Statistics and Its Interface Volume 14 (2021) 33–36

Rejoinder of "The timing and effectiveness of implementing mild interventions of COVID-19 in large industrial regions via a synthetic control method"

Ting Tian, Wenxiang Luo, Jianbin Tan, Yukang Jiang, Minqiong Chen, Wenliang Pan, Songpan Yang, Jiashu Zhao, Xueqin Wang, and Heping Zhang*

We appreciate the valuable comments and insightful discussions contributed by Dr. Chen from University of Connecticut and Dr. Wang from Weill Cornell Medicine; Dr. Chen and Zheng from Peking University; Dr. Kane from Yale University and Dr. Owais Gilani from Bucknell University; Purkayastha and Dr. Song from University of Michigan; Dr. Ray and Bhattacharyya from Johns Hopkins University and Dr. Mukherjee from University of Michigan; Dr. Tang from University of Pittsburgh; and Dr. Zhu from Fred Hutchinson Cancer Research Center. Their comments have enhanced the scope of our work.

1. MATCHING COVARIANTS

We agree with Chen and Wang (2020), Purkayastha and Song (2020), and Ray et al. (2020) that it is critical to explore the essential characteristics to match the real and synthetic Shenzhen in the synthetic control method (SCM) procedure. It is imperative that the characteristics to produce the "highly resembling" region of Shenzhen are not influenced by the human decisions. In addition to population density and latitude, which are associated with the transmission of COVID-19, the weather condition and age distribution may also be useful in constructing a synthetic Shenzhen. By considering the temperature and relative humidity, which are hourly time series data, we computed the average temperature and relative humidity by days and then the averages over 16 days for Shenzhen and 68 counties of the United States, respectively.

For convenience and clarity, we will refer to the synthetic Shenzhen in our original article as Shenzhen-Synthetic-0. The values of the average temperature in Shenzhen and Shenzhen-Synthetic-0 are very close (Table 1), assuring that the temperature is well matched. However, this is not the case for the average relative humidity and the percentage of the population aged over 65 (Table 1). The elderly population is known to be more susceptible to the virus [2]. However, Shenzhen has a relatively younger population, while there are only approximately 0.5% counties with the proportion of elderly (aged over 65) population below 9% in the United States. This disparity makes it difficult to match the age distribution between Shenzhen and Shenzhen-Synthetic-0 by the percentage of the population aged over 65.

Table 1. The real and synthetic values of average temperature, average relative humidity, and the percentage of the population aged over 65 for 68 counties in the pre-intervention period between the real Shenzhen and Shenzhen-Synthetic-0

Features	Shenzhen	
	Real	Synthetic-0
Temperature (°C)	16.33	16.38
Relative humidity	71.0%	63.7%
The elderly population proportion	8.5%	16.1%

Table 2. The real and synthetic values of population density, average temperature, average relative humidity, the percentage of the population aged over 65, the first component and the second component for 68 counties in the pre-intervention period between the real Shenzhen and Shenzhen-Synthetic-1

Features	Shenzhen	
	Real	Synthetic-1
Population density	6670	6667
Temperature (°C)	16.33	16.33
Relative humidity	71.0%	71.0%
The elderly population proportion	8.5%	13.3%
The first component	0.106	0.106
The second component	0.035	0.035

To improve the match, we now consider population density, average temperature, average relative humidity, and the percentage of the population aged over 65 as the basis to reproduce a new "synthetic Shenzhen" in our proposed SCM procedure. This new synthetic Shenzhen will be referred to as Shenzhen-Synthetic-1.

*Corresponding author.

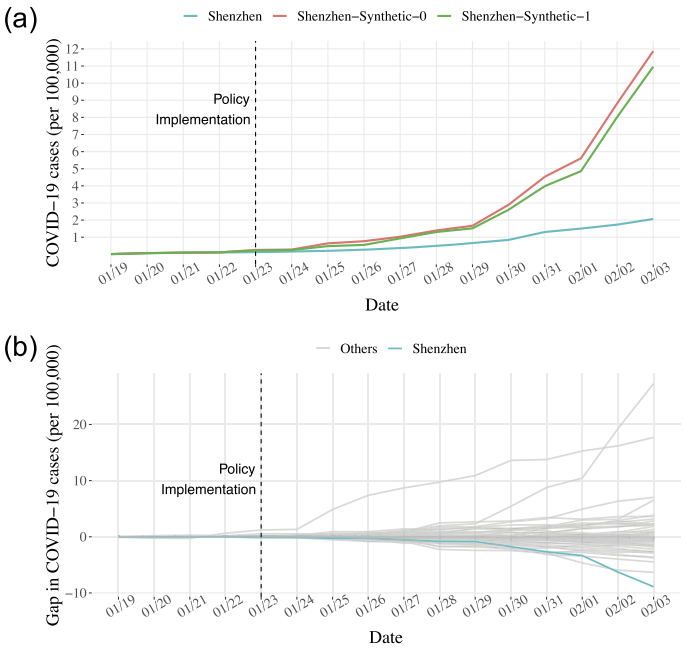

Figure 1. COVID-19 by the number of cases per 100,000 (y-axis) and date (x-axis). (a) The trajectories of Shenzhen and Shenzhen-Synthetic-0 and Shenzhen-Synthetic-1. (b) The permutation test of the treatment effects of implementing the policies in Shenzhen versus the 68 control counties in the United States. A grey curve represents the path of the placebo effect (the difference between a county from the 68 control counties and its "synthetic" control city (a combination of remaining 67 control counties and Shenzhen)), and the blue curve represents the difference between Shenzhen and Shenzhen-Synthetic-1.

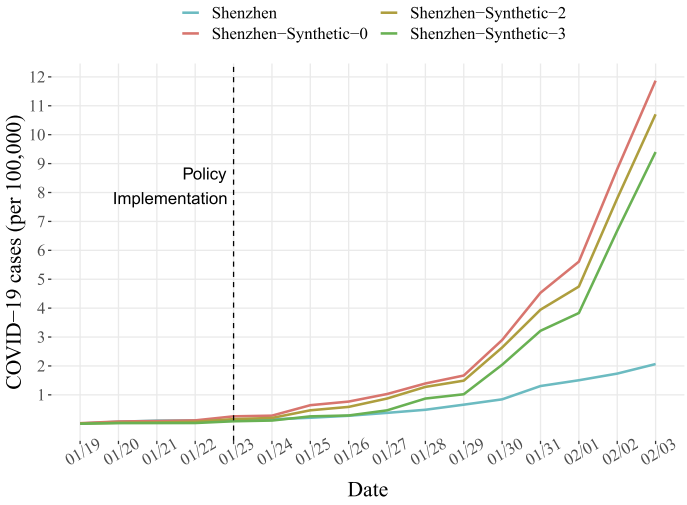

Figure 2. The trajectories of COVID-19 by the number of cases per 100,000. Curves represent the data in Shenzhen, Shenzhen-Synthetic-0, Shenzhen-Synthetic-2, and Shenzhen-Synthetic-3.

We can see from Table 2 that the match is much improved, except the percentage of the elderly population. We then use Shenzhen-Synthetic-1 to estimate the counterfactual result of Shenzhen. Figure 1 (a) depicts the difference between the two versions of the synthetic Shenzhen. The trajectory of COVID-19 in Shenzhen-Synthetic-1 is slightly lower than that of Shenzhen-Synthetic-0 during the post-intervention period. Figure 1 (b) reveals a significant treatment effect of the mild intervention based on a permutation test. The reduction in the COVID-19 cases between Shenzhen and Shenzhen-Synthetic-1 was estimated to be greater than any reduction through the permutation.

Among the confirmed cases in Shenzhen, there were 12.7% cases in those over 65 years old on February 3, 2020 [3]. The number of the confirmed cases by age in the United States warrants further investigation.

2. THE PRE-INTERVENTION PERIOD

Chen and Zheng (2020), Purkayastha and Song (2020), Ray et al. (2020), and Zhu (2020) suggested that the pre-intervention period might not be long enough. To investigate this issue, we took into account the incubation period of COVID-19 and extended the period by 1 and 3 days into the post-intervention period, leading to 5- and 7-day pre-intervention periods in our matching process, respectively. These extended periods were really hybrids of pre- and post-intervention periods. A similar approach was suggested by Abadie [1]. We conducted the principal component analysis (PCA) using the 5- and 7-day trajectories of COVID-19. The first component (explaining over 90% of the variance), population density, and latitude were used to reproduce Shenzhen-Synthetic-2 (using the 5-day period) and Shenzhen-Synthetic-3 (using the 7-day period).

Figure 2 compares the trajectories of COVID-19 in Shenzhen, Shenzhen-Synthetic-0, Shenzhen-Synthetic-2, and Shenzhen-Synthetic-3. The estimated confirmed cases were the largest in Shenzhen-Synthetic-0 on February 3, 2020. Permutation tests were conducted to compare Shenzhen with Shenzhen-Synthetic-2 and then with Shenzhen-Synthetic-3. Both tests gave the same probability of having the lowest placebo gap ($1/69 = 0.014$), indicating statistical significance. We note that although limited days of the post-intervention period were used to construct Shenzhen-Synthetic-2 and Shenzhen-Synthetic-3, the estimation of confirmed cases was based on January 23, 2020, as the intervention starting date. As Figure 2 confirms, not surprisingly, the estimated number of confirmed cases was somewhat affected the number of days by which we extended the pre-intervention period into the post-intervention period. Nonetheless, the overall conclusion is similar.

3. STATISTICAL INFERENCE

Chen and Zheng (2020) noted that two unusually large positive placebo gaps in the permutation test. They were the

results between Westchester and its "synthetic version," and New York and its "synthetic version." Among all 68 counties, Westchester was the county with a sharp increase in the confirmed cases after March 6, 2020, and New York was in a similar situation after March 11, 2020. These two counties were included in the permutation test because their placebo gaps did not deviate from the zero in the pre-intervention period. Thus, we used the 69 placebo runs in our test. As pointed out by Chen and Wang (2020), alternative tools of statistical inference, e.g. "confidence bands" for synthetic Shenzhen, should be explored. We conducted a preliminary analysis using bootstrap samples of 68 counties in constructing a "synthetic region." We found the match between Shenzhen and a combination of bootstrapped selected counties was not adequate. To create "confidence bands" for "synthetic Shenzhen," a modified bootstrap may need to be developed to ensure a reasonable match between Shenzhen and its synthetic version.

4. THE DEVELOPMENT OF SIHR MODEL

We agree with Zhu (2020), Purkayastha and Song (2020), and Ray et al. (2020) that our proposed SIHR model was limited in reflecting the pre-symptomatic and asymptomatic transmissions of COVID-19. Also, our assumption that the time-varying reproduction number tended to zero might be restrictive. We began with our analysis in the early outbreak (late January 2020) of COVID-19, when there was little information on the mechanisms of the transmissions of COVID-19. We have considered other models in our recent research to evaluate the risk of resuming business for the states of New York, New Jersey, and Connecticut [5]. For example, we divided the total population into susceptible (S) individuals, unidentified infectious (I) individuals, self-healing (H) individuals without being confirmed, and confirmed cases (C) (Figure 3).

We assumed the following dynamic system in terms of the numbers of individuals in compartments S, I, H, and C at time t:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta'(t)\frac{I(t)}{N}S(t),\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \beta'(t)\frac{I(t)}{N}S(t) - (\frac{1}{D_C} + \frac{1}{D_H})I(t),\\
\frac{\mathrm{d}H}{\mathrm{d}t} &= I(t)/D_H,\\
\frac{\mathrm{d}C}{\mathrm{d}t} &= I(t)/D_C,
\end{aligned}$$

where $\beta'(t)$ is the time-varying transmission ability and follows the curve in Figure 4 along with four additional parameters. Here $\beta'(t)$ is similar to $\beta(t)$ in the original study, but it is not assumed to tend to zero. D_H is the average duration from catching the virus to self-healing without being confirmed; D_C is the average duration from catching the virus to be confirmed by testing. N is the total number of population.

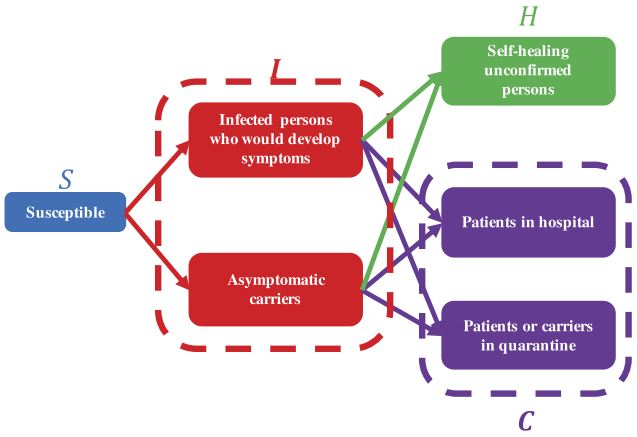

Figure 3. The transition diagram of the SIHC model among compartments S, I, H, and C.

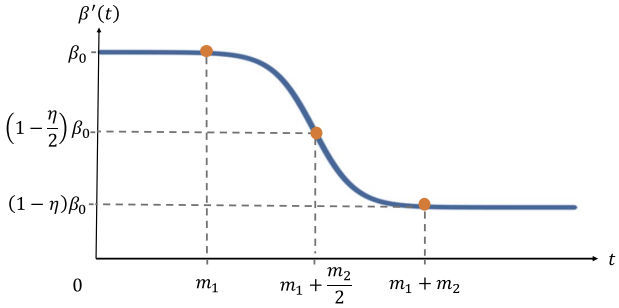

Figure 4. The trend of $\beta'(t)$. β_0 is the transmission potential of early outbreak, m_1 denotes the timing when the intervention begins to take effects, m_2 is the duration of the decreasing process, and $(1-\eta)$ is the reduction of β_0 in the later outbreak.

Our modified SIHC model partially considers the pre-symptomatic and asymptomatic transmissions of COVID-19, the influence of the testing procedure, the lagging effect of the observed cases, the unconfirmed self-healing individuals, and the possibility of repeated outbreaks of COVID-19. This dynamic SIHC model is more realistic in evaluating the transmission of COVID-19. Furthermore, we have also considered a deep learning based model to estimate and predict the trajectories of COVID-19 [4].

Received 31 October 2020

REFERENCES

[1] Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature* 1–19.

[2] Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., 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.
[3] Health Commission of Shenzhen (2020). Outbreak of COVID-19 in Shenzhen.
[4] Tian, T., Jiang, Y., Zhang, Y., Li, Z., Wang, X. and Zhang, H. (2020). COVdNet: A deep learning based and interpretable predication model for the county-wise trajectories of COVID-19 in the United States. *medRxiv*.
[5] Tian, T., Tan, J., Jiang, Y., Wang, X. and Zhang, H. (2020). 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. *medRxiv*.

Ting Tian
School of Mathematics
Sun Yat-Sen University
Guangzhou, 510275
China
E-mail address: tiant55@mail.sysu.edu.cn

Wenxiang Luo
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: luowx8@mail2.sysu.edu.cn

Jianbin Tan
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: tanjb6@mail2.sysu.edu.cn

Yukang Jiang
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: jiangyk3@mail2.sysu.edu.cn

Minqiong Chen
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: mcp04chm@mail3.sysu.edu.cn

Wenliang Pan
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: panwliang@mail.sysu.edu.cn

Songpan Yang
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: yangsp5@mail2.sysu.edu.cn

Jiashu Zhao
School of Mathematics
Sun Yat-Sen University
Guangzhou, GD 510275
China
E-mail address: zhaojsh@mail2.sysu.edu.cn

Xueqin Wang
School of Statistics
Capital University of Economics and Business
Beijing, 100070
China
E-mail address: hawkingwang@gmail.com

Heping Zhang
School of Public Health
Yale University
New Haven, CT 06520-8034
the United States
E-mail address: heping.zhang@yale.edu