

What models can and cannot tell us about COVID-19

Alexander F. Siegenfeld^{1,2} and Yaneer Bar-Yam²

¹*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA and*

²*New England Complex Systems Institute, Cambridge, MA*

(Dated: April 8, 2020)

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has already claimed over 75,000 deaths worldwide [1], and if left unchecked, will claim many, many more. Models can help us determine how to stop the spread of the virus, but it is important to distinguish between that which models can and cannot predict. All models' assumptions fail to describe the details of most real-world systems, but these systems may possess large-scale behaviors that do not depend on all these details [2]. A simple model that correctly captures these large-scale behaviors is useful; a complicated model that gets some details correct but mischaracterizes the large-scale behaviors is misleading at best. Large-scale behaviors of the COVID-19 pandemic include the rate of exponential growth/decay in the number of active infections in each region, as well as the transmission rates between regions. The values of these parameters, both of which can be controlled with interventions, determine whether the large-scale behavior of COVID-19 is that of exponential spread until saturation or exponential decay until elimination. We may not be able to precisely predict the trajectory of the epidemic under any given set of interventions, but we know that a strong enough set of interventions can ensure we are in the latter regime.

Understanding what models cannot predict is sometimes more important than understanding what they can. For example, in a chaotic system such as the weather, only very short-term predictions are accurate because small changes in the present can result in very large changes in the future. The trajectory of the COVID-19 pandemic is another example: because the number of infections depends exponentially on the growth rate of the epidemic, small inaccuracies in the prediction of the growth rate will lead to large changes in the number of deaths after enough time. Furthermore, the growth or decay rate of the epidemic depends on the precise implementation details of interventions, and a very small change in the strength of interventions could be the difference between two hugely different outcomes: exponential growth until saturation versus exponential decay until elimination. Gaining an approximate understanding of the trajectory of the epidemic is important, but, when there is so much uncertainty arising from underlying disease and social dynamics in addition to the uncertainty over exactly how interventions will be implemented, detailed refinements to models are not.

More generally, spending effort trying to pin down de-

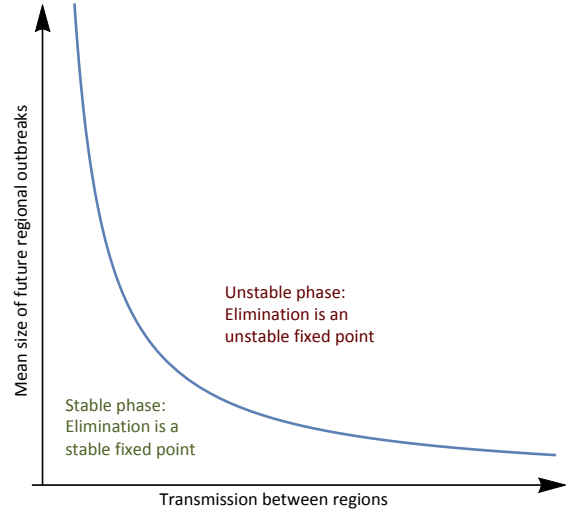


FIG. 1. A collection of geographic regions can exist in one of two phases with respect to COVID-19. If strong enough lockdown measures (which may include testing/contact tracing/quarantine) are imposed, the virus can be eliminated from currently infected regions. The question is then whether this elimination is stable or whether the number of cases will rebound after the lockdown is lifted. Whether or not elimination is stable depends on (1) the average total number of cases that will result from the disease being transmitted to a region, which in turn depends on (among other factors) how quickly regions locally lock down if they are infected or reinfected, and (2) the probability that an infected individual in one region will infect an individual in another, which in turn depends on the rate of travel between regions [4, 5].

tails in models is futile if any accuracy gained is swamped by uncertainty in the measurements or by inaccuracies in the core model assumptions. What is the purpose of refining a model by 10% if there is a 50% uncertainty stemming from other aspects or assumptions of the model, or if there is a relevant behavior of the modeled system that the model fails to capture all together? Models that attempt to capture a system's small-scale detailed behavior (e.g. ref. [3]) will inevitably include some details and leave out others. Depending on which details are included, such models may mischaracterize the system's large-scale behavior, and when they do work, it is often because their specific assumptions are a special case of a simpler, more general model. Thus, sometimes it is not the complicated models but the deceptively simple ones that are most effective for understanding a system's large-scale behavior.

For COVID-19, one large-scale behavior is an exponential increase in infections in the absence of intervention (unless the number of people infected is approaching saturation), with the exact growth rate depending on the location and the precise details of disease transmission. Interventions may change this growth rate and strong enough interventions such as lockdowns may result in an exponential decay rather than exponential growth. Another large-scale behavior is the fact that transmission is predominantly local, with travel creating the possibility of long-range spread. The number of infections does not change uniformly all over the world at once but rather predominantly independently in each region, with the probability that the disease is transmitted from one region to another depending on the number of infections in the first region and the travel rate from the first to the second among contagious individuals. There are many small-scale details to the disease transmission process, but the large-scale dynamics seem to be captured by the rate of increase or decrease within a region and the rates of transmission between regions (both of which may change over time due to interventions, saturation effects, or other variations in external conditions). Depending on the dynamics of these parameters, a collection of regions can exist in one of two phases: a stable phase, in which the disease dynamics tend towards a stable fixed point of elimination (i.e. no infected regions), and an unstable phase, in which the number of infected regions grows until a saturation point is reached (see fig. 1).¹ In order for a collection of regions to be in the stable phase, it is not necessary for regions to be under constant lockdown after they have been cleared of the virus but rather only for each region to be ready to lockdown in the event that it is re-infected [5]. If the virus is introduced or re-introduced into a collection of regions in the unstable regime, the number of infected regions will exponentially grow, but if the virus is introduced or re-introduced into a collection of regions in the stable regime, the system will return to its eliminated state.

If the large-scale behaviors of a system are correctly described, specific details can be understood in terms of their effects on these behaviors. But if a model’s assumptions do not yield the same general large-scale behaviors of the system being modeled, adding additional details to the model will serve only to create a false sense of confidence. For example, models using continuous variables to represent fractions or probabilities of individuals being infected may predict that although a lockdown

can produce an exponential decline in cases, the number of cases will inevitably rebound once the lockdown is lifted. However, the assumption of approximately continuous behavior breaks down for small numbers of infections (which exponential decline will inevitably bring about given sufficient time). If small numbers appear only in the final output of the model, human judgement can correct for the error (e.g. by interpreting a fraction of a case in the model as the virus having been eliminated in reality). But if these small numbers arise as intermediate values in the model, the model will predict exponential growth once the lockdown is lifted, despite the fact that the model is no longer valid in this regime and there may in fact be zero cases.² A rebound in infections after lockdown measures are lifted is a potential large-scale behavior of the system, but it is not inevitable (as predicted by continuous models) but rather depends on our actions: if interventions strong enough to create an exponential decay in the number of active infections are held in place for a sufficient amount of time, the virus will be eliminated.³

It may be objected that even if the fraction of the population infected becomes very small, if the size of the population being considered is large enough, then the number of cases will nonetheless be large enough to be approximated as continuous. However, models often consider the entire population of a country together. In reality, disease transmission is far more local (and can be made even more so with lockdowns and travel reductions) and so the sizes of the populations for which the models apply will be far smaller than that of an entire country. Thus, the locality in the dynamics (the degree of which can be increased by travel restrictions) makes it more likely that a small fraction of the population infected in the model corresponds to the virus being eliminated in reality, and it also allows for the lockdown to be lifted region by region, rather than remaining in all regions until the entire country is cleared of the virus. Ultimately, the specific detailed assumptions of particular models will vary; what matters is not the particular assumptions, but whether or not they appropriately characterize the large-scale behaviors described above.

Finally, “What will happen?” is a different question than “What should we do?” and for COVID-19 the latter question is far easier to answer than the former. In the absence of a full understanding of a system’s details, answering the latter question involves understanding how our potential actions impact the relevant large-scale parameters of the system, which for COVID-19 are the rate of exponential growth or decay in each region and the

¹ A lockdown within a single geographic region can itself be analyzed using fig. 1, if each household is considered as a “region.” In this case, the mean size of an outbreak would be the average number of individuals within a household expected to get COVID-19 if one individual in the household is infected, with the disease transmission between “regions” corresponding to the probability that an infected individual in one household has of infecting an individual in a different household. The primary purpose of a lockdown is to control this probability.

² The virus may still be re-imported, but if elimination is a stable fixed point of a collection of regions (fig. 1), the number of regions with non-zero infections will decrease to zero over time.

³ The elimination of the virus can be hastened by testing, contact tracing, and quarantine, which may become more feasible and/or effective once the number of infections has been sufficiently reduced.

probabilities of transmission between regions. Even if we cannot precisely predict the impact of any given intervention, we know of many interventions that will reduce the rates of transmission within and between regions, and, based on our empirical understanding of COVID-19 transmission and the fact that some countries have seen declines in new cases, we know that combining enough of them together will reduce the rate of transmission sufficiently to achieve exponential decline and stop the out-

break [6]. This, in and of itself, is a simple but powerful formal model that captures the large-scale behaviors of interest. The question of the disease trajectory is less important than the question as to what can be done to (1) cause an exponential decrease rather than increase in new infections and (2) cause this decrease to be as fast as possible. The point is not the specific predictions for each intervention but that together they can eliminate the virus.

-
- [1] Worldometer. COVID-19 coronavirus outbreak. <https://www.worldometers.info/coronavirus/>. Accessed 2020-4-7.
- [2] Yaneer Bar-Yam. From big data to important information. *Complexity*, 21(S2):73–98, 2016.
- [3] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College COVID-19 Response Team*, 2020.
- [4] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, 7(1):46–89, 1997.
- [5] Alexander F Siegenfeld and Yaneer Bar-Yam. Eliminating COVID-19: A community-based analysis. *arXiv:2003.10086*.
- [6] Harvey V. Fineberg. Ten weeks to crush the curve. *New England Journal of Medicine*, 2020.