Analytics and Prediction Modeling During the COVID-19 Pandemic

Douglas W. Challener, MD; Sean C. Dowdy, MD; and John C. O'Horo, MD, MPH

From the Division of Infectious Diseases (D.W.C., J.C.O.), Division of Gynecologic Oncology (S.C.D.), and Division of Pulmonary and Critical Care Medicine (J.C.O.), Mayo Clinic, Rochester, MN.

The ongoing coronavirus disease 2019 (COVID-19) worldwide pandemic has generated substantial interest in mathematical and predictive modeling for infectious diseases. These models have been used to predict the course of the epidemic, inform disaster preparedness planning, forecast the economic outlook, and allocate limited resources, such as personal protective equipment (PPE) and testing supplies. Coronavirus disease 2019 has also made the limitations of modeling evident. It is important to understand potential use cases, current modeling strategies, and underlying assumptions of existing models, as more reliance will be placed on these tools over the coming months.

Since the beginning of the pandemic, multiple groups have published predictive mathematical models for COVID-19, which vary substantially in the method of construction and in predicted outcomes. The Columbia University model, often cited by the *New York Times*, is based on classic epidemiology theory with a susceptible-exposed-infected-recovered framework. This model divides the population into 4 categories: susceptible, preinfectious (or exposed), infectious, and recovered (and presumed immune).[1] Another early model from the Imperial College London, based on traditional epidemiology with Bayesian-based Monte Carlo simulation, forecasted large numbers of British deaths in several projected scenarios and likely prompted a change in UK policy.[2,3] The Institute for Health Metrics and Evaluation (IHME) model is based on matching regional and demographic data within the United States to worldwide locations further along in the epidemic and has thus far led to more optimistic models than either of the 2 mentioned above.[4] The IHME modelers believe that death rates are more accurate than case rates. By using information from countries that have already passed the peak of the pandemic, the IHME model requires fewer starting assumptions. Little agreement exists among most models, and predicted outcomes may fluctuate daily. However, the progression of the pandemic will allow for performance comparisons of many of the currently published models.

Traditional model building relies on many assumptions that should be considered when interpreting a model's predictions. One of the most basic factors to consider when designing models of infectious diseases is an understanding of the infection's natural history. For example, the presence of an asymptomatic infectious period and the possibility of reinfection after recovery will dramatically alter the basic structure of the model. These factors may be difficult to know in an emerging disease and are still being determined for COVID-19. After the basic structure of the model has been determined, more assumptions are made to estimate starting conditions and the flow of patients between the various states (ie, susceptible, preinfectious, infectious, and recovered). In COVID-19, the number of individuals in each group has been difficult to estimate because of a lack of widespread testing and large numbers of asymptomatic infections.[5] The significance of asymptomatic infections has also shifted over time and is now thought to be one of the main factors driving community spread. Other parameter estimations describing the flow of individuals between disease states can

change with public health policy. For example, the flow of people from the susceptible group to the exposed group is a function of the rate at which 2 individuals come into contact per unit time as well as the number of currently infectious individuals within the population. Therefore, a stay-at-home order limiting mobility can change this parameter rapidly. Additionally, population behaviors and policy interventions are not geographically uniform, posing challenges when modeling a large and diverse country such as the United States. Difficulties in detecting asymptomatic and mild cases of the disease have also complicated estimations of COVID-19 mortality rates, adding more uncertainty to the model. Any small errors in the underlying model assumptions are compounded over time, making long-term forecasting often inaccurate.

One of the greatest challenges with any model is determining the current location on the epidemic curve.[6] The initial growth is slow before it reaches an inflection point at which apparent linear growth appears exponential. Determining whether we are in a linear approximation phase of the epidemic vs an exponential phase has substantial implications for the predictive model. Mayo Clinic has worked with internal and external analytics partners and laboratory colleagues to undertake an extensive analysis of current epidemiological data to allow a better understanding of the current state assumptions. Despite the uncertainty in the prediction models, they are of central importance in directing many aspects of the response to COVID-19.

The most important purpose of the models is to inform institutional and nationwide efforts to ensure patient safety. Models of infection cases can inform the implementation of public policy measures designed to reduce the spread of the infection, such as social distancing and closure of nonessential businesses. They can also help guide institutional efforts to care for patients by ensuring adequate hospital beds, PPE, testing resources, and staffing. Predicting surges in infected caseloads can help determine whether outpatient visits and operations should be scaled back to ensure adequate resources to care for the infected or allowed to grow to prepandemic levels.

Because of the substantial variation, Mayo Clinic has used multiple models to estimate the future burden of COVID-19 in our practice. This includes the aforementioned Columbia University and IHME models in addition to internally developed models that blend institution-specific data such as length-of-stay and hospitalization rates with state-based data such as rates of testing. In general, the models have shown a great deal of agreement for the near future despite different underlying assumptions. We have largely relied on models to help plan for a 2-week horizon, as the considerable divergence after this point and the rapid rate of change have made longer-term epidemic modeling less reliable.

We have also used modeling to estimate resource requirements, such as PPE and use of testing. Patient volumes, staffing levels, guidelines for PPE utilization, and observations of actual usage in various clinical scenarios have all helped model future PPE consumption. This strategy ensures adequate supplies across all Mayo Clinic sites and assists in equitable distribution of PPE among sites.

Models are useful tools as long as the underlying assumptions and reasons for substantial divergence are understood. Policymakers and physicians must understand the basic assumptions underlying predictive models to use them effectively. Continued validation, recalculation, and, *critically*, education will allow mathematical modeling to assist in the response to COVID-19.

Correspondence: Address to John C. O'Horo, MD, MPH, Division of Infectious Diseases, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (OHoro.John@mayo.edu; Twitter: @jcohoro).

ORCID
Douglas W. Challener: https://orcid.org/0000-0002-6964-9639; John C. O'Horo: https://orcid.org/0000-0002-0880-4498

REFERENCES

1. Branas CC, Rundle A, Pei S, et al. Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties [published online ahead of print April 6, 2020]. *medRxiv*, https://doi.org/10.1101/2020.04.01.20049759.

2. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*. 2020;368(6490):489-493.

3. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. *Report 9 — Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand.* London: Imperial College London; 2020.

4. IHME COVID-19 health service utilization forecasting team, Murray CJL. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries [published online ahead of print April 26, 2020]. *medRxiv*, https://doi.org/10.1101/2020.04.21.20074732.

5. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic population. *N Engl J Med*. 2020; 382:2302-2315.

6. Ma J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect Dis Model*. 2020;5:129-141.