Statistics and Its Interface Volume 14 (2021) 19–20

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.

Yifan Zhu

This article provides an overview and discussion of the recent published paper from Tian et al. on modeling the differences of COVID-19 outbreak between Shenzhen and a synthetic population constructed from 68 US counties.

Keywords and phrases: Disease transmission model, COVID-19.

The work by Tian et al. on efficacy evaluation of COVID-19 outbreak intervention is very interesting. The authors established synthetic populations that shared several similar key environmental and societal features with population in Shenzhen, China which experienced COVID-19 outbreak from late January to early February 2020. The size of prevented infection numbers was estimated comparing populations whether with immediate or delayed intervention implementation. The authors fitted Susceptible-Infected-Hospitalized-Removed (SIHR) models to the observed case confirmation trends in both Shenzhen and the synthetic populations, and estimated intervention efficacy using a flexible parametric family suitable for the relative mild intervention happened in Shenzhen. They found that these interventions could be effective when implemented early, while being less damaging than rather draconian measures used elsewhere. I would like to provide a few comments on the approach by Tian et al. and discuss several directions for future research.

1. ESTABLISHING SYNTHETIC POPULATIONS FOR CAUSAL INFERENCE

Tian et al. proposed a modified approach using the synthetic control method (SCM) to construct a synthetic version of Shenzhen population during the early COVID-19 outbreak phase from 68 counties in the United States. Their motivation was that after adjusting for factors such as latitude and population density, the important environmental and societal features for community transmission of an infectious disease such as COVID-19 could be well matched between the real and synthetic Shenzhen populations. The "training" step of SCM was performed with the first 4 days since the detection of initial local cases in Shenzhen and 68 US counties. The reason for the 4-day duration was that Shenzhen started intervention afterwards thus creating the gap of transmission dynamics between itself and the US counties. If unlimited by such objective situation, a longer "training" duration could have been more convincing on the quality of the established synthetic population from US counties. The current window limited to 4 days may be a bit short with COVID-19. Considering the commonly recognized R_0 between 2–3 [3, 6, 4] and a serial interval (disease onset time between index and secondary cases) of 4–6 days [1, 2, 3, 5], the overall number of confirmed cases within 4 days after initial detection would be relatively small (after 4 days one could expect about a triple number of new cases compared to day 1). Furthermore, during early phase of a local outbreak, many other limiting factors including insufficient and/or delayed testing, unawareness in both authority and general populations, and longer delay between individual symptom onset and case conformation could all affect the data quality and induce extra uncertainty.

The above weakness from a short "burn-in" window however, might have been partially alleviated by constructing the synthetic Shenzhen population from as many as 68 US counties using SCM with external factors (latitude, population density) and PCA dimension reduction approaches. The SCM process took into consideration the spatial and temporal heterogeneity among the US counties, thus the source of uncertainties from different counties could be partially averaged out in the constructed synthetic population.

2. TRANSMISSION DYNAMIC MODEL FOR COVID-19

After establishing the synthetic "Shenzhen" population using SCM, Tian et al. fitted SIHR models to the case-confirmation curve observed from both real and synthetic Shenzhen and estimated the intervention efficacy with a flexible parametric function family. The proposed SIHR model was claimed to be advantageous over the traditional SIR or SEIR models by considering pre-symptomatic transmission of COVID-19. However, the underlying assumptions of the proposed SIHR model were relatively strict and could be improved to better suit the actual outbreak. In their SIHR model, it is clear from the dynamic operator expression $\tau_{t,\Theta}$ that only a subject in I status would be infectious, and once the subject moves to H status he or she would no longer shed infection hazard. The removal rate L_{in} between I and H status is the exponential rate of time between infection and "hospitalization or isolation", instead of the "incubation period" they claimed. The proper definition of the incubation period is the time between infection and symptom onset. Unless all patients were immediately isolated/hospitalized once they started to show any symptom, L_{in} in the paper doesn't equal to "mean length of incubation period". With the assumption of both pre-symptomatic and symptomatic infections being equally important with COVID-19, L_{in} actually has three components: (1) the incubation period, (2) the time between symptom onset to case confirmation by PCR test, and (3) the time between the positive test result to isolation/quarantine. Note that in practice the situation could be even more complicated as (3) may overlap with either (1) or (2) given different implementation of interventions, proper transmission models for COVID-19 are indeed very interesting and still underdeveloped.

To model the "relatively mild intervention", Tian et al. introduced a flexible logistic function shaped to model the time-varying efficacy of the measures against COVID-19 in Shenzhen. The function parameters m_1, m_2 and λ_{m_2} could be tuned to indicate the timeliness and efficacy of overall intervention implemented locally. As shown in Fig. 1 in the paper, $\beta(t)$ approaches to a vary small value $\epsilon = 0.01$ as $t \to +\infty$. Thus mathematically this time-varying efficacy will eliminate any infection hazard in the population eventually. On the other hand, as shown in expression (6) the intervention function also formats the time-varying effective reproduction number R_t, which eventually will also approach zero. In practice, as long as the effective reproduction number remains below the critical value of one, the transmission would be well under control. Thus the parameter ϵ doesn't need to be close to zero. Interestingly from Fig. 2 it seems m_1 was chosen to be 5 and the estimated m_2, although not shown, should be quite large since until Feb. 3 the number of observed confirmed cases in Shenzhen is still increasing, indicating that at that time R_t had not yet fallen below one. Compared to the estimated effective R_t of Wuhan following the city lockdown on Jan. 23, 2020 [5, 6], the first 15 days of COVID-19 outbreak in Shenzhen indeed showed somewhat the mild intervention was implemented.

Received 1 October 2020

REFERENCES

[1] Du, Z., Xu, X., Wu, Y., Wang, L., Cowling, B. J. and Meyers, L. A. (2020). Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging Infectious Diseases* **26** 1341.

[2] Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J. and Hens, N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **25** 2000257.

[3] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y. et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*.

[4] Liu, Y., Gayle, A. A., Wilder-Smith, A. and Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*.

[5] Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., He, N., Yu, H., Lin, X. et al. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* **323** 1915–1923.

[6] Zhu, Y. and Chen, Y. Q. (2020). On a Statistical transmission model in analysis of the early phase of COVID-19 outbreak. *Statistics in Biosciences* 1–17.

Yifan Zhu
Vaccine and Infectious Disease Division
Fred Hutchinson Cancer Research Center
1100 Fairview Ave N.
Seattle, WA 98109
USA
E-mail address: yzhu2@fredhutch.org