Discussion of "An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China"

YIFAN ZHU[*1] AND YING QING CHEN[1]

[1]*Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle WA*

We found the work by Wang et al. on analyzing the COVID-19 outbreak in China very timely, and interesting. The R package `eSIR` they developed provides a useful tool to understand the ongoing pandemic event. The paper is one of the many contributions from the statistical research communities to fight this invisible but deadly virus. In their model, the classic Susceptible-Infected-Removed (SIR) structure widely used for modelling infectious disease outbreaks served as hyper-parameter generating mechanism and has been adapted for the stochastic time-series type observed data for COVID-19 published by the Chinese Center for Disease Control and Prevention (China CDC). The results from their model provided estimates for R_0 of COVID-19 under various scenarios quantifying the quarantine interventions implemented at various stages by the Chinese government. Additionally, the authors proposed a practical approach to adjusting for the potentially under-reported cases due to progressing understanding of the emerging outbreak. We would like to provide a few comments on the model proposed by Wang et al. and discuss several directions for future research.

1   Incorporating Recent Scientific Knowledge of SARS-CoV-2

The state-space SIR epidemiological model proposed in the paper incorporate several components on top of the basic SIR model to adapt for the early phase COVID-19 outbreak, for the daily confirmed cases both inside and outside Hubei Province. Firstly the stochastic element has been injected through assuming beta-distributed observations generated from the underlying state-space Markov process, further governed by the SIR generating mechanism of the hyper-parameters of prevalence's of each state. Secondly the infection process linking the susceptible (S) and infected (I) states has been modified by either a reduction-of-contact process $\pi(t)$ or an external quarantined (Q) state with rate function $\phi(t)$. Markov-Chain Monte-Carlo (MCMC) algorithms has been developed to fit the proposed model to the daily report of confirmed cases in China.

With the ongoing COVID-19 pandemic across the globe, the scientific community has been working day and night to better understand various aspect of the virus. One crucial feature of SARS-CoV-2 transmission, that differs from other known viruses in the coronavirus family such as SARS and MERS, is that accumulating evidence of infection from pre-symptomatic or fully asymptomatic individuals could be significant (Bi et al., 2020; Gandhi et al., 2020). As mentioned in the discussion section of the current paper, the Susceptible-Exposed-Infected-Removed (SEIR) structure take into consideration the incubation period of COVID-19, which not only affects the transmission dynamic significantly due to its relatively long duration (Guan et al., 2020; Li et al., 2020), but also serves as extremely important factor for implementing proper quarantine strategies. With the SEIR structure, the accompanying processes $\pi(t)$ or $\phi(t)$ will also need to be redefined. It would be interesting to see how the model estimates may change from current findings with the updated model. The exposed state could further develop into sub-states of

---

*Corresponding author. Email: yzhu2@fredhutch.org.

infected such as I^S and I^A, to represent symptomatic and asymptomatic cases. Naturally the link between I^S, I^A and S or R may be very different, and it would be very interesting to quantify the relative infectiousness from cases with or without symptoms, as well as their post-infection prognosis and establishment of future immunity.

Another important aspect of epidemic models, although often limited by the accessible data, is the population infectious contact structure. Homogeneous mixing assumption has been widely used if no clustering structure is available. Generally the R_0 tends be overestimated with such assumption considering clustered transmission events within household/school/workspace and/or super-spreading events are common for respiratory infectious diseases. Demographic, social and environmental factors could be important for the transmission dynamic in similar way to the contact structure and modify the capacity of infection routes. With the available data, these factors could be incorporated into the SIR or SEIR models and provide more detailed description of transmission paths, thus provide guided directions for more efficient intervention efforts.

The authors provided R_0 estimates under various intervention scenarios through sensitivity analyses. The intervention approaches had been modeled with step-wise or exponential functions as the contact-reduction rate, or the transition probability vector with the extra quarantine state. To our interest, although the proposed scenarios reflect probable effect sizes of intervention efforts, could these effect sizes be estimated directly as model parameters? Otherwise the current estimated R_0's under various scenarios showed relative large variations, indicating the model is quite sensitive to intervention effect parameters. We would appreciate certain validation of the presented intervention effect functions.

Nevertheless, mathematical/statistical transmission models always rely on assumptions of disease natural history, social and environment composition, and time-varying intervention/-control efforts to describe and simulate the underlying mechanisms of the emerging outbreaks. During the early-phase of an outbreak caused by a novel virus, simple models with the least required assumption sets such as SIR model provide quick and urgently needed initial answers for understanding and evaluating the level of caution, which serve as the base defence line against larger scale outbreaks. Unfortunately COVID-19 has already become a pandemic, but statisticians must continue to contribute to the global effort to win the war against the virus.

2 Calibrating the Under-Reported Cases

We nevertheless congratulate the authors for providing a simple approach to account for the bias resulted from potential under reported cases. The proposed approach assumes consistent growth rate between January 12 to February 12, 2020 and fitted an exponential curve that minimized one-step ahead extrapolation error. We thus suggest several potential improvements of this approach. First, the consistent exponential growth rate assumption could be affected by the intervention efforts, for instance the city blockade, enhanced quarantines and new hospital openings as modeled in the paper. All these time points with jumping $\pi(t)$ occurred within the calibration duration, thus may affect the growth rate of infected cases. The calibration model should therefore adjust the proposed growth rate to $\pi(t)$ accordingly. Secondly, the sudden spike on February 12, 2020 was mainly due to the change of clinical/diagnostic definition of COVID-19 cases by the Chinese Ministry of Health. These cases were not necessarily purely resulted from the reporting delay. External information might be used to model the delay functions more precisely. Thus, we felt the calibration model could be further improved.

As a side note, the case confirmation time-series data, however, suffer from delayed report

bias compared to the epidemiological curve comprised of individual time of disease onset. The epidemiological curve is more directly related to frequently used features of infectious diseases, such as serial interval (time between symptom onset of an infected case and secondary cases from it), incubation period (time between infection and symptom onset), latent period (time between infection and start of infectiousness) and infectiousness duration (time between infectiousness start and end). These features of the natural disease history provide elevated insights into the transmission dynamics and guidances for intervention practices (Bi et al., 2020; Li et al., 2020; Zhu and Chen, 2020). Of course the epidemiological curve is much more expensive to obtain as it require more detailed clinical information of infected subjects. Ultimately, we would like to point out the importance of testing, contact tracing and epidemiological review. These are the crucial efforts needed not only for public health agencies, policy makers to control the disease spread, but also provide rich and valuable data for modellers to obtain better inference and prediction.

References

Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *The Lancet Infectious Diseases*. Forthcoming, https://doi.org/10.1016/S1473-3099(20)30287-5.

Gandhi M, Yokoe DS, Havlir DV (2020). Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. *New England Journal of Medicine*, 382(22): 2158–2160.

Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*, 382(18): 1708–1720.

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13): 1199–1207.

Zhu Y, Chen YQ (2020). On a statistical transmission model in analysis of the early phase of COVID-19 outbreak. *Statistics in Biosciences*. Forthcoming, https://doi.org/10.1007/s12561-020-09277-0.