



Forecasting of COVID-19: transmission models and beyond

Yang Zhao^{1,2}, Yongyue Wei^{1,3}, Feng Chen^{1,3}

¹Department of Epidemiology & Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China; ²Big Data Center of Nanjing Medical University, Nanjing 211166, China; ³China International Cooperation Center (CICC) for Environment and Human Health, Nanjing Medical University, Nanjing 211166 China

Correspondence to: Feng Chen. Department of Epidemiology & Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China. Email: fengchen@njmu.edu.cn.

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Thoracic Disease*. The article did not undergo external peer review.

Comment on: Yang Z, Zeng Z, Wang K, *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12:165-74.

Submitted Apr 20, 2020. Accepted for publication Apr 29, 2020.

doi: 10.21037/jtd-20-1692

View this article at: <http://dx.doi.org/10.21037/jtd-20-1692>

As of April 7th, 2020, approximately 1,337,749 cases of coronavirus disease 2019 (COVID-19) have been confirmed worldwide, with nearly 74,169 deaths occurred. It is critical to understand the epidemiological features and transmission dynamics of the COVID-19 outbreaks. Compartments models are the most widely used mathematical models to describe the dynamics of infectious diseases. As an example, a susceptible(*S*)-exposed(*E*)-infected(*I*)-removed(*R*) (SEIR) model can be used to modelling the process of an individual experiencing the infection course from susceptible to exposed, then infected and finally recovered or dead. Traditional SEIR model assumes a closed system (no individuals move in or out), and each individual transmits to one of the other states at a fixed rate. However, traditional SEIR model may get challenged when be used to model the epidemic of COVID-9 in China. The size of the overall population in each province of China is not fixed as January of 2020 is the “*Chunyun*” period, with an estimation of around 3 billions trips (1). Mobility information should be included in the model to reflect the impact of *Chunyun* on the populations of susceptible and exposed individuals. Meanwhile, Chinese government has taken intensive interventional policies to control the spread of COVID-19, which definitely changed the transmission dynamics.

SEIR and its extension

Recently, Yang *et al.* reported a modified SEIR model of the

epidemics trend of COVID-19 in China under public health interventions (2). To account for the mitigation during the *Chunyun*, they included several additional parameters in the SEIR model. Their modification made the sizes of compartments *S* and *E* now related to the number of susceptible and exposed individuals moving in and out of the system. Transportation information between Wuhan and other provinces was retrieved from Baidu *qianxi* index. The idea of including additional parameters based on multiple sources of information to reflect the distinct feature of the spread of COVID-19 had been adopted by many studies. Wu *et al.* also reported a modified SEIR model using data from Dec 31, 2019 to Jan 28, 2020. In addition to domestic transportations, they also used air transportation information from Official Aviation Guide (OAG), thus made it possible to make predictions on the global scale (3). Kucharski *et al.* fitted a stochastic transmission dynamic model to multiple publicly available datasets on cases in Wuhan and internationally exported cases from Wuhan (4). Chinazzi *et al.* used a Global Epidemic and Mobility Model (GLEAM), an individual-based, stochastic, and spatial epidemic model, to model the international spread of the COVID-19 out-break by emulating the flow of travelers among over 3,200 sub-populations using real transportation data (5).

Pan *et al.* described the distinct patterns of the effective reproduction numbers at the different stages, characterized by different degrees of interventions, of the epidemic in

Wuhan (6). Yang *et al.* also included an additional parameter $r(t)$ (not the effective regeneration number), which denotes the number of contacts for each individual per day. The authors used $r(t)=15, 3,$ and $10,$ for the days before January 23rd (no intervention), between January 23 and March 1st (fully controls) and after March 1st (mild controls). The inclusion of $r(t)$ made it possible to introduce the effect of interventional and control policies, including travelling and public gathering restrictions. Wei *et al.* further proposed a SEIR^{+CAQ} model which integrated information on the mechanisms of transmission, profiles of infections and quarantine policies (7).

Limitations and applications of dynamic model

Indeed, given the intensive interventional policies, the quick evolution of the virus, the dynamic capacity of diagnosis and treatments, as well as the understanding of the COVID-19 itself, it is almost impossible to precisely predict the future epidemics, especially in a long period. As an example, on February 5th, 2020, the National Health Commission of the People's Republic of China announced the release of the tentative fifth revised edition of the Diagnosis and Treatment Plan for COVID-19 which dramatically changed the curve of confirmed COVID-19 cases in China. It is not possible for any model to predict the exact date of the release, let alone that some studies only used data before January 23rd 2020. Although including more parameters may in theory increase the performance of a predictive model, it is almost impossible to have a stable solution at the early stage of an outbreak with very limited information.

As George E. P. Box, a famous statistician, said "*all models are wrong, but some are useful*". The value of dynamic models, however, lies in the early warning of infectious disease outbreaks, the decision supporting before the implementation of prevention and control measures, and the evaluation of the effects of public health intervention after the outbreak. Yang *et al.* estimated that the effect of initiating the interventions five days after the actual time would increase the number of cases exponentially to 173,372 cases on March 4, 2020. Our estimation based on SEIR showed that the policies constraining public transportations reduced 85% of infections (8).

Causal inference and policy evaluation

However, we propose that it is important to notice that

when declaring some intervention is the "cause" of the reduction of COVID-19 cases, it should be stated in a "causal" language. That is, what would have been if this intervention had not been performed? A common framework of policy evaluation is the "counterfactual", in which the causal intervention effect is defined as the difference between the epidemics in the real world, compared to the epidemics had no interventions been performed, maybe contrary to the fact. Of course, it is impossible as in the real world the interventions are either performed or not. Thus when making policies evaluation, the scientists should bear in mind that whether conditions are "exchangeable" for the epidemics with or without the intervention. It is recommended that the sensitivity analysis should be applied to assess the robustness of the evaluation. Scientists should also carefully evaluate whether the data collected is sufficient to make a causal policy evaluation.

Besides, given that the interventions from the government is a combination of various policies and methods, it is almost impossible to evaluate a single method, such as the "lockdown" of Wuhan city, the recommendation of wearing mask, or the restriction of public transportation, keeping social distance, building mobile cabin hospital, etc.. As an example, the "lockdown" of Wuhan city not only prevented transportations from or into the city, but also raised the awareness of the people in the whole country against the virus. Peoples stayed at home, reduced unnecessary gathering, washed hands more frequently than before. These behaviors also made the curve of the epidemic flatter than that had no public health "level 1" emergency response been taken. One possible solution for the evaluation of individual policy or intervention is to use an agent based simulation of infectious disease (9,10).

Machine learning in epidemic prediction

With the rapid improvement on computational capacity, machine learning methods have been widely used in medical research. The authors used long short term memory (LSTM) model, an improvement of recurrent neural network (RNN), to predict the number of new cases. The RNN and LSTM models had been widely used in the natural language processing to predict the future elements in a word or a sentence given the previous elements. As most of the machine learning methods are data-driven, it is important use an independent sample to validate the present model. The authors used the 2003 SARS data as the

training set, and COVID-19 as the testing set. Their results showed a remarkable agreement between the actual number of cases and the results. Hu *et al.* developed the modified auto-encoder (MAE), an artificial intelligence (AI) based method, for real time forecasting the number of the new and cumulative confirmed cases and deaths of COVID-19 under various intervention scenarios in more than 100 countries (11). While we believe that there will be more and more AI ML based methods applied in the analysis of COVID-19, it should be noted that the ML base methods only relax the assumptions for the form of traditional dynamic models. In agreement with Zhong's opinions, we believe that the integration of deep learning, web scraping, and other big data technologies will hopefully improve the precision of COVID-19 predictions (12). However, policy evaluation based on ML analysis still need the scientists to understand what the estimand is and whether the data can answer a causal question.

To summary, it is important to evaluate the policies that Chinese government had taken at the end stage of the outbreak of COVID-19 in China. As both dynamic modelling and machine learning based methods have their merits, these methods should be adapted so as to reflect the effect of interventions. Ideas of causal inference should also be included when these methods are applied to decision making and policy evaluations.

Acknowledgments

Funding: The study was supported by the Special Program of National Natural Science Foundation of China on tracing, pathogenesis, prevention and treatment for COVID-19 (No. 82041024 to FC).

Footnote

Conflict of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jtd-20-1692>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article

distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* 2020. [Epub ahead of print].
2. Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12:165-74.
3. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020;395:689-97.
4. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020. [Epub ahead of print].
5. Chinazzi M, Davis JT, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020;368:395-400.
6. Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* 2020. [Epub ahead of print].
7. Wei YY, Lu ZZ, Du ZC, et al. Fitting and forecasting the trend of COVID-19 by SEIR(+ CAQ) dynamic model. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020;41:470-5.
8. Wei Y, Zhao Y, Chen F, et al. Principles of dynamics model and its application in forecasting the epidemics and evaluation the efforts of prevention and control interventions. *Chin J Prev Med* 2020;54. doi: 10.3760/cma.j.cn112150-20200315-00340.
9. Zhao Y, Tang S, Peng Z, et al. Simulation of Infectious Disease in a Closed System. *Chinese Journal of Health Statistics* 2010;27:656-8.
10. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Report 9: Impact of non-pharmaceutical interventions

- (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London 2020. doi: <https://doi.org/10.25561/77482>.
11. Hu Z, Ge Q, Li S, et al. Evaluating the effect of public health intervention on the global-wide spread trajectory of Covid-19. medRxiv 2020. doi: <https://doi.org/10.1101/2020.03.11.20033639>.
 12. Zhong N. We are all fighters. *J Thorac Dis* 2020;12:132-3.

Cite this article as: Zhao Y, Wei Y, Chen F. Forecasting of COVID-19: transmission models and beyond. *J Thorac Dis* 2020;12(5):1762-1765. doi: [10.21037/jtd-20-1692](https://doi.org/10.21037/jtd-20-1692)