Journal of Data Science 19 (2), 206–209 DOI: 10.6339/21-JDS994D
April 2021

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"

Shan Yu[1], Guannan Wang[2], and Li Wang[3,*]

[1]Department of Statistics, University of Virginia, Charlottesville, VA, USA
[2]Department of Mathematics, College of William & Mary, Williamsburg, VA, USA
[3]Department of Statistics, Iowa State University, Ames, IA, USA

1 Introduction

The coronavirus outbreak started last December has affected countries all over the world. This global pandemic has not only presented an unprecedented challenge to public health but also has led to a dramatic loss of human life and severe economic consequences worldwide. Many people have been pushed into poverty due to coronavirus pandemic. By middle April, over 20 million Americans filed for unemployment since President Donald J. Trump declared a state of emergency on March 13. Based on the Washington Post, a study by researchers at the University of Illinois, Harvard Business School, Harvard University, and the University of Chicago estimated that over 100 thousand small businesses would be permanently closed, and 5.4 million Americans could lose their health insurance between early March and early May. Therefore, evaluating a reasonable date for resumption of business is essential for combating the COVID-19 and stimulating economic development. Tian et al. (2021) proposed an epidemic model and applied it to assess the real-time risk of the epidemic for the states of New York, New Jersey, California, and Connecticut, and this research is of significant practical meaning.

An essential goal in infectious disease study is to investigate the dynamics of infectious disease spread. The class of mechanistic models, such as the Susceptible-Infected-Removed (SIR) models (Kermack and McKendrick, 1991), plays an essential role in understanding such dynamics. According to WHO (2020), the individuals infected with the virus without significant symptoms, or even the asymptomatic carriers, can still transmit the coronavirus. To better describe the spread mechanism of COVID-19, Tian et al. (2021) modified the traditional SIR model and divided a concerned population into four compartments: susceptible (S), unidentified infectious (I), self-healing without being confirmed (H), and confirmed cases (C).

Under enormous uncertainty about the future of the COVID-19 pandemic, epidemic models are critical planning tools for policy-makers to allocate health resources or plan interventions. Therefore, model interpretability is of crucial importance for policy-makers to understand the underlying process. Tian et al. (2021) proposed a model with strong interpretability, which can assist the decision-making if we need to loosen lockdown measures and reopen the business. For example, d represents the time that takes for the control measures to start their effect and D_C is the average duration from catching the virus to be confirmed by testing. After estimation, results can be used to assess the risk for people to return to work.

*Corresponding author. Email: lilywang@iastate.edu.

Table 1: Point estimates (EST) and confidence intervals (CI) for model parameters in Iowa, California and New York based on data collected from March 13, 2020 to May 24, 2020.

		$\widehat{\alpha}_0$	$\widehat{\eta}$	$\widehat{m}$	$\widehat{D}_c$
IA	EST	0.48	0.31	31.91	4.89
	CI	(0.43, 0.53)	(0.23, 0.38)	(30.64, 33.83)	(4.50, 5.31)
CA	EST	0.45	0.29	24.90	4.86
	CI	(0.43, 0.47)	(0.25, 0.32)	(24.07, 27.57)	(4.32, 5.11)
NY	EST	0.66	0.33	16.35	2.74
	CI	(0.65, 0.68)	(0.32, 0.35)	(15.38, 17.85)	(2.62, 2.88)

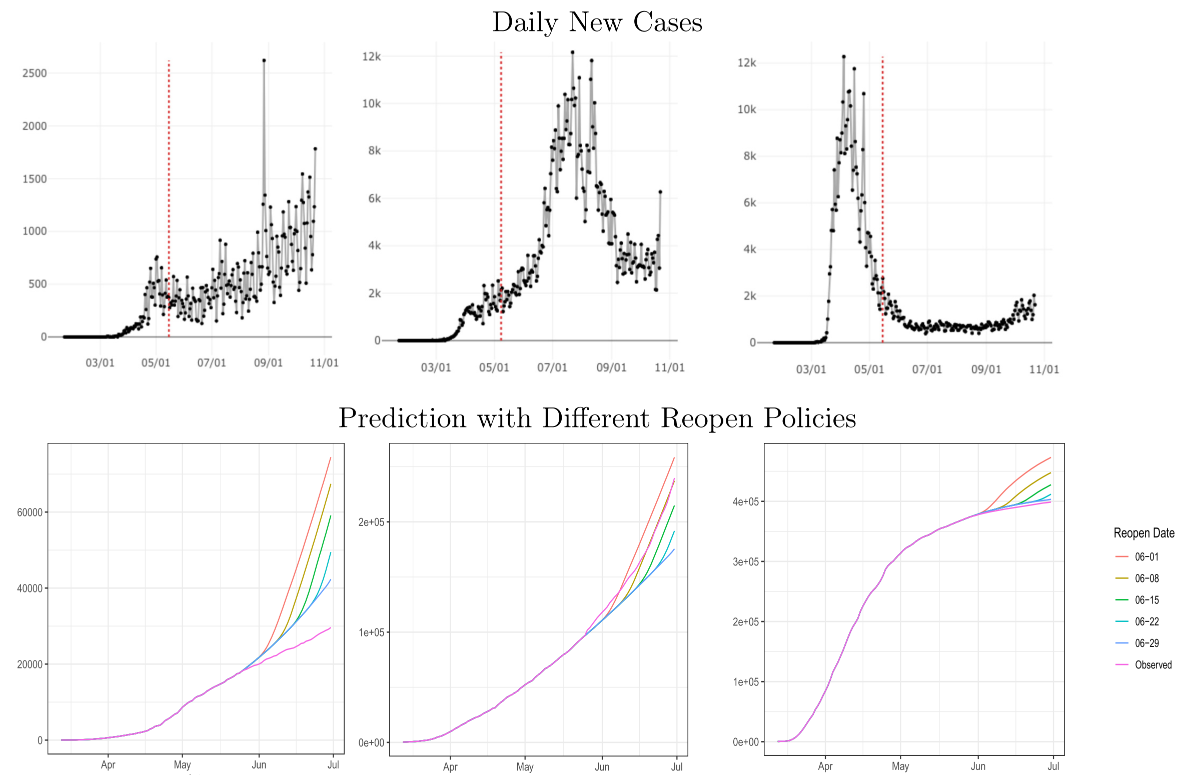
2 Discussion and Future Work

In the epidemic analysis of COVID-19, Tian et al. (2021) trained the model using the data from March 13 to May 24, 2020, in New York, New Jersey, Connecticut, and California and then made predictions for June 2020. The data and the R files relevant to the analysis in Tian et al. (2021) are publicly available. To evaluate the real-time risk of the COVID-19, we apply the method in Tian et al. (2021) to Iowa, California, and New York using the data from the same period. It is worth mentioning that the R functions developed by Tian et al. (2021) are structured well and easy to use. Table 1 shows the estimated parameters and their corresponding 95% confidence intervals. Figure 1 (a)–(c) show the number of observed and predicted cumulative confirmed cases till the end of June 2020.

Based on Figure 1, one can see that the pattern of the spread varies significantly from state to state. To catch the spatial heterogeneity and make accurate predictions, a flexible model might be more appropriate. For example, the number of confirmed cases increased rapidly in New York from middle March to early April, and then the increasing speed started to slow down even after May 15 when counties started to reopen the business. Compared to New York, the spread of COVID-19 in Iowa started slightly later. The number of cases increased fast from late April to early May, then slowed down to some degree in the next two weeks. However, after the reopening on May 15, the number of cases started to rise again. Tian et al. (2021) seems to over-predict the number of cases in both New York and Iowa, especially in a scenario when early June is assumed to be the reopening date. For California, the rapid increase occurred after the reopening on May 15, and the prediction by considering June 8 as the reopening date is accurate with the observations in practice.

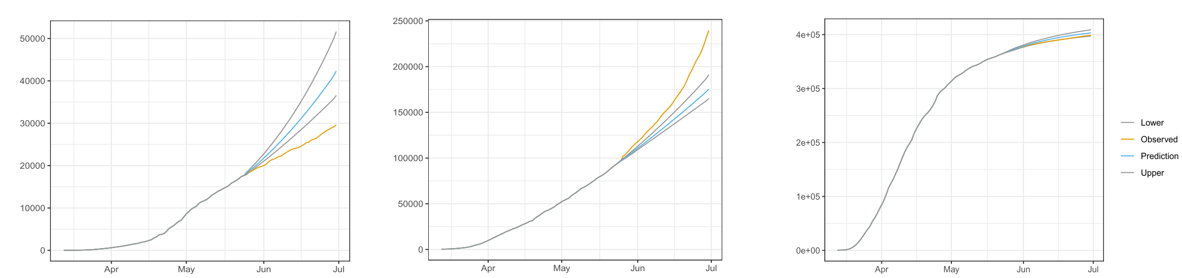
In addition, Tian et al. (2021) provided prediction bands to quantify the prediction uncertainty. Figure 2 (a)–(c) presents the number of observed and predicted cumulative confirmed cases together with the corresponding 95% prediction bands for the states of Iowa, California, and New York. Based on Figure 2, one can see that the prediction bands do not fully cover the observed trajectories. In general, the underlying transmission process of the coronavirus is very complex, especially considering the different strategies to control the spread of COVID-19 at different periods. As the disease processes, there is confounding spatiotemporal heterogeneity in the spread of COVID-19. Therefore, a flexible model allowing for spatiotemporal heterogeneity can help capture the underlying complex process (Lawson et al., 2016).

Months after the lockdown of the US in March, most states are well along the path toward reopening. However, this process is taking place gradually and non-uniformly across the country, depending on local differences in the prevalence of COVID-19. For example, the requirements of



(a) Iowa (b) California (c) New York

Figure 1: The daily new cases and the predicted number of cumulative confirmed cases with different reopening dates in (a) Iowa, (b) California, and (c) New York. Predictions start from May 24, 2020. The dashed line in daily new cases plots indicates the date of businesses reopen.



(a) Iowa (b) California (c) New York

Figure 2: The number of observed and predicted cumulative confirmed cases with 95% prediction band in (a) Iowa, (b) California, and (c) New York. Predictions start from May 24, 2020.

mask-wearing vary across different states. In some states, mask-wearing is required; in others, it is suggested. Therefore, the spread of COVID-19 after reopening economics is different from that before implementing control measures. Some other variables can help to study the spread of COVID-19 after reopening. For example, mobility data, which is used to describe movement trends over time, can be used to model the contact rate in the community. Badr et al. (2020) reveals that mobility patterns are strongly correlated with the growth rates of COVID-19 cases.

The Centers for Disease Control and Prevention recommends that people wear masks in public settings to slow down the spread of COVID-19. Mask-wearing can be another variable that could be used to model the contact rate (Li et al., 2020). Based on the findings in Li et al. (2020) and Badr et al. (2020), we may further improve the method proposed in Tian et al. (2021) by including multiple continuous explanatory variables, such as the mobility and mask-wearing.

For a crisis such as COVID-19, there are many uncertainties in what we have observed. Thus, to better understand the pandemic and take effective public interventions, epidemic modeling plays an essential role. The modeling result can also help medical professionals allocate the medical resources and help policy-makers evaluate interventions. The work by Tian et al. (2021) illustrates the power of epidemic modeling for analyzing and predicting infectious diseases like COVID-19. Although this work focuses on the epidemic data observed in New York, New Jersey, Connecticut, and California from March to May, the proposed model is also applicable to other states and different time periods. COVID-19 cases are surging again in the US after falling from a summer peak, and the spread to new areas of the country suggests the outbreak is far from over. We look forward to more research from the epidemic modeling community to help fight the pandemic and guide us through this pandemic.

References

Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*, 20(11): 1247–1254.

Kermack WO, McKendrick AG (1991). Contributions to the mathematical theory of epidemics-I. *Bulletin of Mathematical Biology*, 53(1–2): 33–55.

Lawson AB, Banerjee S, Haining RP, Ugarte MD (2016). *Handbook of Spatial Epidemiology*. CRC Press.

Li T, Liu Y, Li M, Qian X, Dai SY (2020). Mask or no mask for COVID-19: A public health and market study. *PLOS One*, 15(8): e0237691.

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.

WHO (2020). Coronavirus disease 2019 (COVID-19). Situation Report 73.