Criticality in the COVID-19 pandemic: an introduction to the SIR (susceptible-infected-recovered) model of disease spread

James E. Ferrell, Jr.
Dept. of Chemical and Systems Biology
Stanford Medicine
Stanford CA 94305-5174
James.ferrell@stanford.edu

Summary
Mathematical modeling can help us to understand epidemic illnesses like COVID-19/SARS-CoV-2 and to see how various interventions might affect their course. Here we work through the so-called SIR model (for susceptible-infected-recovered) for COVID-19. One key property of the model is that there is a critical value for R0 (the number of secondary infections per primary infection), above which the disease is expected to ultimately affect much or most of the population, and below which the disease will be extinguished. This underscores the importance of social distancing, which can decrease the number of secondary infections, in the battle against this potentially catastrophic scourge.

The now-standard mathematical model for the time-evolution of an infection, the SIR model, was developed in the 1920's by the Scottish epidemiologists Anderson Gray McKendrick and William Ogilvy Kermack (*1*). In its simplest form, the model divvies up the population (of a city, a county, a country, or the earth) into three compartments or pools (Figure 1). Those who are susceptible to the disease are said to be in the *S* pool, and in our case that is initially everyone, since for a new disease like COVID-19 the expectation is that none of us are immune. Those who have caught the virus and are infectious are in the *I* pool. And those who had caught the virus but are no longer infectious, either because they have recovered to health or have died, are in the *R* pool. The conversions from one pool to another are assumed to be one-way processes.

With these assumptions stated, we can write equations for the rates of infection and recovery. We assume that the rate of infection is directly proportional to the fraction of the population that is susceptible (*S*) and the fraction that is contagious (*I*):

$$rate\ of\ infection = k_1 S \cdot I \,. \qquad \text{Eq. 1}$$

Note that there is positive feedback built into this expression: the rate of infection goes up as the fraction of the population that is infected (and infectious) goes up.

The rate of recovery is proportional to I:

$$rate\ of\ recovery = k_2 I\ . \qquad \text{Eq. 2}$$

We can then combine Eqs 1 and 2 to yield ordinary differential equations (rate equations) for the net rate of change of each of the three time-dependent species (S, I, and R):

$$\frac{dS}{dt} = -k_1 S \cdot I \qquad \text{Eq. 3}$$

$$\frac{dI}{dt} = k_1 S \cdot I - k_2 I \qquad \text{Eq. 4}$$

$$\frac{dR}{dt} = k_2 I\ . \qquad \text{Eq. 5}$$

There are only two adjustable parameters in these equations, the two proportionality constants: k_1, which is a measure of the rate of infection, and k_2, which determines the rate of recovery. The ratio of k_1 to k_2 tells us how effective the virus is in the face of host and societal defenses, and it is traditionally termed R0 (R naught). R0 specifies how many secondary infections, on average, are going to be produced from each primary infection; it can be viewed as the basic reproduction number for the virus. For COVID-19, R0 is estimated to be 2 to 3 (*2*); for comparison, R0 ≈ 1.3 for influenza, a less highly contagious disease.

The easiest way to get a feel for the behavior of the model is to solve the rate equations numerically for various choices of the parameters, and graph the calculated time courses. This is shown in Figure 2. We started the simulation with $I[0]$ = 1/330,000,000, which means that initially there is a single infected individual in a country with a population of 330 million (like the US).

From the linear plot in Figure 2A, it looks like S starts to drop and I and R start to increase at around 12 weeks after the initial infection. However, the semi-log plot (Figure 2B) shows that the number of infections actually increases right from the start, and that the increase is approximately exponential. From Eq. 4, it follows that the doubling time for the exponential increase is initially:

$$\tau = \frac{\mathrm{Ln}[2]}{k_1 - k_2}. \qquad \text{Eq. 6}$$

For the values of k_1 and k_2 we chose here— k_1 = 2.4 infections per week, and k_2 = 1.2 recoveries per week, which corresponds to an optimistic R0 = 2—the calculated doubling time turns out to be 0.693/1.2 = 0.58 weeks or 4 days (*3*), which is the current (as of March 17) doubling time for US deaths. In the US, documented cases are doubling faster than this (t ≈ 2.5 days), but that probably reflects an increase in screening as well as the actual increase in cases.

The modeled number of infections is maximal at ~16 weeks and then begins to drop approximately exponentially. If the first case of COVID-19 in the US began sometime between February 1 and 15 2020, this would put the peak of cases at some time between May 23 and June 6 2020, and the peak of deaths about two weeks later. Again from Eq. 4, the half time for the exponential decrease is:

$$\tau = \frac{\mathrm{Ln}[2]}{S_\infty k_1 - k_2}, \qquad \text{Eq. 7}$$

where S_∞ denotes the fraction of the population that is still susceptible to infection after the epidemic has run its course (calculated to be 0.20 for values of k_1 and k_2 assumed here).

At the peak of the modeled epidemic, about 15% of the population is infected, a huge percentage. And by the end of the epidemic, 80% of the population will have been infected, again a huge percentage (note that *R*, the fraction of the population that has recovered, is the same as the fraction of the population that at some point had been infected) (Figure 2). Assuming a case fatality rate of 1%, ultimately 0.8% of the country's population—2,640,000 people out of 330,000,000—would die. This would be an incredibly catastrophic scenario.

The peak number of infected individuals, and the ultimate number of individuals to become infected, depend only upon R0, not the individual values of k_1 and k_2. This is one of the reasons epidemiologists care so much about what the value of R0 is.

So what would be expected if a country instituted a social distancing policy that reduced the number of secondary cases per primary case from its nominal value of R0 ≈ 2 to something less? A relatively modest decrease in R0, from 2 to 1.5, would "flatten the curve" (*4*), shifting the peak of infections from 16 weeks to 31 weeks and decreasing the peak height from 15% to 6% (Figure 3). This would not only

buy the country some time, but also make it so that the worst days of the infection were less likely to overwhelm the health care system, and therefore improve the ultimate mortality rate.

Moreover, the proportion of the population that would ultimately be infected (and hence the proportion that would ultimately die of the infection) would drop, from 80% to 58% (Figure 3). This is an improvement—58 million infections is certainly better than 80 million—but this hypothetical scenario is still catastrophic; it would result in hundreds of thousands of deaths.

But if R0 can be lowered to 1 or less, something remarkable happens: the peak of infected individuals vanishes, and the proportion of the population that will ultimately become infected drops precipitously. The disease fizzles out instead of growing exponentially. For R0 = 1, our initial case would ultimately result in ~26,000 cases—only 0.008% of the population (Figure 3). And if R0 = 0.5, our initial case would, on average, result in only a single additional case.

The relationship between the assumed value of R0 and the cumulative number of cases that would have occurred after an infinitely long time is plotted in Figure 4. Below R0 = 1, relatively few people get sick, and then once R0 exceeds 1, the number of cases begins to rise sharply. This behavior is really the essence of this type of dynamical system. The positive feedback makes it so that the system has a critical point (a transcritical bifurcation, in the lexicon of nonlinear dynamics) that separates one type of dynamical behavior—the infection fizzles out—from a qualitatively different type of behavior—the infection explodes. This change in behavior is analogous to one type of phase transition seen in biology and physics, which is sort of interesting, but the main thing is that it really matters which side of the bifurcation we are on.

Of course this is all predicated on the assumption that the SIR model is actually applicable to the COVID-19 pandemic. And, to be sure, some of the assumptions built into the model are suspect. For example, the model assumes that the population of the county, state, country, or whatever, is a well-mixed system, where every person interacts with every other person on the time scale of the epidemic. This is clearly not true—there is spatial structure to the evolving epidemic. But the most basic lessons of the model, that the infectiousness of a virus determines whether the infection will fizzle out or explode, and that the infectiousness also determines the fraction of the population that will ultimately become infected, probably are true.

So is there any reason to think that it might be possible to decrease COVID-19's R0 value below 1? The answer is yes, unquestionably, and the evidence comes from the dynamics of the epidemic in China, the first country beset by COVID-19. China is a country of 1.3 billion people, and a virus with R0 = 2 should, in principle, eventually infect 1 billion of them. With a case fatality rate of 1%, that would mean 10 million deaths. But the number of COVID-19 cases in China appears to be leveling off at around 100,000 cases and 3200 deaths (*3*). Assuming these numbers to be accurate, the actual number of cases is 10,000-fold lower than what the model says should be expected, and possibly ~10 million lives have been saved. This is almost certainly the result of the draconian measures taken by the government to prevent person-to-person spread; they must have decreased R0 by a lot, and consequently they have halted the epidemic. The situation in the Republic of Korea is similar: in a country of 51 million people, the epidemic is leveling off at less than 10,000 cases, not the 40,000,000 expected for a disease that is this infectious.

And this is why it is a sign of great hope that around the world, countries, states, businesses, and schools are implementing policies with a real likelihood of substantially decreasing R0. Surely if those individuals known to be infected (through diagnostic screening of people with COVID-19 symptoms and the contacts of those known to be infected) were to be rigorously quarantined, and those who might unknowingly be infected were to decrease their number of daily contacts by, say, a factor of 5 or so, it should be possible to hugely mitigate the millenium's worst global health crisis.

Acknowledgements

We thank Michael Lin, Mingyu Chung, and members of the Ferrell lab for helpful comments. This work ris supported by a grant from the National Institutes of Health (R35 GM131792).

References

1. W. O. Kermack, A. G. McKendrick. A contribution to the mathematical field of epidemics. *Proc Royal Soc A* **115**, 700 (1927).
2. S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, D. Chen. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int J Infect Dis.* **93**, 201 (2020) [Epub ahead of print]

3. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19 2020. Accessed March 13 2020.
4. R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* https://doi.org/10.1016/ S0140-6736(20)30567-5 [Epub ahead of print]

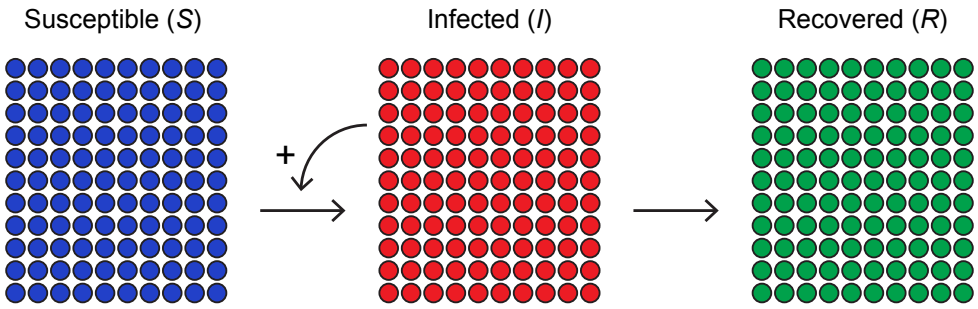

Figure 1. The SIR model of infectious disease. Individuals in the population get transferred from susceptible (*S*) to infected (*I*) to recovered (*R*) pools through one-way processes.

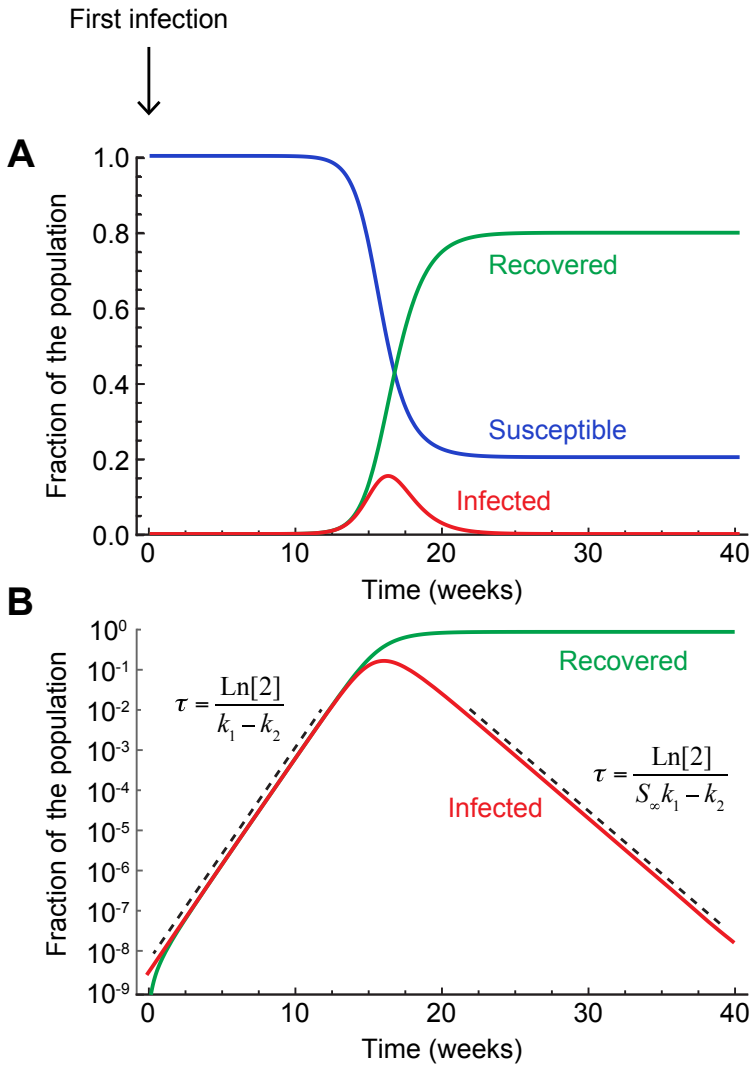

Figure 2. Simulated infection dynamics for the COVID-19 pandemic. We have assumed am infection rate constant k_1 of 2.4 infections per week and a recovery rate constant k_2 of 1.2 per week, which means that R0, the basic reproduction number, is 2. We assumed that at time zero there was one infected individual in a population of 330,000,000.

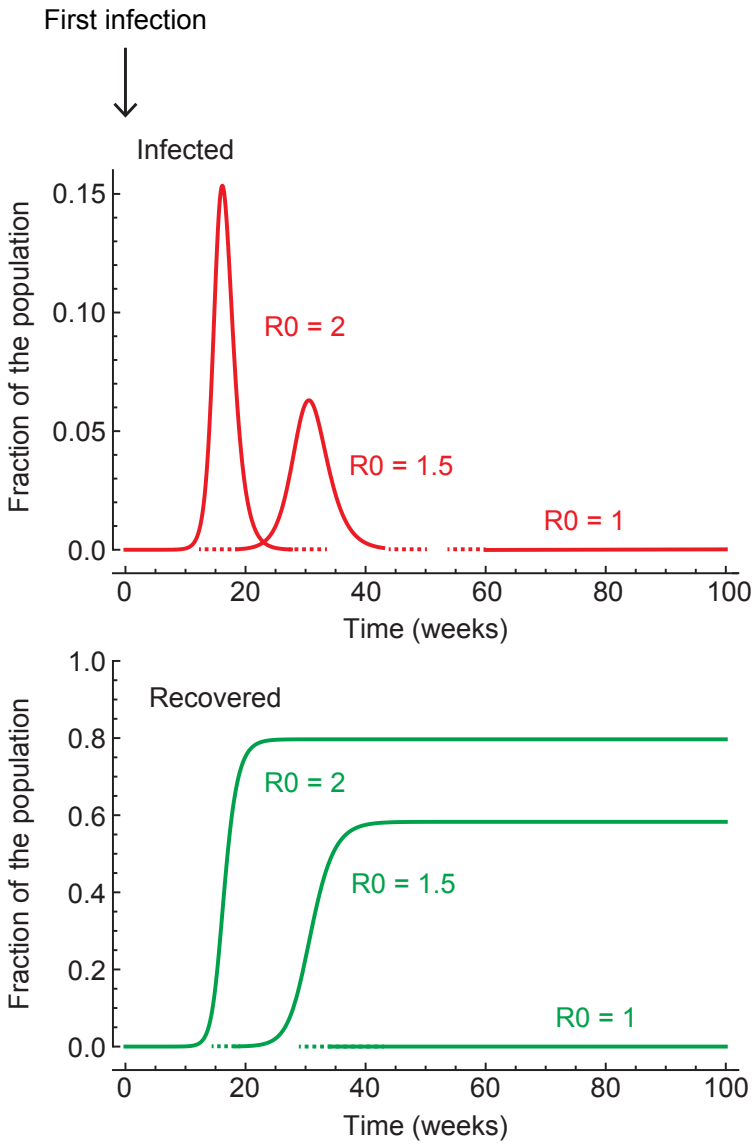

Figure 3. Decreasing R0, for example through social distancing, both "flattens the curve" (top) and decreases the total number of infections (bottom). Note that the fraction of the population that has recovered is the same as the fraction that has had the infection. We have assumed a country of 330,000,000 people with a single individual infected at time zero, a recovery rate constant k_2 = 1.2 week^{-1}, and values of k_1 to make R0 = 2, 1.5, or 1.

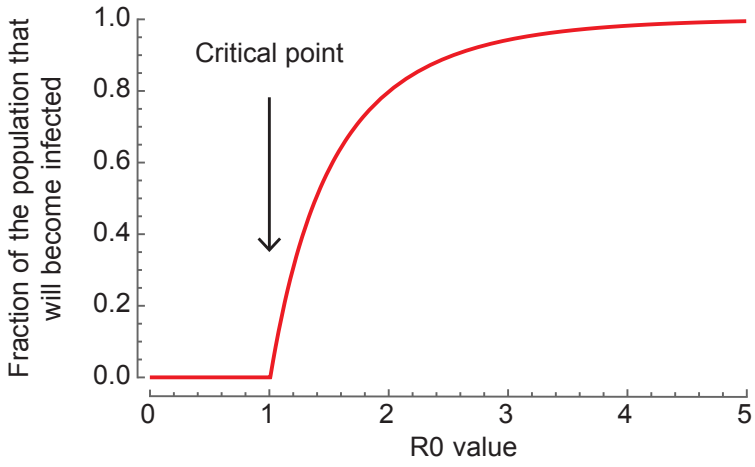

Figure 4. Criticality in the relationship between the ultimate number of infected individuals and the assumed R0 value. Decreasing R0 through social distancing can change an epidemic into an infection that will fizzle out.