Statistics and Its Interface Volume 14 (2021) 15–17

Discussion on "The timing and effectiveness of implementing mild interventions of COVID-19 in large industrial regions via a synthetic control method" by Tian et al.

Kun Chen* and Fei Wang

The ongoing pandemic of the novel coronavirus disease 2019 (COVID-19) has impacted tens of millions of people and caused a huge economic loss. Most of the impacted countries have implemented different non-pharmaceutical interventions (NPIs) to control and prevent the spreading of SARS-Cov-2, which is the virus causing COVID-19. With the coming flu season in the northern hemisphere, many countries are preparing for the potential second or third wave of COVID-19. Therefore it is crucial to understand the differential timing and effectiveness of these NPIs. We congratulate the authors for a very stimulating paper on this timely and crucial topic. The paper tackles several important questions regarding the evaluation of the effects of mild intervention policies for reducing the transmission of SARS-Cov-2, with available observational data amid the ongoing pandemic. The proposed approach combines a variety of statistical tools and practical wisdom in an intriguing manner. There's no doubt about the importance and the potential impact of this paper. In the following, we would like to further discuss several related aspects.

ON THE CONSTRUCTION OF SYNTHETIC SHENZHEN

Constructing a synthetic Shenzhen by combining U.S. counties to enable causal inference is a novel strategy with good rationales. The problem is challenging because the pre-intervention period in Shenzhen was very short (4 days) and yet there are many city/county characteristics that could be considered in the matching process. Nevertheless, this paper successfully constructed a synthetic Shenzhen from 68 U.S. counties that highly resembled Shenzhen in several key characteristics as well as the pre-intervention transmission dynamics of COVID-19. The authors should be applauded for offering such a pragmatic and elegant solution to best utilizing the very limited data.

Here we would like to share our thoughts on this topic and make some suggestions.

*Corresponding author.

On the selection of matching variables

Due to the limitation of the data, only a few city characteristics could be used in the matching. The authors rightfully chose to use population density and latitude, which are both important characteristics relating to the transmission of COVID-19 and would not be affected by the intervention. While some other important characteristics, such as temperature and humidity [5], are correlated to and thus could be well represented by these two characteristics, it is unclear whether there remain a few "underrepresented" characteristics. One example is the population age structure. Davies *et al.* [2] showed that the susceptibility to COVID-19 infection is age-dependent, and the per capita incidence of clinical cases is lower in countries with a younger population structure.

We humbly make two suggestions. First, it could be informative to check how the real Shenzhen and the synthetic Shenzhen agree or differ on a few characteristics that are known to be important to the transmission of COVID-19 but not used in the matching. Second, inspired by the authors' use of PCA on reducing the dimensionality of the daily incidence data, we are wondering whether PCA can also be effective for summarizing the information from multiple community characteristics.

On within-city heterogeneity and resolution of analysis

Shenzhen is a vibrant, large, and heterogeneous city, consisting of several administrative districts with varying population densities and socioeconomic compositions. Moreover, Shenzhen is a Special Economic Zone of China and has several ports directly connecting to Hong Kong, making the city unique on population dynamics. We think all these facts make the construction of a synthetic Shenzhen challenging. In particular, since there is substantial heterogeneity within the city, we are wondering whether it is possible and beneficial to consider such within-city heterogeneity when constructing or evaluating the synthetic Shenzhen under the proposed framework.

In a recent study, Huang *et al.* [3] demonstrated that population stratification could enable better modeling of the effects of reopening policies on mortality and hospitalization rates. They validated their hypothesis and showed the superiority of their approach on the forecasting performance in Harris County, TX (the most populated county in the Greater Houston area) during the Phase I and Phase II reopenings. These findings trigger the question of whether one should "divide and conquer" Shenzhen to allow more precise modeling (or synthetic construction) at different resolutions when such data are available, and if yes, what is an appropriate resolution.

On assessing variability of synthetic Shenzhen

We are wondering whether it is possible to assess the variability of the trend of COVID-19 in the synthetic Shenzhen (shown in Figure 2 of the paper) and consequently that of the estimated causal effects. One immediate thought is to perform a bootstrap analysis: each time draw a bootstrap sample of the 68 U.S. counties and follow the proposed procedure to construct a synthetic Shenzhen. All these bootstrap samples of the synthetic Shenzhen could then be used to construct a "confidence band" of its trend of COVID-19. Is this a valid approach? We think this question may be of interest to the authors.

On uncontrollable factors

There may exist some uncontrollable factors between Shenzhen and the U.S. counties, which may affect the interpretation of the estimated causal effects and especially the effect size of the intervention. In particular, public attitudes to COVID-19 and public compliance to intervention policies could be quite different between China and the U.S. populations. A survey among adults in New York City, Los Angeles, and broadly across the U.S. conducted on May 5–12 found that in the three cohorts, the proportions of respondents supported stay-at-home orders and nonessential business closures were 86.7%, 81.5%, and 79.5%, and the proportions of respondents reported always or often wearing cloth face coverings in public areas were 89.5%, 89.8%, and 74.1%, for New York City, Los Angeles, and the U.S. cohorts, respectively [1]. We speculate that such statistics for Shenzhen, if were collected, would be quite different. The implication is that even without any intervention policy from the local government, overall Shenzhen citizens may tend to take more precautions against COVID-19 than those in its synthetic counterpart from the U.S. We understand that this is merely our speculation, but we think it is worthwhile to consider such general differences between China and the U.S., whatever the causes might be, when interpreting the estimated effects of the intervention.

ON THE EXTENSION OF COMPARTMENT MODELS

The current compartment models used in epidemiology, such as SIR or SEIR, are based on specific assumptions of disease transmission dynamics and cannot capture all complexities of COVID-19. In this work, the authors used an extended compartment model, SIHR (Susceptible, Infectious, Hospitalized, Removed), to better capture the dynamics of COVID-19.

Recently, researchers have also been making modifications to these compartment models from machine learning perspectives directly targeting on improving forecasting performance. For example, Huang *et al.* [3] proposed to use neural networks to model the hidden variables in the SIR-HCD model to augment the epidemiological estimation process. Zou *et al.* [6] proposed the SuEIR model, which is a variant of the SEIR model by taking into account the untested/unreported cases; the model was trained with a gradient-based optimizer. In another paper, Li *et al.* [4] proposed a transfer learning model which transfers knowledge among the models trained on cities with similar characteristics, where the model was pre-trained on the data from one city (referred to as the source city) and fine-tuned in another city (referred to as the target city). These research works demonstrated strong potentials on the incorporation of state-of-the-art machine learning techniques to improve the epidemic model performance. We are eager to hear the authors' opinions on such directions.

Received 29 September 2020

REFERENCES

[1] Czeisler, M. Ã., Tynan, M. A., Howard, M. and et al. (2020) Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance — United States, New York City, and Los Angeles, May 5–12, 2020. *MMWR Morb Mortal Wkly Rep*, **69**, 751–758. URL http://dx.doi.org/10.15585/mmwr.mm6924e1.

[2] Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., Pearson, C. A. B., Quilty, B. J., Kucharski, A. J., Gibbs, H., Clifford, S., Gimma, A., van Zandvoort, K., Munday, J. D., Diamond, C., Edmunds, W. J., Houben, R. M. G. J., Hellewell, J., Russell, T. W., Abbott, S., Funk, S., Bosse, N. I., Sun, Y. F., Flasche, S., Rosello, A., Jarvis, C. I., Eggo, R. M. and CMMID COVID-19 working group (2020) Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, **26**, 1205–1211. URL https://doi.org/10.1038/s41591-020-0962-9.

[3] Huang, T., Chu, Y., Shams, S., Kim, Y., Allen, G., Annapragada, A. V., Subramanian, D., Kakadiaris, I., Gottlieb, A. and Jiang, X. (2020) Population stratification enables modeling effects of reopening policies on mortality and hospitalization rates. *arXiv preprint* arXiv:2008.05909.

[4] Li, Y., Jia, W., Wang, J., Guo, J., Liu, Q., Li, X., Xie, G. and Wang, F. (2020) ALeRT-COVID: Attentive lockdown-aware transfer learning for predicting COVID-19 pandemics in different countries. *medRxiv*.

[5] Wang, J., Tang, K., Feng, K., Lin, X., Lv, W., Chen, K. and Wang, F. (2020) High temperature and high humidity reduce the transmission of COVID-19. *SSRN*. URL https://ssrn.com/abstract=3551767.

[6] Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W. and Gu, Q. (2020) Epidemic model guided machine learning for COVID-19 forecasts in the United States. *medRxiv*. URL https://doi.org/10.1101/2020.05.24.20111989.

Kun Chen
Department of Statistics
University of Connecticut
Storrs, CT 06269
USA
E-mail address: kun.chen@uconn.edu

Fei Wang
Department of Population Health Sciences
Weill Cornell Medicine
New York, NY 10065
USA
E-mail address: few2001@med.cornell.edu