Leveraging Rich Data and Machine Learning to Facilitate Policymaking on Public Health Interventions in Fighting to Contain the COVID-19 Pandemic

Robin G. Qiu

Penn State

robinqiu@psu.edu

The COVID-19 pandemic has caused painful social and economic disruption around the world, including the largest global recession since the Great Depression. On June 18, 2020, over 150,000 new cases as a new daily record reported worldwide. The World Health Organization at its briefing on June 19, 2020, stated that the COVID-19 pandemic has entered a "new and dangerous phase". Epidemiologists have warned and suggested that public health interventions, like social distancing, travel restriction, face covering mandate, and lockdown as intervention measures would be the only "weapon(s)" to fight COVID-19 effectively before a COVID-19 vaccine or effective treatment becomes widely available. The voice is crystal clear, but hardly heard or simply/partially ignored in places where politics sets the tone, unfortunately. Therefore, identifying effective means of fighting to contain the COVID-19 pandemic becomes extremely imminent.

The compartmental SEIR (**S**usceptible, **E**xposed, **I**nfectious, and **R**ecovered) model (Fig. 1) and the like in epidemiology prove effective in facilitating policymaking on intervention measures enforcement and healthcare planning to fight a pandemic at different control levels, regionally, nationally, or internationally (1). However, the notorious reproduction number (R_0), an epidemiologic metric used to describe the contagiousness or transmissibility of infectious agents, in an SEIR model has been challenging to get accurately estimated due to the complexity of measuring and estimating R_0 (2-3). R_0 for COVID-19 is a function of all the parameters (e.g., λ, δ, μ, γ, p, and time) in Fig. 1. Without a vaccine or effective treatment, R_0 dictates the curve shape of a pandemic, while public health interventions determine the change of R_0 over time, i.e., when and how much R_0 drops (1-3).

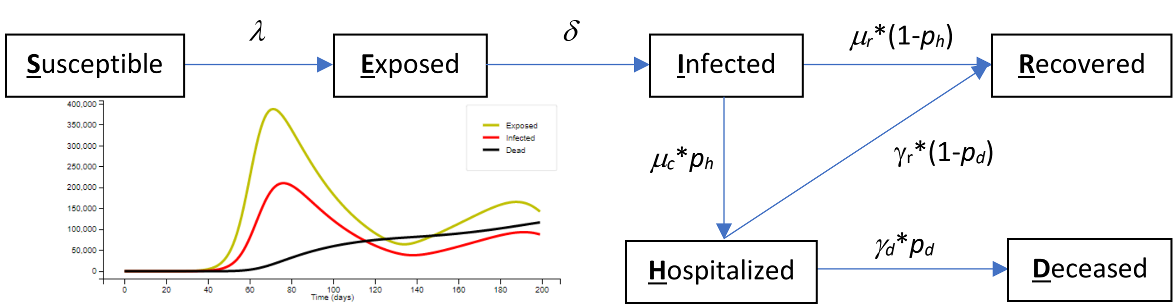

Fig. 1 An Extended SEIR model (λ: expected number of persons an infected person infects per day, δ: 1/days to get infected, also called incubation time, μ_r: 1/days to recover without hospitalization, μ_c: 1/days to be hospitalized from infection, p_h : percentage of Infected persons to be hospitalized, γ_r : 1/days to recover from hospitalization, p_d : percentage of Infected persons to die, γ_d: 1/days of hospitalized patients to die. R_0 is contextual, a function of time and these above-mentioned parameters.)

The recent study from Dehning et al. (2) has clearly shown that appropriate public health interventions enacted and enforced promptly in Germany proves effective in containing the COVID-19 pandemic in Germany. They show that their approach using Bayesian inference with Markov-Chain Monte Carlo (MCMC) by leveraging the available rich pandemic data can help quantify R_0 and other key parameters in SEIR modeling, which further helps assess the effectiveness of public health interventions. It is noteworthy that the proposed approach proves to be practically useful in terms of quantifying the time and rates of R_0 dropping owing to different restrictive levels of intervention measures that have been adopted over their study period.

Flattening the new case curve with intervention while increasing healthcare capacity over time and considering social and economic impacts has become the essential strategy around the world (1). R_0 for COVID-19 is highly related to the spreading rate of the coronavirus and the recovering rate of COVID-19 patients. The former rate changes along with the mutation of the virus and the timing, enforcement, and types of public health interventions as time goes; the latter rate varies with the communities where population and healthcare systems differ considerably (1-3). Thus, weaponizing and implementing public health interventions face tremendous challenges, substantially varying with communities, locally, nationally, and internationally (1-3). Unfortunately, this is particularly true in the US. As of today, the outcomes of fighting COVID-19 across the US are substantially different from state to state as witnessed since the beginning of March when the outbreak started in the US. The northeastern corridor centered at New York city had been the epidemic center since March until the end of May. Gratifyingly, we have seen how the COVID-19 across the region has been contained due to the enforcement of strict intervention measures. However, since this past Memorial Day, new coronavirus infections and hospitalizations are rising in many other states. The latest data clearly shows that COVID-19 hospitalizations have begun to swell in places like Texas, Arizona, Florida, and California. On June 18, 2020, more states, including Arizona, North Carolina, Texas, and Florida finally mandated face coverings in public to get a grip on spiraling coronavirus cases as those states continue to set daily records since this past Memorial Day.

Although the reasons for this unfortunate consequence might be associated with political and economic considerations and sometimes individual beliefs, the lack of data-driven approach that can better measure and estimate R_0 in developing effective SEIR and corresponding social and economic models that can accurately show the consequences in a scientific and timely manner can be a key contributing factor. Even worse, R_0 has been quite misleading due to too much unknown about the COVID-19. Consequently, policymakers tend to be influenced by their partisan stance rather than the science and reality because of lacking the needed scientific evidence and support.

We understand that inadequate intervention could result in a resurgence of COVID-19 pandemic locally and nationally (Fig. 1). One-size-fits-all lockdown might not be an optimal policy either. In the US, COVID-19 hot spots have shifted to the south where population is with more retired people who tend to have some chronic underlying conditions. As a result, more infected persons would become more seriously ill, leading to more hospitalizations and heavily taxed hospital systems. This will be surely more "worrisome" in the fall and winter when flu viruses circulate routinely, jeopardizing the healthcare systems across the country.

"History never repeats itself but it rhymes," said Mark Twain. Dehning et al. (2) reveal that intervention measures have been the greatest "weapon(s)" to fight COVID-19 effectively in Germany. We believe that more research on improving SEIR modeling will play a critical role in facilitating the policymaking on public health interventions. As the COVID-19 pandemic data is daily updated state by state, the currently available rich COVID-19 pandemic data and machine learning algorithms can thus be leveraged to extend and enhance SEIR modeling with parameter auto-adjustment capability weekly if not daily. R_0, a time- and location-dependent epidemiologic metric, along with other parameters (Fig. 1), can be more accurately estimated at the state level and the local community level if needed. More promisingly, this kind of study can be further enhanced and used to facilitate policymaking on the timing and restrictive levels of intervention measures, types and lifting or hardening, needed for a controlled area, at the state, region, or country level as time goes. Rich data and machine learning work well in reducing discrepancies. Because of knowing all truth that is highly context-dependent, promisingly and confidently, we can win this "war" and be well prepared for the future.

References

1. M. Enserink, K. Kupferschmidt. With COVID-19, modeling takes on life and death importance. Science **367**. 1414-1415 (2020).

2. J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, V. Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* 10.1126/science.abb9789 (2020).

3. K. Linka, M. Peirlinck, E. Kuhl. The reproduction number of COVID-19 and its correlation with public health interventions. *medRxiv* (2020). May 10, 2020: https://doi.org/10.1101/2020.05.01.20088047