Article:

by Ronald E. Mickens, Department of Physics, and Talitha M. Washington, Department of Mathematics, Clark Atlanta University

The Roles of SIR Mathematical Models in Epidemiology[1]

Abstract: Many human based diseases are analyzed using so-called SIR mathematical models. Our major goals are to examine the structure of these models, discuss what useful information can be derived from them, and indicate how they may be used to make general predictions on the possible courses of the associated diseases when particular types of actions are taken. We conclude that the simplest SIR models are valuable as tools for deriving critical qualitative features of the spread of disease. Various issues are also considered relative to the successes and failings of these models.

1 Introduction

Currently, October 2020, the world is engaged in a pandemic caused by "severe acute respiratory syndrome coronavirus-2", better known as ARS-CoV-2 [2]. This not previously seen disease in humans can spread easily from person to person. It is estimated that the median time from onset to clinical recovery for mild cases is approximately two weeks and is on the range of 3 to 6 weeks for patients with severe or critical disease [3]. A major non-laboratory tool used to analyze, understand, and predict the course of this disease in humans is mathematical modeling [4]. The most basic of such models is the SIR model where the total population is assumed to be constant is divided into three sub-populations:

$S(t)$ – susceptible and individuals who are uninfected;
$I(t)$ – infected individuals who upon contact with susceptible individuals can infect the susceptible;
$R(t)$ – recovered and/or removed individuals who have either recovered from their infection or have removed themselves by dying, etc.

The task of such models is to predict the trajectory of the epidemic as transitions are made from one population class to another, i.e.,

$$S(t) \longrightarrow I(t) \longrightarrow R(t). \tag{1}$$

SIR models are explicitly constructed to provide this information or estimates of it. These results are important since "there is a critical need to understand both the likely number of infections and their time course to inform both public health and health care system responses" [5].

Mathematical modeling is important because we wish to both understand and manipulate the universe, so that predictions can be made of its future states or conditions. Mathematical modeling allows for a partial resolution of this goal. However, different models may, in general, only allow the probing of different aspects of our original system. Thus, it should be kept in mind that the models are not the actual system. They are abstract mathematical representations of some of its features, the ones (hopefully) of relevance for our needs.

The purposes of this paper are several:

(1) Show the construction of a SIR model where the total population is assumed to be constant. This model consists of three coupled ordinary differential equations and is perhaps the simplest of possible SIR models.

(2) Provide appropriate interpretations of the parameters appearing in the equations and their connection to epidemiology data.

(3) Derive, using elementary mathematics, a number of the significant aspects or features of the solutions.

Figure 1: Talitha Washington and Ronald Mickens (Photo by Talitha Washington)

(4) Give a direct geometrical explanation of the effects of "stay-at-home-orders" and their relaxation.

(5) Provide "an abundance of evidence," as a consequence of the above results, that a simple SIR differential equation mathematical model allows detailed predictions to be made for all the major qualitative feature for he spread of a disease such as COVID-19.

This paper is organized as follows: In the next section, we present the general methodology for the construction of a SIR mathematical model with the total population constant. Section 3 presents on explicit SIR model and show how it may be used to calculate quantities such as the conditions necessary for an epidemic to occur, the general time dependent behaviors expected for $S(t)$, $I(t)$ and $R(t)$, and the number of total persons infected. Section 4 gives a geometrical argument to show the consequences of stay-at-home orders. In Section 5, a SIR model, satisfying all the conditions given in Section 2 is presented and discussed briefly. An important aspect of this model is that an exact, explicit solution to it can be calculated. Finally, we summarize our general results in Section 6 and briefly indicate how the simple SIR model can be generalized to include an exposed, but not infectious population class, how to include vaccination, and how to model vaccination with limited immunity.

2 Methodology of SIR Models

The most elementary SIR models are based on certain (simplistic) assumptions [6, 7]:

(i) The total population is composed of only three sub-populations, i.e., susceptibles, $S(t)$; infected, $I(t)$; and recovered, $R(t)$. Susceptibles are uninfected and susceptible to the disease; the infected population is, by definition, infected and they can in turn infect susceptibles; and recovered individuals have recovered from the disease and are now immune to re-infection.

(ii) The total population is taken to be constant, i.e.,

$$S(t) + I(t) + R(T) = N = \text{ constant}. \tag{2}$$

This constraint means that over the time interval for which the model is relevant, the birth and death rates are equal.

(iii) It is assumed that there is homogeneous mixing of the three populations. This means that all individuals in the total population have exactly the same probability of coming into contact with each other and interacting [5]. Further, it is assumed that the disease can be transmitted between any two individuals regardless of their location and age.

The explicit construction of a SIR model begins when a framework is formulated for determining how the three populations transfer from one population to another; see Eq. 1. For our purposes, we use the scheme

$$\frac{\Delta S}{\Delta t} = -T_1(S \to I), \tag{3}$$

$$\frac{\Delta I}{\Delta t} = T_1(S \to I) - T_2(I \to R), \tag{4}$$

$$\frac{\Delta R}{\Delta t} = T_2(I \to R), \tag{5}$$

where

$$\Delta V \equiv V(t_2) - V(t_1), \quad \Delta t = t_2 - t_1, \tag{6}$$

and

$$\begin{aligned} T_1(S \to I) &= \text{ transition rate from the } S \text{ population to the } I \text{ population,} \\ T_2(I \to R) &= \text{ transition rate from the } I \text{ population to the } R \text{ population.} \end{aligned}$$

Note that adding Eqs. 3 through 5 gives

$$\frac{\Delta}{\Delta t}\left[S(t) + I(t) + R(t)\right] = 0, \tag{7}$$

which is equivalent to the result expressed in Eq. 2. Also, observe the placement of the negative signs in Eqs. 4 and 5. This convention allows us to define the transition functions to be non-negative, i.e.,

$$\begin{cases} T_1(S \to I) > 0, \ T_2(I \to R) > 0, \\ S > 0, \ I > 0. \end{cases} \tag{8}$$

So what about the mathematical structure of $T_1(S \to I)$ and $T_2(I \to R)$? To be definitive in what follows, we will only examine deterministic, ordinary differential equation models and doing this gives for Eqs. 3 to 5 the three coupled equations

$$\frac{dS}{dt} = -T_1(S \to I), \tag{9}$$

$$\frac{dI}{dt} = T_1(S \to I) - T_2(I \to R), \tag{10}$$

$$\frac{dR}{dt} = T_2(I \to R). \tag{11}$$

This particular structure is easy to understand. First, take the transition rate functions, T_1 and T_2, to be non-negative. It follows that there must be a negative sign on the right-side of Eq. 9 since each S that gets infected gets transferred into the I-population. Thus, the S-population decreases and the I-population increases. Similarly, the first term on the right-side of Eq. 10 corresponds to additions to the I-population coming from newly infected members of the S-population. The second term on the right-side of Eq. 10 represents those members of the I-population that have recovered from the disease and now get transferred to the R-population.

A deep, careful consideration of the general dynamics of the systems to be modeled by an SIR representation allows us to conclude that the simplest transition functions must have the following

mathematical properties:

$\mathbf{T_1(S \to I) \equiv T_1(S, I)}$

(a) $T_1(S, I) > 0$, for $S > 0$, $I > 0$;

(b) $T_1(0, I) = T_1(S, 0) = 0$;

(c) $T_1(S, I)$ is a monotonic increasing function of S and I.

$\mathbf{T_2(I \to R) \equiv T_2(I)}$

(a) $T_2(I) > 0$, for $I > 0$;

(b) $T_2(0) = 0$;

(c) $T_2(I)$ is a monotonic increasing function of I.

$T_1(S_1, I)$ and $T_2(I)$ indicate that the respective transition amplitudes are functions of S and I, and I. There are many possible selections of functions $T_1(S, I)$ and $T_2(I)$ which satisfy these restrictions. The papers of Korobeinikov and Maini [9], and Hethcote and Driessche [8] provide some choices.

3 Standard SIR Model

The standard, most used, SIR model is constructed using the following choices for the transition functions [6, 7]

$$T_1(S, I) = \beta S \left(\frac{I}{N}\right), \quad T_2(I) = \gamma I, \tag{12}$$

or

$$\frac{dS}{dt} = -\beta S \left(\frac{I}{N}\right), \tag{13}$$

$$\frac{dI}{dt} = \beta S \left(\frac{I}{N}\right) - \gamma I, \tag{14}$$

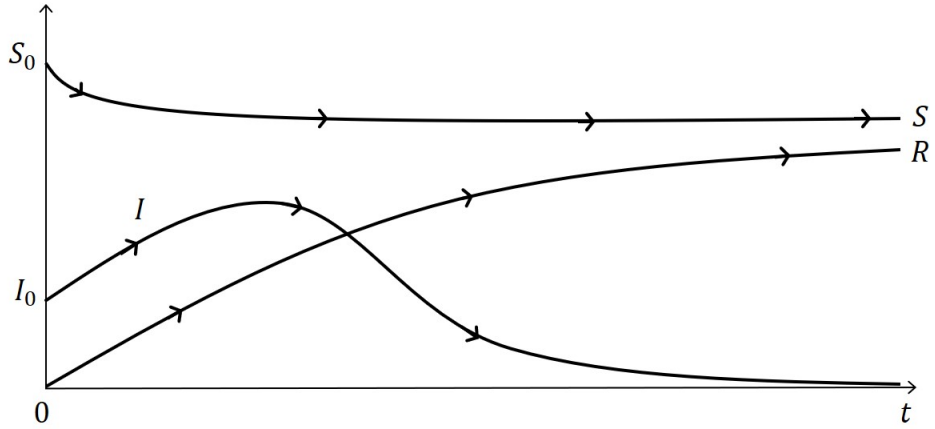
$$\frac{dR}{dt} = \gamma I. \tag{15}$$

The β and γ are constant parameters and are usually interpreted as being related to important aspects of the progression of the disease as seen in the following argument.

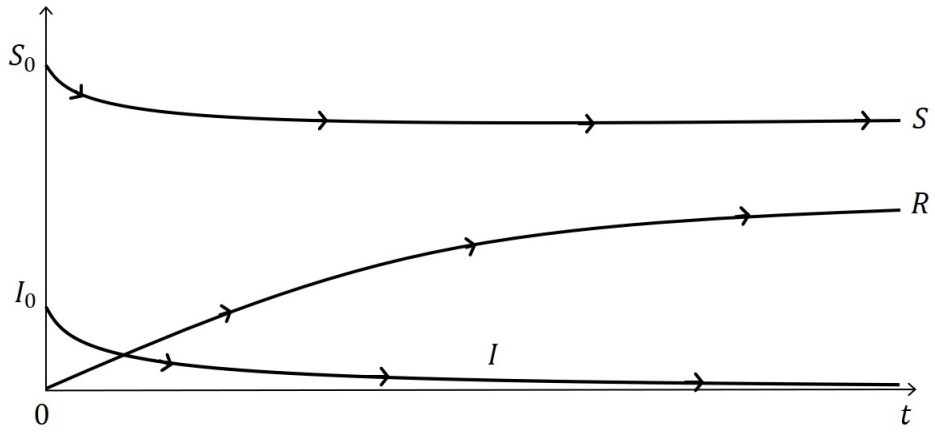
The left-hand sides of Eqs. 13 to 15 have the physical units of populations number over unit time. Therefore, all the right-hand side terms must also have these physical units. Examination of $T_1(S, I)$ and $T_2(I)$ allows the conclusion that the two parameters, β and γ, have the physical unit of inverse time. From this result, they are then given the interpretations

$t_c = \frac{1}{\beta} =$ average time between contacts of the S and I populations,
$t_r = \frac{1}{\gamma} =$ average time a member of the infected population stays infected and then gets transferred to the removed population.

It is important to observe that Eqs. 13 and 14 do not involve R. A major consequence of this fact is that in the analysis of this SIR system of equations only Eqs. 13 and 14 need to be considered since R can be determined from Eq. 2, i.e., $R = N - S - I$.



(a) $S_0 > S^*$



(b) $S_0 < S^*$

Figure 2: General time dependence of $S(t)$, $I(t)$, and $R(t)$ for (a) an epidemic $S_0 < S^*$, $r_0 > 1$) and (b) no epidemic ($S_0 < S^*, r_0 < 1$).

Note that the equation for dI/dt can be algebraically manipulated into the following expression

$$\frac{dI}{dt} = \beta \left(\frac{I}{N} \right) (S - S^*), \tag{16}$$

where

$$S^* = \frac{\gamma N}{\beta} = \left(\frac{t_c}{t_r} \right) N = \text{constant}. \tag{17}$$

Keep in mind the fact that $dI/dt > 0$ implies that $I(t)$ is increasing, while $dI/dt < 0$ means that $I(t)$ is decreasing. Also, both $S(t)$ and $I(t)$ are always non-negative. With this information, the following conclusions can be immediately drawn from an examination of Eq. 16:

(i) Let at time $t = 0$, $S(0) = S_0 > 0$ and $I(0) = I_0 > 0$, with $I_0 \ll S_0$. If $S_0 > S^*$, then $I(t)$, for $t > 0$, initially will increase. But from Eq. 13, $S(t)$ will decrease. At some future time, $S(t)$ will fall below S^* and $I(t)$ will begin to fall. This situation is depicted in Figure 2(a). For this case, we have a classical epidemic.

(ii) For the same initial conditions as in (i), but with $S(0) < S^*$, $I(t)$ decreases to zero; see Figure 2(b). No epidemic takes place.

Thus, we are led to the famous threshold theorem of epidemiology [4, 6, 7]. The placement of a single infective in a susceptible population will only initiate an epidemic if the number of susceptibles

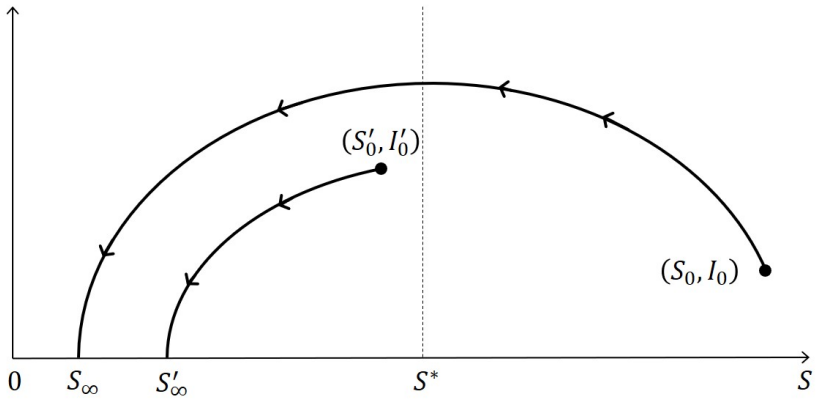


Figure 3: Plots of I versus S. The upper curve depicts an epidemic ($S_0 > S^*$), while the lower curve ($S_0^* < S^*$) does not lead to an epidemic.

in larger than a certain threshold value, in our case it is S^*. Another way of stating this result is to note that this is equivalent to the condition that the rate at which susceptibles become infectives must be larger than the rate at which infectives are eliminated from the population; see Eq. 16.

An alternative way to proceed is to introduce r_0, the so-called "basic reproduction number" [5, 6]. We define r_0 as

$$r_0 \equiv \left(\frac{\beta}{\gamma}\right)\left(\frac{S_0}{N}\right) = \frac{S_0}{S^*}. \tag{18}$$

The parameter r_0 plays a fundamental role in SIR based epidemiology and is generally interpreted as the average number of secondary infections caused by the introduction of a single infective individual into a susceptible population. It is easily seen that if $r_0 > 1$, then an epidemic will occur, while for $0 < r_0 < 1$, the infective population decreases from the start and no epidemic takes place.

The curve of I versus S provides further insights into the dynamics of the SIR model. It can be shown (with just a knowledge of first-year calculus) that the slope of the $I(s)$ versus S curve is given by the expression [6, 7]

$$\frac{dI}{dS} = -1 + \frac{S^*}{S}, \tag{19}$$

and this equation can be integrated to give the result

$$I + S = I_0 + S_0 + S^* \ln\left(\frac{S}{S_0}\right), \tag{20}$$

where $S = S(t)$, $I = I(t)$ and $S_0 = S(0)$, $I_0 = I(0)$, and $R(0) = R_0 = 0$.

Figure 3 gives a plot of the I versus S curve for two situations: (i) $S_0 > S^*$, leading to an epidemic; and (ii) $S_0 < S^*$, for which no epidemic occurs. These two cases correspond, respectively, to whether $r_0 > 1$ or $r_0 < 1$. Note that in terms of time behaviour, the motion along these curves goes from the right to the left sides of the graph and exactly correspond to the plots presented in Figure 2.

For the remainder of this section, only the case of an epidemic will be considered. Of interest is a determination of $I_{\max}$, S_∞, and I_{total} where

$I_{\max}$ = the maximum value of the number of infectives during the course of the epidemic;
S_∞ = the number of susceptibles who do not succumb to the epidemic;
I_{total} = the total number of susceptibles who become infected.

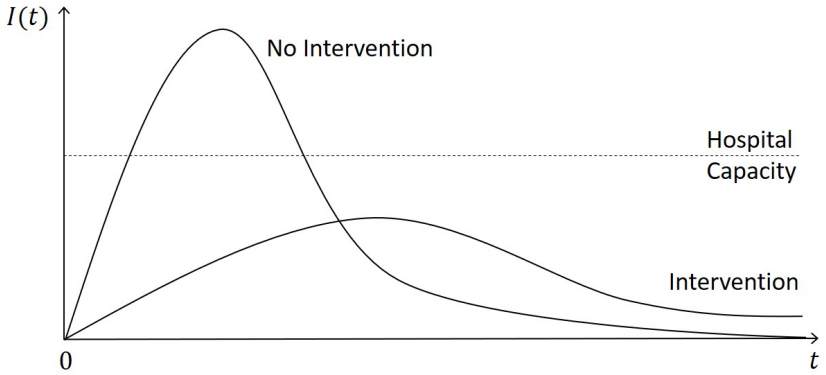


Figure 4: Flattening the curve: plot of $I(t)$ versus t for no intervention and for the outcome of good interventions.

Using Eq. 20, it follows that these quantities are given by the expressions

$$I_{\max} = I_0 + (S_0 - S^*) + S^* \ln\left(\frac{S^*}{S_0}\right), \tag{21}$$

$$S_\infty = S_0 + I_0 + S^* \ln\left(\frac{S_\infty}{S_0}\right), \tag{22}$$

$$I_{\text{total}} = (S_0 - S_\infty) + I_0. \tag{23}$$

Note that we are "given" the parameters β and γ, and the total population N, and this allows S^* to be calculated, see Eq. 17. In addition, the initial numbers of the infectives and susceptibles, i.e., I_0 and S_0, are also specified. Thus, $I_{\max}$, S_∞, and I_{total} may be determined, respectively, from Eqs. 21, 22, and 23.

Comment: The equation to be solved for S_∞, given in Eq. 22, allows for it to be expressed in terms of a "new elementary function." the so-called Lambert-W function [10].

In summary, given the elementary SIR model where the parameters β and γ are known, given the size of the total population, $N = S_0 + I_0$, it follows that all of the general qualitative features can be determined, and values for $I_{\max}$, S_∞, and I_{total} can be calculated. This realization is important since for this model explicit, exact, close-form solutions do not exist for $S(t)$ and $I(t)$, expressible in terms of a finite combination of the elementary functions.

4 Flattening the Curve

The curve that is being talked about is the plot of $I(t)$ versus t, i.e., the number of infectives as a function of time, t.

Hospitals generally have a maximum capacity for treating acute illnesses in terms of the number of beds and available care teams. Further, some hospitals and emergency facilitates may already be operating close to their maximum capacity under normal circumstances.

If the sharply peaked curve (see Figure 3) could be changed into a broader, flatter, lower curve, lying below the capacity curve of the hospital, then this would help hospitals provide better care for their patients and allow some relief to their emergency care staff. Figure 4 illustrates this situation [11].

Note that what we wish to achieve is a lowering of the total number of cases who are in the hospital at any given time to a number smaller than the maximum capacity and spread them out over a longer period of time [11]. This lowering and spreading can be achieved, in the absence of an actual effective vaccine through the use of physical distancing, stay-at-home orders, the appropriate

wearing of face masks and a number of other measures. All of these actions may be characterized as non-pharmaceutical interventions [12].

We now study the consequences of a non-pharmaceutical intervention on the outcomes of the simple SIR model. This can be done without any new mathematical effort. The purpose of the intervention is to "flatten the curve" and for ease of interpretation and explanation, we consider a stay-at-home order. In the immediate discussion, all of our comments and observations will refer to the curves in Fig. 5.

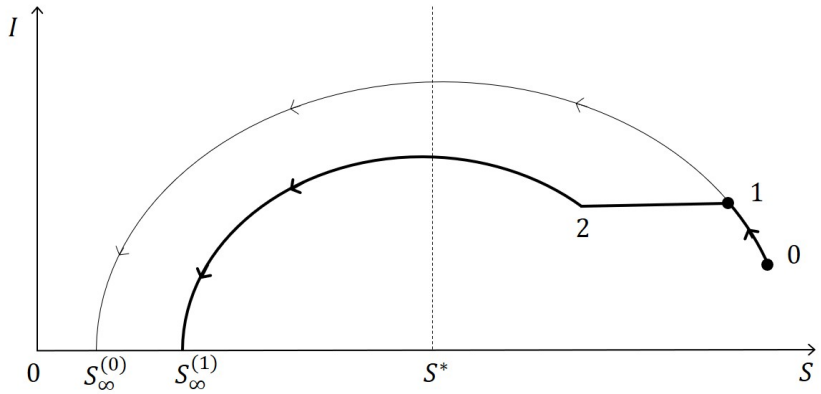


Figure 5: Consequences of the initiation of a stay-at-home order (beginning at 1; follow the heavy lines).

The initiation of the epidemic is at point 0, where $P_0 = (S_0, I_0)$. The system then evolves to point $P_1 = (S_1, I_1)$ where

$$S_1 < S_0, \quad I_1 > I_0. \tag{24}$$

At P_1, the stay-at-home order is made and we assume that some fraction of the susceptible population obeys this edict. (The exact value is not important; it just needs to remain constant for the arguments to be valid.) The system now goes from P_1 to $P_2 = (S_2, I_1)$, i.e., S changes (decreases) from S_1 to S_2, with $S_2 < S_1$, but the number of infectives remains constant at the value I_1.

At P_2, the SIR system evolves in the usual manner, with $S(t)$ steadily decreasing to the value $S_\infty^{(1)}$, and $I(t)$, at first increasing to a peak and then decreasing to zero.

A close inspection of Figure 5, allows the following conclusions to be made:

(i) The evolution of the system with the stay-at-home order gives a smaller value for the peak number of infectives as compared with having no such order.

(ii) Because $S_\infty^{(1)} > S_\infty^{(0)}$, the total number of infections during the epidemic is reduced with a stay-at-home order.

(iii) Since the graphs in Figure 5 do not contain temporal information, we cannot directly show the peak in infectives occurs later for the situation where a stay-at-home order is in place. However, this does turn out to occur [13].

The graphs in Figure 6 illustrate what can happen when a stay-at-home order is released. Assume that the path connecting point 0 (P_0) and point 1 (P_1) corresponds to a stay-at-home situation. Let the order be released at $P_1 = (S_1, I_1)$. The system now goes to $P_2 = (S_2, I_1)$ and continues along the upper curve. Examining this graph gives the following results:

(a) The second peak in the infective curve is now larger after the stay-at-home order is dropped.

(b) After the epidemic has reached its course, the total infective population is also larger.

(c) In summary, a stay-at-home order, followed by letting the SIR system evolve and then releasing the order may lead to a situation where the second peak and spread of the disease increases. This is due mainly to the sudden infusion of new susceptibles into the population and they can now potentially be infected.

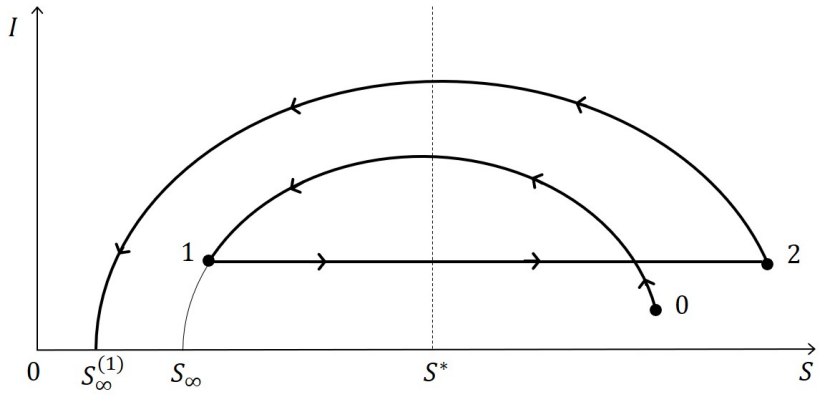


Figure 6: One possibility of halting a stay-at-home order (beginning at 1).

5 Model with Exact Analytical Solution

It turns out that an SIR model can be constructed satisfying all the requirements in Section 2, such that its explicit solutions may be calculated in terms of a finite combination of the elementary functions, namely, the trigonometric functions [14]. For our purposes, this model takes the form

$$\frac{dS}{dt} = -\beta\sqrt{S}\sqrt{I}, \quad \frac{dI}{dt} = \beta\sqrt{S}\sqrt{I} - \gamma\sqrt{I}, \tag{25}$$

$$\frac{dR}{dt} = \gamma\sqrt{I}, \tag{26}$$

$$S(0) = S_0 > 0,\ I(0) = I_0 > 0,\ R(0) = 0 \tag{27}$$

Note that the total population is constant, i.e.,

$$S(t) + I(t) + R(t) = N = \text{ constant }, \tag{28}$$

and further, only the differential equation involving $S(t)$ and $I(t)$ need be investigated.

The transformation

$$u(t) = \sqrt{S(t)}, \quad v(t) = \sqrt{I(t)}, \tag{29}$$

gives

$$\frac{du}{dt} = -\left(\frac{\beta}{2}\right)v, \quad \frac{dv}{dt} = \left(\frac{\beta}{2}\right)u - \left(\frac{\gamma}{2}\right), \tag{30}$$

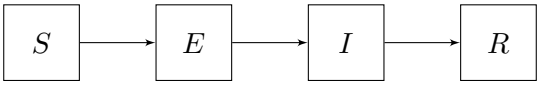
a pair of linear, coupled equations which may be easily solved [14]. This model does have the interesting feature that $I(t)$ goes extinct in a finite time. See Mickens [14] for the full details.

6 Discussion

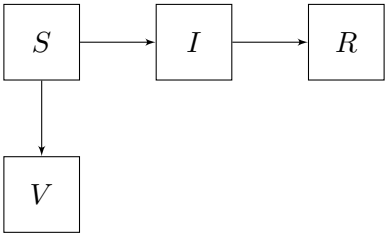
An important feature of our investigation of the elementary SIR model is that no explicit knowledge of the solutions for $S(t)$ and $I(t)$ are needed to either analyze or understand the essential details of the evolution of these solutions. This means that all of the basic qualitative properties can be determined

without the use of advanced mathematical techniques. Further, instances where exact solutions do exist, by means of a proper selection of $T_1(S,T)$ and $T_2(I)$, these solutions are in full agreement with the expectations obtained from the qualitative investigations. Consequently, we may expect that our modeling process can be applied to actual epidemics even if we only wish to know the various aspects of their major qualitative features as they evolve. Thus, in spite of the fact that we only have incomplete knowledge, there is enough value obtained from the use of these qualitative techniques to help policy makers make certain general but valid decisions to limit disease spread.

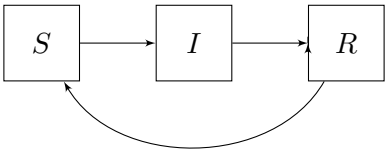
Also, it is critical to understand that simple SIR models can be readily generalized to include other components such as adding an exposed class (infected, but not infectious individuals)



the inclusion of vaccination



and putting in limited immunity



As a further complexity, these features can be combined in a mega-model, with time-dependent parameters [4, 6, 7, 12].

Finally, it should be observed that the modeling of the evolution of a disease is not dependent on a mathematical representation by a set of coupled, ordinary differential equations. There are many mathematical structures available for use: continuous versus discrete time, deterministic versus stochastic, qualitative versus quantitative methods, etc. Generally, the particular mathematical structure selected is the choice of the modeler(s). What this article indicates is that even elementary mathematical models can be used to provide important insights into how to limit the spread of disease.

References

[1] [Editor's Note: This is an invited contribution, serving as an introduction to epidemic mathematical modeling, by two experts – APS Fellow Ronald Mickens and AMS Fellow Talitha Washington.]

[2] World Health Organization (WHO), Coronavirus disease (COVID-19) technical guidance. Accessed at: `https://www.who.int/westernpacific/emergencies/covid-19/technical-guidance`

[3] Centers for Disease Control and Prevention, Coronavirus disease 2019 (COVID-19), National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Accessed at: https://www.cdc.gov/ncird/dvd.html

[4] R.M. Anderson and R.M. May, *Infectious Diseases of Humans*, Oxford University Press, Oxford, 1991.

[5] J. Tolles and T-B. Luong, *Modeling epidemics with compartmental models*, Journal of the American Medical Association (JAMA). Vol. 323 (24), June 23/30, 2020, pp. 2515–2516.

[6] H.W. Hethcote, *The mathematics of infectious diseases*, Society for Industrial and Applied Mathematics Review, Vol. 42 (2000), pp. 599–653.

[7] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Springer, New York, 2001.

[8] H. Hethcote and P. Driessche, *Some epidemiological models with nonlinear incidence*, Journal of Mathematical Biology, Vol. 29 (1991), pp. 271–287.

[9] A. Korobeinikov and P. K. Maini, *Non-linear incidence and stability of infectious disease modes*, Mathematical Medicine and Biology, Vol. 22 (2005), pp. 113–128.

[10] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, and C.W. Clark, editors, *NIST handbook of Mathematical Functions*, Cambridge University Press, 2010.

[11] K. Gavin, Flattening the curve for COVID-19: What does it mean and how can you help?. Accessed at: https://healthblog.uofmhealth.org/wellness-prevention/flattening-curve-for-covid-19-what-does-it-mean-and-how-can-you-help

[12] C.N. Ngonghala, E. Iboi, S. Eikenberry, M. Scotch, C.R. MacIntyre, M.H. Bonds, and A.B. Gumel, *Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus*, Mathematical Biosciences, Vol. 325 (July 2020), 108364.

[13] Z. Feng, J.W. Glasser, and A.N. Hill, *On the benefits of flattening the curve: A perspective*, Mathematical Biosciences, Vol. 326 (August 2020), 108389.

[14] R.E. Mickens, *An exactly solvable model for the spread of disease*, The College Mathematics Journal, Vol. 43 (2012), pp. 114–121.