COMMENTARY



Individual model forecasts can be misleading, but together they are useful

Caroline O. Buckee^{1,3} · Michael A. Johansson^{2,3}

Received: 11 July 2020 / Accepted: 17 July 2020 © Springer Nature B.V. 2020

The broad use by media and governments of model forecasts to inform the COVID-19 response has been a prominent and controversial feature of the pandemic so far. In this issue, Chin et al. compare the accuracy of four high profile models that, early during the outbreak in the US, aimed to make quantitative predictions about deaths and Intensive Care Unit (ICU) bed utilization in New York [1]. They find that all four models, though different in approach, failed not only to accurately predict the number of deaths and ICU utilization but also to describe uncertainty appropriately, particularly during the critical early phase of the epidemic. While overcoming these methodological challenges is key, Chin et al. also call for systemic advances including improving data quality, evaluating forecasts in real-time *before* policy use, and developing multi-model approaches.

The authors reveal substantial variability in "ground truth" data; epidemiological surveillance data used for both building and evaluating forecasting models. Coupled with uncertainty about basic epidemiological parameters of SARS-COV-2 as well as limitations in model frameworks, it is not surprising that such models have the potential to generate inaccurate forecasts. Improved data quality can certainly help improve model predictions, but forecasts are often needed in moments where surveillance systems that generate key data are new, imperfect, and rapidly changing.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Caroline O. Buckee cbuckee@hsph.harvard.edu

- ¹ Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard TH Chan School of Public Health, Boston, USA
- ² Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico
- ³ COVID-19 Modeling Team, Centers for Disease Control and Prevention, Atlanta, GA, USA

These uncertainties need to be integrated into the forecast itself. Moreover, the additional sources of uncertainty associated with the model—parameter uncertainty and structural uncertainty—also need attention, and are often dealt with superficially. Taken together, these challenges may lead some to question the use of forecasts for policy making in the first place.

But what the model comparison by Chin et al. highlights is an important principle that many in the research community have understood for some time: that no single model should be used by policy makers to respond to a rapidly changing, highly uncertain epidemic, regardless of the institution or modeling group from which it comes. Due to the multiple uncertainties described above, even models using the same underlying data often have results that diverge because they have made different but reasonable assumptions about highly uncertain epidemiological parameters, and/or they use different methods. While there are clear red flags indicating potential problems-for example, as Chin et al. point out, the fact that some forecasts had uncertainty estimates that decreased into the future-it can be challenging, and indeed sometimes impossible, to know a priori which model approaches will work best.

One way noted by the authors to overcome these challenges is by using ensemble modeling approaches, in which estimates from multiple models are combined in a single forecast. Combining multiple models, either in an ensemble or in side-by-side comparisons, has been shown to provide more robust forecasts that account for different aspects of uncertainty across multiple pathogens, multiple outbreaks, and multiple years [2, 3]. Even simple ensemble approaches provide more robust forecasts than the majority of individual models. This is particularly important at the start of outbreaks, when forecasts are most useful to policy makers, and it is unclear what the "best" individual model is.

As the authors argue, the rapid deployment of this approach requires pre-existing infrastructure and evaluation systems now and for improved response to future epidemics. Many models that are built to forecast on a scale useful for local decision making are complex, and can take considerable time to build and calibrate, not to mention the fact that the more epidemiological uncertainties exist, the more possibilities must be modeled. It is no surprise that a group with a history of successful influenza forecasting in the US (Los Alamos National Lab (4)) was able to produce early COVID-19 forecasts and had the best coverage of uncertainty in the Chin et al. analysis (80-100% of observations fell within the 95% prediction interval for most forecasts). In contrast, the new Institute for Health Metrics and Evaluation statistical approach had low reliability; after the latest analyzed revision only 53% of reported death counts fell with the 95% prediction intervals.

If the rapid development of multiple models is a reasonable solution to the enormous uncertainties associated with emerging epidemics, how can the necessary coordination occur? Building on years of work with forecasting other pathogens, progress is being made by many forecasting teams in a collaborative effort coordinated by the Centers for Disease Control and Prevention and the Reich Lab at the University of Massachusetts, Amherst (https://www.cdc. gov/coronavirus/2019-ncov/covid-data/forecasting-us.html, http://github.com/reichlab/covid19-forecast-hub). However, barriers remain, including the incentive structures that continue to pervade academic science. Academia generally recognizes individual scientists, not collaborative groups, and rewards novelty, which does not include recreating the models of other scientists or even making incremental yet important improvements to them by increasing their accuracy. A model that debunks or contradicts the main findings of an established model may be more likely to be considered for publication in a high-end journal or for competitive grant funding. This is antithetical to making real progress; new, innovative models are needed, but not at the cost of halting development, and deployment, when necessary, of existing models. Moreover, teams of scientists must be incentivized to produce multiple models of the same outbreak, and work with policy makers during outbreaks.

There have been long-standing and more recent discussions in the modeling research community about how to fund and structure institutions within or outside academia that prioritize disease forecasting, in recognition of these challenges. It remains to be seen whether this pandemic will galvanize efforts to develop such institutions. What is clear from Chin et al.'s analysis is that public health is not well served by the promotion of individual forecasting models. Incentive structures that drive the development of multiple models, with different data, assumptions, and structures, can enable science to move beyond the single model approach. Multi-model approaches drive the advancement of forecasting science, but also, more importantly, provide more robust information to public health decision makers when they need it most.

References

- Chin V, Samia NI, Marchant R, et al. A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York State. Eur J Epidemiol. 2020. https://doi. org/10.1007/s10654-020-00669-6.
- Reich NG, McGowan CJ, Yamana TK, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US. PLoS Comput Biol. 2019;15(11):e1007486.
- Johansson MA, Apfeldorf KM, Dobson S, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc Natl Acad Sci. 2019;116(48):24268–74.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.