Statistics and Its Interface Volume 14 (2021) 21–22

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.

Soumik Purkayastha and Peter X. K. Song*

It is a pleasure to have the opportunity to comment on this contribution by [6]. The authors use a matching technique called the synthetic control method (SCM) [1] to compare the spread of the COVID-19 pandemic in Shenzhen, China with a synthetic reference population in the USA that matches certain characteristics of Shenzhen, as chosen by the authors. The primary goal of this analysis is to examine the effectiveness of early interventions in the containment of the infectious disease. The basic idea of the SCM is to create and compare a 'control region' versus the 'treatment region', which in this case is Shenzhen, where a policy change has taken place. The invocation of the SCM in a counterfactual framework to the study of intervention policies for the infectious disease is interesting, although there are certain challenging technical issues involved. In this discussion, we will focus on the following domains of challenges in this type of 'case-and-control' analysis.

COUNTERFACTUAL MATCHING

The validity of the SCM depends critically on the suitability of counterfactual controls, which mirrors the disease spread in the absence of comparable intervention policies in order to reach a proper evaluation of interventions of interest. To construct said counterfactuals, the authors consider sixty eight counties with population density values, latitude and COVID-counts similar to that of Shenzhen. A weighted linear combination of the COVID-counts of the sixty eight counties is formed as the counterfactual to the observed time series data of COVID-counts observed in Shenzhen. Much of the inferential comments made in this paper rely on the success of synthetic region matching with Shenzhen in terms of transmission dynamics of the virus. A natural point of thought is how to assess the success of this matching. How an inadequately matched counterfactual might affect inferential statements is of importance as well. For example, the significant gap highlighted in Figure 2 may be attributed to racial and age disparities – disparities which have not been addressed in the article. It has been revealed that US Black and Hispanic residents were experiencing several-fold greater incidence of infection and increased rates of hospitalization [3, 4]. It is known that the average age of the population is much younger in Shenzhen (a migration-type industrial city with a nickname 'silicon valley in China') than that of the population of sixty eight counties, so that the number of asymptomatic people in Shenzhen is expected to be higher. In other words, the gap shown in Figure 2 may be attributed to the higher vulnerability of the USA population. Thus, accounting for proper confounding factors like racial makeup, age distribution and weather is of critical importance to define proper controls and to investigate the causal effects. Although the authors do consider latitude as a feature in their matching process (and claim that latitude and temperature are correlated), we feel a direct matching of weather conditions would have been better received. To fulfill the rigor of the SCM method requires an evaluation of the success of the matching proposed in the paper.

MULTI-COMPARTMENT INFECTIOUS DISEASE MODEL

The authors consider a compartmental Susceptible-Infectious-Hospitalized-Removed (SIHR) model to infer about the pandemic situations in Shenzhen. The SIHR model is a special type of multi-compartment model for infectious disease [5], in which the removed compartment is assumed to be the only outlet for those individuals in the hospitalized compartment. In our view, this assumption oversimplifies the underlying infectious dynamics; for example, since a significant number of infected individuals have mild or no symptoms they are going to recover at home, instead of being hospitalized.

The variable nature of under-reporting has been a cause of major concern across countries. As such, the surveillance data used in the analysis are typically much lower than the actual number of infections [8]. In this case serological antibody surveys are the method of choice to understand the population-level infection rate. In our opinion, it would be interesting to analyze the pandemic using compartmental models which account for pre-symptomatic [7] and asymptomatic [2] individuals as well, given their importance in the transmission dynamics of COVID-19. The under-reporting

*Corresponding author.

issue for the number of infected cases in the surveillance data can yield misleading conclusions.

SENSITIVITY ANALYSIS

Once again, we would like to emphasize the need of a diagnosis of the matching methodology, which would greatly support any specific procedures taken to handle potential features confounding the inferential process. The authors declare that the trend of actual infection cases (per 100,000) in Shenzhen and its synthetic counterpart are similar prior to the implementation of mild intervention policies. This period of 'similarity' between Shenzhen and synthetic Shenzhen is fixed at four days. We feel, given that the median incubation period of the virus is five days, this time period is not adequate to support the authors' claim that the synthetic region could be used to estimate the "counterfactual" results of Shenzhen.

Further, it is tricky to claim the counterfactual effect in a community public health intervention because some of the fundamental counterfactual assumptions may not be satisfied. For example, in Shenzhen where the intervention took place it is likely to be the case that some residents in the city did not comply with the intervention, and the compliance of the intervention was correlated within families and organizations. Further investigation to address these issues might reveal interesting insights, especially since the period of comparison (prior to the implementation of mild intervention) was just twelve days. At least, a sensitivity analysis, both on the choice of the sixty eight counties chosen, as well as the features used in the construction of the synthetic region should be interesting exercises to generate additional supporting evidence for the reported findings to achieve the maximal rigor of the work. It is vital to address the limitations of the SCM method for the infectious disease modeling and analyses should so as to motivate readers to pursue some further research in this direction.

CLOSING REMARKS

We congratulate the authors' attempt to bring a torrent of empirical data to bear on this important field of research and appreciate how their work might help move our knowledge and understanding of policy interventions forward with this contribution. Additionally, we would like to extend our gratitude to the editors for their comments, which helped improve this discussion.

Received 6 October 2020

REFERENCES

[1] Abadie, A., Diamond, A., and Hainmueller, J. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010. MR2759929

[2] Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.-Y., Chen, L., and Wang, M. Presumed asymptomatic carrier transmission of COVID-19. *JAMA*, 323(14):1406–1407, 2020.

[3] Boulware, L. E.. Race disparities in the COVID-19 pandemic—solutions lie in policy, not biology. *JAMA Network Open*, 3(8):e2018696–e2018696, 2020.

[4] Centers for Disease Control and Prevention. COVID-19 hospitalization and death by age, 2020. URL https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html.

[5] Tang, L., Zhou, Y., Wang, L., Purkayastha, S., Zhang, L., He, J., Wang, F., and Song, P. X.-K. A review of multi-compartment infectious disease models. *International Statistical Review*, 88(2):462–513, 2020.

[6] Tian, T., Luo, W., Jiang, Y., Chen, M., Pan, W., Zhao, J., Yang, S., Zhang, H., and Wang, X. The timing and effectiveness of implementing mild interventions of COVID-19 in large industrial cities. *Available at SSRN 3619819*, 2020.

[7] Tong, Z.-D., Tang, A., Li, K.-F., Li, P., Wang, H.-L., Yi, J.-P., Zhang, Y.-L., and Yan, J.-B. Potential presymptomatic transmission of SARS-COV-2, Zhejiang Province, China, 2020. *Emerging Infectious Diseases*, 26(5):1052, 2020.

[8] Zhou, Y., Wang, L., Zhang, L., Shi, L., Yang, K., He, J., Zhao, B., Overton, W., Purkayastha, S., and Song, P. A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States. *Harvard Data Science Review*, 2020.

Soumik Purkayastha
Department of Biostatistics
University of Michigan
Ann Arbor
MI 48109-2029
USA
E-mail address: soumikp@umich.edu

Peter X. K. Song
Department of Biostatistics
University of Michigan
Ann Arbor
MI 48109-2029
USA
E-mail address: pxsong@umich.edu