

## STUDENT VERSION

### Pandemic Modeling – Ebola, COVID-19, and Many More

Jue Wang

Union College

Schenectady NY 12308 USA

wangj@union.edu

**Abstract:** The recent coronavirus outbreak has infected millions of people worldwide and spread to over 200 countries. The virus is spreading rapidly. Infectious disease modeling is an essential part of the effort to minimize the spread and help predict the likely course of an epidemic. How can we use differential equations to study the spread of coronavirus? In this activity, students are engaged in the real life challenge to examine the dynamics of the pandemic using systems of nonlinear differential equations. The SIR model, force of infection and basic reproduction number are introduced to guide students to set up the equations and analyze the results. Students obtain estimates of model parameters from the real data and visualize the behavior of the solutions.

#### SCENARIO DESCRIPTION

On December 31, 2019, the city of Wuhan in China reported an outbreak of a novel coronavirus (COVID-19). It is the seventh member of the coronavirus family, together with MERS-CoV and SARS-CoV, that can spread to humans [1]. The virus is spreading rapidly. As of April 20, 2020, over 2.4 million people worldwide have been infected, spanning 210 countries and territories. The World Health Organization (WHO) has declared the outbreak as a pandemic.

The word ‘pandemic’ comes from the Greek word ‘pandemos’, meaning “pertaining to all people”. The Greek word ‘pan’ means “all” and ‘demos’ means “people”. Pandemics are usually caused by an infectious agent that is capable of rapidly spreading. An epidemic is specific to one city, region, or country, but a pandemic spreads beyond national borders, possibly worldwide. The 1918 Spanish flu was the worst pandemic in modern human history. It infected 500 million people around the world and resulted in the deaths of over 50 million people [2]. Increased travel and mobility have increased the likelihood of new diseases spreading. Figure 1 lists the major epidemics and pandemics that have occurred over time. The recent ones are SARS, swine flu, Ebola, MERS, and COVID-19.

Infectious disease modeling is an essential part of the effort to minimize the spread. A well-designed model not only can help predict the likely course of an epidemic, but also can reveal the most promising and realistic strategies for containing it.

**How can we use mathematical models and differential equations to study the spread of coronavirus?**

Name	Time period	Type / Pre-human host	Death toll
Antonine Plague	165-180	Believed to be either smallpox or measles	5M
Japanese smallpox epidemic	735-737	Variola major virus	1M
Plague of Justinian	541-542	Yersinia pestis bacteria / Rats, fleas	30-50M
Black Death	1347-1351	Yersinia pestis bacteria / Rats, fleas	200M
New World Smallpox Outbreak	1520 – onwards	Variola major virus	56M
Great Plague of London	1665	Yersinia pestis bacteria / Rats, fleas	100,000
Italian plague	1629-1631	Yersinia pestis bacteria / Rats, fleas	1M
Cholera Pandemics 1-6	1817-1923	V. cholerae bacteria	1M+
Third Plague	1885	Yersinia pestis bacteria / Rats, fleas	12M (China and India)
Yellow Fever	Late 1800s	Virus / Mosquitoes	100,000-150,000 (U.S.)
Russian Flu	1889-1890	Believed to be H2N2 (avian origin)	1M
Spanish Flu	1918-1919	H1N1 virus / Pigs	50M+
Asian Flu	1957-1958	H2N2 virus	1.1M
Hong Kong Flu	1968-1970	H3N2 virus	1M
HIV/AIDS	1981-present	Virus / Chimpanzees	25-35M
Swine Flu	2009-2010	H1N1 virus / Pigs	200,000
SARS	2002-2003	Coronavirus / Bats, Civets	770
Ebola	2014-2016	Ebolavirus / Wild animals	11,000
MERS	2015-Present	Coronavirus / Bats, camels	850
COVID-19	2019-Present	Coronavirus – Unknown (possibly pangolins)	More than 165,000 (as of April 20, 2020)

**Figure 1.** Major epidemics and pandemics in human history, modified from [3].

## Mathematical modeling – the SIR model

Let's first examine differential equation models in epidemiology. Epidemiology is the study of the distribution and causes of disease in populations. The spread of infectious diseases, such as measles

and malaria, can be modeled by a system of nonlinear differential equations. The simplest model is the **SIR** model (Susceptible-Infected-Recovered). The model makes several basic assumptions:

- A1. The total population is constant.
- A2. Once an individual has been infected and subsequently recovered, that individual cannot be re-infected, such as in the case of measles, mumps, and smallpox.
- A3. The rate of transmission of the disease is proportional to the number of encounters between susceptible and infected individuals.
- A4. The rate at which infected individuals recover is proportional to the number of infected.

As the first step in the modeling process, we identify the independent and dependent variables. The independent variable is time  $t$ , measured in days. The dependent variables are the numbers of individuals in each group, each as a function of time.

$s(t)$  = the number of susceptible individuals

$i(t)$  = the number of infected individuals

$r(t)$  = the number of recovered individuals

$s(t) + i(t) + r(t) = N$  (the total population)

It may seem more natural to work with population counts, but some of our calculations will be simpler if we use the fractions instead. Therefore, we normalize the dependent variables by the total population to represent the fraction of the total population in each group. The two sets of dependent variables are proportional to each other, so either set will give us the same information about the progress of the epidemic.

$S(t) = s(t)/N$  = the fraction of susceptible individuals

$I(t) = i(t)/N$  = the fraction of infected individuals

$R(t) = r(t)/N$  = the fraction of recovered individuals

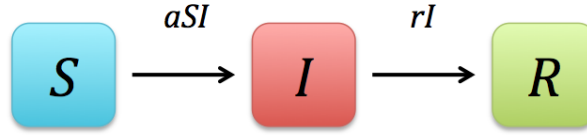
$S(t) + I(t) + R(t) = 1$

The SIR model is illustrated schematically in Figure 2, and described by the nonlinear system of ODEs below, where  $a$  is the transmission rate and  $r$  is the recovery rate.

$$S'(t) = -aSI \quad (1)$$

$$I'(t) = aSI - rI \quad (2)$$

$$R'(t) = rI \quad (3)$$



**Figure 2.** The SIR model.

To understand this model, let's analyze the differential equations and parameters. The **transmission rate**  $a$  is the product of two factors: the rate of contact and probability of transmission.  $a$  describes the fraction of those contacts that may result in infection, or it can be interpreted as the expected number of people an infected person infects per day. The function  $\lambda(I) = aI$  is the rate at which susceptible individuals become infectious, called the **force of infection**. The rate of change of the susceptible population is negative the force of infection times the number of susceptible, leading to (1).

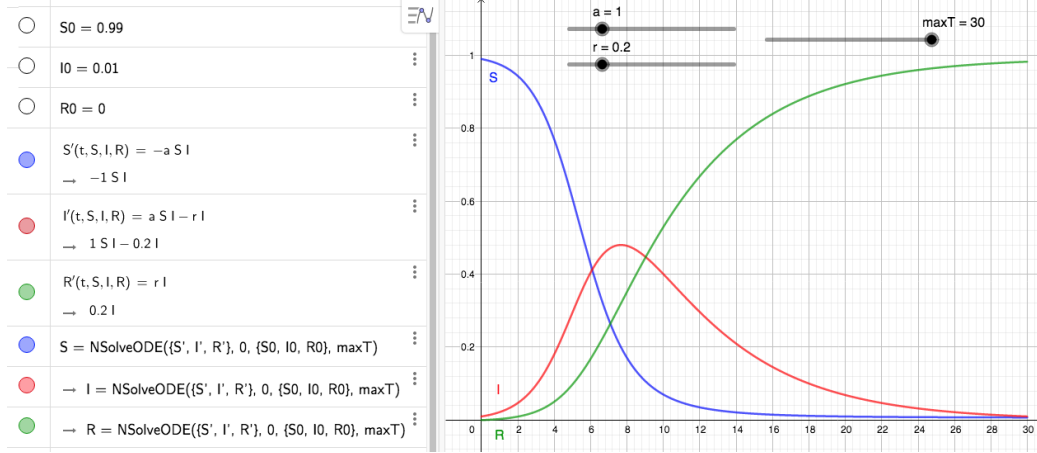
The **recovery rate**  $r$  is the rate at which infected individuals get over the disease, that is the proportion of infected recovering per day, thus the rate of change of the recovered population is  $rI$ , leading to (3). Importantly, the parameter  $r$  is measured in  $\text{time}^{-1}$  and  $1/r$  can be interpreted as the **average time to recover**, or the number of days an infected person has and can spread the disease.

Can you explain (2)?

In the GeoGebra App [4], let's solve the nonlinear system (using 'NSolveODE'), plot the solutions, and visualize the interaction of the three groups. The GeoGebra tutorials are available online [5]. Consider the initial population fractions  $S(0) = 0.99, I(0) = 0.01, R(0) = 0$  ( $S + I + R = 1$ ),  $a = 1, r = 0.2$ , and a time period of 30 days. An example demonstration is provided in [6] with a screenshot in Figure 3. Adjust the parameters values to examine different behaviors and view the changes in action. What do you observe by increasing or decreasing  $a$ ?

### Basic Reproduction Number $R_0$

The *basic reproduction number*  $R_0$ , " $R$  naught", is introduced to measure the transmission potential of a disease. It is the expected number of secondary infections produced by a single infective. For

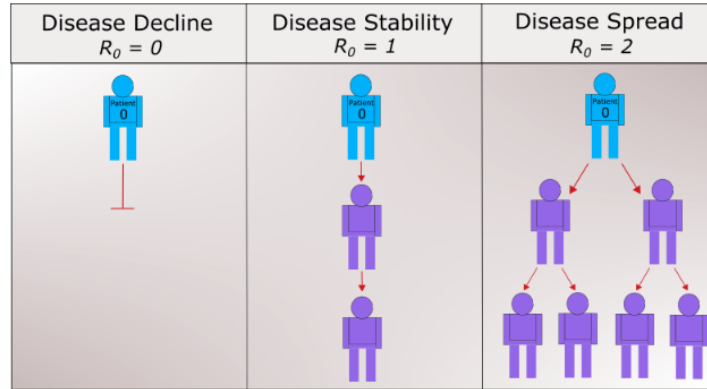


**Figure 3.** A screenshot of GeoGebra dynamic constructions.

example, if the  $R_0$  for measles in a population is 16, then we would expect each new case of measles to produce 16 new secondary cases over the period of time during which the infected individual can actually spread the disease.  $R_0$  is affected by several factors: the rate of contacts in the host population, the probability of infection being transmitted during contact, and the duration of infectiousness. In the SIR model described above,

$$R_0 = \frac{a}{r}. \quad (4)$$

$R_0$  is essentially a metric of how contagious a disease is. It can capture three basic scenarios, illustrated in Figure 4.



**Figure 4.** Basic reproduction number  $R_0$  [7].

If  $R_0 < 1$ , on average, an infected person infects less than one person. The disease is expected to stop spreading. If  $R_0 = 1$ , an infected person infects an average of one person. The disease spread is stable, or endemic. If  $R_0 > 1$ , on average, an infected person infects more than one person. The

disease is expected to increasingly spread in the absence of intervention. Here is the estimated  $R_0$  for several diseases [3]. Measles tops the list, being the most contagious with a  $R_0$  range of 12-18. Early estimates of  $R_0$  for COVID-19 vary in the range of 2-3.

**The following activities will guide you through some deeper investigations.**

1. Examine (1) and (2) again

$$\frac{dS}{dt} = -aSI, \quad \frac{dI}{dt} = aSI - rI$$

- (a) According to the differential equation for  $dI/dt$ , under what conditions will  $I(t)$  be increasing? decreasing? What does this mean about the spread of the disease? Using this result, explain whether quarantine will be effective against the disease.
- (b) Use the chain rule to show that

$$\frac{dI}{dS} = -1 + \frac{1}{R_0 S} \quad (5)$$

Two features of this new equation are particularly worth noting:

- The only parameter that appears is  $R_0$ , the reproduction number.
- The equation is independent of time. That is, what we learn about the relationship between  $S$  and  $I$  must be true for the entire duration of the epidemic.

Compute  $d^2I/dS^2$ . Determine when the number of infected will begin to decrease.

Compare this to your result from (a).

- (c) Show that  $I$  must have the form

$$I = -S + \frac{1}{R_0} \ln S + C \quad (6)$$

where  $C$  is a constant.

- (d) For a disease such as COVID-19,  $I(0)$  is approximately 0 and  $S(0)$  is approximately 1. A long time after the onset of the epidemic, we have  $I(\infty)$  approximately 0 again, and  $S(\infty)$  has settled to its steady state value, observable as the fraction of the population that did not get the disease. Explain why

$$R_0 = \frac{\ln S_\infty}{S_\infty - 1}.$$

- (e) Describe the meaning and significance of “herd immunity”. How can vaccination lead to herd immunity?

2. Recovered individuals may lose their immunity and become re-infected, such as in the case of malaria and tuberculosis, as well as COVID-19. We add a new assumption.

A5. Re-infection occurs at a rate proportional to the population of recovered individuals. Call this proportional parameter  $b$ .

Sketch a schematic diagram similar to Figure 2. Modify (1)-(3) and set up the new model – the **SIRS** model, with parameters  $a$ ,  $r$ , and  $b$ .

3. In reality, deaths may occur due to the disease. For simplicity, let's only consider deaths from the infected group  $I$ , with a death rate  $c$ .
  - (a) Modify your previous diagram and model in 2, including four differential equations with the fourth being  $D'(t) = \dots$ , and parameters  $a$ ,  $r$ ,  $b$ , and  $c$ .
  - (b) Pick one pandemic from the list in Figure 1 and a region, such as a city or a country. Research into the data available and obtain reasonable estimates of the parameters  $a$ ,  $r$ ,  $b$ , and  $c$ , as well as initial conditions  $S(0)$  and  $I(0)$ . Set  $R(0) = D(0) = 0$ . Note that  $S + I + R + D = 1$ .
  - (c) Solve your model with these parameter values and visualize the solutions in GeoGebra. Include a sharable link to your work. Describe the behavior of the solutions. Compare your results with the real data. Does your model seem reasonable? Explain.

Other models include SEIR, SEIS, SEAIR, MSIR, MSEIR, and MSEIRS. For many infections, there is an incubation period during which individuals have been infected but are not yet infectious themselves. During this period the individual is in class E (for exposed), leading to the SEIR model. The SEIS model is like the SEIR except that no immunity is acquired at the end. In the E class, a significant number of persons may never develop symptoms, but they are capable of transmitting the disease, introducing an additional compartment A, and thus SEAIR. For many infections, such as measles, babies are not born into the susceptible compartment but are immune to the disease for the first few months of life due to protection from maternal antibodies. The M class contains infants with maternally derived immunity (or passive immunity). The MSEIRS model is similar to the MSEIR, but the immunity in the R class would be temporary, so that individuals would regain their susceptibility when the temporary immunity ended.

## REFERENCES

- [1] Centers for Disease Control and Prevention. 2020. Human Coronavirus Types. <https://www.cdc.gov/coronavirus/types.html>. Accessed 10 May 2020.
- [2] Centers for Disease Control and Prevention. 2019. 1918 influenza pandemic (H1N1 virus). <https://www.cdc.gov/flu/pandemic-resources/1918-pandemic-h1n1.html>. Accessed 10 May 2020.
- [3] A visual history of pandemics. 2020. <https://www.weforum.org/agenda/2020/03/a-visual-history-of-pandemics>. Accessed 10 May 2020.
- [4] GeoGebra Math Apps. <https://www.geogebra.org>. Accessed 10 May 2020.
- [5] GeoGebra Tutorials. <https://www.geogebra.org/m/XUv5mXTm>. Accessed 10 May 2020.
- [6] Wang, J. 2020. GeoGebra example: <https://www.geogebra.org/classic/zumeuffq>. Accessed 10 May 2020.

- [7] How COVID-19 and other infectious diseases spread: mathematical modeling. 2020.  
<https://triplebyte.com/blog/modeling-infectious-diseases>. Accessed 10 May 2020.