AN ENDEMIC DISEASE MODEL AND ITS APPLICATION TO MALARIA
Charles S. Peskin
April 9, 2012

In these notes we describe a generic endemic disease model, which we then elaborate into a model for the evolution of the prevalence of the sickle-cell hemoglobin gene in the presence of malaria.

# 1   Endemic Disease Model

We consider a chronic disease, from which recovery does not occur. The population thus involves two categories of people, susceptible and infected. We track only females, but make the assumption that there are equal numbers of males and females in each of the two categories. Thus

$$S = \text{number of susceptible females}$$
$$I = \text{number of infected females}$$
$$N = S + I = \text{total number of females}$$

All three of the above variables are functions of time.

The parameters of the model are as follows (see also further explanation and more precise definition below):

$$a = \text{infectivity of the disease}$$
$$\beta = \text{female birth rate per susceptible female}$$
$$\beta' = \text{female birth rate per infected female}$$
$$\delta = \text{death rate for a susceptible person}$$
$$\delta' = \text{death rate for an infected person}$$

The "infectivity" $a$ is defined by the statement that $a(I/N)$ is the probability per unit time that a given susceptible person becomes infected. The birth rate $\beta$ is the probability per unit time that a given susceptible female in the population will give birth to a female offspring. The birth rate $\beta'$ is the probability per unit time that an infected female will give birth to a female offspring. This raises the question whether the offspring of an infected female will be born infected. Although this is certainly possible, we assume here that it is not the case. Thus, in the present model, everyone is born susceptible. The death rate $\delta$ is the probability per unit

time that a given susceptible person dies, and the death rate $\delta'$ is the probability per unit time that a given infected person dies.

Note that $2\beta/\delta$ is the expected number of offspring that a female will have in her lifetime, counting both male and female offspring, assuming that she remains susceptible, i.e. uninfected, for her entire life. Similarly, $2\beta'/\delta'$ is the expected number of offspring that an infected female will have in her lifetime, assuming that she becomes infected so early that she is effectively infected for her entire life.

The following inequalities are important:

$$\beta > \beta' \quad \text{(infection lowers fertility)} \tag{1}$$
$$\delta < \delta' \quad \text{(infection increases death rate)} \tag{2}$$
$$\beta > \delta \quad \text{(with no infection, population grows)} \tag{3}$$
$$\beta' < \delta' \quad \text{(with everyone infected, population declines)} \tag{4}$$

The equations of the model are as follows:

$$\frac{dS}{dt} = -a\left(\frac{I}{N}\right)S + \beta S + \beta' I - \delta S \tag{5}$$

$$\frac{dI}{dt} = +a\left(\frac{I}{N}\right)S - \delta' I \tag{6}$$

$$N = S + I \tag{7}$$

The following properties of the model can be verified by analysis or simulation, and will simply be stated here.

First, regardless of initial conditions, the long-time behavior of the model is that the population settles into exponential growth or decay, with some fixed proportion of infected people. That is

$$S(t) \sim S_0 \exp(\lambda t) \tag{8}$$
$$I(t) \sim I_0 \exp(\lambda t) \tag{9}$$
$$N(t) \sim N_0 \exp(\lambda t) \tag{10}$$

where

$$N_0 = S_0 + I_0 \tag{11}$$

It is important to note that $S_0$, $I_0$, and $N_0$ are not initial conditions, but simply constants that describe the longterm behavior of the system. The growth rate $\lambda$ may be positive or negative (or zero in borderline cases).

It is quite remarkable that exponential behavior of this kind emerges in a nonlinear model. The fundamental reason is that the equations, despite being nonlinear, are scale-invariant. If we have a solution $(S(t), I(t), N(t))$, we can multiply all three variables by any constant, and the result will again be a solution.

The second property of the model is that there are two critical values of the infectivity $a$, which we denote here by $a_0$ and $a_1$ with $0 < a_0 < a_1$. These critical values are functions of the other parameters of the model, that is, they depend on $\beta$, $\beta'$, $\delta$, and $\delta'$.

When $a < a_0$, the only longterm exponential solution of the type described above has $I_0 = 0$. Thus, the infectivity is too low for the disease to become endemic in the population. No matter what its initial prevalence (infected fraction of the population), the disease will be gradually be "diluted out" of the growing population. The absolute number of cases may continue to grow, but the fraction of the population that is infected declines and approaches zero as time increases. The longterm growth rate of the population is therefore the maximum possible, namely $\beta - \delta$, since in the long run essentially nobody is infected.

When $a_0 < a$, on the other hand, the disease does become endemic, and the prevalence settles down to a fixed nonzero value as time increases. The growth rate $\lambda$ also settles down to a fixed value, which may be positive or negative (or zero). This is where the critical value $a_1$ comes in. When $a_0 < a < a_1$, the growth rate $\lambda > 0$, so the population grows, albeit at a slower rate than $\beta - \delta$. When $a_1 < a$, on the other hand, $\lambda < 0$, which means that the population is headed for extinction. In the limit $a \to \infty$ the steady-state prevalence of the disease approaches 1, and $\lambda \to (\beta' - \delta')$, which is as negative as the growth rate can be.

It should be noted that only in the middle range of infectivity, namely $a_0 < a < a_1$, is it possible for there to be a mutually successful outcome, both for the pathogen that causes the infection and also for the host population. If the infectivity is too low, the pathogen fails to become endemic in the population, and if the infectivity is too high the pathogen kills off the host population, and in the process destroys itself. Thus it is important that $a_0 < a_1$, so that there is a nontrivial interval of infectivity in which endemic disease can occur. In the present model, this turns out to be a consequence of the nonzero fertility of infected females. If we set $\beta' = 0$, then we find that $a_0 = a_1$ which means that there is only one value of the infectivity at which a mutually successful outcome is possible for both pathogen and host. In practice, such a fine-tuning of infectivity would be impossible for nature to achieve, so the outcome would either be the extinction of the host or the diluting out of the pathogen.

## 2 Malaria and the Evolution of the Prevalence of the Sickle-Cell Hemeoglobin Gene

The gene for sickle-cell hemeoglobin has the peculiar feature that it confers resistance to malaria but also causes a potentially fatal condition, sickle-cell anemia. Moreover, one copy of the sickle-cell gene is enough to confer resistance to malaria, but two are needed to cause sickle-cell anemia. That is, the gene is dominant with respect to malaria resistance but recessive with respect to its cause of sickle cell anemia. This means that natural selection will operate on the sickle-cell gene both in a positive and in a negative way, with the strength of the positive selection depending on the prevalence of malaria.

The purpose of this section is to describe a model in which one can observe the evolution of the prevalence of the sickle-cell gene in response to the above selective pressures. The model is an elaboration of the endemic disease model presented above. The genetic details will make the model of this section look much more complicated than the previous one, but remarkably, it has exactly the same parameters, since all of the genetic effects follow unambiguously from Mendelian laws and do not therefore introduce any additional parameters. To achieve this, however, we have to idealize the model in various ways that will become clear as we proceed.

We use the symbol "0" to represent the wild-type hemeoglobin gene, and "1" to represent the sickle-cell hemeoglobin gene. The genotypes that will appear in our model are therefore "00" and "01" (which is the same as "10"). The genotype "11" does not appear explicitly because it represents the phenotype of sickle-cell anemia, which we regard as fatal. We are concerned here with evolution that happened over the long span of human history before the intervention of modern medicine. For the same reason, we do not consider the medical treatment of malaria.

The possible matings that can occur in the population, and the probabilities of

the resulting genotypes are as follows:

$$00 \times 00 \rightarrow 00 \text{ with probability } 1$$

$$00 \times 01 \rightarrow 00 \text{ with probability } 1/2$$
$$\rightarrow 01 \text{ with probability } 1/2$$

$$01 \times 01 \rightarrow 00 \text{ with probability } 1/4$$
$$\rightarrow 01 \text{ with probability } 1/2$$
$$\rightarrow 11 \text{ with probability } 1/4$$

We do not show any matings involving 11 as the genotype of a parent, since we are making the assumption that genotype 11 (sickle-cell anemia) is fatal.

People with genotype 00 can be in one of two categories: susceptible, or infected with malaria. People of genotype 01=10 are in a different category that we shall call "resistant". In our model, they cannot get malaria, but they are otherwise just like susceptibles, i.e., their birth and death rates are the same as those of the susceptibles. As in the simpler endemic disease model of the foregoing section, we track only females, but we assume that there are equal numbers of males as females in each category. This will be a valid assumption if the male/female birth ratio is 1:1; if malaria affects males and females equally except for any effect on fertility, which we assume operates on the females only; and if the population is large enough for differences that arise by chance between the numbers of males and females in a category are small enough to be negligible.

Thus, the variables of the model are

$$S = \text{number of susceptible females (genotype 00)}$$
$$I = \text{number of infected females (genotype 00)}$$
$$R = \text{number of resistant females (genotype 01)}$$
$$N = S + I + R = \text{number of females in the population}$$

All four of these variables are functions of time.

Although we use the letter "R" for resistant, this should not be confused with the "R" in the "SIR" model of an epidemic. In particular, in the present model, it is not possible to move from the infected category to the resistant category, as the people in these categories have different genotypes. Also, biologically, the

resistance to malaria that is conferred by the sickle-cell gene has nothing to do with the immune system and is not the result of having experienced malaria.

The processes that cause changes in the numbers of people in the different categories are infection, death, and birth. We discuss each of these in turn (saving birth for last because it is by far the most complicated of the three).

We model infection exactly as in the endemic disease model of the previous section. That is, the probability per unit time that a given susceptible person becomes infected is $a(I/N)$ where $a$ is the "infectivity" of malaria. This is, of course, a drastic oversimplification of the biology of malaria. In fact, malaria is not directly contagious from one person to another, and has a complicated life cycle involving the mosquito and its larvae as intermediate hosts. Nevertheless, if we view the mosquito population merely as the means by which the malaria parasite gets from one human to another, then it does not seem completely unreasonable to use a generic infection model as a first, crude approximation. The way that mosquitos appear in our model, then, is simply through their influence on the infectivity parameter $a$. In warm, wet climates favorable to the reproduction of mosquitos, $a$ is larger than in cooler, drier places. Similarly, public health measures such as the use of mosquito netting can drastically reduce the infectivity parameter $a$. This is one of the most important tools in the ongoing efforts to control malaria worldwide.

The death rates in our model are $\delta$ for susceptible people and also for resistant people, and $\delta'$ for infected people. As before, $\delta$ is the probability per unit time that a given susceptible or resistant person will die, and $\delta'$ is the probability per unit time that a given infected person will die.

Similarly, the female birth rates in our model are $\beta$ for susceptible and also for resistant females, and $\beta'$ for infected females. Thus, $\beta$ is the probability per unit time that a susceptible or resistant female will give birth to a female child, and $\beta'$ is the probability per unit time that an infected female will give birth to a female child. We assume in this case that the child is born uninfected, and hence either susceptible or resistant, depending on her genotype.

The complication regarding birth concerns the genotype of the offspring, which depends not only on that of the mother but also on that of the father. We assume here random mating, so that for each birth a father is chosen at random from the population, without regard to genotype or infection status. Now the probability that a randomly chosen male is of genotype 00 is $(S + I)/N$, and the probability that a randomly chosen male is of genotype 01 is $R/N$. Note that these two probabilities add up to 1, since $S + I + R = N$.

Thus, if a mother is of genotype 00, her offspring are of genotype 00 with

probability

$$P_{00,00} = \frac{S+I}{N}(1) + \frac{R}{N}\left(\frac{1}{2}\right) \tag{12}$$

since 1 is the probability that the offspring has genotype 00 when the mother and father both have genotype 00, whereas $1/2$ is the probability that the offspring has genotype 00 when the mother has genotype 00 and the father has genotype 01.

Similarly, if the mother has genotype 00, her offspring are of genotype 01 with probability

$$P_{00,01} = \frac{R}{N}\left(\frac{1}{2}\right) \tag{13}$$

since in this case the only way that the offspring can be of type 01 is if the father is of type 01, and then the probability of that outcome is 1/2. Note that

$$P_{00,00} + P_{00,01} = 1 \tag{14}$$

since 00 and 01 are the only possible offspring genotypes when the genotype of the mother is 00.

In exactly the same way, we can calculate the probabilities that the offspring will be of type 00, 01, and 11, when the genotype of the mother is 01. These are given by

$$P_{01,00} = \frac{S+I}{N}\left(\frac{1}{2}\right) + \frac{R}{N}\left(\frac{1}{4}\right) \tag{15}$$

$$P_{01,01} = \frac{S+I}{N}\left(\frac{1}{2}\right) + \frac{R}{N}\left(\frac{1}{2}\right) \tag{16}$$

$$P_{01,11} = \qquad\qquad + \frac{R}{N}\left(\frac{1}{4}\right) \tag{17}$$

Note that the fractions that multiply $(S+I)/N$ in the above three equations add up to 1, and likewise the fractions that multiply $R/N$ in the above three equations add up to 1. Therefore,

$$P_{01,00} + P_{01,01} + P_{01,11} = \frac{S+I}{N} + \frac{R}{N} = 1 \tag{18}$$

With the help of the above probabilities we can now say that the number of susceptible female births per unit time is given by

$$B_{\mathrm{S}} = (\beta S + \beta' I)P_{00,00} + \beta R P_{01,00} \tag{19}$$

7

and the number of resistant female births per unit time is given by

$$B_R = (\beta S + \beta' I)P_{00,01} + \beta R P_{01,01} \tag{20}$$

Finally, the number of females born with sickle-cell anemia per unit time is

$$B_{\text{fatal}} = \beta R P_{01,11} \tag{21}$$

Note that no one is born infected with malaria in our model.

Finally, we can write the differential equations of the model as follows:

$$\frac{dS}{dt} = B_S - \delta S - a(I/N)S \tag{22}$$

$$\frac{dI}{dt} = \qquad -\delta' I + a(I/N)S \tag{23}$$

$$\frac{dR}{dt} = B_R - \delta R \tag{24}$$

$$N = S + I + R \tag{25}$$

As advertised, the parameters are exactly the same as in the simple endemic disease model of the previous section, and we assume that they obey the same inequalities.

Like the simpler endemic disease model, this malaria model is scale invariant. Its longterm behavior is that all solutions settle down to exponential solutions of the form

$$S(t) \sim S_0 \exp(\lambda t) \tag{26}$$

$$I(t) \sim I_0 \exp(\lambda t) \tag{27}$$

$$R(t) \sim R_0 \exp(\lambda t) \tag{28}$$

$$N(t) \sim N_0 \exp(\lambda t) \tag{29}$$

where

$$N_0 = S_0 + I_0 + R_0 \tag{30}$$

Recall that $S_0$, $I_0$, $R_0$, and $N_0$ are not initial data, but simply constants that characterize the large-time behavior.

As before, there is a threshold value of $a$ denoted $a_0$. In fact, $a_0$ depends on the parameters of the model in exactly the same way as before. When $a < a_0$, no matter what the initial conditions, malaria cannot become endemic — it is diluted out of the population. Here, however, we have a new phenomenon, which is that

8

the sickle-cell gene is diluted out of the population as well. Thus, with $a < a_0$ the longterm behavior has $I_0 = R_0 = 0$, and therefore $\lambda = \beta - \delta > 0$.

When $a > a_0$, malaria, and with it the sickle-cell gene, can become endemic in the population. Here, however, it is not necessarily the case that $\lambda$ eventually becomes negative for $a$ sufficiently large. Because of the protective effect of the sickle-cell gene, it is possible for the growth rate of the population to remain positive no matter how large the infectivity of malaria may be. See whether you can find a sufficient condition on the parameters $\beta$ and $\delta$ which will indeed insure the survival of humanity no matter how great the threat from malaria. Interpret your condition in terms of the number of children a typical woman has in a lifetime.