Research | May 01, 2020

Modeling the Spread of COVID-19

By David I. Ketcheson

In the last several months, the rapid spread of COVID-19—the disease caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—has turned our daily lives upside down. As of early April, the number of confirmed cases was well over one million, with the number of total infections certainly much higher. But these numbers are small compared to what we know will come. This is because the virus's spread is a classic example of exponential growth, in which the increase of some quantity (in this case, the infected population) is proportional to its current size. Infectious diseases grow exponentially because each newly-infected person becomes another source of infection.

One of the simplest epidemiological models is the SIR model [3]. It divides the population into three groups:

- Susceptible individuals S, who have not yet been infected
- Infectious individuals I, who are infected and can infect others
- Recovered individuals R, who were previously infected and are now immune.

The model takes the form of three differential equations that describe the rates at which individuals transition between these groups:

$$\frac{dS}{dt} = -\beta I S$$
$$\frac{dI}{dt} = \beta I S - \gamma I$$
$$\frac{dR}{dt} = \gamma I.$$

$S(t)$, $I(t)$, and $R(t)$ are the respective fractions of the population in the susceptible, infectious, and recovered groups. We assume that people randomly come into contact with each other at a rate of β encounters per person per day. Each time an infectious person encounters a susceptible person, the virus spreads. Meanwhile, $1/\gamma$ is the average infectious period. Thus, if a single individual is initially infectious, he/she will pass the disease to $R_0 = \beta/\gamma$ other people on average. R_0 is called the *basic reproduction number* and is a property of both the disease and the behavior of the population in which it propagates.

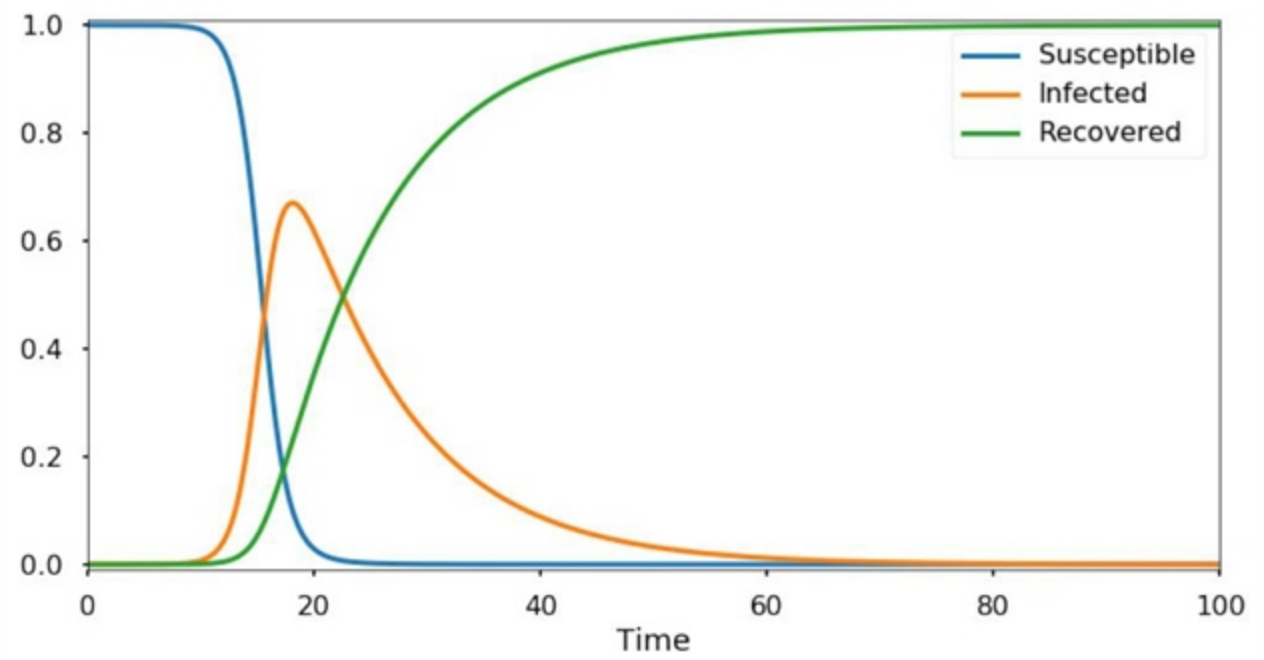

Figure 1. Typical SIR (susceptible-infectious-recovered) model behavior.

Figure 1 depicts the typical spread of a new disease in a completely susceptible population. Since initially $S \approx 1$ and $I, R \ll 1$, we observe exponential growth as predicted by the second differential equation:

$$I'(t) \approx (\beta - \gamma)I(t),$$

whose approximate solution is $I(t) = \exp((\beta - \gamma)t)$. If $\beta < \gamma$, the infectious population will decrease; this makes sense because $R_0 < 1$ in this case, meaning that each infected individual passes the disease to less than one other person on average. But if $\beta > \gamma$, the number of infected people will double over each time interval of length

$$t_{\mathrm{d}} = \frac{\ln(2)}{\beta - \gamma},$$

which is known as the *doubling time*. This exponential growth will eventually slow as the susceptible population S decreases. We reach the peak in the infectious population when $I'(t) = 0$, i.e., when $\beta S = \gamma$ or $S = 1/R_0$.

Although the SIR model is simple, it is a surprisingly powerful tool for both qualitative and quantitative predictions in the real world [2]. However, it is notoriously difficult to determine the model parameters β, γ —or equivalently, R_0 and t_d —for a new disease while the epidemic is still growing. Early approximations of R_0 for COVID-19 ranged from about 2.5 to as low as 1.4, but more recent estimates put R_0 closer to 4. Measured values of the doubling time have ranged from two days to one week, with typical values of roughly three-four days.

Figure 2 displays the results of modeling COVID-19's spread on a global scale with these parameters. The natural consequence of rapid proliferation within a completely susceptible population is a pandemic that peaks during the summer and infects the great majority of humankind. Although most cases of COVID-19 are mild or asymptomatic, it is clear that this scenario would completely overwhelm the healthcare systems in every country for at least several weeks during the peak of the crisis.

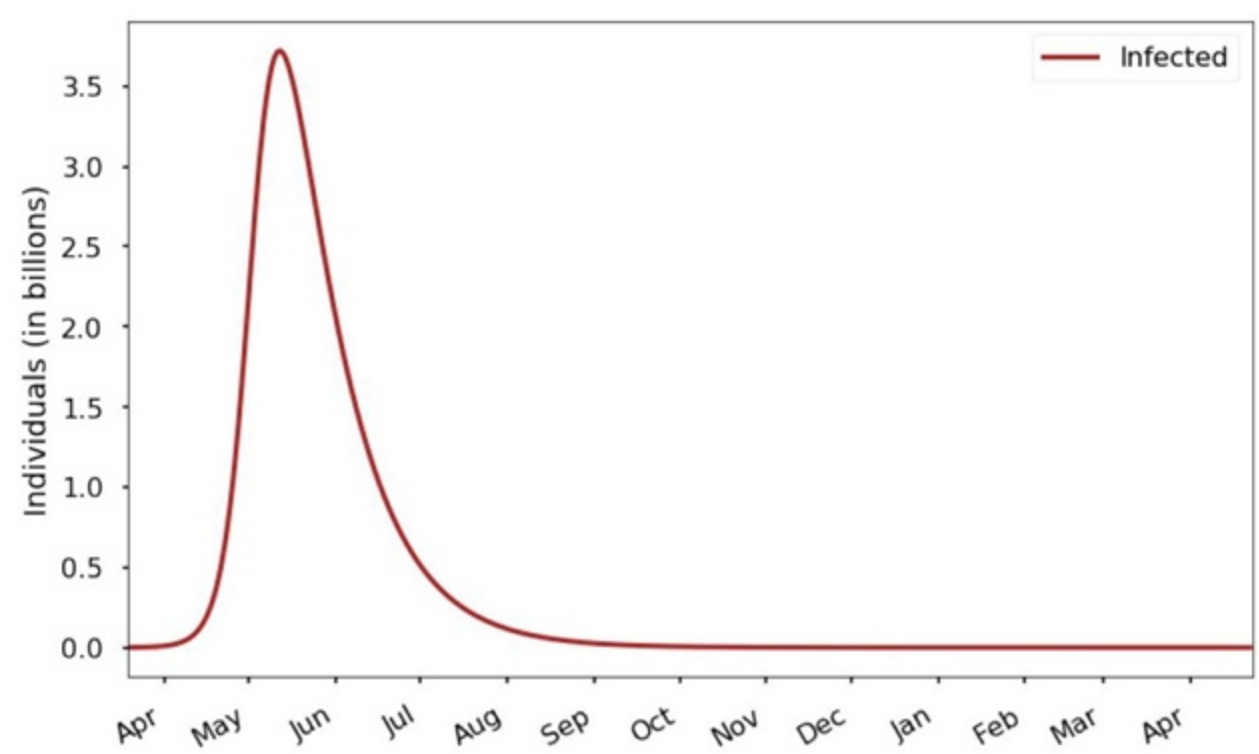

Figure 2. Predicted spread of COVID-19 in the absence of intervention.

Equilibrium

The SIR model is at equilibrium if and only if $I = 0$, since all terms on the right-hand side of the system are proportional to I. If this equilibrium is stable, a small local outbreak will not spread. But if it is unstable, any small outbreak will become a global epidemic. Eigenvalue analysis confirms what we have already observed: the zero-infection equilibrium is stable if and only if $S \leq 1/R_0$. This means that a certain fraction of the population *must* catch the disease before it will die out; that fraction is $(R_0 - 1)/R_0$. This intuitively makes sense. For instance, suppose that R_0 is 4. In a fully susceptible population, each infected person will infect about four other people. But if three-fourths of the population is already recovered and thus immune, then one infectious individual will only pass the disease to one new susceptible individual on average. This notion is called *herd immunity*. An individual who is still susceptible but lives in a society that is largely immune to a given illness is very unlikely to catch that illness because of the stability of the zero-infection equilibrium. In fact, the fraction of people who will eventually become infected is greater than $(R_0 - 1)/R_0$, since that is simply the point at which the equilibrium stabilizes.

The system still must reach this equilibrium, and along the way many more cases will emerge. The number of excess infections is known as *epidemiological overshoot*.

Intervention

The basic SIR model assumes that the contact rate β is constant in time. But as with COVID-19, a threatened population may change its behavior in an effort to reduce the rate of contact and slow or halt a disease's spread. This is the goal of the current widespread school and work closures, stay-at-home directives, and other restrictive social distancing measures. The medical literature refers to such phenomena as *non-pharmaceutical interventions*. If we denote the fraction of contact prevented through intervention as $q(t)$, we can include it in the SIR model as follows:

$$\frac{dS}{dt} = -(1 - q(t))\beta IS$$
$$\frac{dI}{dt} = (1 - q(t))\beta IS - \gamma I$$
$$\frac{dR}{dt} = \gamma I.$$

In the absence of intervention, we have $q = 0$, In contrast, putting each person in complete isolation yields $q = 1$. Real-world interventions lie somewhere between these two extremes.

Increasing the value of q has the same effect as decreasing β, in that it both slows the rate of exponential growth—stretching the epidemic over a longer time period—and lowers the peak. We now see that the zero-infection equilibrium is stabilized when $(1 - q(t))S(t) \leq 1/R_0$. Considering our example value of $R_0 = 4$, reducing contact by 50 percent means that only half of the population must now get infected before the epidemic will begin to subside. Figure 3 illustrates these effects, which reflect the now-well-known phrase of "flattening the curve."

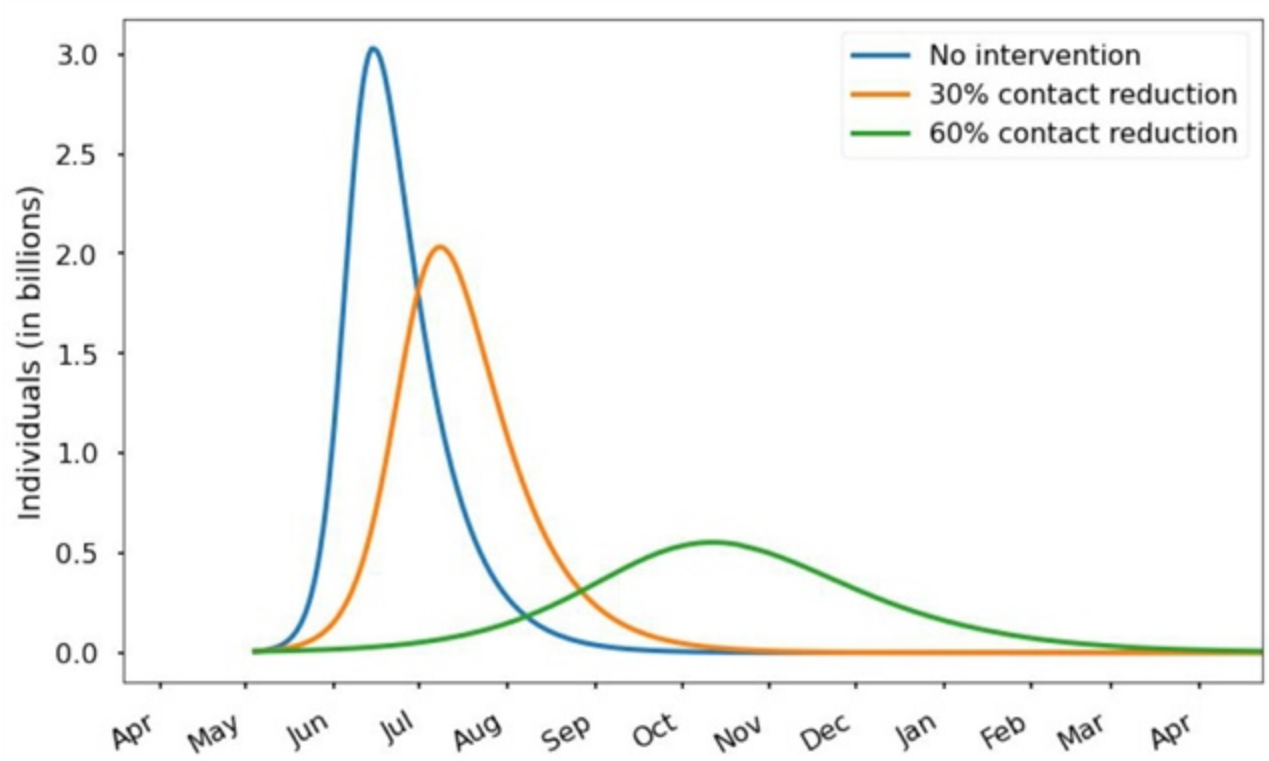

Figure 3. Flattening the curve through contact reduction.

One consequence of exponential growth is that interventions are most effective when they are imposed before an epidemic becomes widespread. For instance, a study of the 1918 Spanish flu pandemic found that reduced peak rates of infection strongly correlated with how early a community imposed intervention measures [1].

Exit Strategy

We must remember that the aforementioned reduced criterion for equilibrium holds only as long as the intervention persists. When a community reverts to its pre-intervention lifestyle, $q(t)$ returns to 0. It is extremely difficult to completely eliminate a virus on a global scale, and a new epidemic will emerge if the susceptible fraction of the population is higher than $1/R_0$. This type of resurgence occurred in many U.S. cities during the 1918 Spanish flu epidemic — especially in locations where strong, early interventions were imposed. Figure 4 depicts an example of this scenario for COVID-19

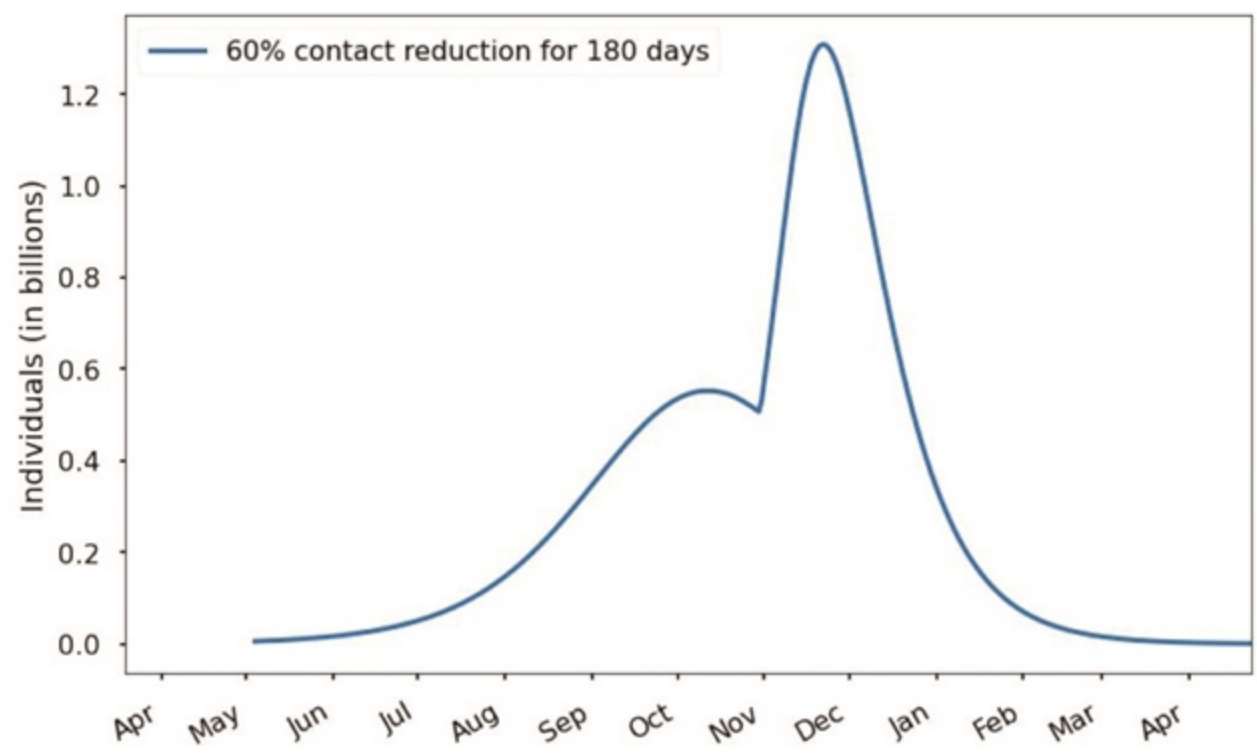

Figure 4. Scenario with a second outbreak after intervention is lifted.

At the moment, humankind is more or less unified in a colossal effort to drive the spread of COVID-19 to an artificial and temporary equilibrium. In the short term, this effort seems necessary to prevent the disease from completely overwhelming healthcare systems. But the long-term strategy is unclear. A natural question is whether intervention is effective in reducing the eventual toll of an epidemic. Such a reduction is limited by the fact that in the absence of permanent and broad lifestyle changes, we must eventually reach at least the herd immunity threshold in which a majority of the population has been infected (and gained immunity). Interventions can lessen the amount of epidemiological overshoot, but even this effect is limited. For the 1918 flu pandemic, studies reveal only a very weak—and not statistically significant—correlation between interventions and eventual death toll.

Many refinements of the SIR model exist, and researchers are applying several more detailed mathematical models to the current crisis. But the broad strokes of this article hold true for any reasonable model. While we cannot completely avoid the far-ranging consequences of this viral outbreak, mathematical modeling will help us know what to expect and how to prepare for and handle it.

The figures in this article were provided by the author.

References

[1] Hatcher, R.J., Mecher, C.E., & Lipsitch, M. (2007). Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Nat. Acad. Sci., 104*(18), 7582-7587.

[2] Hethcote, H.W. (2000). The mathematics of infectious diseases. *SIAM Rev., 42*(4), 599-653.

[3] Kermack, W.O., & McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. London. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700-721.

David I. Ketcheson is an associate professor of applied mathematics and computational science at King Abdullah University of Science and Technology (KAUST). He has written several blog posts about the mathematical modeling of COVID-19, which are available here.