

Mathematical Modeling of Infectious Disease

Guest Lecture, PH253B

23 February 2009

Travis C. Porco

FI Proctor Foundation for Research in Ophthalmology

UCSF

Scope and role of modeling

In the most general sense, we may consider modeling as an effort to understand one aspect of the world by its similarity to something else. For instance, researchers may develop animal models of the pathogenesis of various disease agents. Studying the disease process in the animal model may then lead to insights that can be developed further. To be useful, the animal model must resemble the disease in humans in ways the researchers consider to be important. To be believable, the results from animal models must be tested.

Another type of model is the statistical model, which attempts to describe relevant aspects of data or a data generation process. Such models frequently do not attempt to describe the mechanism yielding the data, but rather form a basis for determining, for instance, if an estimated difference between a treatment and control group could have plausibly resulted from chance. Statistical models based on real data play an important and vital role in all areas of epidemiology and lead to essential insights. However, when such models are not based on a representation of the underlying medical or epidemiologic processes, such models may not be generalizable beyond the circumstances under which the data were collected.

While statistical models play a fundamental role in research, including research in analytical epidemiology, models based on representing the mechanism of disease transmission have played a role in epidemiology for decades. Some of the earliest mathematical models were developed by Sir Ronald Ross, who showed that mosquitoes were vectors of the malaria plasmodium. Ross wished to describe the factors responsible for spread and dissemination of the plasmodium in a quantitative way, and published papers describing what was called *a priori pathometry*.

Mathematical models are attempts to gain insight into the processes of disease transmission and persistence using mathematical representations of the mechanisms of disease transmission. Models may be used to address a variety of theoretical and practical questions; these applications may include forecasting for planning, designing interventions, or simply improving our understanding. For instance, modelers may wish to estimate the number of hospital beds needed to prepare for SARS, or to estimate how many deaths of hospitalizations might result from pandemic influenza. Modelers may wish to determine whether quarantine is warranted during an influenza outbreak, or whether ring vaccination would be sufficient to control bioterrorist smallpox. As examples of improving understanding, modelers may wish to estimate the contribution of superspreaders to the invasion of a pathogen into a new region, estimate the best tradeoff of virulence and transmissibility a pathogen should seek if public health control measures change, or to determine how the mortality rate may affect the number of people who are ultimately infected by a pathogen. In general, the most important strength of mathematical models of the disease process is the ability to explore counterfactual scenarios or conditions for which no data is available, for instance, to examine the consequences of

untested control strategies or the spread of a novel pathogen.

Mathematical models are most credible when the mathematical analogy which constitutes the representation are clear and plausible—when the model representation resembles the epidemic in ways the researchers consider most important. Understanding what the most important features of a disease transmission process are leads modelers at times into vigorous debate, frequently manifesting a tension between the need for insight through elegant simplicity on the one hand, and realism through increased detail on the other. Finally, note that as is the case with other types of models in science, the insights gained from modeling must ultimately be tested in some way.

In this brief lecture, we will discuss some key concepts and simple models that have been used in mathematical epidemiology, and discuss how researchers applied some of these principles to an emerging pathogen.

The Reed-Frost Model

A classic model of infectious disease transmission was developed during the 1930s by Lowell J. Reed and Wade Hampton Frost of Johns Hopkins. Because the model is simple to explain and provides valuable insights, we will discuss it at this time.

In the classical Reed-Frost model, we assume a fixed population of size N . At each time, there are a certain number of cases of disease, C , and a certain number of susceptibles, S . We assume each case is infectious for a fixed length of time, and ignore the latent period; when individuals recover, we assume that they are immune to further infection. During the infectious period of each case, we assume that susceptibles may be infected, so that the disease may propagate further. This constitutes an idealized, or abstract, model, exhibiting some features of an epidemic system.

Because we assume a fixed length infectious period and neglect the latent period, the generations of infection stay separate. At the beginning, we have only the generation of cases that starts the disease transmission. After the recovery of this generation, the new cases that resulted from transmission constitute the second generation of cases. These, in turn, recover, but may give rise to a third generation. Let C_1 be the number of cases in the first generation, and S_1 be the number of susceptibles that the first generation may place at risk of new infection. Similarly, let C_2 be the number of cases in the second generation, and S_2 the number of susceptibles present for the second generation of cases to potentially expose to disease; in general, the number of cases at generation t is denoted C_t and there are S_t susceptibles at that time.

The basic Reed-Frost model assumes homogeneity of risk of infection throughout the population. In particular, we assume that each susceptible has a risk p of being infected by any of the infectives in the population. In a more realistic model, we might assume that this probability depends on the population size. We might also assume that each susceptible does not have the same risk of being infected by each infective.

Assuming, however, that each susceptible has a risk p of being infected by each infective, we can then find the probability distribution of the number of cases in the second generation. In other words, we can find the probability that the number of cases in the second generation takes various values. For instance, if none of the susceptibles gets infected, we would have no cases in the second generation; we could determine the probability that C_2 is zero, i.e. $P(C_2 = 0)$.

If each susceptible has a risk p of being infected by each infective, what is the chance that a particular susceptible will be infected? Since there are C_1 infectives, a sufficient exposure from any infective would cause the susceptible to become infected. There are thus many ways to be infected—a susceptible could be infected by the first infective, or the second, or the third, etc., or may receive a sufficient exposure from more than one infective or even from all. On the other hand, there is only one way to escape infection, and that is to escape infection (to fail to receive a sufficient exposure) from all the infectives in the population.

The next assumption we will make for the basic Reed-Frost model is that the exposures are independent. Each infective constitutes an independent risk for each susceptible, and whether any susceptible is infected or not is independent of all the other susceptibles. With this assumption, we can compute the chance of escaping infection. Considering one particular susceptible individual, this susceptible has a chance $1 - p$ of escaping infection from the first case. There is also a chance $1 - p$ of escaping infection from the second case, and so on through all C_1 cases we have at the beginning. Since we must determine the chance of escaping infection from the first and from the second and so forth, we may use independence to compute this by multiplying the probabilities. The probability of escaping infection from the first and second individuals is the probability of escaping infection from the first, times the probability of escaping infection from the second, and so forth. Thus, the chance of escaping infection from all susceptibles is $(1 - p) \times (1 - p) \times \cdots \times (1 - p)$, where there are C_1 terms being multiplied. This is simply $(1 - p)^{C_1}$ for the probability of escaping infection. Thus, the probability of being infected is simply $1 - (1 - p)^{C_1}$.

As a brief digression, models of this form, known as *binomial risk models*, are frequently used to analyze data for many real diseases. For instance, the application of binomial risk models to data from the San Francisco Men’s Health Study led to the estimate that there is a 10% chance, per partnership, of the transmission of the HIV virus from an infected partner to an uninfected partner, for MSMs in San Francisco in the 1980s. Such models have been used to analyze HIV transmission data per act, rather than per partnership, to estimate potential declines in HIV infectivity due to the widespread use of antiretrovirals, or to analyze the cost-effectiveness of HIV prevention interventions in sub-Saharan Africa.

Returning to the Reed-Frost model, we have shown that the probability that any of the S_1 susceptibles will be infected is $1 - (1 - p)^{C_1}$. We must now determine the probability distribution of the number of new cases, C_2 . Mathematically, we may consider each susceptible to be a *Bernoulli trial*, a random experiment with two outcomes, conventionally known as *success* and *failure*. Since these trials are independent of one another, and since the “success” probability is the same for each trial, we may use the *binomial distribution* to determine the probability distribution of the number of successes. For simplicity, let us denote the infection probability (“success” probability) by r ; we have shown that $r = 1 - (1 - p)^{C_1}$. Then the binomial probability distribution gives us the probability distribution of the number of cases in the next generation:

$$P(C_2 = x) = \binom{S_1}{x} r^x (1 - r)^{S_1 - x}.$$

Here, the notation $\binom{S_1}{x}$ is the number of ways to choose the supposed x new cases out of the S_1 number of susceptibles. The number of susceptibles at time 2 is simply $S_2 = S_1 - C_2$.

The same reasoning can be applied at each time. Of course, the risk r changes if the number of cases changes, so we may write the risk at time t as $r_t = 1 - (1 - p)^{C_t}$. Therefore, the probability distribution of

the number of cases at time t is

$$P(C_t = x) = \binom{S_{t-1}}{x} r_{t-1}^x (1 - r_{t-1})^{S_{t-1} - x},$$

and $S_t = S_{t-1} - C_t$.

Before exploring the behavior of this model further, observe that the number of cases over time is *random*. The randomness arises because of our assumptions about the nature of transmission. The quantities predicted by the model, i.e. C_t and S_t , are random variables whose distribution is specified by the model. In the same way that a statistician might model the height of a randomly selected person as having a particular distribution, say a normal distribution with some specific mean and variance. Just as the statistician may consider the height of a specific person as being a realization of this random distribution (a random variate drawn from the specified distribution), we may think of a particular realization of this process as leading to a particular number of cases and susceptibles over time. In general, even if our parameters p , N , and C_1 remain the same, the number of cases may be different, just as the height of a second person drawn from the same population may be different. The Reed-Frost model is an example of a *stochastic model*, and the sequence of numbers of cases and susceptibles constitutes a *stochastic process*.

Models similar to the Reed-Frost model have been analyzed carefully by mathematicians, and these models are called *chain binomial* models. Rather than explore the formal analysis, we will explore the dynamics of the Reed-Frost model using computer simulation. The exciting open-source statistical package R <http://www.r-project.org> will provide us with an excellent platform for such exploration.

We first write a *function* in R which provides us with a simulation of the Reed-Frost model. Without going into details of the programming language, the function does three things. First, it checks that the preconditions for the computation are met. For instance, a negative transmission probability is simply meaningless, so the function checks to make sure the transmission probability is not negative. The function checks other conditions as well, such as the requirement that the number of cases not be larger than the population size. Second, the function computes the risk for each susceptible as $1 - (1 - p)^{C_t}$, and uses the built-in random number generator for the binomial, called `rbinom`, to compute the random number of secondary cases. Finally, the function returns the results to us.

```
> reed.frost <- function(pp, nn, c1, t.end, cumul.only = FALSE) {
+   if (t.end > 1000) {
+     stop("t.end too big")
+   }
+   if (t.end <= 0) {
+     stop("negative or zero ending time")
+   }
+   if (c1 < 0 || abs(round(c1) - c1) > 1e-07) {
+     stop("invalid starting number of cases")
+   }
+   if (nn < 0 || abs(round(nn) - nn) > 1e-07) {
+     stop("invalid population size")
+   }
+ }
```

```

+   }
+   if (pp > 1 || pp < 0) {
+     stop("invalid transmission probability")
+   }
+   if (nn < c1) {
+     stop("more cases than people")
+   }
+   ss <- rep(0, t.end)
+   cc <- rep(0, t.end)
+   cumul <- 0
+   current.cc <- c1
+   current.ss <- nn - c1
+   if (!cumul.only) {
+     cc[1] <- current.cc
+     ss[1] <- current.ss
+   }
+   for (ii in 2:t.end) {
+     rr <- 1 - (1 - pp)^current.cc
+     current.cc.new <- rbinom(1, size = current.ss, prob = rr)
+     current.ss.new <- current.ss - current.cc.new
+     if (!cumul.only) {
+       ss[ii] <- current.ss.new
+       cc[ii] <- current.cc.new
+     }
+     current.ss <- current.ss.new
+     current.cc <- current.cc.new
+     cumul <- cumul + current.cc.new
+     if (current.cc.new == 0) {
+       if (!cumul.only) {
+         for (jj in (ii + 1):t.end) {
+           ss[jj] <- current.ss
+         }
+       }
+       break
+     }
+   }
+   if (cumul.only) {
+     cumul
+   }
+   else {

```

```

+       list(susc = ss, cases = cc, cumul.new.cases = cumul)
+     }
+ }
> reed.frost.average <- function(pp, nn, c1, t.end, nreps = 1, cumul.only = FALSE) {
+   if (cumul.only) {
+     cumul <- reed.frost(pp, nn, c1, t.end, cumul.only)
+     for (ii in 2:nreps) {
+       cumul <- cumul + reed.frost(pp, nn, c1, t.end, cumul.only)
+     }
+     cumul/nreps
+   }
+   else {
+     run <- reed.frost(pp, nn, c1, t.end)
+     ss <- run$susc
+     cc <- run$cases
+     for (ii in 2:nreps) {
+       run <- reed.frost(pp, nn, c1, t.end)
+       ss <- ss + run$susc
+       cc <- cc + run$cases
+     }
+     list(susc = ss/nreps, cases = cc/nreps)
+   }
+ }

```

Let us use this function to create a random epidemic. Assume that the population size is 100, and the transmission probability is 0.02. We will run this model five times, and plot them all.

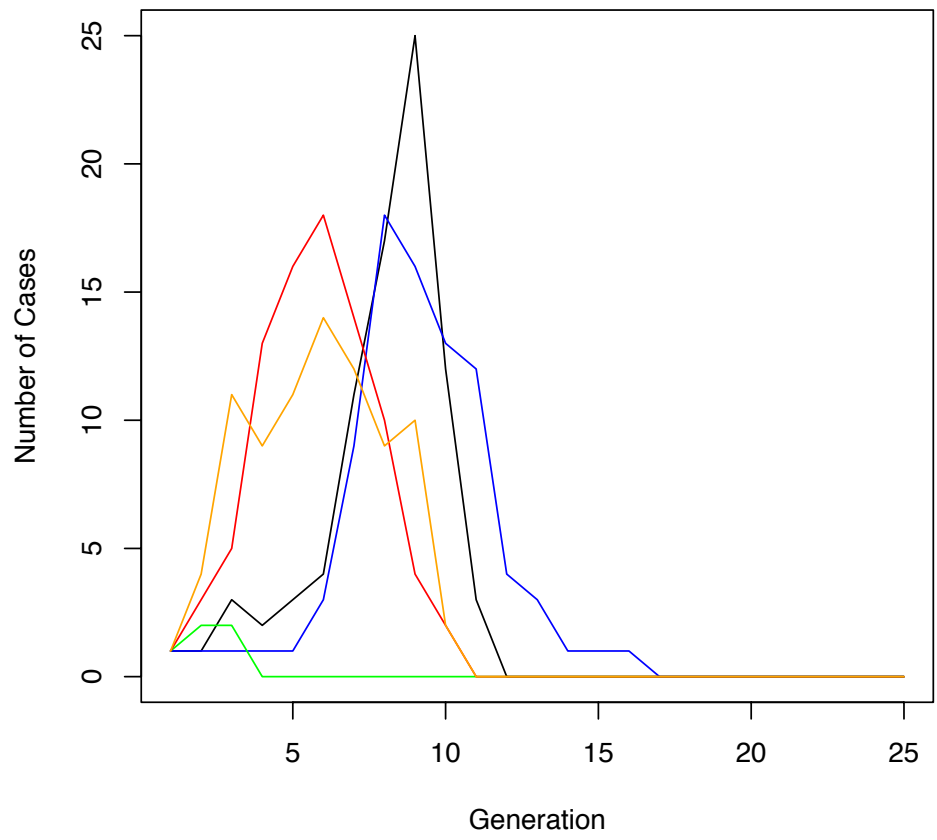
```

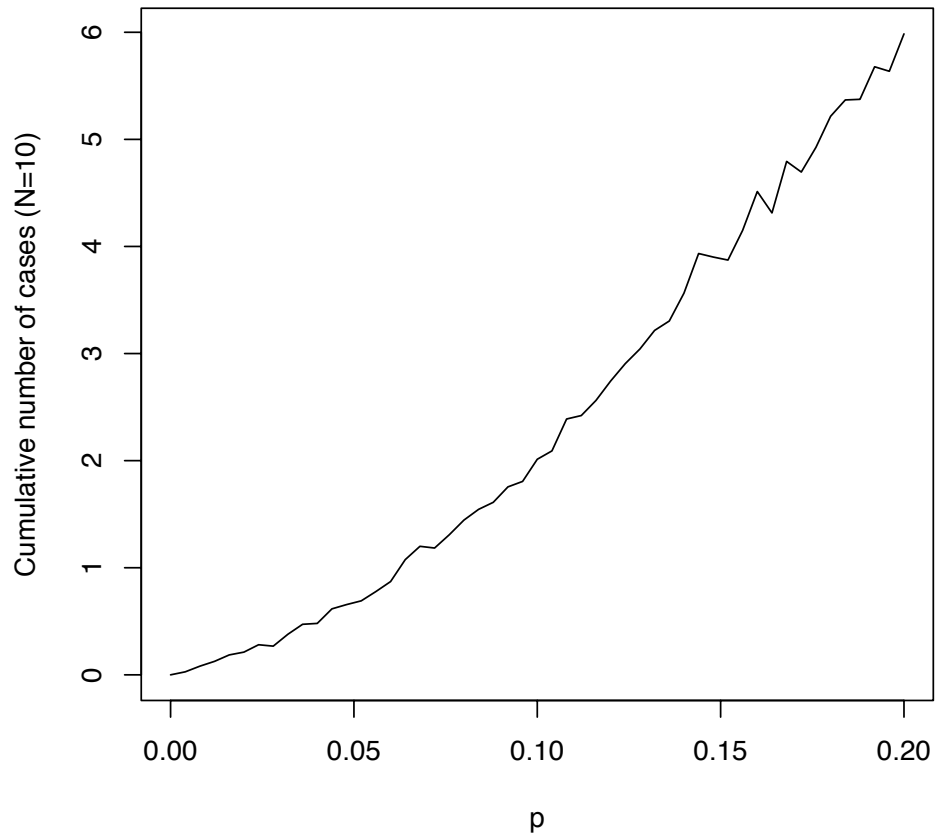
> end.time <- 100
> pp <- 0.02
> sim1 <- reed.frost(pp = pp, nn = 100, c1 = 1, t.end = end.time)
> sims <- list(sim1)
> tot.reps <- 5
> for (ii in 2:tot.reps) {
+   new.sim <- reed.frost(pp = pp, nn = 100, c1 = 1, t.end = end.time)
+   sims[[ii]] <- new.sim
+ }
> times <- 1:end.time

```

As the figure illustrates, each epidemic is somewhat different.

It will be interesting to plot the average number of cumulative cases for different values of the probability p . Let us begin for a small population, of size ten. We will average the results of 1000 repetitions for each value of the probability.





```

> nseps <- 50
> ps <- seq(0, 0.2, by = (0.2)/nseps)
> ans <- rep(0, length(ps))
> nreps <- 1000
> nn <- 10
> for (ii in 1:length(ps)) {
+   ans[ii] <- reed.frost.average(pp = ps[ii], nn = nn, c1 = 1, t.end = 50, nreps = nreps, cumul.only
+ }

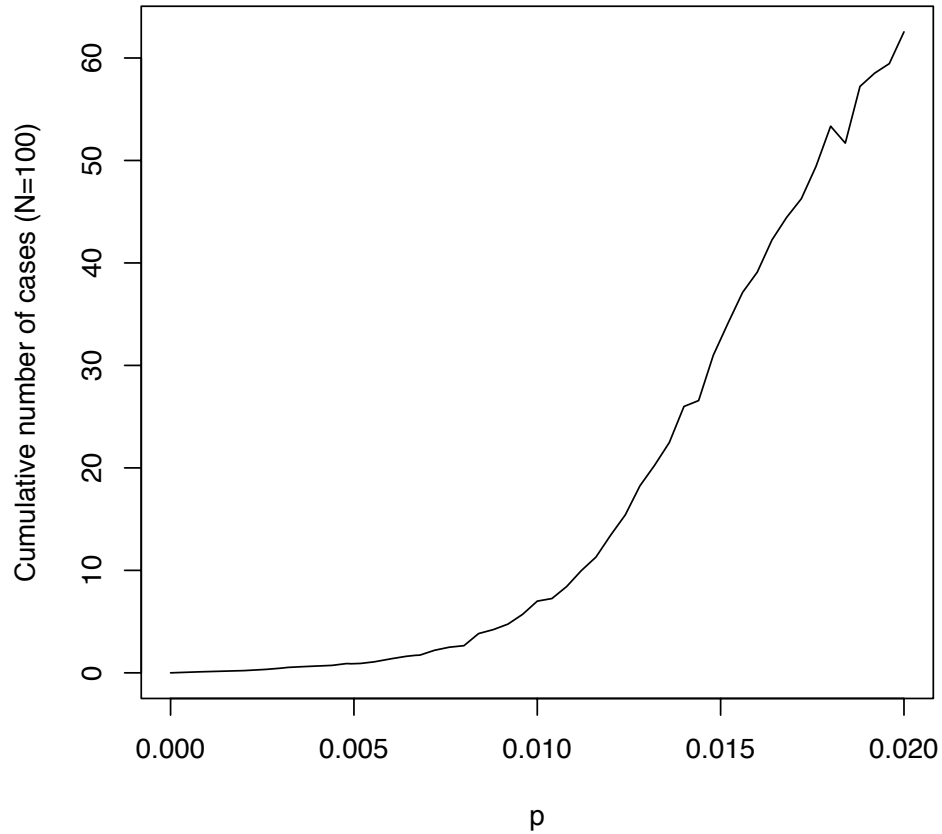
```

Next, we consider a population of size 100:

```

> ps <- seq(0, 0.02, by = (0.02)/nseps)
> ans <- rep(0, length(ps))
> nn <- 100

```

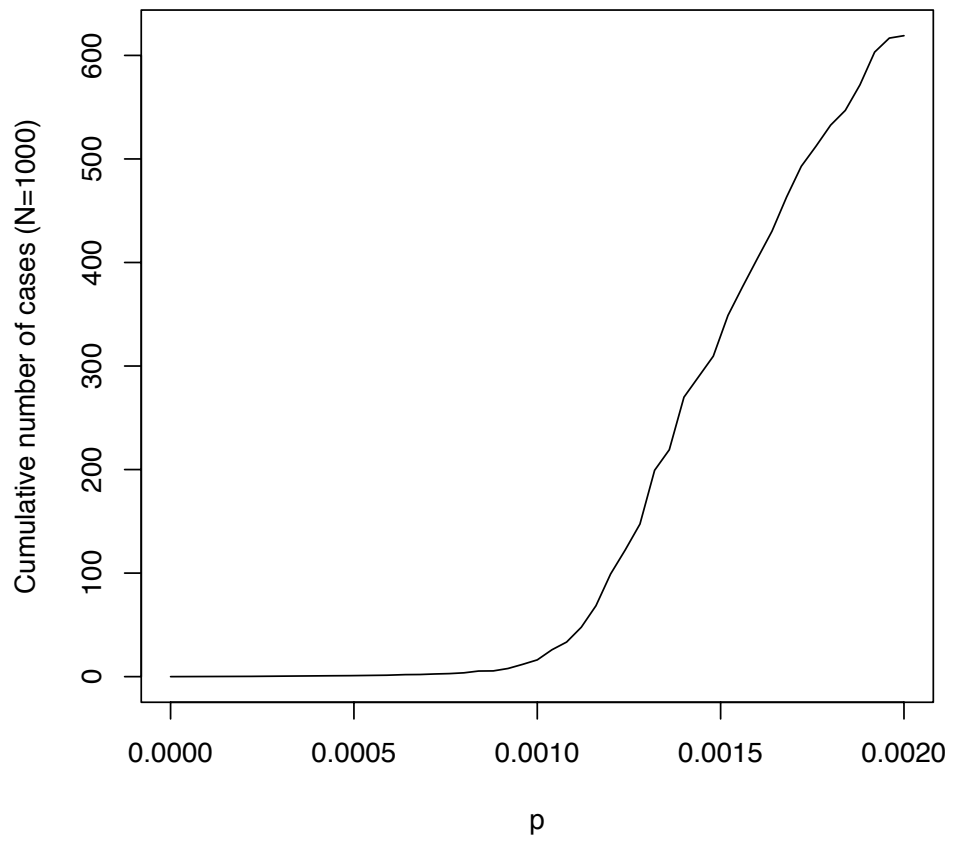


```
> for (ii in 1:length(ps)) {
+   ans[ii] <- reed.frost.average(pp = ps[ii], nn = nn, c1 = 1, t.end = 200, nreps = nreps, cumul.onl
+ }
```

For a population of size 1000:

```
> ps <- seq(0, 0.002, by = (0.002)/nseps)
> ans <- rep(0, length(ps))
> nn <- 1000
> for (ii in 1:length(ps)) {
+   ans[ii] <- reed.frost.average(pp = ps[ii], nn = nn, c1 = 1, t.end = 1000, nreps = nreps, cumul.onl
+ }
```

To help us understand these results, we will compute the expected number of cases that will result from



a single case at the beginning of the epidemic. At the beginning of the epidemic, we have S_1 susceptibles, and one single case. The risk that each susceptible has of becoming infected is $1 - (1 - p)^1 = p$.

For a binomial distribution with N trials and p the probability of success per trial, the expected value is Np . This is what the average of a very large number of repetitions should be close to. Intuitively, imagine that you have a very large population, much larger than any possible sample. Suppose the prevalence of a risk factor is 20% in this population. If you take a sample of size 100, you expect 20 people, or 100×0.2 , to have the risk factor.

For the Reed-Frost model, we expect the number of cases in the second generation to be S_1p . If $S_1p > 1$, we expect the epidemic to initially increase, and if $S_1p < 1$, we expect the initial case to not even, on average, replace itself. In this latter circumstance, we expect a small cluster of cases, perhaps, but no large-scale epidemic.

In the examples we examined, we saw fairly small numbers of cases when we were below the critical value of p . Above this value, we began to see a substantial fraction of the population begin, on average, to become infected.

The expected number of secondary cases at the beginning of an epidemic, when everyone is susceptible, is called the **basic reproduction number** or basic reproduction ratio or basic reproductive rate, and is usually denoted R_0 . For most epidemic models, we find that when the basic reproduction number is less than one, conditions do not favor epidemic spread when the disease is introduced, and conditions do not favor the endemic persistence of the disease.

Many authors reserve the term “basic reproduction number” to refer to a hypothetical population in which no disease control measures are present. When control measures are in place, the expected number of secondary cases an initial case can cause in a susceptible population is referred to by some other expression. Thus, if the basic reproduction number is greater than one, but the, say, realized reproduction number is less than one, the measures are sufficient to control a disease that would otherwise invade.

One way to understand the effect of a control measure is to consider a perfect vaccine administered to a large fraction of the population. Suppose that we are considering a disease for which the basic reproduction number is two. If we vaccinate, say, 80% of the population, then we have deprived the pathogen of most of its potential hosts. If we are assuming a homogeneously mixing model, then we may imagine that on average, only 20% of the contacts of the initial case are actually susceptible. The initial case may only produce 20% of the number of new cases that it would have produced without the vaccination program, or 0.4 on average. We expect the disease to not spread far in the population. Thus, it is conceivable that we could control a disease by vaccination even without complete coverage; the remaining individuals are said to be protected by *herd immunity*. However, it should be noted the crucial role of our assumption of homogeneity; we needed to conclude that if 80% of the population was vaccinated, that 80% of the contacts of a case are vaccinated. In practice, this is not often realized, and claims regarding herd immunity may need to be tempered with a careful consideration of how the vaccinated cases are distributed.

0.1 The Kermack-McKendrick Model

Models similar to the Reed-Frost model were developed based on chemical analogies, and analyzed by W. O. Kermack and A. G. McKendrick (and later extended by H. W. Hethcote, R. M. May, R. M. Anderson, and others). An interesting feature of such models is that, in the absence of renewal of susceptibles, epidemics die out when the number of susceptibles has become so low that on average each case can no longer replace itself. In the past, it had occasionally been argued that epidemics died out because the pathogen always lost its virulence during an epidemic. However, virulence changes are not necessary to end an epidemic. More general models include the renewal of the susceptible pool due to birth, and in such models, the renewal of susceptibles may eventually balance the rate of new infection, leading to sustained endemic equilibrium and the persistence of the infection.

In models of Kermack-McKendrick type, the basic reproduction number may be expressed as the rate at which a case produces new infections, times the duration of infectivity, times the chance a newly infected person will become a case. The basic reproduction number may also be computed for vector-borne diseases, sexually transmitted diseases, or for more complex epidemic structures involving mixing of subpopulations.

We may begin a discussion of such models by first considering a different, and simpler, process: exponential decay. Suppose we have a large population of individuals subject to a constant hazard μ of death. In other words, if dt is a small unit of time, the probability or risk of death is μdt in this small unit of time—given that you were alive at the beginning of the interval. If we had $N(t)$ living individuals at time t , then we expect that $N(t)\mu dt$ will die during the interval (just using the binomial distribution). So at the end of the interval, which is to say at time $t + dt$, we have how many individuals? It must be the number you had at the beginning, $N(t)$, minus the mortality: $N(t + dt) = N(t) - N(t)\mu dt$. We could write

$$N(t + dt) - N(t) = -N(t)\mu dt,$$

or

$$\frac{N(t + dt) - N(t)}{dt} = -N(t)\mu.$$

If we let dt be very small, or go to zero, we could then observe that

$$\lim_{dt \rightarrow 0^+} \frac{N(t + dt) - N(t)}{dt} = \frac{dN}{dt}(t).$$

Thus,

$$\frac{dN}{dt}(t) = -N(t)\mu.$$

This says the rate of decline of N is proportional to N itself. This is the equation for exponential decay, and it turns up in many different places than population biology. For instance, the decay of radioactive elements follows such a relationship.

What we have is a linear differential equation. We can solve it:

$$\frac{dN}{N} = -\mu dt.$$

We can integrate both sides from 0 to some T :

$$\int_{t=0}^{t=T} \frac{dN}{N} = - \int_0^T \mu dt.$$

to get

$$\log N(t) - \log N(0) = -\mu T,$$

so that

$$\log N(t)/N(0) = -\mu T.$$

We can exponentiate both sides:

$$N(t)/N(0) = e^{-\mu T},$$

or

$$N(T) = N(0)e^{-\mu T},$$

an exponential decay function. If we were to plot $N(T)$ versus T , we would be graphing the number surviving. Note that $N(T)/N(0)$ is the fraction surviving, or the survival curve. Because the survival function is a decaying exponential, we can also see that a constant hazard implies an exponential waiting time until death. The average waiting time is $1/\mu$. Of course, if we modeled any other process, such as recovery, where we had a constant hazard, we would also obtain exponential decay.

Let's now begin thinking about an epidemic model. Instead of the waiting time until death, suppose we considered the waiting time until infection. If we had an environmental infection, such as coccidiodomycosis, we could imagine that the force or hazard for infection was independent of the number of people who were infected. If we assumed the hazard were constant, we would then have an exponential waiting time until infection.

Now, if we were to begin thinking about epidemic models for transmissible infections, we can see that the more people have the infection, the greater you would expect the hazard of infection to be, just like in the Reed-Frost model. One simple assumption is to make the force of infection proportional to the number of infectives. Let's let X be the number of uninfected people, and Y be the number of infectives. Then, $\mu = \beta Y$. Because Y is changing over time, the force of infection is changing over time. Let's first consider an infection from which there is no recovery, such as HIV. We could begin with this equation for the number uninfected:

$$\frac{dX}{dt} = -\mu X,$$

and

$$\frac{dY}{dt} = \mu X.$$

We'll suppose there is no birth, immigration, death, or emigration, so the total number of people stays constant, and we'll call this constant N . So $N = X(t) + Y(t)$; X and Y change, but N stays fixed. If we

knew $X(t)$, we could just write $Y(t) = N - X(t)$, so we really only have one equation to worry about. Let's go ahead and substitute the hazard in:

$$\frac{dX}{dt} = -\beta X(t)Y(t) = -\beta X(t)(N - X(t)).$$

This is a nonlinear differential equation, but it can still be analytically solved, because it is *separable*:

$$\frac{dX}{X(N - X)} = -\beta dt.$$

Let's work on $1/(X(N - X))$. It turns out that we can break this up:

$$\frac{1}{X(N - X)} = \frac{A}{X} + \frac{B}{N - X} = \frac{A(N - X) + BX}{X(N - X)}.$$

If $A = B = 1/N$, this equality holds. So

$$\frac{1}{X(N - X)} = \frac{1}{NX} + \frac{1}{N(N - X)}.$$

So

$$\frac{dX}{X(N - X)} = \frac{dX}{NX} + \frac{dX}{N(N - X)} = -\beta dt.$$

Now we can integrate:

$$\int_{t=0}^{t=T} \frac{dX}{NX} + \int_{t=0}^{t=T} \frac{dX}{N(N - X)} = -\beta \int_0^T dt.$$

This gives

$$(1/N) \log X(T)/X(0) - (1/N) \log (N - X(T))/(N - X(0)) = -\beta T.$$

Simplifying:

$$\log \frac{X(T)(N - X(0))}{X(0)(N - X(T))} = -\beta NT.$$

Then

$$\frac{X(T)(N - X(0))}{X(0)(N - X(T))} = e^{-\beta NT}.$$

You can solve this for $X(T)$ and find

$$X(T) = \frac{NX(0)e^{-\beta NT}}{N - X(0) + X(0)e^{-\beta NT}}.$$

Then $Y = N - X(T)$ is

$$Y(T) = \frac{N(N - X(0))}{N - X(0) + X(0)e^{-\beta NT}}.$$

The prevalence fraction $y(T) = Y(T)/N$ is

$$y(T) = \frac{N - X(0)}{N - X(0) + X(0)e^{-\beta NT}}.$$

This is called a *logistic* or sigmoid curve. Does it make sense? First, look at what $X(T)$ is when $T = 0$. You get

$$\frac{NX(0)e^{-\beta N0}}{N - X(0) + X(0)e^{-\beta N0}} = \frac{NX(0)}{N - X(0) + X(0)} = X(0)$$

as it should. And as T is large, we have $X(T)$ going to zero—if you wait long enough, everyone gets the disease.

Now suppose that we begin with

$$\frac{dX}{dt} = -\beta XY.$$

but now, we assume that infected individuals can recover. Let's assume a constant recovery hazard ρ ; if you started with a closed cohort of infected individuals, the number still infected would be an exponential decay function just like before. But we actually have new people getting infected, so

$$\frac{dY}{dt} = \beta XY - \rho Y.$$

What happens to the recovered people? If they stay immune and never get infected again, we could write

$$\frac{dZ}{dt} = \rho Y$$

with $N = X + Y + Z$ is constant. This particular form is a special case of a model first analyzed by Kermack and McKendrick back in the 1920s. There is no clean analytic solution to these equations. But let's suppose that at the beginning, $Z = 0$, and $X \approx N$, with $Y(0)$ very small, i.e. $Y(0) \ll N$. At the very beginning, we have

$$\frac{dY}{dt} \approx (\beta N - \rho)Y.$$

As long as $Y \ll N$, we might write $k = \beta N - \rho$. If $\beta N - \rho < 0$, we just have exponential decay—there is no epidemic. The initial cases don't replace themselves, and the epidemic fizzles out and never takes off. But if $\beta N - \rho > 1$, we have exponential growth at the outset; this is just like exponential decay, but without the minus sign.

$$\frac{dY}{dt} = kY.$$

If we solve this, we get $Y(t) = Y(0)e^{kt}$, suggesting that the epidemic begins exponentially. But of course it can't stay this way. Eventually you run out of susceptibles as X declines, and eventually the growth rate of Y turns negative and the epidemic is over. Transmission continues even in the declining phase, but eventually the epidemic burns out leaving many susceptibles uninfected.

We can get some insight into what $\beta N - \rho$ means if we look at how many new infections an infective can cause over time. Suppose $Y = Y_0$ is some constant; somehow imagine we fix the number of infectives at this constant. Now if X is so large that it doesn't change much ($X \approx N$ at the beginning of the epidemic):

$$\frac{dX}{dt} = -\beta NY_0.$$

The incidence rate is just a constant. The total incidence rate is βNY_0 , so in time t you would infect $\beta NY_0 t$ new people, or $\beta N t$ per infective. Now, how long is an infective present on average? If you have an exponential waiting time before recovery, the average waiting time is $1/\rho$. So in this amount of time, you would infect

$$\beta N/\rho$$

people. The condition $\beta N - \rho > 0$ is equivalent to $\beta N/\rho > 1$. As long as each infective can generate on average more than one new infective at the beginning (when there are plenty of susceptibles), the epidemic can take off. The quantity $\beta N/\rho$ is called the basic reproduction ratio or the basic reproduction number of this epidemic. Different epidemic models lead to different expressions for R_0 , the basic reproduction number, but this behavior is a common feature of epidemic models.

Beyond simple models

Models of realistic diseases extend the Reed-Frost model in include many features that the infection is known to have. Such extensions may include a more complex representation of the natural history. No disease really has a fixed duration of infectivity, as we assumed in the Reed-Frost model. Rather, the duration of infectivity may follow a distribution that could be estimated, in principle, from data. Moreover, all diseases have a latent period, between the time of infection and the time of infectiousness. Infectiousness itself may vary over the course of the illness, and may precede the appearance of specific symptoms. Infections may also differ considerably from person to person, depending on nutritional status, prior immunity, age or other factors. Diseases may be more likely to be transmitted to individuals in the same household, or to other individuals in the same risk group. For some infections, host immunity effectively prevents ever having the same disease again, but for others, such protection may not be realized. Diseases may also be transmitted by different routes, and different methods may be required to model a vector borne disease, a water borne disease, a sexually transmitted disease, or an airborne disease.

Construction of convincing models requires such considerations be examined in detail. Frequently, simplifying assumptions are made in modeling, and sometimes these assumptions serve to enhance understanding with little quantitative effect. For instance, simplifying the shape of the incubation period may have a small effect on the predicted epidemic curve. On the other hand, if a model assumes that a certain vaccine is more effective than the data indicate, or that the disease is transmissible before it can be detected through symptoms, the results may markedly change. Effective critique of a model requires understanding of both the biology and epidemiology of the pathogen, as well as the sensitivity of the mathematical conclusions to the assumptions.

Concluding remarks

Epidemic transmission models provide a valuable way to gain insights into the nature of an epidemic. Mathematical models frequently depend on data that may be difficult to collect or validate. Construction of effective models requires collaboration between medical experts and modelers; effective critique of them requires an understanding of both the biomedical assumptions as well as the mathematical details. However, models remain a valuable tool for enhancing our understanding of epidemic mechanisms, and are perhaps

most valuable in examining counterfactual scenarios for disease control, where no data are yet available. When an emerging pathogen threatens to cause an epidemic, it is unlikely that a controlled trial of different intervention strategies will be available in time.