

Discussion of “Evaluate the Risk of Resumption of Business for the States of New York, New Jersey and Connecticut via a Pre-Symptomatic and Asymptomatic Transmission Model of COVID-19”

JEFFREY S. MORRIS^{1,*} AND JING HUANG¹

¹*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*

We congratulate Tian, Tan, Jiang, Wang and Zhang for their well-written article and rigorous effort to model COVID-19 data and evaluate the potential effects of mitigation strategies and reopening, which is crucial to provide policymakers with empirical guidance for their decision-making. This effort is substantial and produces interesting insights. We agree that modeling effects of interventions is important and understudied in existing models, and also agree with the statements in the motivation that it is important to weight the potential cost of mitigation strategies against the benefits they promise. We acknowledge some of the fundamentally difficult challenges the authors had to address in order to develop such a model, including the nontrivial issue of computing the specified differential equations and the estimation of time-varying parameters, and that the Bayesian modeling approach chosen by the authors does a nice job of overcoming these limitations, and making computations feasible. We focus our discussion on evaluating several key assumptions underlying the modeling and describing how they may impact results, and suggest certain characteristics of the pandemic that we believe might be important to incorporate into the model to further improve it.

1 Novel Statistical-Epidemiological Modeling of the Pandemic

There are numerous modeling efforts for the pandemic that involve either classic epidemiological approaches using generative compartment models such as the SIR model or SEIR model while making homogeneity assumptions, or statistical models of the time series as curve data. The classic epidemiological models capture the dynamics of infectious disease spread but lack flexibility to model complex features of the data that might be important, e.g. time varying effects. The statistical models, although possessing great flexibility to capture the features of the curves, suffer from lack of interpretability and failure to capture the actual dynamics of infectious disease spread. We would characterize the author’s model as a hybrid statistical-epidemiological model, based on underlying compartment models so generative in nature yet incorporating statistical modeling components to capture the effect of time and intervention plus incorporating stochastic components and uncertainty quantification, in some sense striking a balance between the extremes and capturing the best of both worlds.

2 Interesting Novel Features of the Model

There are a few key components we find novel and interesting. First, we agree that modeling the effects of mitigation strategies is fundamentally important and understudied, and appreciate

*Corresponding author. Email: jeffrey.morris@pennmedicine.upenn.edu.

the efforts of the authors in doing this. Second, the incorporation of a hidden infection component into the modeling has advantages over the standard SIR/SEIR frameworks and reflects a fundamentally important characteristic of SARS-CoV-2 that many infected individuals remain asymptomatic, yet can potentially spread the disease. Any modeling that only acknowledges confirmed cases underestimates the scope of the pandemic, and this fundamental flaw can limit the performance of the models. Third, we agree that the idea of considering tradeoffs of infectious disease suppression and collateral damage of strict mitigation strategies is a good one. We discuss more about each of these ideas, and also mention critical points and suggestions for how we believe it might be improved.

3 Modeling Effect of Interventions to Support Policy Decisions

The authors specify a parametric model for the effect of interventions on the viral spread that we find to be well-justified and parsimonious, which stabilizes the estimation and provides interpretable parameters while being flexible enough to capture the most likely curve shapes. The function $\alpha(t)$ in Eq (2) imposes a constraint on the shape of $\alpha(t)$, parameters of $\alpha(t)$ have strong interpretability and can be used to assess the effects of policy interventions.

However, the proposed form only allows modeling of a single policy assuming the change of intervention occurs at one time point. A more realistic scenario is that multiple policy interventions are used to mitigate the spread of virus, and the relaxation of restriction is implemented gradually in step with the gradual reopening of the economy. It would be interesting to see extensions of the proposed form to accommodate such scenarios. In addition to the method of assuming parametric forms of $\alpha(t)$ that correlates disease transmission with policy interventions, the modeling of intervention effects can also be made using regression models. For example, the work by Rubin et al. (2020) from the Children’s hospital of Philadelphia PolicyLab incorporates a measure of social distancing as a covariate to gain knowledge on the impact of public policies on disease transmission. The major strength of using regression model is its flexibility to study multiple interventions, interactions of interventions and even the change of intervention effects over time. Different types of policies affect individuals, businesses, and communities differently and alter measures of social distancing differently, which form a continuum of the nation’s mitigation plan that controls the spread as well as balance against individual rights and economic outcomes. It could be challenging to handle these complexities in a parametric function of $\alpha(t)$ and regression may be a better option.

4 Modeling the Unreported Infections

A major challenge of modeling disease transmission based on the traditional SIR model is that the data of infections are often under-reported. As the SIR model assumes all infected individuals are recorded, the inference could be biased. The unreported infections of COVID-19 can be substantial due to the narrow case definition, constrained testing criteria and limited testing capacity, particularly in the initial stage of the outbreak (Wu et al., 2020; Chow et al., 2020; Hortaçsu et al., 2021). In this paper, Tian et al. (2021) extended the SIR model to add a compartment of the unreported infections. Such an extension is significant as it provides a convenient approach to gain a deeper understanding of the spread of the virus. It also provides an opportunity to monitor community transmission and study herd immunity even with the under-reported data. To our best knowledge, this is one of the first efforts to address the under-

reporting issue of the of COVID-19 data. Because information that can be used to study the unreported infections is limited, the proposed model and inference are based on an assumption that the average duration from being infected to recovered among the unreported cases is known, which is a fair assumption. In practice, the value of such a duration can be specified using knowledge learned from epidemiology studies of infected individuals.

However, the proposed approach also assumes that all infections are equally likely to be unobserved which may not be true. We thus suggest a potential improvement of the model. Specifically, we suggest adding an additional parameter, w , which indicates probability of an infection stays unreported. Then the differential equation of $H(t)$ will be related to $wI(t)$ and the equation of $C(t)$ will be related to $(1 - w)I(t)$. One would imagine that this parameter is related to the proportion of asymptomatic infections in the population, as asymptomatic infections are most likely to be unreported. This parameter itself cannot be estimated from the data of reported infections but could be informed by testing capacity and eligibility requirements for testing which could be time variant. Data from large-scale geographic seroprevalence studies could be used to infer this parameter, and perhaps it could be allowed to vary over time based on the increased testing practices that likely capture an increasing proportion of infections over time. Another idea would be to model separate dynamics for the asymptomatic and symptomatic infected cases, and then with the model making assumptions about the relative proportions of these, which might be allowed to vary over time.

Also, the proportion of hidden infections seems determined by the D_H and D_C parameters, with D_H fixed and D_C effectively pinned down to a limited interval via an informative prior, and these parameters essentially determine the number of hidden infections. It may be true that these assumptions are made to stabilize the challenging calculations, but it seems that if the number of hidden infections is too restricted, this could bias the model and compromise the quality of projections. Based on the results, it appears that the values used assume that most infectious cases are detected as confirmed cases by the end of the modeling period, which is almost certainly a false assumption based on serology test results. At a minimum, it seems like sensitivity analyses to these parameters and hyperparameters should be done, or alternatively more care taken to try to match the characteristics of case detection proportions based on serology studies and confirmed case counts. Further, it would be instructive to consider how model projections would have changed sans the novel hidden infection component to demonstrate the realized benefit of this innovation in the modeling.

5 Modeling Tradeoffs of Disease Suppression and Collateral Damage

We strongly agree with the authors’ point that it is important to acknowledge the tradeoffs of infectious disease spread suppression, which adjusting for the cost of each mitigation strategy, in order to mitigate the collateral damage while maximizing disease suppression. This is the key to sustainable strategies for viral control. While we agree with the strong discussion, we were disappointed that there was not more effort to substantively address this question using the modeling results. The quantification of damage via GDP loss provides a well-defined, if incomplete, assessment of one of the key types of collateral damage severe mitigation strategies can cause, the financial ones. How does one decide when the damage is too much to justified a specific intervention; i.e. how is the tradeoff determined? If you open too early, it is clear that more cases will occur and indirectly reduce GDP via hospital burden and the future strict

mitigation strategies that may become necessary, and if you open too late, unnecessary damage can be induced on the GDP. After setting up this important point in the introduction, it would be interesting to see a more detailed analysis and discussion of how the model results inform the assessment of these tradeoffs.

6 Conclusions

The authors have pulled together a detailed hybrid statistical-epidemiological modeling approach that captures several elements central to the pandemic that is not rigorously addressed in most existing models, including a hidden infection compartment and modeling specific effects of intervention in a parsimonious and interpretable fashion. These contribute interesting insights into the pandemic and have general usefulness in continuing to assess effects over time. We have also raised several questions and suggested improvements of how the modeling framework might be adapted and extended to capture the pandemic dynamics even more accurately, which may help it realize its impact potential.

References

- Chow CC, Chang JC, Gerkin RC, Vattikuti S (2020). Global prediction of unreported SARS-CoV2 infection from observed COVID-19 cases. MedRxiv preprint: <https://doi.org/10.1101/2020.04.29.20083485>.
- Hortaçsu A, Liu J, Schwieg T (2021). Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19. *Journal of Econometrics*, 220(1): 106–129.
- Rubin D, Huang J, Fisher BT (2020). Association of social distancing, population density, and temperature with the instantaneous reproduction number of SARS-CoV-2 in counties across the United States. *JAMA Network Open*, 3(7): e2016099.
- Tian T, Tan J, Jiang Y, Wang X, Zhang H (2021). Evaluate the timing of resumption of business for the states of New York, New Jersey, and California via a pre-symptomatic and asymptomatic transmission model of COVID-19. *Journal of Data Science*, 19. In this issue.
- Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. (2020). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, 11(1): 4507.