medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

# Evaluating the secondary transmission pattern and epidemic prediction of the COVID-19 in metropolitan areas of China

Na Hong[1,2], Jie He[2], Yingying Ma[2], Huizhen Jiang[3], Lin Han[2], Longxiang Su[1], Weiguo Zhu[3,4], YunLong[1]

[1]Department of Critical Care Medicine, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[2]Digital China Health Technologies Co. Ltd., Beijing 100080, China

[3]Department of Information Management, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[4]Department of General Internal Medicine, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

## Abstract

Understanding the transmission dynamics of COVID-19 is crucial for evaluating the spread pattern of it, especially in metropolitan areas of China which may cause secondary outbreaks outside Wuhan, the center of the new coronavirus disease outbreak. We used reported data from Jan 24, 2020, to Feb 23, 2020, fitted the model of infection, and on the number of cases reported to estimate likely number of infections in four high risk metropolitan areas, as well as facilitate understanding the COVID-19's spread pattern. A group of SERI model statistical parameters were estimated using Markov Chain Monte Carlo (MCMC) methods, and our modeling integrated the effect of the official quarantine regulation and travel restriction of China. As a result, we estimated that the basic reproductive number R0 is 3.11 in Beijing, 2.78 in Shanghai, 2.02 in Guangzhou, and 1.75 in Shenzhen. In addition, we inferred the prediction results and compared the results of different level of parameters, For example, In Beijing, the predicated peak number of cases is around 466 at the peak time Feb 29, 2020; however, when the city conducts different levels (strict, mild, or weak) of travel restrictions or regulation measures, the estimation



NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

results show that transmission dynamics will change and the peek number of cases shows the changing proportion is between 56%~159%. We concluded that public health interventions would reduce the risks of COVID-19 spreading and more rigorous control and prevention measures will effectively contain its further spread, but risk increases when businesses and social activities returning back before the ending date. Besides, the experiences gained and lessons learned from China are potential to provide evidences supporting for other metropolitan areas and big cities with emerging cases outside China.

## Introduction

In December 2019, since a novel coronavirus, outbreak in Wuhan, China, the epidemic center, it has been quickly spreading out to other provinces, autonomous regions, municipalities and countries. The World Health Organization (WHO) has named it 2019 novel coronavirus disease (COVID-19) and the novel virus severe acute respiratory syndrome coronavirus 2 (SARS-COV-2). The new coronavirus is a strain that has never been found in human before. This virus can cause an acute respiratory disease, common signs of infection include respiratory symptoms, fever, cough, shortness of breath, and dyspnea. In more severe cases, infection can cause pneumonia, severe acute respiratory syndrome, kidney failure, and even death[1].

According to the WHO situation reports, the outbreak of COVID-19 has led to global 79407 confirmed cases and 2622 deaths in 32 countries as of 24th February 2020, of which 64287 were from Hubei, China. Cases have been reported in other areas outside Hubei with numerous even hundreds of cases, including metropolitan areas Beijing (n=399) and Shanghai (n=335) inside China, as well as other countries outside China, such as South Korea (n=833), Japan (n=144), Italy (n=124), With the continuously increased cumulative cases, understanding the COVID-19'sspreadpattern and monitoring its burst trends is crucial by informing and inferring evidences for public health intervention and health care policy making.

Several mathematical models and data analysis approaches were recently reported attempting to estimate the transmission of COVID-19 [2-4]. Public health interventions and transportation restriction effect for the disease transmission has also been evaluated in some studies[5,6]. Researchers have published many different results for forecasting when the outbreak will peak in different areas[7,8]. These models are



2

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

certainly useful to understand the emerging trends of COVID-19. However, there are several challenges to such timely analysis and forecasting. Due to the barriers bought from disease incubation period, asymptomatic infectious, diagnosis testing capacity, overloaded medical stuff and complicated reporting processes, there can be a delay or missed reporting from real evolving situation to confirmation of cases. Furthermore, the adopted models are mostly complicated with many pre-settings or assumptions or parameters values which are likely not accuracy. Although some modeling approaches can estimate parameters values through statistical methods, it is can only contribute a rough simulation for the modeling. As a result, those studies actually achieved different prediction results by using different methods and datasets.

To achieve a relatively objective judgment giving such a new disease and complicated situation with many unknown knowledges, we used mathematical modeling methods to disclose the COVID-19 transmission with four datasets for cross-validation. Since individual data sources may be biased or incomplete, according to the related studies, multiple data sources rather than a single dataset can enable more robust estimation of the underlying dynamics of transmission[9].Therefore, we investigated and collected data from four sources, including released data and official daily reports from commercial technologies company, academic institutes, authorities or local healthcare commissions, and World Health Organization, to minimize the result errors caused by single biased data. The data were obtained from the reports published by Beijing Municipal Health Commission (BMHC)[10] , Shanghai Municipal Health and Family Planning Commission(SMHFPC)[11],Health Commission of Guangdong Province[12] , National Bureau of Statistics of China (NBSC)[13] , Baidu Migration Big Data Platform[14], Center for Systems Science and Engineering (CSSE) of Johns Hopkins University[15],and the WHO coronavirus disease (COVID-2019) situation reports[16]. Considering the cases detected in these 4 cities were all imported cases and their secondary transmission cases, as well as the reported data are available after Jan 20$^{th}$, 2020, during this time, Chinese authorities have implemented prevention measures on these cities to contain the outbreak and prevent the disease spreading, we regarded that the secondary transmission pattern of the COVID-2019 is different compared with the early spread pattern in Wuhan, where the virus is strongly transmitted without any prevention measures. Therefore, we collected data from Jan 24, 2020 (Chinese New Years Eve), to Feb 23, 2020 to give an overall objective estimation forCOVID-19 development in 4 high risk metropolitan



3

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

areas of China: Beijing, Shanghai, Guangzhou, and Shenzhen. We estimated how COVID-19 human-to-human transmitted in these big cities, which have developed considerable cases and with large city population size. We further used these estimates to forecast the potential risks and development trends of these four metropolitan areas inside China.

## Results

**Data Characterization**. To give an overall characterization of epidemic size and dynamics, Figure 1 shows the epidemic curve of COVID-19 cases identified in Beijing, Shanghai, Guangzhou and Shenzhen, from Jan 24 to Feb 23 2020.

**Adjusted SERI Model Estimation.** We summarize and interpret the transmission dynamics of COVID-19 in the 4 municipal areas. Adjusted SEIR model were used to give a prediction for Beijing, Shanghai, Guangzhou and Shenzhen, and the following Figure 2 show the comparisons between predicted results and actual results. The results are under the assumption with no further imported cases to these cities since China implemented strong regulation measures during the observation time.

Based on our observations, inferring from below Table 1 and Figures 3, we also find that the number of infected individuals will change with different levels of public health interventions, strict interventions could decrease the peak number of infected individuals compared with the scenario of weak interventions, where we use different levels of contact rate and to measure the different levels of interventions. The baseline contact rate was derived by MCMC method, and the results show that reducing the contact rate persistently decreases the peak value but may either delay the peak, besides, with the strict public intervention, the number of infected individuals eventually will decrease but may bring forward the peak; with large amount of inflow population after Feb 3 2020, the returning back to work time after holiday leaving, many people returns back to these cities, which is inferred from Baidu transportation index, we added risk factors for contact rate (1.5c, 2c), accordingly, the number of infected individuals will increased, compared with the scenario of decreased contact rate(0.8c, 0.6c).

**Table1**.The effects of contact rate on the peak time and peak value with estimated c=6.2.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

| Areas | Parameter c | 2c | 1.5c | c | 0.8c | 0.6c |
|---|---|---|---|---|---|---|
| Beijing | Days to Peak | 16 | 19 | 25 | 30 | 41 |
| | Peak Value | 605 | 560 | 466 | 398 | 294 |
| Shanghai | Days to Peak | 16 | 19 | 26 | 30 | 42 |
| | Peak Value | 588 | 551 | 469 | 412 | 320 |
| Guangzhou | Days to Peak | 16 | 18 | 26 | 30 | 41 |
| | Peak Value | 515 | 481 | 403 | 353 | 269 |
| Shenzhen | Days to Peak | 17 | 20 | 28 | 34 | 47 |
| | Peak Value | 616 | 572 | 476 | 410 | 308 |

In addition, we compared the transmission dynamics with different quarantined rate of exposed individuals, which reflects the contact tracing capability and management efforts of local governments (Table 2 and Figure 4), and the results show that reducing the quarantined rate of exposed individuals (0.8q, 0.6q) leads to increase of the peak value and delay the peak time. And vice versa, the peak value decreases and bring forward the peak time with higher quarantined rate of exposed individuals (2q, 1.5q).

**Table 2**. The effects of quarantined rate of exposed individuals on the peak time and peak value.

| Areas | Parameter q | 2q | 1.5q | q | 0.8q | 0.6q |
|---|---|---|---|---|---|---|
| Beijing | Days to Peak | 18 | 21 | 25 | 27 | 30 |
| | Peak Value | 259 | 325 | 466 | 566 | 740 |
| Shanghai | Days to Peak | 21 | 24 | 26 | 28 | 29 |
| | Peak Value | 290 | 373 | 469 | 674 | 892 |
| Guangzhou | Days to Peak | 19 | 23 | 26 | 28 | 29 |



5

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

| | | | | | | |
|---|---|---|---|---|---|---|
| | Peak Value | 268 | 352 | 403. | 622 | 821 |
| Shenzhen | Days to Peak | 21 | 25 | 28 | 29 | 31 |
| | Peak Value | 263 | 336 | 476 | 607 | 801 |

**R0 Estimation Results**. We used the MCMC method to fit the model and adopted an adaptive Metropolis-Hastings (MH) algorithm to carry out the MCMC procedure, as a result, we inferred the $R_0$= 3.11, 2.78, 2.02, and 1.75 for Beijing, Shanghai, Guangzhou and Shenzhen, respectively.

## Discussions and Conclusions

From our analysis results, it is strongly verified that reducing secondary infections among close contacts would effectively limit human-to-human transmission, public health measures, such as rapid identification of the cases, tracing and follow up the contacts, infection prevention and control in health care settings, implementation of health measures for travelers, are greatly prevented further spread.

The documented COVID-19 reproduction numbers ranging from 2.0 to 4.9[6,7,17], which based on cases developed in different transmission phases and areas. For instance, the R0 in Wuhan is obviously higher than other cities outside Wuhan. Furthermore, the dataset used in this study is after implementing prevention measures by China government and local authorities, we regarded that the inferred R0 results of four cities are reasonable and interpretable.

In this study, we tried to monitor trends of the COVID-19 after cases were imported into other cities, and estimated the spread pattern by mathematical modeling which could be helpful for evaluating the potential risk and severity of new outbreaks. The results of our study show that, for four metropolitan areas of China, they contain the spreading in effective control so far, however, it still needs to raise awareness in



6

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the population and prevent the potential outbreak risk before the epidemic completed ends, early May 2020, as our predications.

The study has limitation. Current reported data are insufficient to understand the full spread pattern of the transmission and its potential new outbreaks. For example, estimates in this manuscript have a certain extend of uncertainty and delay due to the limitations in reporting mechanism over the course of the natural history of cases, the impact of other potential asymptomatic cases and some hidden unreported cases. However, the outbreak may still come with time when there is a super spreading case, and the outbreak may not come because strict public health interventions, travel restrictions and peoples sell-protection efforts.

As the conclusion from report of WHO-china Joint Mission[18], the COVID-19 transmission dynamics of any outbreak are inherently contextual, people all the world need to work together to defense this disease, including: 1) To enhance understanding of the evolving COVID-19 outbreak in China and the nature and impact of ongoing containment measures; 2) To share knowledge on COVID-19 response and preparedness measures being implemented in countries affected by or at risk of importations of COVID-19; 3)To generate recommendations for adjusting COVID-19 containment and response measures in China and internationally; 4) and To establish priorities for a collaborative program of work, research and development to address critical gaps in knowledge and response and readiness tools and activities.

The public health interventions would reduce the risks of COVID-19 spreading and more rigorous control and prevention measures will effectively contain its further spread. In the face of the global epidemic and with the facilitating of mathematical models, China's experience may help providing epidemiological suggestions for the other countries with rapidly emerging cases.

## Methods

To evaluate spread pattern and estimate the COVID-19 transmission in 4 metropolitan areas, we adopted adjusted SEIR model for modeling the data, respectively. We only considered human-to-human transmission in our models.

**Adjusted SEIR model for COVID-19**. SEIR is a deterministic metapopulation transmission model, and in the SEIR model definition, the population is divided into



7

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

four classes: Susceptible (all the population are likely infected), exposed (people are exposed), infectious (people are infected) and removed (recovered). We assumed the real risk situation started with infectious cases on Feb 3 2020 which is authorities announced returning back to work time after Chinese Spring Festival Leaving. Therefore, we modeled the period from Feb 3 2020. The SEIR model state transaction is as shown in Figure 4. In our estimation, the entire population was initially susceptible since COVID-19 is an emerging new infectious disease and all the people don't have immunization of it, we estimated the initial exposed population using MCMC methods based on the reported data between Jan 24 2020 and Feb 16 2020. We assumed that the median incubation period are 5-6 days (ranging from 0-14 days) based on the report of WHO[19].

Based on the basic SEIR model, we further considered the influence of multiple factors on the transmission pattern during the actual situation of Beijing, including the public health intervention measures, people's self-protection behaviors, diagnosis rate, population flow, etc.

Having the assumption that public health intervention contributed to the control of epidemic dynamics, we imported an index that indicates the changes of contact rate. According to the urban daily-adjusted index of travel intensity from Baidu migration big data platform, for period Jan 24 2020 to Feb 23 2020, we inferred that people's contact rate is obviously lower than normal level of same period last year. Furthermore, considering the Spring Festival population flow and returning back to work time after holiday leaving (Officially announced time is Feb 3, 2020), we regarded that the real risk of 4 metropolitan areas grows with the inflow population increasing started from Feb 3, 2020.

We also estimated the contact rate within Beijing, and the initial number of human infections present in Beijing using the MCMC method. Cases in BMHC and other sources reported between Jan 24 2020 and Feb 23 2020 were used for fitting. Considering the possible complex influence factors, we proposed an adjusted SERI model for COVID-19 estimation, as displayed in Figure 5.

In the adjusted SEIR model, we consider the inflow of city's population, so the total number of people is not fixed and the population is divided into seven classes: Susceptible (all the population are likely infected), exposed (people contacted with



8

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

infection), infectious (people are infected), removed (recovered and dead) , Sq (quarantined susceptible) , Eq (isolated exposed) and H (isolated infected). Besides, we consider the inflow population of Beijing, which we denote P, and the average daily inflow is M. The transmission dynamics are governed by the following system of equations :

$$\begin{cases} \frac{dS}{dt} = -[c\beta + cq(1-\beta)]S(I+\theta E) + \lambda S_q + (1-\tau)M, \\ \frac{dP}{dt} = M, \\ \frac{dE}{dt} = c\beta(1-q)S(I+\theta E) - \sigma E + \tau M, \\ \frac{dI}{dt} = \sigma E - (\delta_I + \alpha + \gamma_I)I, \\ \frac{dS_q}{dt} = (1-\beta)cqS(I+\theta E) - \lambda S_q, \\ \frac{dE_q}{dt} = \beta cqS(I+\theta E) - \delta_q E_q, \\ \frac{dH}{dt} = \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H, \\ \frac{dR}{dt} = \gamma_I I + \gamma_H H \end{cases}$$

Where q is the quarantined rate of exposed individuals, β is the transmission probability and c is the contact rate. Then, the quarantined individuals, if infected (or uninfected), move to the compartment $E_q$ (or $S_q$) at a rate of βcq(or (1 − β)cq). Those who are not quarantined, if infected, will move to the compartment E at a rate of $\beta c(1-q)$. Among them,τ is the exposed ratio of inflow population, and the susceptible ratio is $1-\tau$, θ is the ratio of the transmission ability between the latent and the infected population. We assume that the transmission ability of the people in incubation period and the diagnosed infected patients are same, that is, $\theta = 1$. λ is the isolation contact rate, σ is the transformation rate from the latent to the infected, α is the mortality rate, $\delta_I$ is the isolation rate of the infected, $\gamma_I$ is the recovery rate of the infected population. $\delta_q$ is the conversion rate from isolated latent to isolated infected, and $\gamma_H$ is the recovery rate of isolated infected population.

**Parameter estimates methods.** The MCMC method is one of the commonly used algorithms in modern statistical calculations. This algorithm provides an effective tool

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

for establishing actual statistical models and is widely used in Bayesian calculations of complex statistical models[20]. We use the MCMC method and MH algorithm sampling, use the normal distribution as the recommended distribution, estimate the parameters of the modified SEIR model, apply the data collected from infectious diseases to the above statistical inference, and simulate the process of infectious disease transmission to obtain the parameters c,β,q,$\sigma$,$\delta$I, etc., Use Beijing as example, the parameter estimates and initial values of SEIR model is listed in Table 3.

**Table3**. Parameters and initial values for adjusted SEIR model (Beijing)

| Parameter | Definitions | Referenced Value | Methods |
|---|---|---|---|
| c | Contact rate | 6.2 | MCMC |
| β | Probability of transmission per contact | 1.75302e-9 | MCMC |
| q | Quarantined rate of exposed individuals | 3.80967e-5 | MCMC |
| $\sigma$ | Transition rate of exposed individuals to the infected class | 1/6 | Source: WHO |
| $\lambda$ | Rate at which the quarantined uninfected contacts were released into the wider community | 1/14 | Source: NHC |
| $\delta$I | Transition rate of infected individuals to the quarantined infected class | 0.130180 | MCMC |
| $\delta$q | Transition rate of quarantined exposed individuals to the quarantined infected class | 0.128036 | MCMC |
| $\gamma_1$ | Recovery rate of symptomatic infected individuals | 0.004695 | MCMC |
| $\gamma_H$ | Recovery rate of quarantined infected individuals | 0.009158 | MCMC |
| $\alpha$ | Disease□induced death rate | 0.2% | Source: WHO ( 2-20 |



10

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

| | | | |
|---|---|---|---|
| | | | report ) [21] |

In addition, to simulate the contact rate for model estimation, we used urban travel index data from Baidu, a major internet company of China that hosts popular navigator APP Baidu maps, which indirectly monitor real time urban travel intensity and population flow. The Baidu impact index of travel intensity and population flow is converted into corresponding coefficient of contact rate and quarantined susceptible population.

**The basic reproduction number R0 estimates.** At the beginning of the onset, when all people are susceptible, R0 is the average number of new infections directly caused by a case in a population of people who are all susceptible. Given the model structure with quarantine and isolation, we used the next generation matrix to derive a formula for the control reproduction number when control measures are in force, as follows:

$$R_0 = [\frac{\beta c(1-q)}{\delta_I + \alpha + \gamma_I} + \frac{\beta c\theta(1-q)}{\sigma}]S_0$$

## References


1 *Coronavirus*, <https://www.who.int/health-topics/coronavirus> (2020).

2 Hermanowicz, S. W. Forecasting the Wuhan coronavirus (2019-nCoV) epidemics using a simple (simplistic) model. *medRxiv* (2020).

3 Liu, T. *et al.* Transmission dynamics of 2019 novel coronavirus (2019-nCoV). (2020).

4 Imai, N. *et al.* Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. *Imperial College London* (2020).

5 Tang, B. *et al.* Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions. *Journal of Clinical Medicine* **9**, 462 (2020).

6 Zhao, S. *et al.* Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases* (2020).




11

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .


7 Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, doi:10.1016/S0140-6736(20)30260-9. (2020).

8 Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A. & Jewell, C. P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* (2020).

9 Funk, S. & Eggo, R. M. Early dynamics of transmission and control of 2019-nCoV: a mathematical modelling study. doi:10.1101/2020.01.31.20019901.

10 Decker, C. in *SAS Global Forum 2010 Conference. Seattle, WA, USA.* 14.

11 *Shanghai Municipal Health and Family Planning Commission*, <http://wsjkw.sh.gov.cn/yqfk2020/> (2020).

12 *Health Commission of Guangdong Province*, <http://wsjkw.gd.gov.cn/zwyw_yqxx/> (2020).

13 *National Bureau of Statistics of China* <http://data.stats.gov.cn/> (2020).

14 *Baidu Migration Big Data Platform*, <http://qianxi.baidu.com/> (2020).

15 Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30120-1 (2020).

16 Organization, W. H. *Coronavirus disease (COVID-2019) situation reports*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (2020).

17 Shen, M., Peng, Z., Xiao, Y. & Zhang, L. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv* (2020).

18 Organization, W. H. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)* <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> (2020).

19 Organization, W. H. *Coronavirus disease 2019 (COVID-19) Situation Report*, <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200219-sitrep-30-covid-19.pdf> (2020).

20 Endo, A., van Leeuwen, E. & Baguelin, M. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics* **29**, 100363, doi:10.1016/j.epidem.2019.100363. (2019).

21 *Coronavirus disease 2019 (COVID-19) Situation Report – 31*, <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200220-sitrep-31-covid-19.pdf> (2020).




12

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Acknowledgments

This study is supported by National Key R&D Program of China, project 2018YFC0116905

## Author Information

### Affiliations

Peking Union Medical College, Beijing, China
*Na Hong, Huizhen Jiang, Longxiang Su, Weiguo Zhu and Yun Long*

Digital China Health Technologies Co. Ltd., Beijing, China
*Jie He, Yingying Ma, Lin Han and Na Hong*

### Contributions

N. H, J. H and Y. M contributed equally. L. S, W. Z and Y. L take responsibility for the integrity of the work as a whole, from inception to published article; N. H and L. S were responsible for study design and conception; J. H, Y. M and H. J were responsible for data modeling and analysis; W. Z and L. H interpreted the results; N. H, J. H, Y.M and L. S drafted the manuscript. All authors revised the manuscript for important intellectual content.

### Corresponding authors

Longxiang Su, Weiguo Zhu, and Yun Long

## Ethics declarations

Competing interests

The authors declare no competing interests.



13

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

## Additional Information

None

## Legends legends

Figure 1: Cumulative and daily reported cases of 4 metropolitan areas of China

Figure 2: The comparison between the predicted number and the actual number of infected people and cured people, 4 cities.

Figure 3: Infected population curve with different contact rate, 4 cities.

Figure 4: Infected population curve with different quarantined rate of exposed individuals, 4 cities.

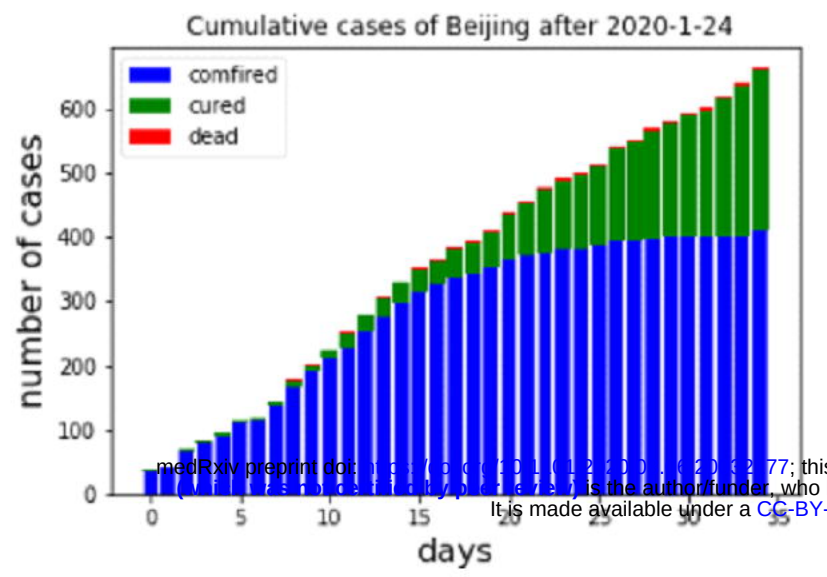
Figure 5: SEIR Model

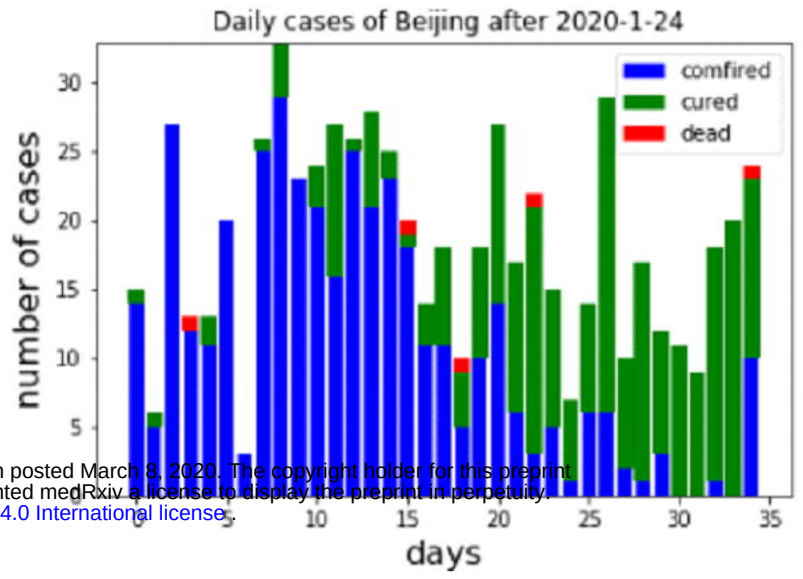
Figure 6: Adjusted SEIR model for COVID-19

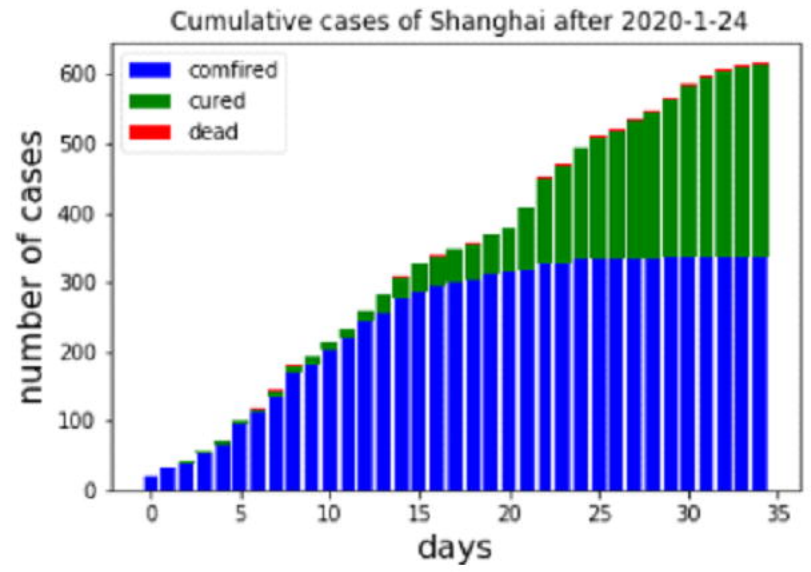


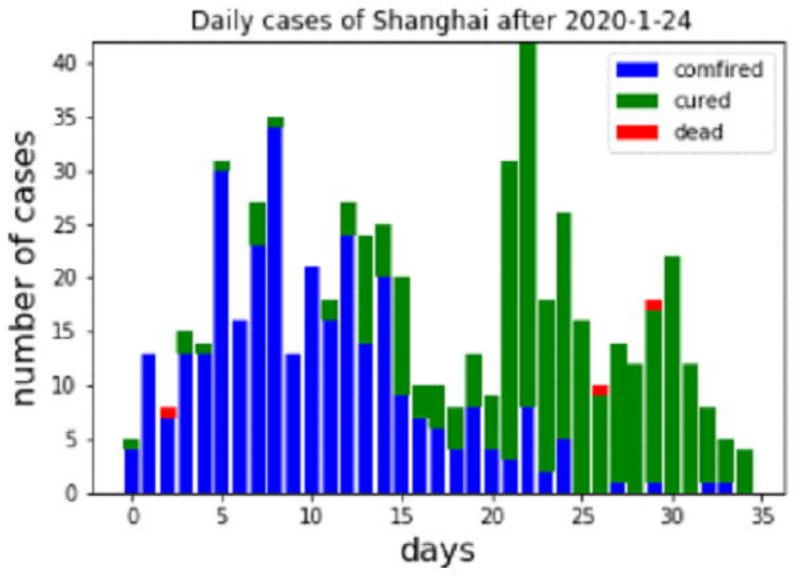
14

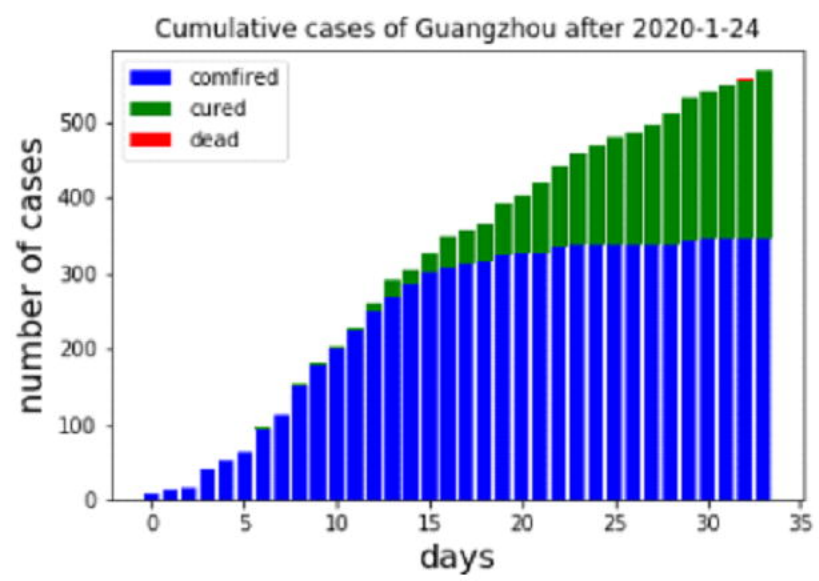
medRxiv preprint doi: https://doi.org/10.1101/2020.02.27; this version posted March 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license.

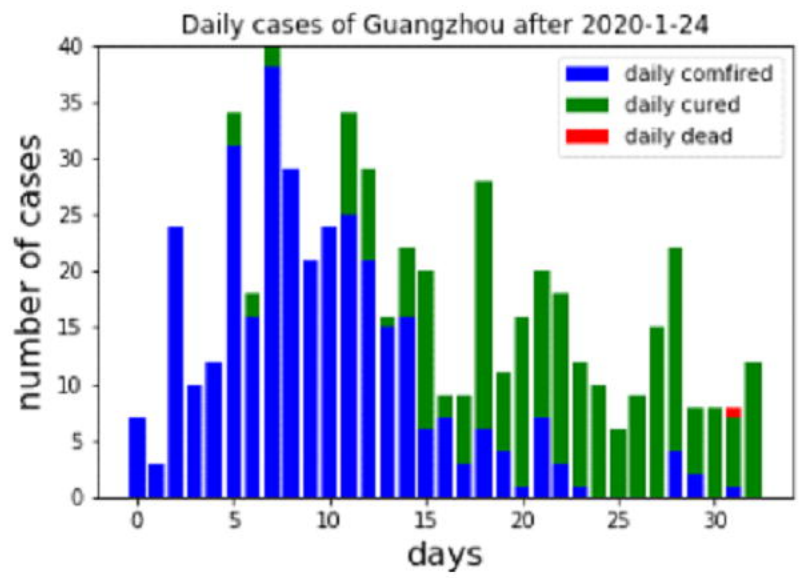


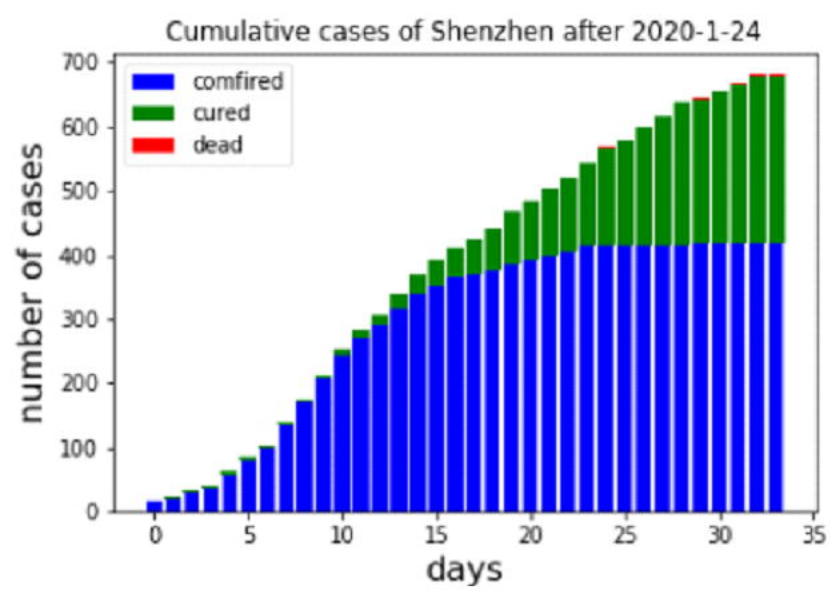



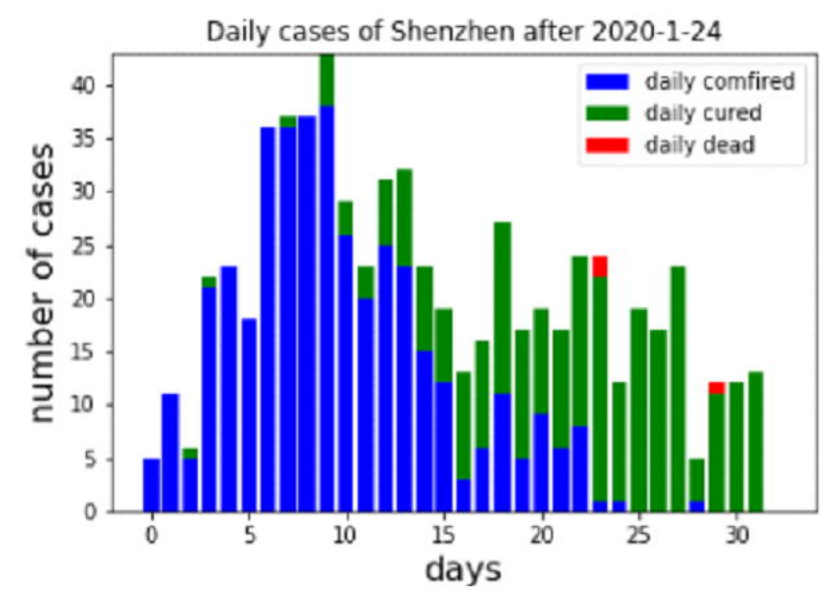



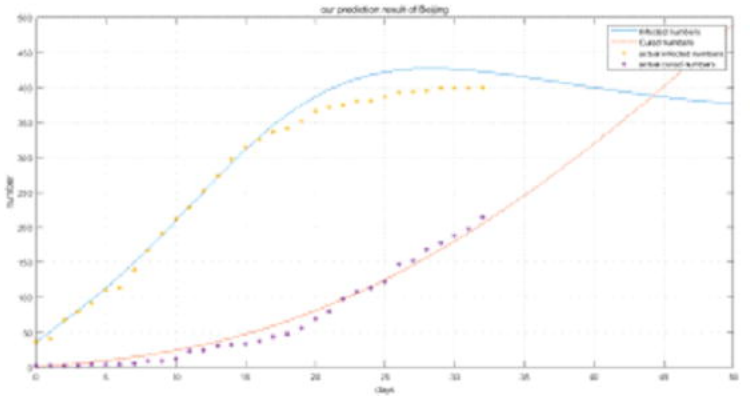



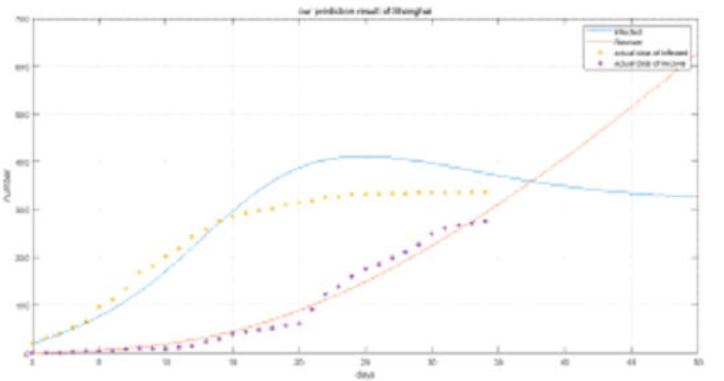



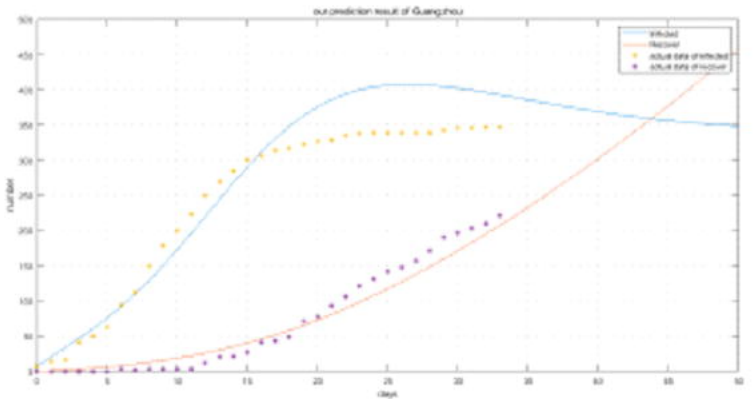



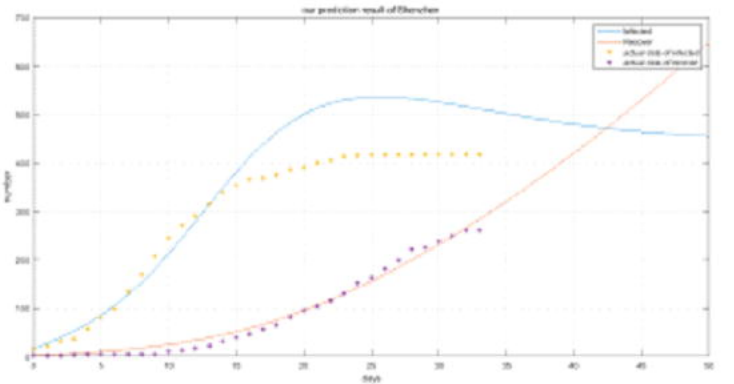



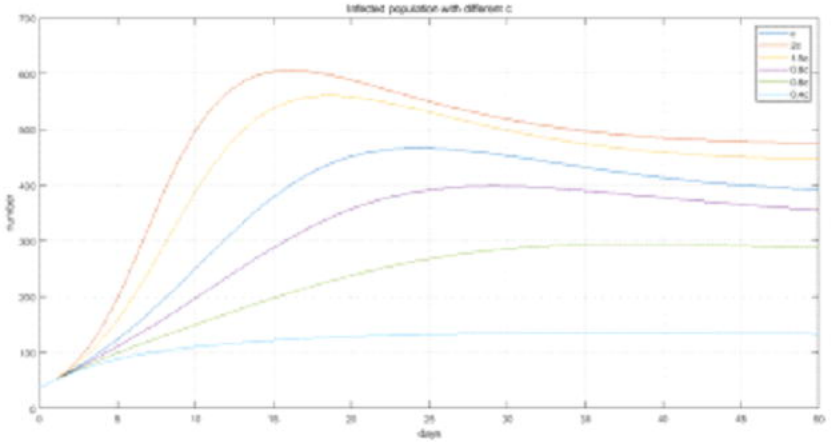



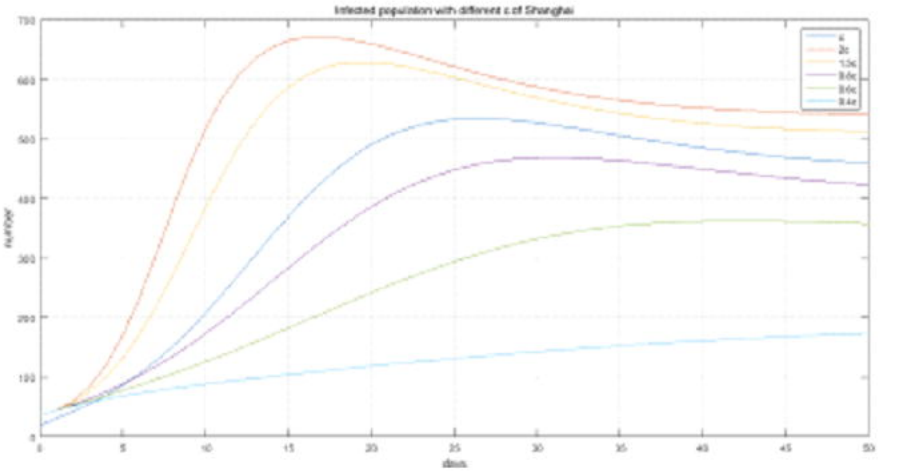

our prediction result of Beijing
Infected numbers
Cured numbers
actual infected numbers
actual cured numbers
number
days

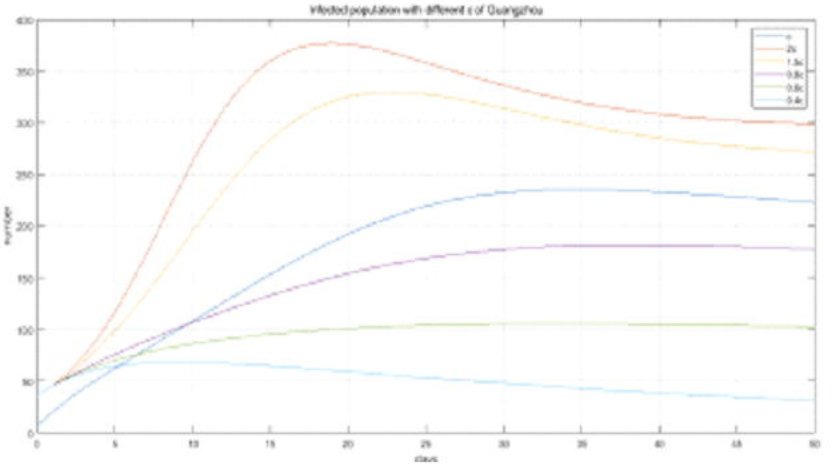

Infected
Recover
number
days

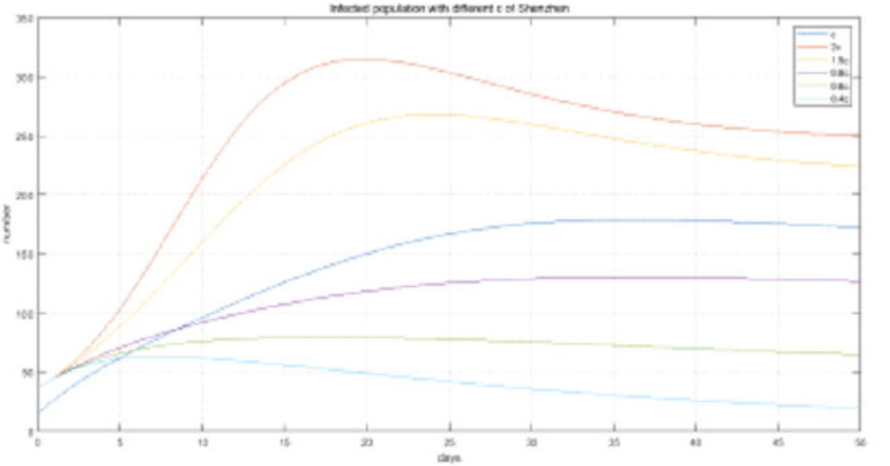

our prediction result of Guangzhou
Infected
Recover
actual data of infected
actual data of recover
number
days

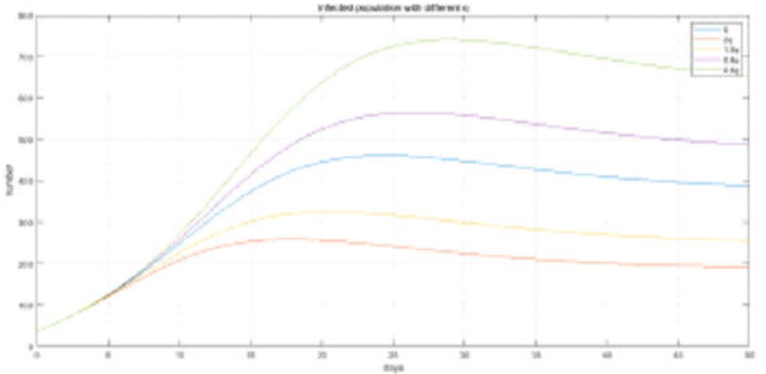

Infected
Recover
number
days

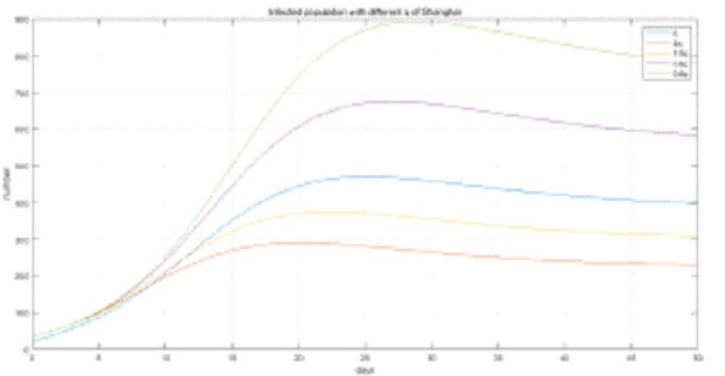

Infected population with different c
c
2c
1.5c
0.8c
0.6c
0.4c
number
days

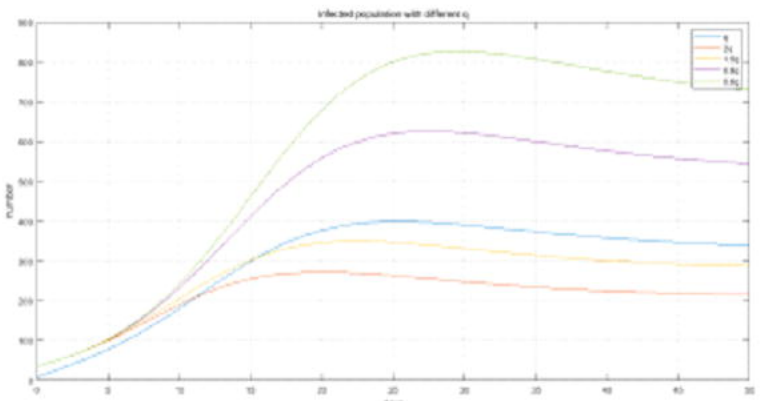

Infected population with different c of Shanghai
c
2c
1.5c
0.8c
0.6c
0.4c
number
days

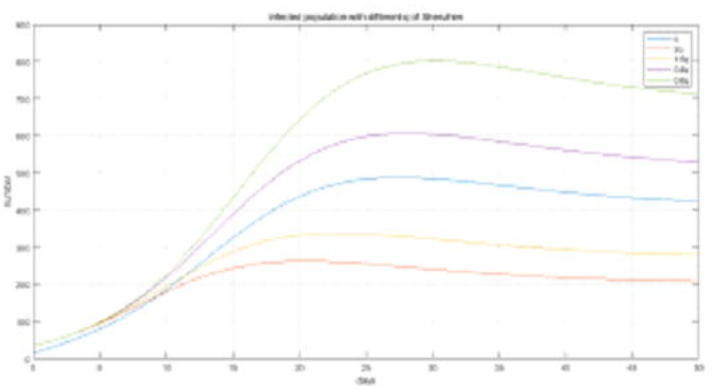

Infected population with different c of Guangzhou
c
2c
1.5c
0.8c
0.6c
0.4c
number
days

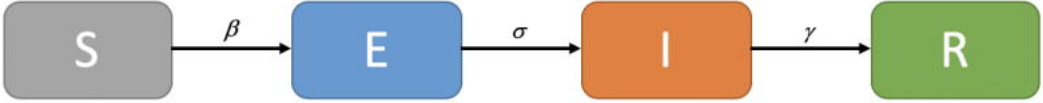

Infected population with different c of Shenzhen
c
2c
1.5c
0.8c
0.6c
0.4c
number
days

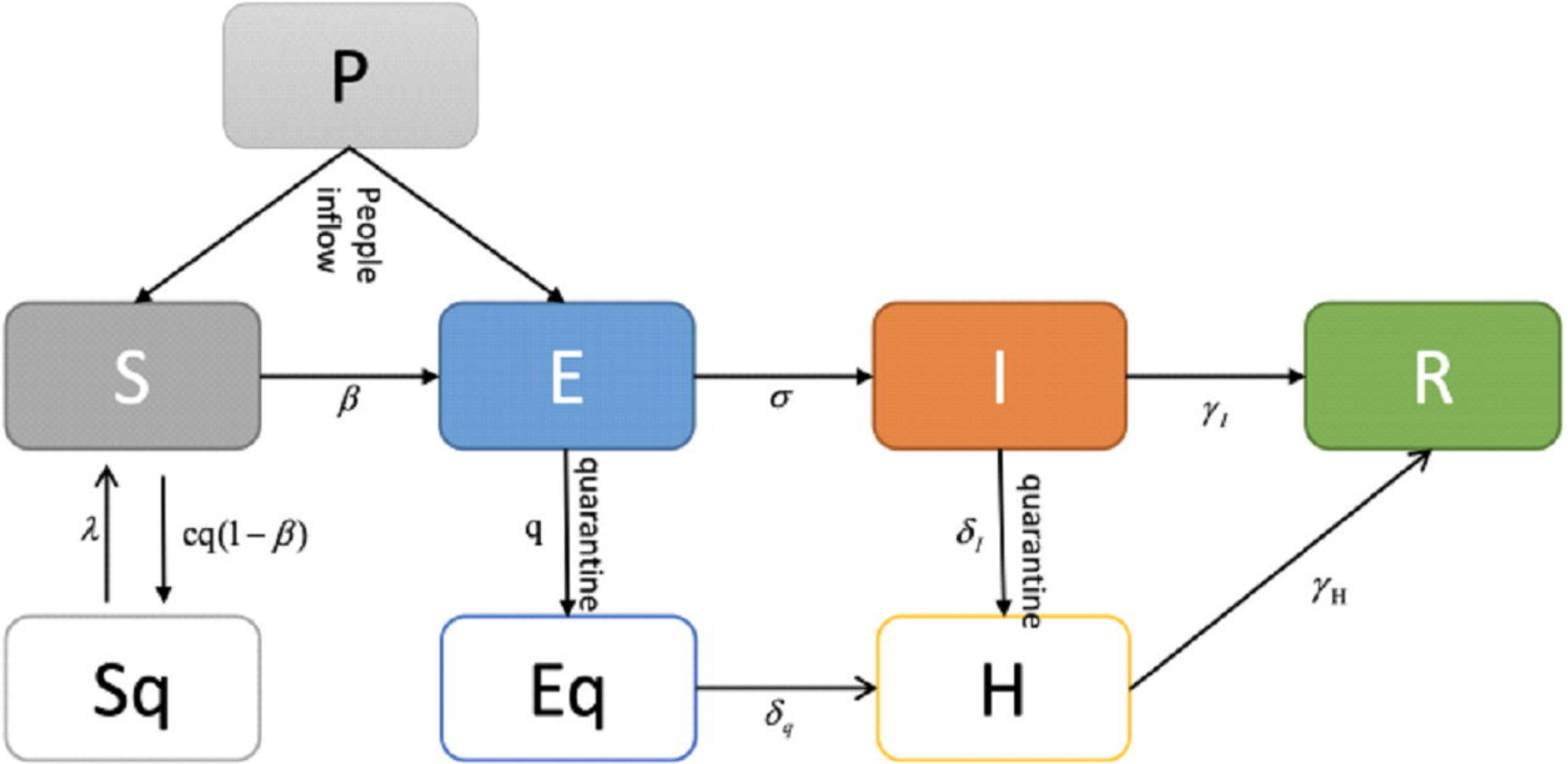









S
β
E
σ
I
γ
R


P
People inflow
S
E
I
R
$\beta$
$\sigma$
$\gamma_I$
$\lambda$
$cq(1-\beta)$
q
quarantine
$\delta_I$
quarantine
$\gamma_H$
Sq
Eq
H
$\delta_q$