

Discussion of “An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China”

DEBANGAN DEY¹ AND VADIM ZIPUNNIKOV*¹

¹*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Predictions during the early stage of an epidemic are essential to inform quarantine protocols, plan medical resources, and implement economic strategies. One of the major obstacles for making accurate predictions is imperfect available data on the current state of the disease dynamics typically summarized via the number of infected and recovered cases, and disease-related deaths. In the case of rapidly evolving COVID-19 pandemic, major challenges include under-reporting of infected and recovered cases due to the shortage of available tests, non-uniform performance and not sufficiently high sensitivity/specificity of currently available tests, a long incubation period, and a significant number of asymptomatic/unconfirmed cases, inconsistent accounting practices in death classification. All these make it quite challenging to accurately predict the future trajectory of this pandemic.

Putting aside the imperfections in the currently available data, choosing an appropriate model is another essential component of the modelling endeavor. Modifications of the classical SIR (Susceptible-Infectious-Removed) and SEIR (Susceptible-Exposed-Infectious-Removed) models are among the most popular modelling frameworks used during the early stage of this pandemic. Ironically, even the notorious Institute for Health Metrics and Evaluation at the University of Washington that originally proposed an misleadingly inflexible curve-fitting model ([The Institute for Health Metrics and Evaluation, 2020](#)) that received vigorous criticism within statistical and epidemiological communities (see ([Brookmeyer, 2020](#); [Jewell et al., 2020](#); [Jin et al., 2020](#); [Etzioni, 2020](#)) and many others), has eventually switched to SEIR-based modelling (<http://www.healthdata.org/covid/updates>) right around the time of writing this discussion (May 5th, 2020).

In their paper, Wang et al. have extended SIR (eSIR) model to incorporate changes based on the observed sequence of lockdown protocols implemented by local authorities. Although, the proposed eSIR model is relatively simple, we believe the model structure can accommodate simple, but realistic scenarios and provide a reasonable starting point in prediction modelling. We have applied eSIR model to Maryland data right after the model has been posted online and would like to discuss a few practical considerations. First, this model is sensitive to the pre-defined transmission rate modifiers, which are fixed hyper-parameters. In practice, these hyper-parameters have to be estimated from observed data using validation techniques such as minimizing deviance criterion or prediction mean squared error. It would be great to see this step implemented directly within the R package that accompanies the paper. Second, although, the turning-point prediction (peaks of the daily and cumulative incidence curve) is a nice component of the proposed eSIR model, their estimates varied significantly and were quite sensitive to the choice of quarantine or transmission modifiers. The authors extensively discussed the challenges of accurate peak prediction in the current first wave of the pandemic. Varied and often quite different predictions of the peaks in recent COVID-19 analyses are a testament to that fact. This likely implies that the actual peak will depend on the compliance with the lockdown protocols and, thus, can only be reasonably predicted with a continuous time period accounting for this uncertainty.

*Corresponding author. Email: vzipunni@jhsphe.edu.

It is worth mentioning many other extensions of SIR among which SEIR model that additionally includes an “Exposed” compartment and is considered to be more flexible and potentially more accurate due to taking into account a time lag between the time of being exposed and the time becoming infectious. Time-varying implementations of SEIR models have also been proposed (for example, (Teles, 2020)) which consider additional “asymptomatic” and “hospitalized” compartments. One practical suggestion for these dynamic models would be to calibrate their parameters using available testing data. This calibration step can help alleviate the problem of under-reporting cases to some extent, which the eSIR model does not address. In addition, the contact probability between a susceptible person and an infectious person can vary across ages and other demographics and properly accounting for that could help with more accurate prediction (Global Epidemic and Mobility Project, 2020). Finally, publicly available mobility reports across states and counties can provide local proxies for the lockdown compliance that can help with more accurate estimation of transmission rate modifiers under various social distancing measures (Google LLC, 2020; Apple Inc, 2020; Unacast, 2020).

From our perspective, a major contribution of the authors is the R package that implements the proposed eSIR model. We found it to be very helpful in exploring various hypothetical scenarios as done by us or in modifying the proposed model as done by others Jin et al. (2020). With a rapid "on-demand" model development happening almost real-time, transparency of proposed models, open access to used data and developed software are critical for external and independent reproducibility of the results and speedy validation of the proposed models. As an illustration of the practicability of the proposed eSIR model (time-varying transmission rate), we considered different scenarios for the State of Maryland and present projections based on different transmission rate modifiers. Governor Larry Hogan issued stay-at-home order for Maryland on the March 30th, 2020. Using a minimum deviance criterion, we estimated the transmission rate modifier as 0.38 between March, 30th and May, 1st of 2020. We assume that transmission rate varied as (1, 0.9, 0.6) before that date with change-points being March 12th (closure of bars, restaurants, gyms, movie-theaters, and sporting venues), March 23rd (closure of non-essential businesses). We considered a hypothetical reopening scenario starting at June 1st and ran the eSIR model across three different scenarios assuming post-reopening transmission rate modifiers to be 0.3 (“Strictly continuing”), 0.4 (“Slightly loosening”) and 0.5 (“Severely loosening”). Across the three scenarios, the estimated models, shown at Figures 1 and 2, project on average anywhere between 200K+ to 400K+ to 500K+ of the cumulative number of infected cases by the middle of September of 2020 compared to about 27K confirmed infected cases on May 5th, 2020.

Given on-going limitations with the currently available data on the number of exposed, infected, and recovered cases, epidemiologists started looking at more robust population level summaries such as excessive mortality. Arguably, even though there is inconsistent accounting of COVID-19 related deaths across regions and countries, the number of reported COVID-19 related deaths still may be a reliable measure to estimate the current state of the dynamic disease process. This is why we provide daily updates of the time-varying doubling times for number of deaths across different US states as well as different countries around the world (<https://bit.ly/dtlivecovid>). Our goal is to track state and country-level death trajectories to better understand local dynamics of the disease spread.

Now, when most of the countries, states, and cities are actively considering various scenarios of reopening, accurate statistical predictions updated in real-time are essential to inform critical decisions about public health and economics. When a large number of new modelling proposals are posted daily on open access pre-print repositories such as <https://arxiv.org/>, <https://www.biorxiv.org/>, <https://www.medrxiv.org/> and published in peer-reviewed jour-

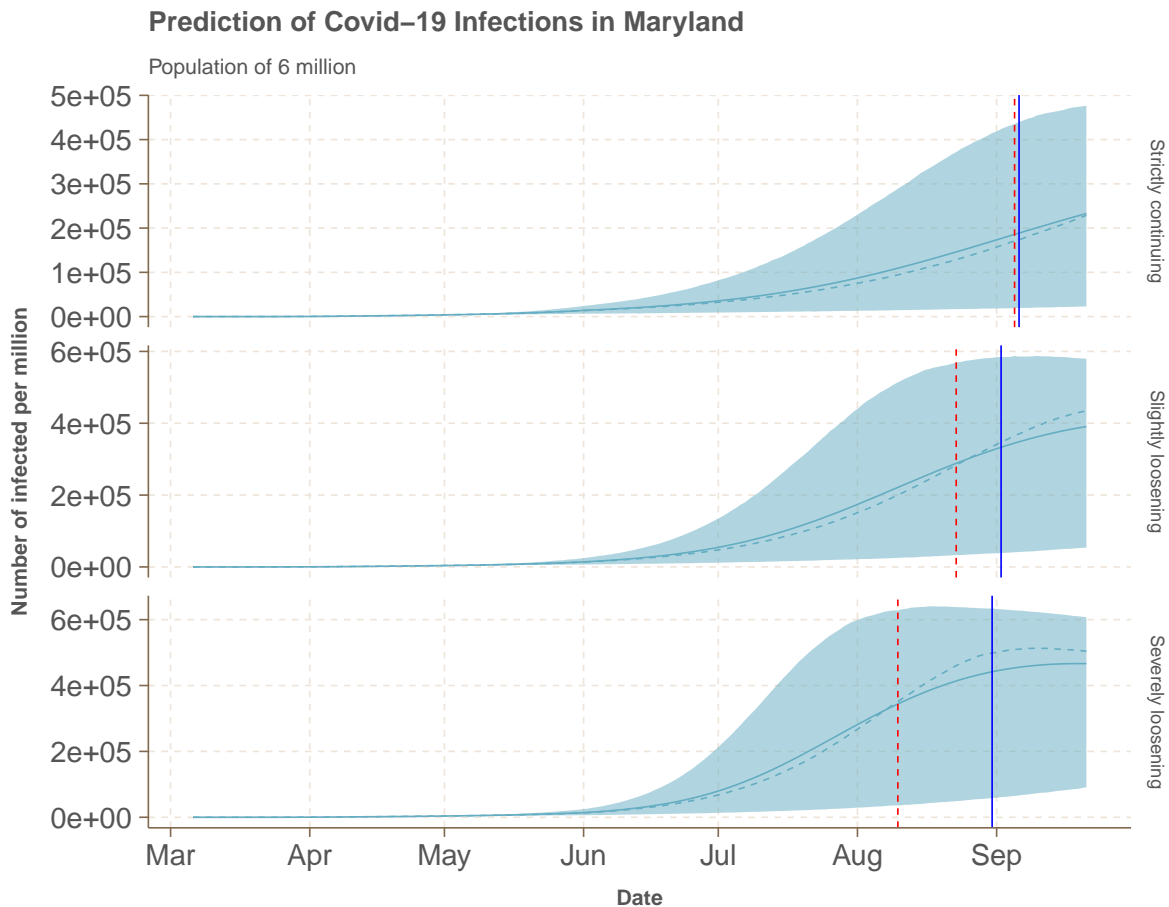


Figure 1: The cumulative number of predicted infected cases in Maryland across the three scenarios. The solid line indicates mean and the dotted line indicates median. The shaded area shows the credible intervals. Vertical red and blue lines indicate two turning points: the peaks for daily and cumulative incidence curves, respectively.

nals, the statistical and epidemiological modelling community needs to create widely supported open platforms to efficiently aggregate, validate, and disseminate broadly accepted by the community model predictions. One exemplary effort of such an interactive hub is the ensemble of COVID-19 forecast models developed by Reich Lab (2020) and co-hosted by CDC at Center for Disease Control and Prevention (2020). While social distancing is essential in mitigating the pandemic in communities across the country, we need modelling community coming together to provide a coordinated and harmonized modeling response to the challenges of COVID-19 pandemic.

References

- Apple Inc (2020). Mobility reports. <https://www.apple.com/covid19/mobility>.
 Brookmeyer R (2020). Op-Ed: Predictions about where the coronavirus pandemic is going vary

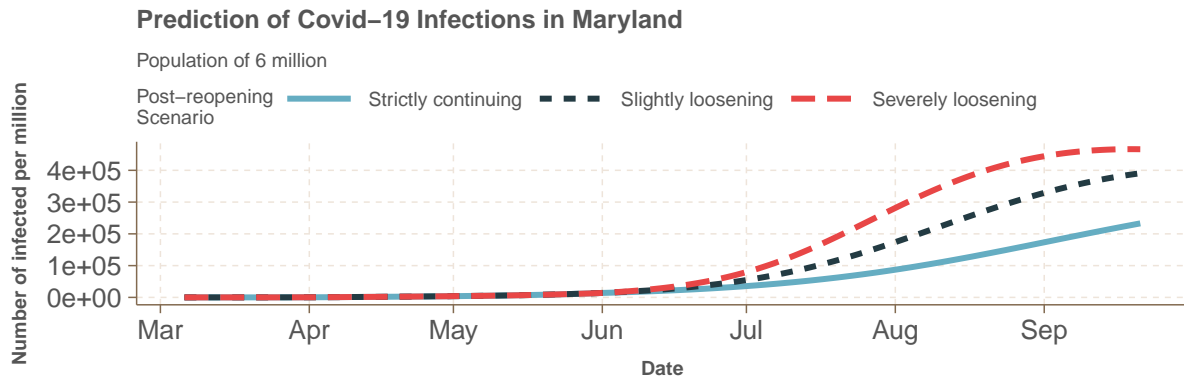


Figure 2: The number of predicted infected cases in Maryland across the three scenarios.

widely. Can models be trusted? Los Angeles Times, <https://www.latimes.com/opinion/story/2020-04-22/models-modeling-coronavirus-covid-19>.

Center for Disease Control and Prevention (2020). COVID-19 mortality forecast. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>.

Etzioni R (2020). Giving models and modelers a bad name. <https://timmermanreport.com/2020/05/giving-models-and-modelers-a-bad-name/>.

Global Epidemic and Mobility Project (2020). Modeling COVID-19 in the United states. https://uploads-ssl.webflow.com/58e6558acc00ee8e4536c1f5/5e8bab44f5baae4c1c2a75d2_GLEAM_web.pdf.

Google LLC (2020). COVID-19 community mobility reports. <https://www.google.com/covid19/mobility/>.

Jewell NP, Lewnard JA, Jewell BL (2020). Caution warranted: Using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. *Annals of Internal Medicine*. Forthcoming, <https://www.acpjournals.org/doi/10.7326/M20-1565>.

Jin J, Agarwala N, Kundu P, Wang Y, Zhao R, Chatterjee N (2020). Transparency, reproducibility, and validity of COVID-19 projection models. <https://www.jhsph.edu/covid-19/articles/transparency-reproducibility-and-validation-of-covid-19-projection-models.html>.

Reich Lab (2020). COVID-19 forecast hub. <https://reichlab.io/covid19-forecast-hub/>.

Teles P (2020). A time-dependent SEIR model to analyse the evolution of the SARS-CoV-2 epidemic outbreak in Portugal. ArRxiv preprint: <https://arxiv.org/abs/2004.04735>.

The Institute for Health Metrics and Evaluation (2020). COVID-19 projections. <https://covid19.healthdata.org/united-states-of-america>.

Unacast (2020). Mobility reports. <https://www.unacast.com/covid19/social-distancing-scoreboard>.