Statistics and Its Interface Volume 14 (2021) 25–28

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.

Debashree Ray, Rupam Bhattacharyya, and Bhramar Mukherjee*

Tian et al. ought to be commended for their approach of using synthetic control methodology (SCM) to evaluate effectiveness of mild intervention strategies (e.g. wearing masks, isolation of overseas travelers, etc.) in controlling the spread of COVID-19 in industrial regions. The authors use Shenzhen in the Guangdong province of China as an example and compare it with several control counties in the United States. While SCM is often used for causal inference based on observational data in economics and social science literature, it is a relatively new tool in public health research (Bouttell et al., 2018; Rehkopf & Basu, 2018). In this discussion article, we comment on the imperfect data and the resultant biases one needs to be mindful of; and briefly describe the inferential framework of this new epidemiologic tool, its usefulness and potential concerns. We also comment on what could have been done differently.

SAMPLING AND MISCLASSIFICATION

The authors used the reported cumulative number of confirmed COVID-19 cases per 100,000 individuals as their outcome. Any statistical model with reported case, hospitalization and death count data is wrinkled with the issue of differential testing strategies across time and across location. From recent literature, we already know that the degree of covert infections is large in Wuhan (Hao et al., 2020) and in many other places in the world (Rahmandad et al., 2020). This ascertainment rate is different in different places of the world based on availability of tests and strategies for testing. Assumptions regarding this latent unobservable compartment of unreported cases in the susceptible-exposed-infected-recovered (SEIR) model greatly affects the estimate of the unascertained cases, but the estimates of the compartments with reported counts remain fairly stable. Consequently, conclusions on the relative scale still remain valid for the reported number of cases, hospitalizations, and deaths. When analyzing Shenzhen on its own, this ascertainment bias will lead to valid conclusions on a relative scale for the reported/observed compartments. However, when combining several controls and comparing with the average, this

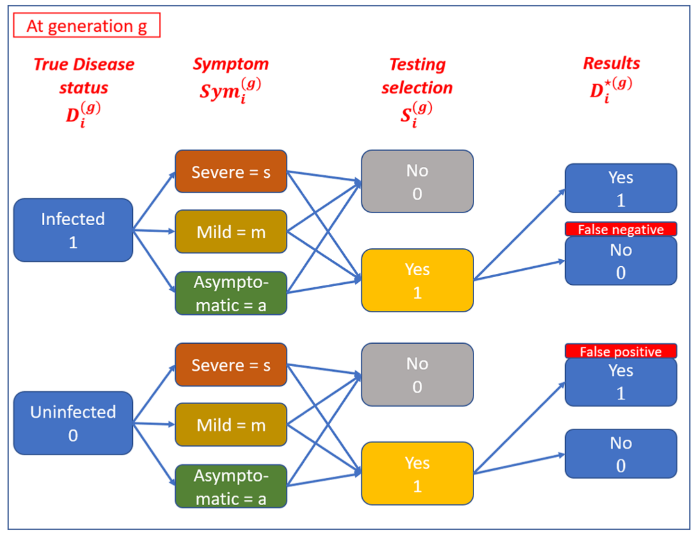

Figure 1. Accounting for false negatives and false positives of viral tests and selection of subjects who are offered testing.

potential difference in ascertainment rate across the control counties may cause some spurious conclusions as one control group may report half of its cases while the other may report one out of 40. One would also like to know the types of tests, particularly as the pandemic is continuing for a long time. Initially the gold standard RT-PCR tests using nasopharyngeal swabs were used but many countries switched to large scale rapid antigen tests with a higher false negative rate. In a recent work, we have considered false negatives and symptom-based testing in the SEIR model (Bhaduri et al., 2020) that may prove to be a useful tool in this context to put the case-counts on a comparable platform.

Figure 1 shows a schematic explaining symptom-dependent testing. With this formulation, one can establish the relationship between the true case counts and the reported case counts using symptom prevalence in infected and uninfected groups, and sensitivity and specificity of the tests. It was not clear how time lags were taken into account in control counties to ensure the generation of the pandemic in Shenzhen aligns with the control counties, and the con-

*Corresponding author.

trols align among themselves. The most sensible thing will be to start at a common zero in terms of when the number of cases exceeds a certain threshold but with counties of different sizes this threshold may have quite different meanings. It will also be interesting to compare Shenzhen with countries where there were severe lockdowns and to Wuhan itself – an important public health question of how much one can gain by imposing draconian lockdowns versus mild interventions. Lockdowns come at a huge cost and since the methodology will be similar, perhaps one can examine this broader question.

FRAMEWORK OF CAUSAL INFERENCE

The research question considered by the authors boils down to the estimation of average causal effect in a population, which requires potential outcomes in the absence of intervention, formally referred to as "counterfactual outcomes". In randomized experiments, individuals are exchangeable between the treated and the untreated groups, and causal effects can be directly estimated. However, as in the current context, randomizing interventions is often not possible, and one needs to rely on observational data. Similar to regression discontinuity or difference-in-differences (DiD) or instrumental variable approaches, SCM allows estimation of causal effect of an intervention or treatment using observational data. These methods rely on the comparison group reproducing the counterfactual outcome trajectory in the absence of intervention. SCM is particularly useful when there is no instrumental variable that influences only the intervention but not the outcome (this problem is exacerbated when there is only 1 treated unit as in the current context), or when treated and control units do not have parallel pre-intervention trends (a crucial assumption for DiD). Much like DiD, SCM considers a treated unit and an aggregate control unit to infer causal effect based on post-intervention difference between treated and control units. Unlike DiD, SCM aggregates the control units from an eligible pool of untreated units (referred to as a "donor pool") using a data-adaptive weighted combination. The weights are selected such that the resultant weighted control has the closest pre-intervention match with the treated unit. Consequently, SCM safeguards against estimation of outlying or extreme counterfactuals (Abadie et al., 2010).

There are 3 key assumptions behind SCM (Bouttell et al., 2018): treated unit and untreated units in the donor pool are sufficiently similar across different variables known to influence the outcome; the donor pool is not affected by the intervention on the treated group or exposed to a similar intervention; and neither the treated unit nor any unit in the donor pool has experienced an event that may differentially affect the outcome either in the pre- or the post-intervention period. To satisfy these assumptions, the authors incorporated 68 US counties from March 1-16, 2020 in their donor pool by assessing similarity to Shenzhen in terms of two predictors of COVID-19 case counts: latitude and population density. For predictors, the authors focused on natural characteristics not influenced by human decisions. They mentioned that the availability of only 4 days in their short pre-intervention period precludes inclusion of related predictors, like latitude and temperature, due to possible issues of collinearity in the construction of "synthetic Shenzhen". Other variables that potentially influence risk to COVID-19 include socio-demographic factors (e.g. age, sex, race/ethnicity), behavioral factors (e.g. smoking status) and predisposing conditions (e.g. asthma, chronic respiratory disease, chronic heart disease, diabetes, cancer, hypertension, chronic kidney disease) (Jin et al., 2020). However, the authors did not consider any of these other important COVID-19 risk factors in choosing their donor pool. Control units could have been better chosen if the authors additionally assessed similarity based on, for instance, proportion of population above 65 years of age or proportion of population with an underlying predisposing health condition. We acknowledge that considering all COVID-19 risk factors for donor pool selection can result in a small donor pool, which can make it hard to achieve a match when deriving the data-driven weights.

After determining weights for the synthetic controls, the authors constructed a counterfactual trend for their outcome over time in the post-intervention period, and assessed if the difference in trends between Shenzhen and the synthetic Shenzhen is statistically significant. The authors employed a well-established "placebo test" that enables appropriate statistical inference to be drawn under SCM design despite the small-sample nature of the data, the absence of randomization, and the absence of probabilistic sampling (Abadie et al., 2010, 2015; Rehkopf & Basu, 2018). In particular, the authors used the concept of "in-space placebos" where the intervention is re-assigned to units of the donor pool with the premise that our confidence on any observed intervention effect would disappear if estimates as large as the observed effect arose when intervention is artificially assigned to any unit in the donor pool not directly exposed to the intervention as per SCM design (Abadie et al., 2015). The whole process of obtaining intervention effect is repeated by assigning treatment status to one unit of the donor pool at a time, thus creating a distribution of placebo effects against which the observed intervention effect is to be compared. This, however, is unclear from the description of the procedure in Section 2.2, and the statement "…assigning an arbitrary sample of K counties in the United States as a "treated unit"…" begs for further clarification. Using this permutation-based placebo test, at 5% level, the authors found significant evidence of reduction in COVID-19 cases per 100,000 due to the early, mild intervention policies of Shenzhen. Note, the success of SCM critically depends not only on the construction of an appropriate control pool but also on having a sizable number of time points in the pre-intervention period (Abadie et al., 2015). In Tian et al. article, "there were only 4 days in the pre-interventions period",

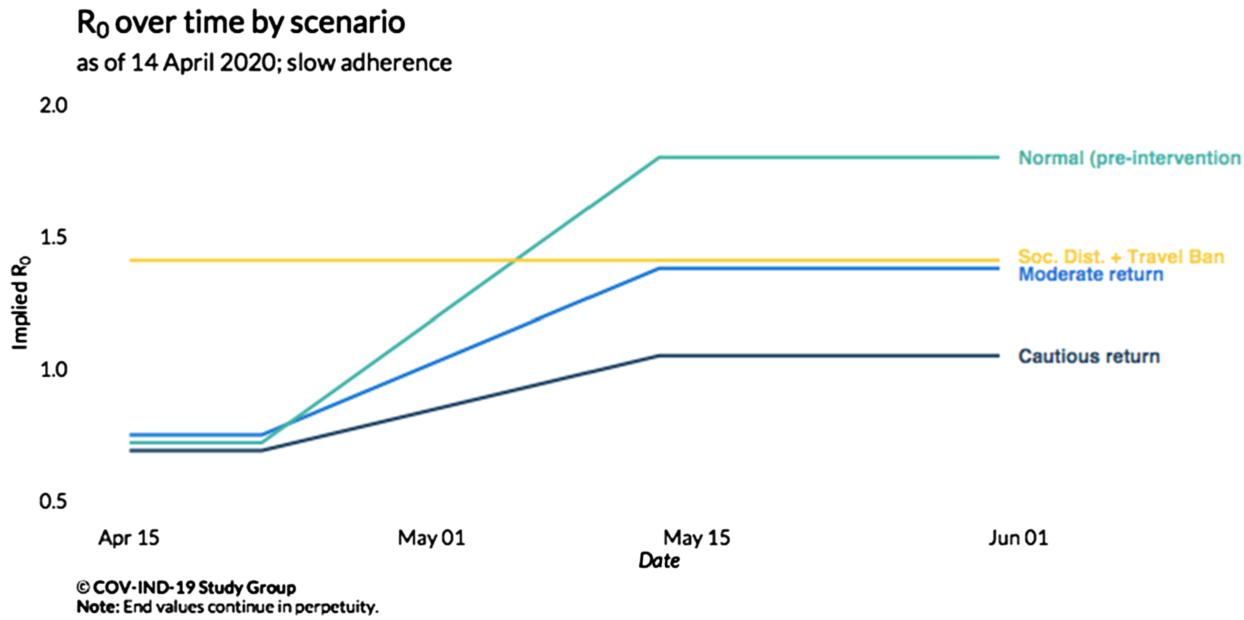

Figure 2. Time-varying reproduction number schedules under hypothetical scenarios. Reproduced from Ray et al., 2020.

which raises some concern about credibility of causal inference using SCM in this context.

COMPARTMENTAL MODELS

Tian et al. additionally explored the effect of delay on the efficacy of mild interventions by first proposing a modified compartmental epidemiologic model incorporating hospitalized individuals, namely, the susceptible-infected-hospitalized-removed (SIHR) model. This model attempts to separate the infected individuals into non-isolated (and therefore able to spread infection) and isolated (therefore not contributing to new infections) groups. This distinction, as mentioned by the authors, overcomes the "lack of consideration for the incubation period and the pre-symptomatic transmission pattern during the incubation period" in the traditional SIR or SEIR models. The hidden Markov structure of the transmission follows a standard multinomial conditional model with the cell probabilities controlled by a dynamic transmission operator determined by the differential equations governing the transmissions. The time-varying transmission parameter from the susceptible to the infected compartment (termed as the "contact rate" by the authors) is controlled by additional parameters via a logistic function. Combining this with the mean length of the incubation period, the authors obtain an expression for the time-varying reproduction number. The estimation of the base parameters is performed via standard MCMC procedures, and prediction for future time points is performed by computing the median of the posterior distribution for the conditional mean of the compartmental proportions. The effect of delay is then simulated by shifting the intercept parameter in the logistic function by the length of delay (in days). As expected,

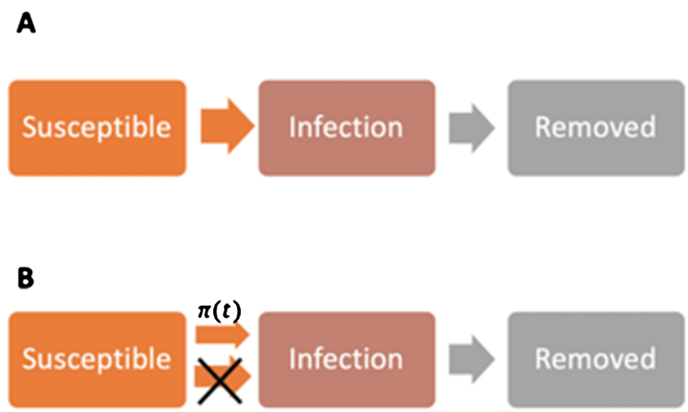

Figure 3. The SIR model with (A) or without (B) considering human intervention by introducing a transmission rate modifier $\pi(t)$.

the authors found that 4-5 days delay in implementing mild interventions could have led to 2-3 times higher cumulative confirmed case count per 100,000 individuals. The authors have not focused on the estimated parameters in their discussion. Given that the model inherently relies upon a time-varying reproduction number (which can be calculated based on the estimated parameters), it would be interesting to look at this quantity over time as estimated by the model for the different delay lengths, as in Figure 2. A ratio of the time-varying reproduction number at a given time point t with that at time point 0 (R_0) can be interpreted as a time-varying transmission rate modifier π_t (Figure 3), which has been used in recent works on SIR-type models of COVID-19

(Ray et al., 2020). Summarizing this time-varying parameter for the varying delays would offer a useful interpretation of the efficacy of the interventions applied.

SUMMARY

Overall, the role of study designs including synthetic controls, test negative designs and variations thereof are incredibly needed for valid inference in COVID studies, and the Tian et al. article does a thorough job of exposing this idea and igniting future ideas.

Received 20 October 2020

REFERENCES

Abadie, A., Diamond, A., & Hainmueller, J. (2010). 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. MR2759929

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2), 495–510.

Bouttell, J., Craig, P., Lewsey, J., Robinson, M., & Popham, F. (2018). Synthetic Control Methodology as a Tool for Evaluating Population-level Health Interventions. *Journal of Epidemiology and Community Health*, 72(8), 673–678.

Jin, J., Agarwala, N., Kundu, P., Harvey, B., Zhang, Y., Wallace, E., & Chatterjee, N. (2020). Assessment of Individual- and Community-level Risks for COVID-19 Mortality in the US and Implications for Vaccine Distribution. *medRxiv*, 10.1101/2020.05.27.20115170.

Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., Purkayastha, S., Halder, A., Rix, A., Barker, D., Kleinsasser, M., Zhou, Y., Bose, D., Song, P., Banerjee, M., Baladandayuthapani, V., Ghosh, P., & Mukherjee, B. (2020). Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the COVID-19 Pandemic: Data Science Call to Arms. *Harvard Data Science Review*, 2020(Suppl 1), 10.1162/99608f92.60e08ed5.

Rehkopf, D. H., & Basu, S. (2018). A New Tool for Case Studies in Epidemiology-the Synthetic Control Method. *Epidemiology*, 29(4), 503–505.

Debashree Ray
Department of Epidemiology
Johns Hopkins University
Baltimore, MD 21205
USA
E-mail address: dray@jhu.edu

Rupam Bhattacharyya
Department of Biostatistics
University of Michigan
Ann Arbor, MI 48109
USA
E-mail address: rupamb@umich.edu

Bhramar Mukherjee
Department of Biostatistics
University of Michigan
Ann Arbor, MI 48109
USA
Center for Precision Health Data Science
University of Michigan
Ann Arbor, MI 48109
USA
E-mail address: bhramar@umich.edu