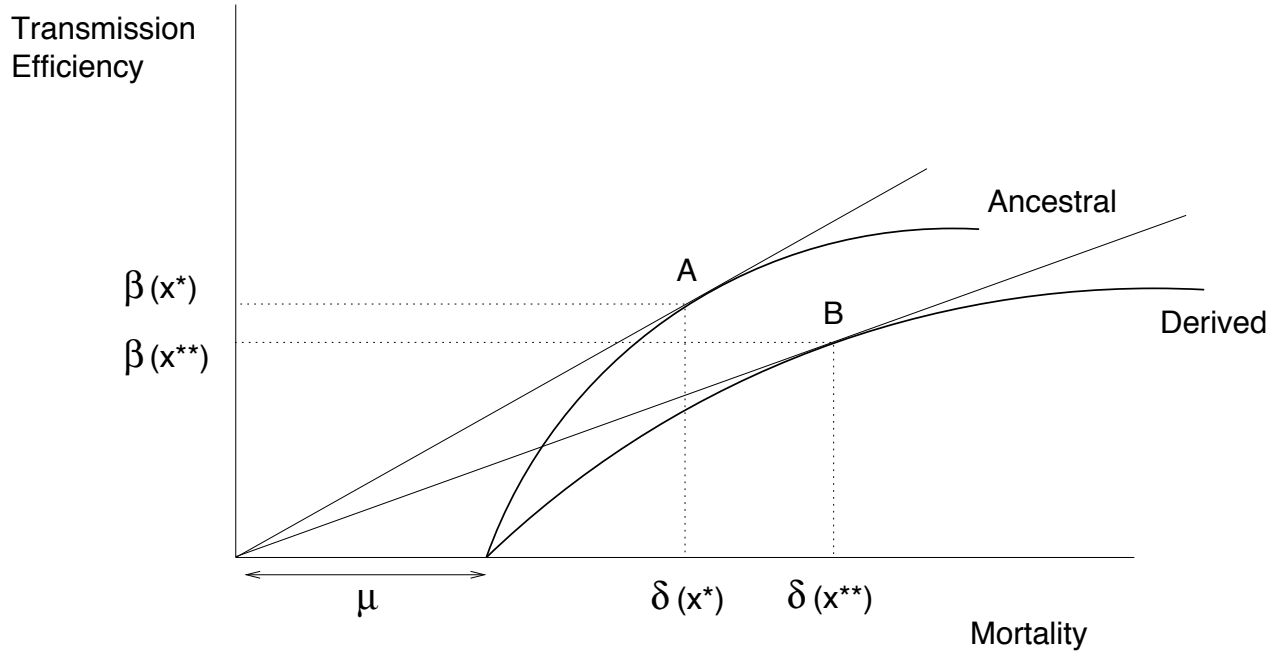
Figure 5: Marginal value theorem for optimal virulence. The ESS virulence occurs where a line rooted at the origin is tangent to the curve that relates β to δ. Two curves are depicted. The first curve shows a pathogen in which transmissibility increases relatively rapidly with mortality. Point A indicates the optimal balance between $\beta(x)$ and $\delta(x)$ under this case, and the optimal virulence is indicated x^*. For the second curve, relative transmission is less efficient. Therefore, the tangent line from the origin to the curve hits further out (B) along the mortality axis and the optimal virulence is higher (x^{**}).
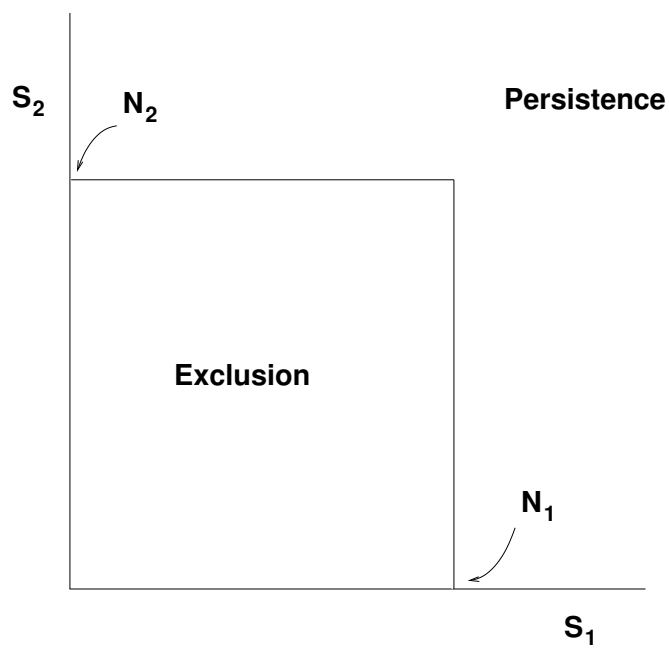
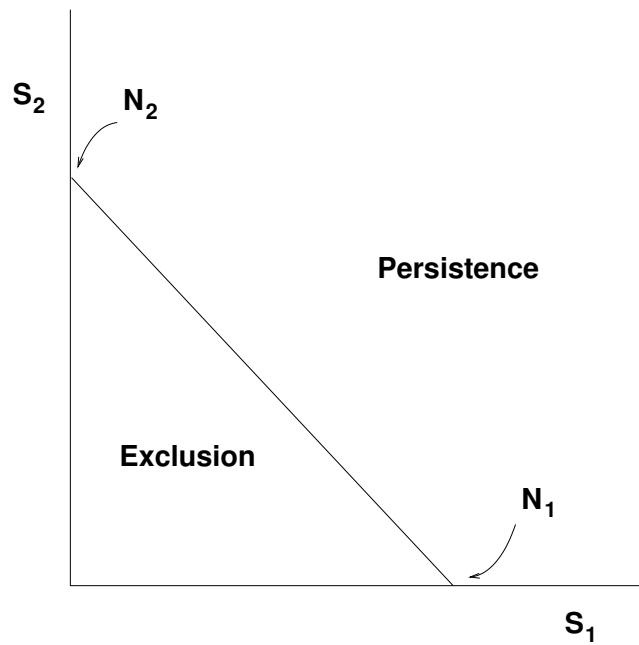Figure 6: Noninteracting hosts
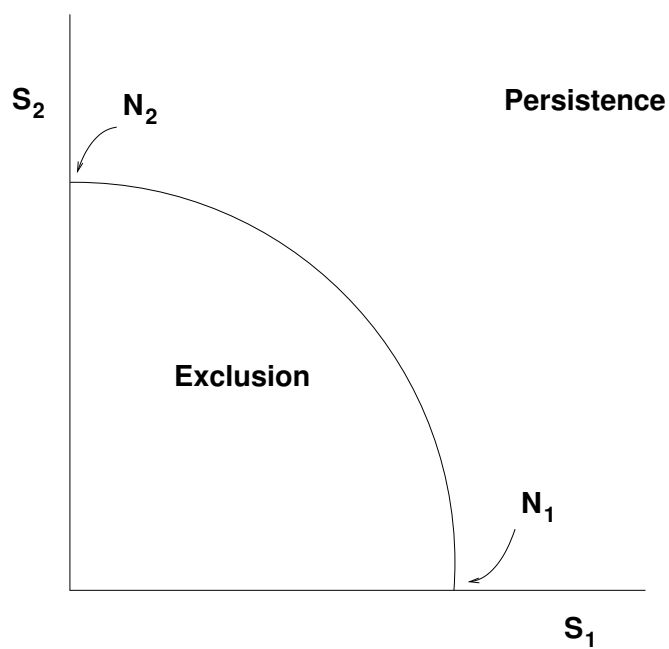
Figure 7: Weakly interacting hosts
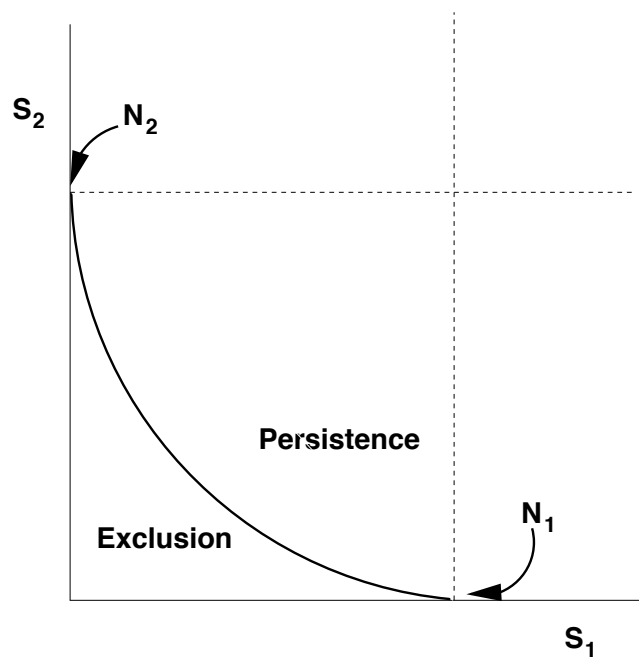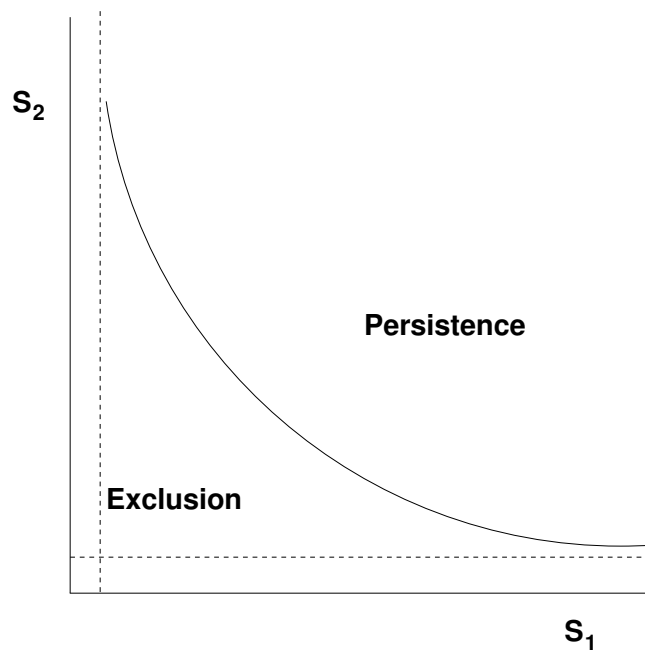
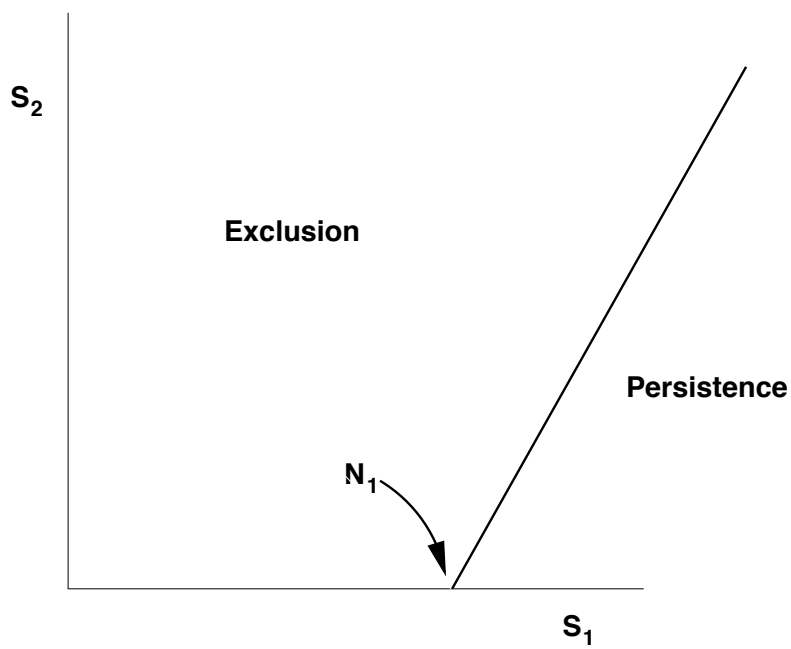Figure 8: Substitutable hosts

Figure 9: Complementary hosts

Figure 10: Alternating hosts

Figure 11: Inhibitory hosts