

MODELING DISEASE

FINAL REPORT

5/21/2010

SARAH DEL CIELLO, JAKE CLEMENTI, AND NAILAH HART

ABSTRACT

This paper models the progression of a disease through a set population using differential equations. Two cases were examined. Model 1 models an asymptomatic disease within a static population while accounting for a constant recovery without the chance of immunity. The equation for model 1 was first order, separable, and linear. Its resulting logistic solution entailed a scenario where the disease had no chance of dying out. Model 2 is a more realistic yet complicated model; it modeled a disease being transmitted through a population with both asymptomatic and symptomatic carriers while accounting for a constant recovery rate, constant death rate from both the infected and healthy population, and constant birth rate within the healthy population. The model for model 2 resulted in a system of nonlinear homogeneous equations that we were unable to solve explicitly. Instead their solutions were graphed.

PROBLEM STATEMENT

We seek to model the transmission of a disease through a population. Such modeling is very important to the study of epidemiology and the practice of medicine, since examining the relative effect of factors that govern the spread of a disease can help communities and health workers better prepare for and combat an outbreak.

Every disease has certain characteristics that affect the manner in which it spreads through a population. For instance, the virulence of a disease affects the chance that the disease will be transmitted from a sick person to a healthy person when they meet one another. That is, diseases with high virulence could be called very contagious. In addition, the recovery time will affect how many people the infected individual can come into contact with before he or she is healthy again. If a disease is contagious for a long time, it will obviously affect proportionally more people. Some diseases have incubation periods, where the carrier will not exhibit symptoms but is still able to transmit it to uninfected individuals. If a disease has a significant incubation period, it would be very difficult to detect and quarantine infected individuals. For this reason, the disease would likely spread much faster than it would if all carriers exhibited symptoms. In addition, diseases with very clear symptoms, such as vomiting and diarrhea in the case of e. bola, are likely to spread much slower than diseases with subtle symptoms. For some diseases, such as chicken pox, people who recover develop immunity and are no longer susceptible to the disease. Our goal is to create a model that takes as many of these different characteristics into account as the scope of this project allows.

MODEL DESIGN 1

We begin with a very simple model that gives us a starting point in our investigation. This model works on a few basic assumptions. We confine our population to an area where it is essentially static. This population might be similar to a dormitory at University of Chicago—taken in a small enough span of time, essentially no new people arrive, no one leaves, and no one is born or dies.

We also assume that our entire population has equal susceptibility to the disease, including those who have already been infected and recovered. The result looks something like Figure 1, where we have two subpopulations and people moving between them at different rates. Let us call the number of people infected I , the number of susceptible people S , and the rates of infection and recovery r_I and r_R respectively. Let t represent time.

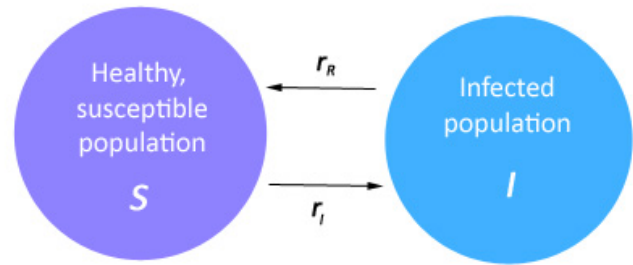


Figure 1: The total population is divided into two groups, susceptible and infected. People can move between these groups by becoming infected or recovering.

IMPLEMENTATION AND ANALYSIS

Now we may begin constructing a model. First, we know that the change in the number of people infected over time is going to be a function of the number of people who are newly infected less the number of people newly recovered, that is

$$\frac{dI}{dt} = \text{infected} - \text{recovered}. \quad (1)$$

In order for someone to become infected, they must be healthy and interact with someone who is infected. Out of each of these interactions, people become infected at the rate of infection, so

$$\text{Infected} = SIr_I. \quad (2)$$

Recoveries, however, do not depend in any way upon the healthy, susceptible population. They are merely a function of the number infected and the recovery rate:

$$\text{Recovered} = Ir_R. \quad (3)$$

Substituting equations (2) and (3) into equation (1), we get

$$\frac{dI}{dt} = SIr_I - Ir_R. \quad (4)$$

Note that in this equation, r_I and r_R are constants, but S is not. This means equation (4) is not linear in I . To solve this problem, we can call P the total population, which will be constant under our assumptions. This means that $P = I + S$, so we can say that

$$\frac{dI}{dt} = (P - I)I r_I - I r_R. \tag{5}$$

There is a trivial solution when $I = 0$, but this isn't very interesting, and we can hardly hope to model an infectious disease that doesn't exist. Equation (5) is first order, nonlinear, and separable. We can therefore conclude that a solution exists, and is given by

$$\int \frac{dI}{I[(P r_I - r_R) - I r_I]} = \int dt. \tag{6}$$

Equation (6) is tricky but possible to integrate, and the solution is given by

$$\frac{1}{P r_I - r_R} \ln \left| \frac{I^{r_I}}{[(P r_I - r_R) - I r_I]^{1/r_I}} \right| = t + c. \tag{7}$$

where c is an arbitrary constant. This solution cannot tell us much, however, since it cannot be solved for I . However, we may be able to find something useful if we graph the solution using Euler's method.

Using an online Java applet that calculates values using Euler's method, we were able to analyze our model [1]. In order to get an idea for how solutions to this equation behave, it was necessary to assign values to our constants (P , r_I , and r_R) as well as an initial value. All manipulations were carried out with a step size of .01 days and a population of 100 people. The rates of infection (unit $\text{people}^{-1}\text{days}^{-1}$) and recovery (unit days^{-1}) were varied, as well as the initial value at $t_0=0$, to determine their relative effects on the behavior of the solution. Results are summarized in table 1 and sample values from the applet can be found in the appendix.

As our messy-looking solution suggests, the graph takes on a logistic growth shape, complete with a limiting value. As the value of the infection rate decreases, so does the limiting behavior; the time taken to reach equilibrium increases,

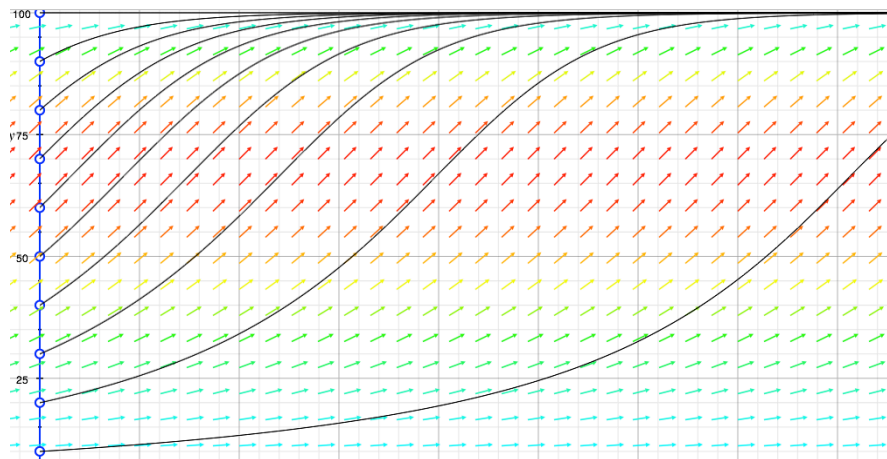


Figure 2: Slope field and solutions of model 1, with several curves drawn.

however. The influence of the recovery rate is reversed, with larger values causing the limit to decrease and the time taken to get there to increase. The initial value has no effect on the limiting behavior of the solution, but larger values will cause the function to reach its limit faster. See Figure 2 for a slope field and several possible graphs.

Trial	r_I ($\text{people}^{-1}\text{days}^{-1}$)	r_R (days^{-1})	Initial value (infected people)	Stabilization value (infected people)	Time reached (days, approximate)
1	.1	.1	1	99	1.1
2	.01	.1	1	90	11
3	.005	.1	1	80	30
4	.01	.01	1	99	10
5	.01	.05	1	95	10.5
6	.01	.5	1	50	17
7	.01	.05	3	95	9.1
8	.01	.05	5	95	8.6
9	.01	.05	10	95	7.8

Table 1: Variation in constant values to determine their effect on solution behavior. All solutions increased monotonically before eventually reaching some constant value. Times are approximate because, although the applet gives decimal values for both t and I , I is a measure of people and would in reality take on only integer values. Times given are at or around where the value for I came to a value which would be rounded up to the limiting value.

DISCUSSION

This model is very simple, but accurate for small, confined populations over relatively short periods of time when we would not expect the disease to “die out” and where people have a high amount of contact with each other. It might, for example, accurately predict the trajectory of a disease such as the common cold as it is introduced into a dormitory or other relatively confined populated space.

Recall that one of the important assumptions we made in this model is that the disease confers no immunity after recovery. This means that the limiting behavior of the solution is a dynamic equilibrium. If, for example, the limit is 80 out of 100 people, that means that after some time, there will constantly be 80 infected people, because over any given time period after that, the number of people newly infected and the number of people newly recovered will be the same. Note that this does not mean that after that time there are 80 people who are and remain infected and 20 people who are and remain healthy, but rather that there is a constant “shuffle” of who is sick and who healthy.

Our plot will have equilibrium points when $dl/dt = 0$. Using this, we found our equilibrium points to be $I = P - (R_r/R_i)$. We can see this is verified most easily by looking at trial 3 in Table 1 where $R_r = .1$ $R_i = .015$ $R_r/R_i = 20$, $P - 20 \Rightarrow 100 - 20 = 80$. This was indeed the stable number of infected people.

This model has a major flaw: we do not take into account the fact that people become immune, or at least more resistant to an infection once they've recovered from it. Once the disease propagates through the population, we would expect the number of infected people to peak, and then begin to decline. Rather than staying at a stable equilibrium of sick people, an actual population would acquire immunities and resistance so that the disease dies out. Because we have no method of accounting for immunity in our model, our stable equilibrium, the points where the disease stops spreading, are likely to represent peaks of the spread in an actual population.

MODEL DESIGN 2

Having completed a preliminary model, we will now try to design a model fit for more realistic circumstances. In our first model, we operated with the assumptions that all infected people were asymptomatic (giving others no reason to differentially avoid them) and the population was static. This time, we will take into account people's tendency to avoid symptomatic carriers and a

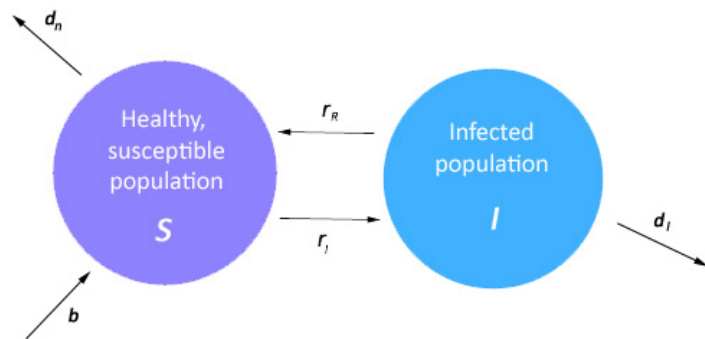


Figure 3: The population is still divided into two groups, but in addition to moving between the groups, birth and death allow for the population to change.

dynamic population with healthy people giving birth and both sick and healthy people dying (Figure 3). In addition to the variables used for the first model, let q represent the percentage of infections that become symptomatic, d_n be the natural mortality rate, d_I be the mortality rate due to disease, and b be the birth rate. Also, we introduce k , the sociability constant, which incorporates the idea that people may avoid others or not depending on the social situation.

IMPLEMENTATION AND ANALYSIS

For this model, we will end up with two different equations. First, we have the model for the infected population, which is of the general form

$$\frac{dI}{dt} = \text{infected} - \text{recovered} - \text{dead}. \quad (8)$$

Second, we have the model for the change in the healthy population over time, which has the general form

$$\frac{dS}{dt} = \text{recovered} - \text{infected} + \text{born} - \text{dead}. \quad (9)$$

While the transmission rate will be much more complicated for this model, the recovery rate, along with the other linear factors, remain relatively simple. The amount of people recovered per unit time is still given by

$$\text{Recovered} = I\tau_R. \quad (3)$$

The amount of people infected per unit time must now take into account the presence of both symptomatic and asymptomatic carriers. The proportion of symptomatic versus asymptomatic carriers, coupled with the way our population deals with each, will have a significant effect on the rate at which our disease spreads.

To factor in the presence of symptomatic and asymptomatic carriers, we first assume that a certain portion of the infected population is symptomatic at any given time. This could be due to either symptomatic and asymptomatic strains, or due to an asymptomatic incubation period. We will assign a constant $q < 1$ so that q times the infected population is equal to the number of symptomatic carriers. Likewise, $(1-q)$ times the infected population is equal to the number asymptomatic carriers.

We will assume that people exhibiting symptoms come into contact with less people than those who are asymptomatic. Therefore, the transmission rate from symptomatic and asymptomatic carriers will be different. We will assign 'sociability' constants K_s and K_a to symptomatic and asymptomatic carriers respectively, where $K_s < K_a$ to reflect the difference in the transmission rates from the two groups.

The overall transmission rate will be the transmission rates of each of these groups added together,

$$\text{newly_infected} = r_i I S q k_s + r_i I S (1 - q) k_a. \quad (10)$$

Factoring out the common terms, we are left with

$$\text{newly_infected} = r_i I S (q k_s + (1 - q) k_a), \quad (11)$$

which is a modified version of equation (2).

Death and birth also follow linear models. We will assume only healthy people give birth, so that the births per time are given by

$$\text{births} = bS \tag{12}$$

Death will afflict both sick and healthy people, but the sick will, because of their illness, die more frequently than the healthy ones. We therefore assign each condition—death from disease, d_i , and death from everything else d_n (natural causes)—a separate mortality rate. We therefore have

$$\text{deaths in } S = d_n S \quad \text{deaths in } I = d_i I \tag{13}$$

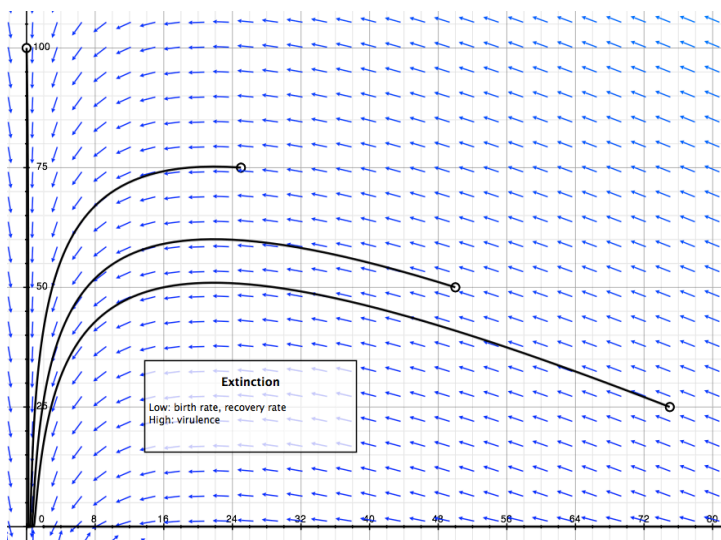
This gives us the tools we need to assemble our two equations. Substituting equations (3), (10), and the infected population term of equation (12) into equation (8), we get

$$\frac{dI}{dt} = r_i IS(qk_s + (1 - q)k_a) - Ir_R - Id_i \tag{14}$$

For the healthy population, we substitute equations (11) and (12) into equation (9) and get:

$$\frac{dS}{dt} = Ir_R - r_i IS(qk_s + (1 - q)k_a) - Sd_n + Sb \tag{15}$$

Equations (13) and (14) constitute a nonlinear system of equations. These are very difficult to solve, but we can at least say: Because the functions and their derivatives are continuous, there exists a unique solution to the system [3].



In order to approximate a solution, we used a graphing application. Figures 4, 5, and 6 show the number of healthy people as the x-axis value, and the number of infected people as the y-axis value. The initial population is 100, and in each figure there are four initial conditions corresponding to different initial amounts of sick people.

Figure 4: A disease with a high infectivity and low mortality. Each line represents a different initial value of sick and healthy people.

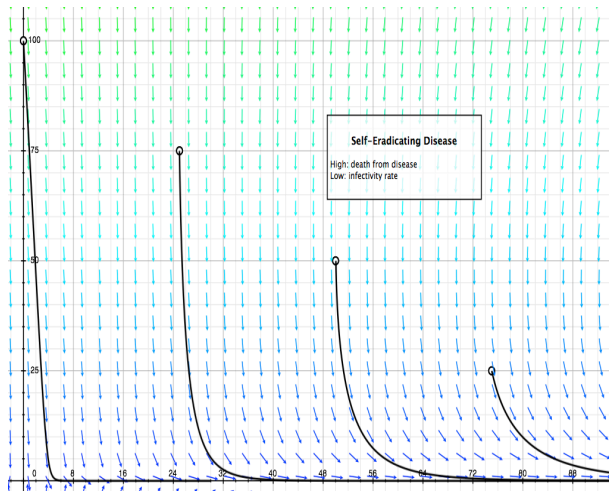


Figure 5: A disease with a high mortality rate.

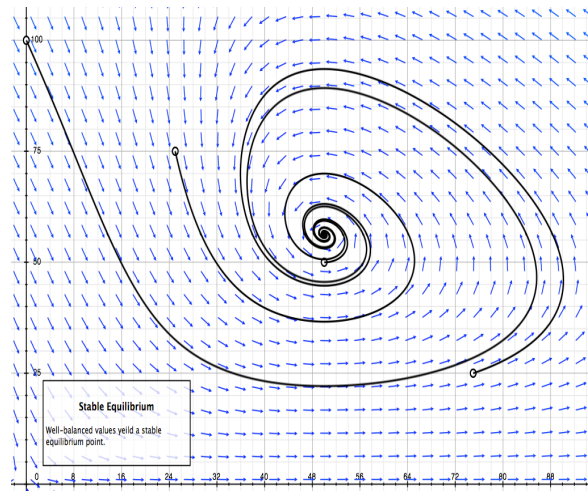


Figure 6: A population with a high birth rate subjected to a moderately infectious and only slightly deadly disease.

Figure 4 shows a scenario where the disease has a high infectivity rate relative to its mortality and recovery rate. Figure 5 shows a scenario where the mortality rate from the disease is relatively high. Figure 6 shows a population with a high birth rate and a disease with a moderate infectivity and low mortality.

DISCUSSION

This model's complexity, though it makes it difficult to solve, also makes it applicable to very large populations over long spans of time. This is because it takes into account the changes in a population that are only applicable at a large scale. It might model an epidemic which spreads worldwide or country-wide.

The scenario approximated in Figure 4 is something of a worst case scenario. Even in a large population, a disease which is particularly infectious and has a low enough mortality rate to allow carriers to spread it before dying will end up infecting and subsequently killing the entire population.

In Figure 5, we see the effects of a disease with a very high mortality rate. It dies out quickly no matter how infectious it is, because carriers die before they can spread it. The population then continues to multiply exponentially toward infinity. Technically, it will reach some sort of carrying capacity eventually, but human population growth occurs so rapidly that it seems we are nowhere near the limit, so this model is fairly realistic.

Figure 6 is much like Figure 4, but includes a very high birth rate. The disease is stuck at the same dynamic equilibrium we encountered in our first model, where diseased and healthy

individuals are being exchanged at the same rate. This is a particularly interesting case that might prove relevant in third world countries where the birth rate is particularly high.

By setting each of our differential equations to zero and then solving, we found the equilibrium points in terms of the constants.

$$I = \frac{(d_I + r_R)(b - d_n)}{r_I(qk_s + (1 - q)k_a)d_I} \quad (16)$$

$$S = \frac{(d_I + r_R)}{r_I(qk_s + (1 - q)k_a)} \quad (17)$$

If we plug the constants used to obtain the graph in figure 6, we can easily verify the validity of these equations. Examining these equations, we can see that there are some values of constants that will not yield equilibrium points. For example, if d_I in the denominator of eq. 16 is 0, (this would mean our disease confers immortality on its victims) there will be no equilibrium. Likewise, if r_I is 0, (this would mean its not possible to transmit the disease from person to person), then these formula both blow up to infinity. This obviously cannot be correct, because it's obvious that if the disease cannot be transmitted, the population of infected people would go to 0 while the healthy people would continue as normal. It therefore seems appropriate to add the constraint that r_I must be greater than 0 to use these formulas. Overall, though it does not account for resistant populations, this model is much more realistic than model 1.

CONCLUSION

The cases presented in this paper were an attempt to model the spread of a disease through both static and dynamic populations. Model 1 was simple and left out many important factors, yet it served both as a fair model for small populations and, more importantly, as a template for the more complicated second model. The second model presented is more realistic because it considers the differing interactions of infectious people with and without symptoms. Model 2 is also useful because it incorporates the effects of a changing population by accounting for death and birth. Although model 2 includes many relevant variables that affect the spread of a disease, because it leaves out factors such as population density and the chance of immunity after recovery, we cannot expect it to give an entirely accurate model for the spread of diseases that occur in areas with variable population density, or diseases that confer immunity upon recovery. Including these variables would have resulted in an even more complicated model which was outside the scope of this report. On the whole, these models serve as an excellent starting point for the modeling of disease.

GROUP CONTRIBUTION

Sarah Del Cielo was responsible for helping to design and solving the first model, including Euler's method approximations, their analysis, and the writing of its discussion, and the non-graph figures. She also wrote most of the second discussion.

Nailah Hart was responsible for helping write the abstract, problem statement, and conclusion, as well as brainstorming model designs.

Jake Clementi was responsible for designing and implementing the second model. He also responsible for using graphing applications to obtain figures.

CITATIONS

1. Schwarz, Susan, Scott Rankin, and Richard Williamson. "Euler Java Applet." *Interactive Math Programs*. N.p., 26 June 2000. Web. 19 May 2010. <<http://www.dartmouth.edu/~rewn/euler.html>>.
2. "Online LaTeX Equation Editor." *CodeCogs*. N.p., 2007-2010. Web. 19 May 2010. <<http://www.codecogs.com/latex/eqneditor.php>>.
3. William E. Boyce, Richard C Dprima,. Elementary Differential Equations and Boundary Value Equations. New Jersey: John Wiley and Sons, Inc: 2009 p. 358

