medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Evaluation of the secondary transmission pattern and epidemic prediction of COVID-19 in the four metropolitan areas of China

Longxiang Su[1], Na Hong[2], Jie He[2], Yingying Ma[2], Xue Wang[3], Qun Xu[4,5], Huizhen Jiang[7], Lin Han[2], Fengxiang Chang[2], Weiguo Zhu[6,7], Yun Long[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]Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[4]Department of Epidemiology and Biostatistics, Institute of Basic Medicine Sciences, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China

[5]Center of Environmental and Health Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100005, China

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

[7]Department of Information Management, 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 its spread pattern, especially in metropolitan areas of China, as its spread can lead to secondary outbreaks outside Wuhan, the center of the new coronavirus disease outbreak. In addition, the experiences gained and lessons learned from China have the potential to provide evidence to support other metropolitan areas and large cities outside China with emerging cases. We used data reported from January 24, 2020, to February 23, 2020, to fit a model of infection, estimate the likely number of infections in four high-risk metropolitan areas based on the number of cases reported, and increase the understanding of the COVID-19 spread pattern. Considering the effect of the official quarantine regulations and travel restrictions for China, which began January 23~24, 2020, we used the daily travel intensity index from the Baidu Maps app to roughly simulate the level of restrictions and estimate the proportion of the

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 April 7, 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 .

quarantined population. A group of SEIR model statistical parameters were estimated using Markov chain Monte Carlo (MCMC) methods and fitting on the basis of reported data. As a result, we estimated that the basic reproductive number, R_0, was 2.91 in Beijing, 2.78 in Shanghai, 2.02 in Guangzhou, and 1.75 in Shenzhen based on the data from January 24, 2020, to February 23, 2020. In addition, we inferred the prediction results and compared the results of different levels of parameters. For example, in Beijing, the predicted peak number of cases was approximately 466 with a peak time of February 29, 2020; however, if the city were to implement different levels (strict, mild, or weak) of travel restrictions or regulation measures, the estimation results showed that the transmission dynamics would change and that the peak number of cases would differ by between 56% and ~159%. We concluded that public health interventions would reduce the risk of the spread of COVID-19 and that more rigorous control and prevention measures would effectively contain its further spread but that the risk will increase when businesses and social activities return to normal before the end of the epidemic. 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.

Keywords：COVID-19, novel coronavirus，secondary transmission，epidemic prediction，SEIR，basic reproduction number

Introduction

The World Health Organization (WHO) named the virus 2019 novel coronavirus disease (COVID-19) and the novel virus severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), which has attracted worldwide attention. The new coronavirus is a strain that has never been found in humans before. This virus can cause an acute respiratory disease, and 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 WHO situation reports, the outbreak of COVID-19 has led to 79407 confirmed cases worldwide and 2622 deaths in 32 countries as of February 24, 2020, of which 64287 were from Hubei, China. Numerous cases have been reported in other areas outside Hubei, including metropolitan areas of Beijing (n=399) and

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Shanghai (n=335) inside China, as well as other countries outside China, such as South Korea (n=833), Japan (n=144), and Italy (n=124). With the continuously increasing number of cases, understanding the spread pattern of COVID-19 and monitoring spikes in the number of cases are crucial steps to provide evidence that can guide public health intervention strategies and healthcare policy making.

Several mathematical models and data analysis approaches attempting to estimate the transmission of COVID-19 have been recently reported[2-4]. Public health interventions and transportation restriction effects for disease transmission have also been evaluated in some studies[5,6]. Some studies indicated that public intervention measures greatly mitigate the final size of the epidemic, and shift the turning point about 24 days before the turning point without these measures[7]. Some noted that travel restrictions would not affect much unless combined with a 50% or higher reduction of transmission in the community[8]. And a report from Imperial College COVID-19 Response Team concluded that the intensive intervention or something equivalently effective, such as combining home isolation of suspect cases, home quarantine of those living in the same household as suspect cases, and social distancing of the elderly and others at most risk of severe disease, could reduce transmission; however, it will need to be maintained until a vaccine becomes available, and the team also predicted that transmission will quickly rebound if interventions are relaxed so it requires multiple interventions to be combined to have a substantial impact on transmission[9]. Besides, to predict the outbreak size and time, researchers have published many different results for forecasting when the outbreak will peak in different areas[10,11]. These models are certainly useful to understand the emerging trends of COVID-19. However, there are several challenges to such timely analyses and forecasting. Due to barriers, such as the disease incubation period, asymptomatic infection, diagnosis testing capacity, overloaded medical staff and complicated reporting processes, there can be a delay or missed reporting in this evolving situation regarding the confirmation of cases. Furthermore, the adopted models have mostly been complicated with many pre-settings or assumptions or parameter values that are likely not accurate. Although some modeling approaches can estimate parameter values through statistical methods, they can only contribute a rough simulation for the modeling. As a result, those studies achieved different prediction results by using different methods and datasets.

To achieve a relatively objective judgment, given that that this new disease and complicated situation has many unknown factors, we used mathematical modeling methods to characterize COVID-19 transmission and used multiple datasets for

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

ensuring the data reliability. Since individual data sources may be biased or incomplete, according to related studies, the use of multiple data sources rather than a single dataset can enable a more robust estimation of the underlying dynamics of transmission[12]. Therefore, we investigated and collected data from four sources, including released data and official daily reports from commercial technology companies, academic institutes, authorities or local healthcare commissions, and the World Health Organization, to minimize the resulting errors caused by potentially biased single data sources. The data were obtained from the Beijing Municipal Health Commission (BMHC)[13], Shanghai Municipal Health and Family Planning Commission (SMHFPC)[14], Health Commission of Guangdong Province[15], National Bureau of Statistics of China (NBSC)[16], Baidu Migration Big Data Platform (BMBDP)[17], Center for Systems Science and Engineering (CSSE) of Johns Hopkins University[18], and WHO coronavirus disease (COVID-2019) situation reports[19]. Considering that the cases detected in these four cities were all imported and secondary transmission cases, and based on the reported data available after January 20, 2020, Chinese authorities have implemented prevention measures in these cities to contain the outbreak and prevent the disease from spreading; thus, we considered the secondary transmission pattern of COVID-2019 to be different than the early spread pattern in Wuhan, where the virus was rampantly transmitted without any prevention measures. Therefore, we collected data from January 24, 2020 (Chinese New Year's Eve) to February 23, 2020 to give an overall objective estimation of COVID-19 development in 4 high-risk metropolitan areas of China: Beijing, Shanghai, Guangzhou, and Shenzhen. We estimated how COVID-19 human-to-human transmission occurred in these large cities, which have developed considerable cases. We further used these estimates to forecast the potential risks and development trends of these four metropolitan areas inside China.

Methods

To evaluate the COVID-19 spread pattern and estimate its transmission in 4 metropolitan areas, we used an adjusted SEIR model with data. We only considered human-to-human transmission in our models.

Adjusted SEIR model for COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 SEIR model is a deterministic metapopulation transmission model in which the population is divided into four classes: S (susceptible, people who are likely to be infected), E (exposed, people who are exposed), I (infectious, people who are infected) and R (removed, recovered and dead persons). We assumed that the epidemic risk started with infectious cases on February 3, 2020, when authorities announced that people were returning to work after the Chinese Spring Festival holiday. Therefore, we modeled a period beginning on February 3, 2020. The SEIR model state transition is shown in Figure 5. In our estimation, the entire population was initially susceptible since COVID-19 is an emerging new infectious disease and not all people have immunity against it. In January (before Chinese New Year) , there were estimated 3.246, 2.847, 3.430, 3.271 million people flow out from Beijing, Shanghai, Guangzhou and Shenzhen, respectively. We took this outflow number out from these four cities initial population and assume they inflow back after Chinese New Year by February 17, 2020. We estimated the initial exposed population using number of confirmed cases during the next 7 days. We assumed that the median incubation period was 5-6 days (ranging from 0-14 days) based on the WHO report[20].

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

Assuming that public health interventions contributed to the control of the dynamics of the epidemic, we incorporated an parameter that indicates the changes in the population flow into the model. According to the inflow index, outflow index and urban daily adjusted index of the travel intensity from the Baidu Migration Big Data Platform, for the period from January 24, 2020 to February 23, 2020, we inferred that the people's activity was obviously lower than the normal level for the same period last year. Furthermore, considering the Spring Festival population flow and those returning to work after the holiday (officially announced as February 3, 2020), we regarded that the risk for these 4 metropolitan areas grows with the inflow population increase starting on February 3, 2020, an average introduced number of cases were counted into the model.

We also estimated the parameter values within these cities using the MCMC method. Cases in the reported data and other sources reported between January 24, 2020 and February 23, 2020 were used to adjust the model. Considering the possible

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

complex influencing factors, we proposed an adjusted SEIR model for COVID-19 estimation, as displayed in Figure 6.

In the adjusted SEIR model, we considered the inflow of the city's population, so the total number of people was not fixed, and the population was divided into seven classes: S (susceptible, people who are likely to be infected), E (exposed, people who are exposed), I (infectious, people who are infected), R (recovered and dead persons), Sq (quarantined susceptible persons), Eq (isolated exposed persons) and Iq (isolated infected persons). The transmission dynamics are governed by the following equations：

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

where q is the quarantined proportion of exposed individuals, β is the transmission probability per contact and c is the contact rate which defines how many people are contacted with an infected person per day, i is the estimated infected people within the inflow population each day. The quarantined infected people moved to the compartment E_q at a rate of βcq, with the quarantined uninfected people moved to the compartment S_q at a rate of (1–β)cq. Those who were not quarantined, if infected, moved to compartment E at a rate of $\beta c(1-q)$. θ is the transmission capability between the latent and the infected population. According to the reported results of related work, the transmission capability of the people in the incubation period and the diagnosed infected patients are similar[21], we assume that $\theta = 1$. λ is the transition rate from the quarantined to susceptible, σ is the transition rate from the exposed to the infected population, α is the mortality rate, δ_I is the transition rate from the infected population to quarantined infected population, and γ_I is the recovery rate of the infected population. δ_q is the transition rate from the quarantined

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

exposed population to the quarantined infected population, and γ_H is the recovery rate of the quarantined infected population.

Parameter estimate methods

The MCMC method is a commonly used algorithm in modern statistical calculations. This algorithm provides an effective tool for establishing statistical models and is widely used in Bayesian calculations of complex statistical models[22]. We used the MCMC method and Metropolis-Hastings(MH) algorithm sampling[23], with a normal distribution as the recommended distribution, estimated the parameters of the modified SEIR model, to obtain the baseline estimation of parameters, and incorporated the data collected from infectious disease reports into the above statistical inference, and simulated the process of infectious disease transmission to further fix some parameters on the basis of fitting reported data. Using Beijing as an example, the parameter estimates and initial values of the SEIR model are listed in Table 3.

In addition, to simulate the contact rate for model estimation, we used urban travel index data from Baidu, a major internet company in China that hosts the popular navigator app Baidu Maps, which indirectly monitors the real-time urban travel intensity and population flow. The Baidu index of travel intensity and population flow was converted into the corresponding coefficients for the contact rate and the quarantined susceptible population. In terms of Baidu index, we simulated people's activity level by comparing our observed period (under strict interventions) with normal level in same period last year, we also consider the assumption scenario that when people returning back to work (limited interventions), accordingly, we added the coefficients (0.6c, 0.8c, c, 1.5c, 2c) for the baseline contact rate to compare different effectiveness of interventions. Similarly, the coefficients were added to baseline quarantine proportion (0.6q, 0.8q, q, 1.5q, 2q).

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Basic reproduction number R_0 estimates

At the onset, when all people are susceptible, R_0 is defined as the average number of new infections directly caused by a case in a population of people who are all susceptible. Given the model structure includes quarantine and isolation, we used the next generation matrix to derive a formula for the basic reproduction number after public health interventions were executed, the principal eigenvalue of the next generation matrix is the expectation of population growth and the equation is as follows and the parameter definition is same with adjusted SEIR model.

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

Results

Data Characterization

To characterize the overall epidemic size and dynamics, Figure 1 shows the epidemic curve of COVID-19 cases identified in Beijing, Shanghai, Guangzhou and Shenzhen from January 24, 2020 to February 23, 2020.

Adjusted SEIR Model Estimation

We summarized and interpreted the transmission dynamics of COVID-19 in the 4 municipal areas. The adjusted SEIR model was used to predict cases in Beijing, Shanghai, Guangzhou and Shenzhen, and Figure 2 shows the comparisons between the predicted results and actual results. The results are based on an assumption of no further imported cases to these cities since China implemented strong regulation measures during the observation period.

Based on our observations from the data shown in Table 1 and Figure 3 below, we also found that the number of infected individuals changed with different levels of public health interventions and that strict interventions could decrease the peak number of infected individuals compared with the scenario of weak interventions; accordingly, we used different contact rates to reflect the different levels of interventions. The baseline contact rate was derived by the MCMC method, and the results show that reducing the contact rate persistently decreased the peak value or could delay the peak. In addition, with strict public intervention, the number of infected individuals eventually decreased and the peak appeared sooner than it would

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

with weak intervention methods. After February 3, 2020, as people returned to work after a holiday, many people returned to these cities, which was inferred from the Baidu transportation index. We added this information into the risk factors for the contact rate (1.5c, 2c). Accordingly, the number of infected individuals increased compared with the scenario of a decreased contact rate (0.8c, 0.6c).

In addition, we compared the transmission dynamics with different quarantined proportion of exposed individuals, which reflects the contact tracing capability and management efforts of local governments, and the results show that reducing the quarantined proportion of exposed individuals (0.8q, 0.6q) led to an increase in the peak value and delayed the peak time. Conversely, the peak value decreased and an earlier peak time occurred with a higher quarantined proportion of exposed individuals (2q, 1.5q).

R_0 Estimation Results.

We used the MCMC method to fit the model and adopted an adaptive MH algorithm to carry out the MCMC procedure. As a result, we inferred R_0= 2.91, 2.78, 2.02, and 1.75 for Beijing, Shanghai, Guangzhou and Shenzhen, respectively.

Discussions

Our analysis results strongly demonstrate that reducing secondary infections among close contacts would effectively limit human-to-human transmission, and public health measures, such as the rapid identification of cases, tracing and following up with people who had contact with an infected person, infection prevention and control in health care settings, and the implementation of health measures for travelers, can greatly prevented further spread of the disease.

The documented COVID-19 reproduction numbers range from 2.0 to 4.9[6,10,24], which are based on cases that developed during different transmission phases and in different areas. For instance, the R_0 in Wuhan was obviously higher than that in other cities outside Wuhan during the timeframe analyzed. Furthermore, after implementing the prevention measures employed by the Chinese government and local authorities, we regarded the inferred R_0 results of the four cities as reasonable and interpretable.

In this study, we aimed to monitor COVID-19 trends after cases were imported into other cities and countries and estimate the spread pattern by mathematical modeling, which can be helpful for evaluating the potential risk and severity of new

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

outbreaks. The results of our study show that, for four metropolitan areas of China, the containment measures were effective control at that time; however, it is imperative to raise awareness in the population and prevent potential outbreak risks any time. The study has limitations. The present reported data are insufficient to understand the full epidemiological pattern of COVID-19 transmission and new potential outbreaks. For example, the estimates in this manuscript have a certain extent of uncertainty and delays due to the limitations in reporting mechanisms over the course of the natural history of the cases, the impact of other potential asymptomatic cases and some unreported cases. Some studies conducted assumption that a small fraction, 20%, were not reported[7] and others reported the estimated asymptomatic proportion was 17.9%[25] or 60%[21]. Evidently, such asymptomatic infectious cases are not fully reported by authorities. In the future, we will evaluate how will the number of unreported cases influence the severity of the epidemic and consider involving these unreported data to our mathematical model.

As concluded from the WHO-China Joint Mission report[26], the COVID-19 transmission dynamics are inherently contextual, as are the dynamics for any outbreak, and people worldwide need to work together to defend against this disease, including 1) to enhance the understanding of the evolving COVID-19 outbreak in China and the nature and the impact of ongoing containment measures; 2) to share knowledge on the 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; and 4) to establish priorities for a collaborative program of work, research and development to address critical gaps in knowledge, responses, readiness tools and strategies.

As a consequence of our study, we concluded that the outbreak could be greatly reduced by strict public health interventions. The public intervention strategies and implemented protection measures conducted in these four areas may help provide epidemiological suggestions to governments that guide measures for the international cases that are rapidly emerging.

References

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

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).

7 Liu, Z., Magal, P., Seydi, O. & Webb, G. Understanding Unreported Cases in the COVID-19 Epidemic Outbreak in Wuhan, China, and the Importance of Major Public Health Interventions. *Biology* **9**, 50 (2020).

8 Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020).

9 Ferguson, N. M. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College, London. DOI: https://doi. org/10.25561/77482* (2020).

10 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).

11 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).

12 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.

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

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

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

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

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

18 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).

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

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

20 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).

21 Qiu, J. Covert coronavirus infections could be seeding new outbreaks. *Nature* (2020).

22 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).

23 Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. (1970).

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

25 Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25** (2020).

26 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).

27 *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).

Author Contributions

L. S and N. H contributed equally. 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; F. C collected and cleaned the data; J. H, Y. M and H. J were responsible for data modeling and analysis; W. Z, Q. X and L. H interpreted the results; L. S and N. H drafted the manuscript. All authors revised the manuscript for important intellectual content.

Corresponding authors

Weiguo Zhu(zhuwg@pumch.cn), and Yun Long(ly_icu@aliyun.com)

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Declarations

Author Na Hong, Jie He, Yingying Ma, Lin Han and Fengxiang Chang were employed by the company Digital China Health Technologies Co. Ltd., Beijing, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Legends

Figure 1. Cumulative and daily reported cases in 4 metropolitan areas in China.

Figure 2. Comparison of the predicted and actual numbers of infected and cured people in 4 cities.

Figure 3. Infected population curves with different contact rates for 4 cities.

Figure 4. Infected population curve with different quarantined proportion of exposed individuals in 4 cities.

Figure 5. SEIR Model

Figure 6. Adjusted SEIR model for COVID-19

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Tables

Table 1. The effects of the contact rate on the peak time and peak value with an estimated value.

Areas	Parameter c	2c	1.5c	c	0.8c	0.6c

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Beijing	Days to Peak	16	19	27	31	41
	Peak Value	642	608	476	399	286
Shanghai	Days to Peak	16	19	25	29	40
	Peak Value	592	545	473	406	309
Guangzhou	Days to Peak	16	18	25	28	40
	Peak Value	515	481	403	353	279
Shenzhen	Days to Peak	17	20	25	32	45
	Peak Value	688	542	487	478	377

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

Table 2. The effects of the 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	20	22	27	29	32
	Peak Value	259	325	476	576	742
Shanghai	Days to Peak	20	22	25	27	29
	Peak Value	290	378	473	662	886
Guangzhou	Days to Peak	20	23	25	27	29

medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 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 .

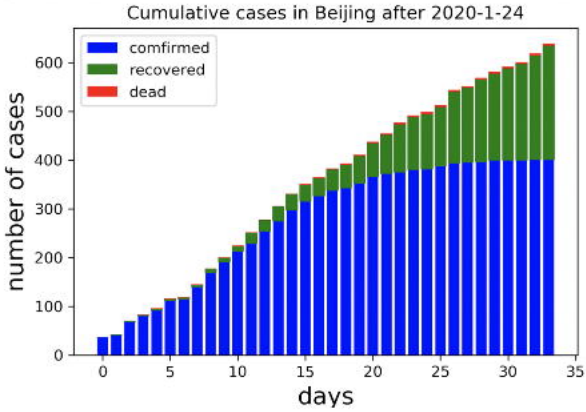
	Peak Value	389	352	403	643	842
Shenzhen	Days to Peak	18	21	25	25	27
	Peak Value	272	329	487	598	789

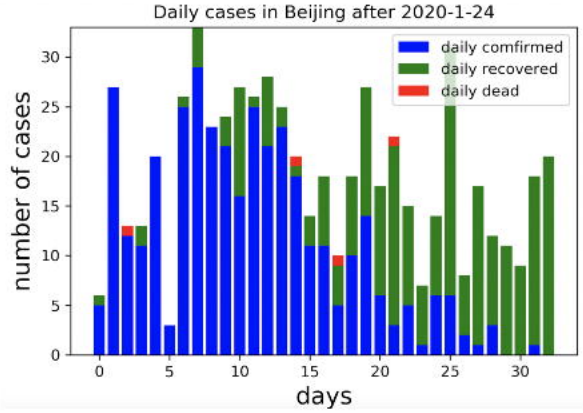
medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

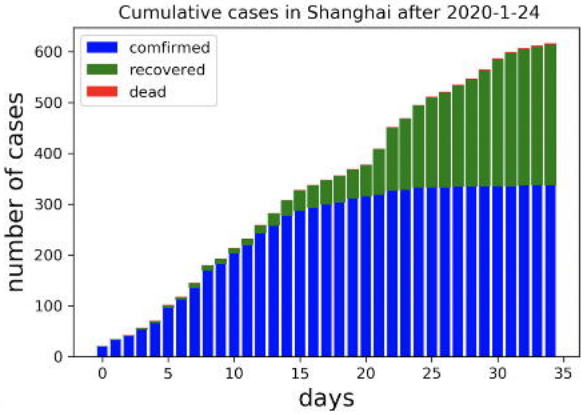
Table 3. Parameters and initial values for the adjusted SEIR model (Beijing)

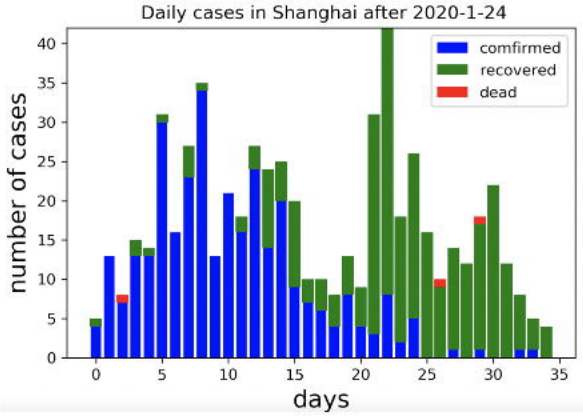
Parameter	Referenced Value	Methods
c	5.4	MCMC and data fitting
β	2.18e-9	MCMC and data fitting
q	3.4e-5	MCMC and data fitting
σ	1/6	Source: WHO
λ	1/14	Source: NHC
δI	0.13	MCMC and data fitting
δq	0.13	MCMC and data fitting
γ_1	0.0046	MCMC and data fitting
γ_H	0.0092	MCMC and data fitting
α	0.2%	Source: WHO (2-20 report) [27]

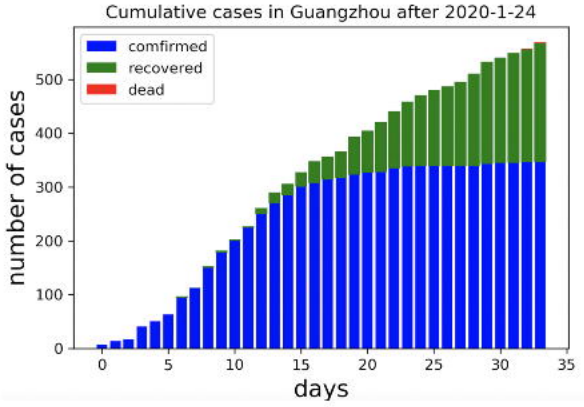
medRxiv preprint doi: https://doi.org/10.1101/2020.03.06.20032177; this version posted April 7, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

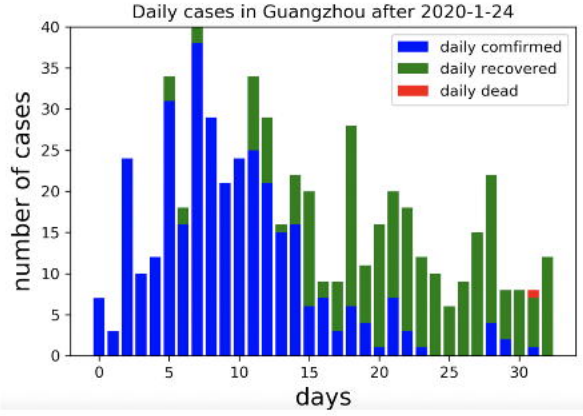

Cumulative cases in Beijing after 2020-1-24
comfirmed
recovered
dead
number of cases
days

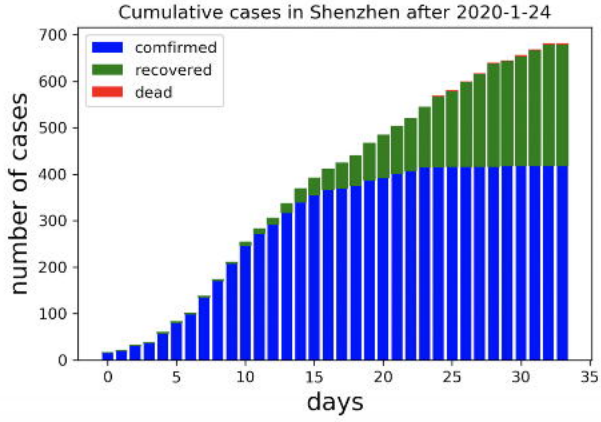

Daily cases in Beijing after 2020-1-24
daily comfirmed
daily recovered
daily dead
number of cases
days

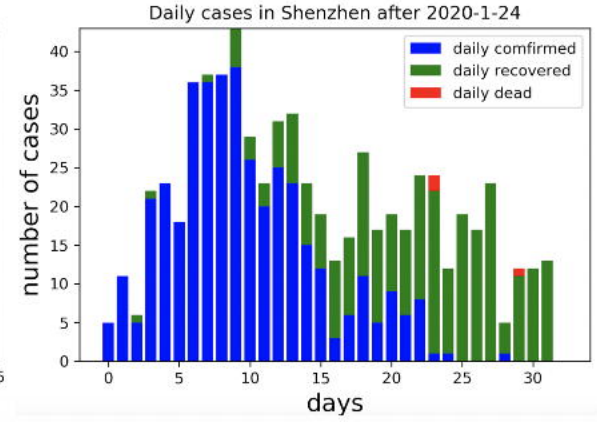

Cumulative cases in Shanghai after 2020-1-24
comfirmed
recovered
dead
number of cases
days

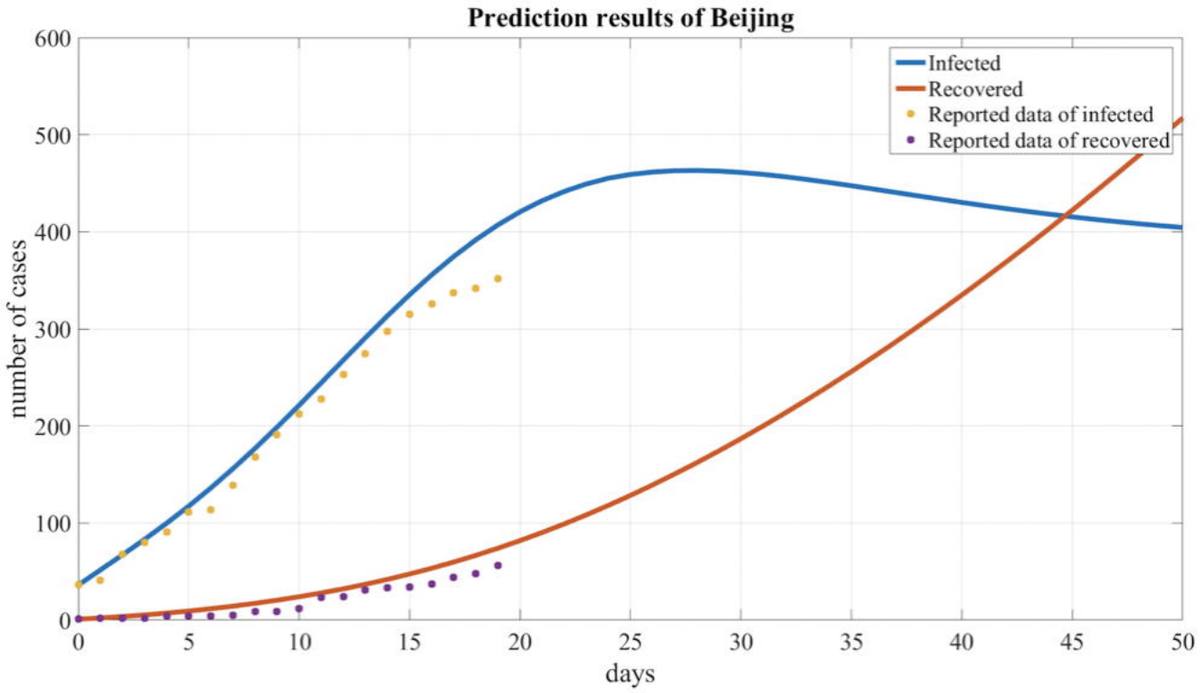

Daily cases in Shanghai after 2020-1-24
comfirmed
recovered
dead
number of cases
days

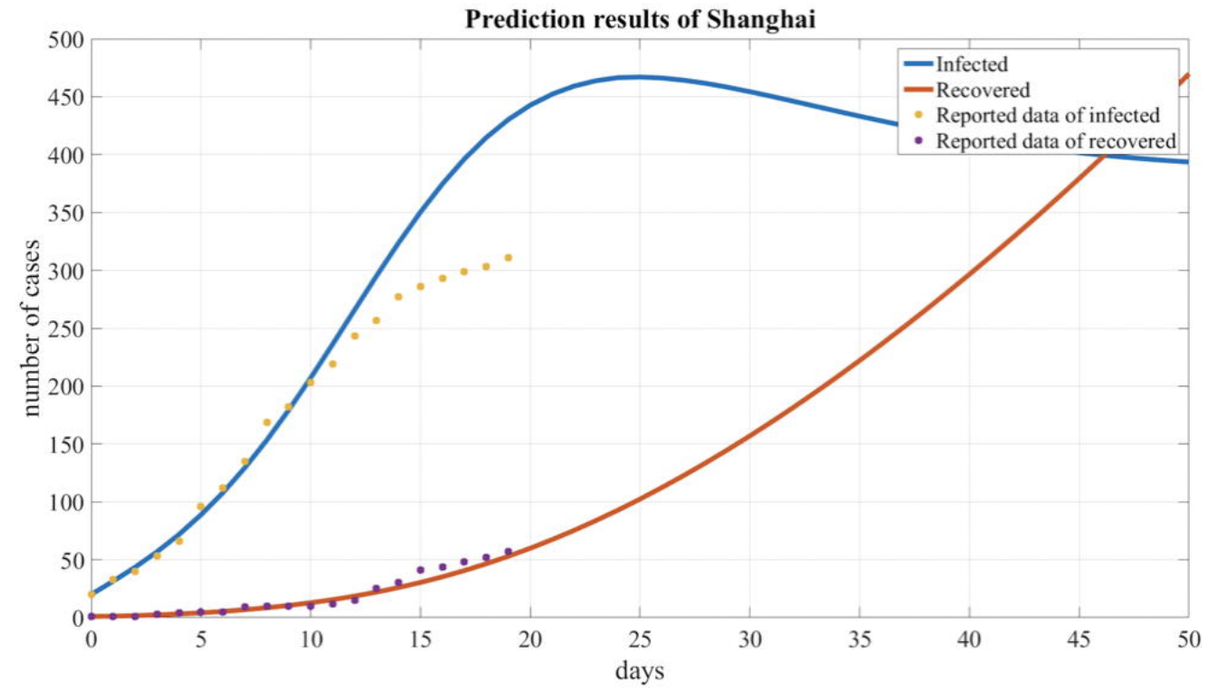

Cumulative cases in Guangzhou after 2020-1-24
comfirmed
recovered
dead
number of cases
days

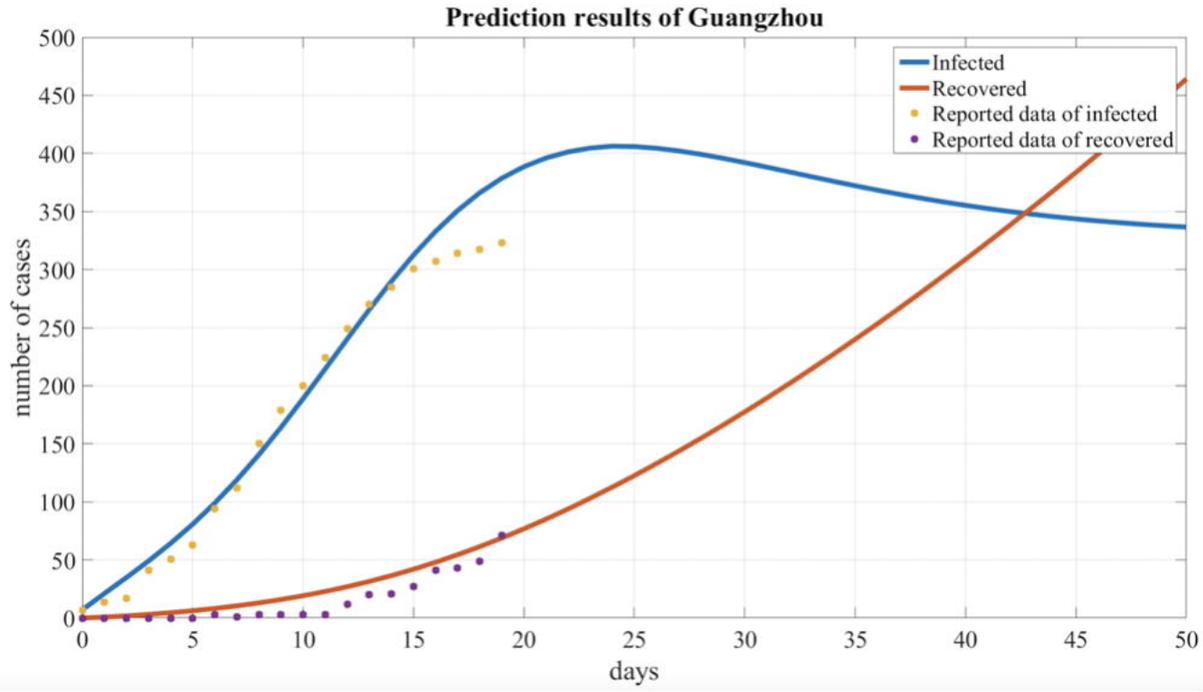

Daily cases in Guangzhou after 2020-1-24
daily comfirmed
daily recovered
daily dead
number of cases
days

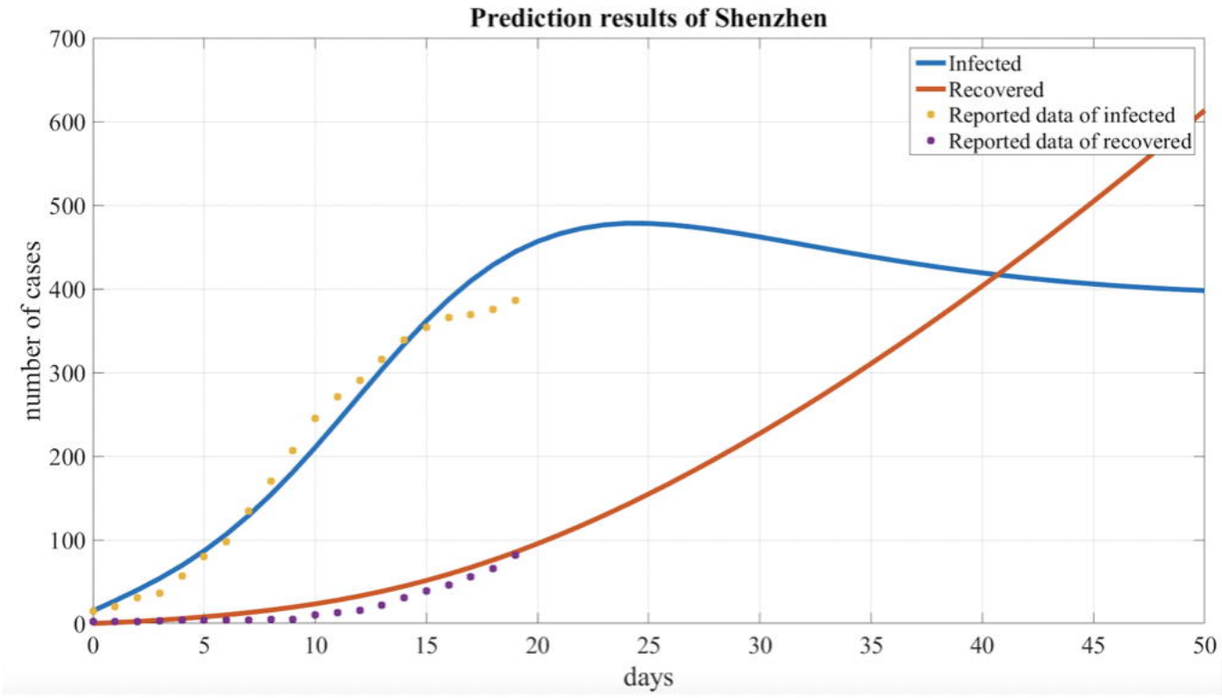

Cumulative cases in Shenzhen after 2020-1-24
comfirmed
recovered
dead
number of cases
days

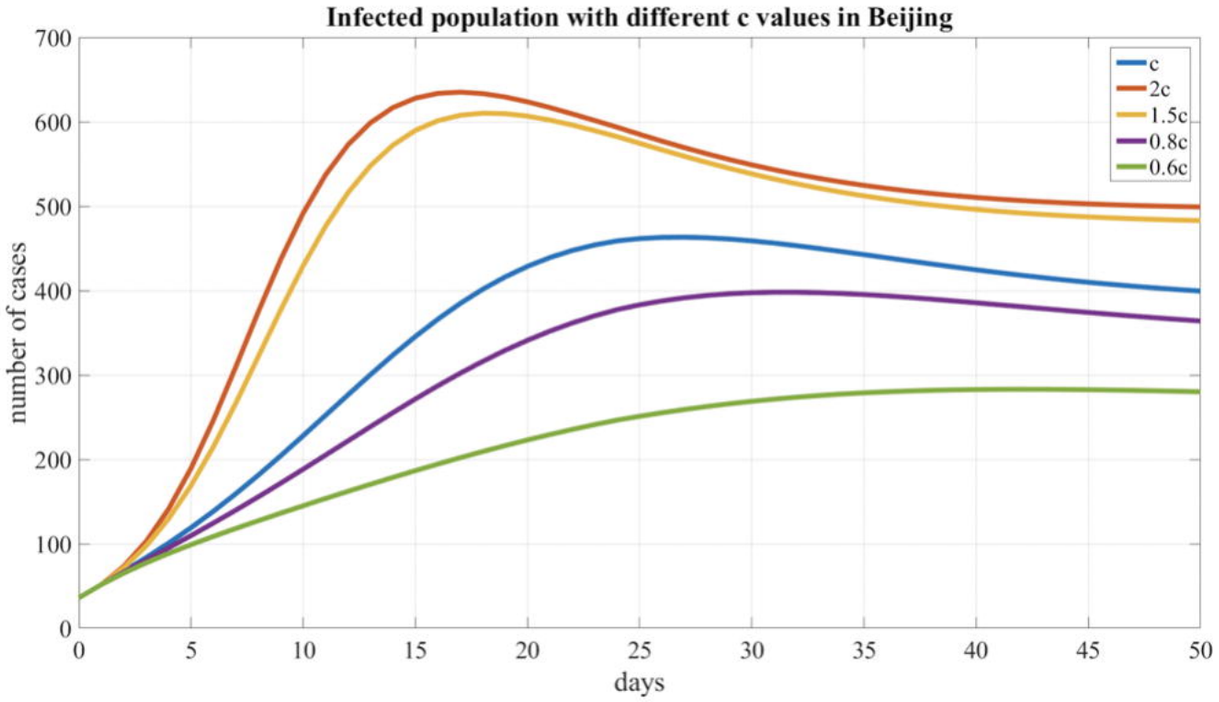

Daily cases in Shenzhen after 2020-1-24
daily comfirmed
daily recovered
daily dead
number of cases
days

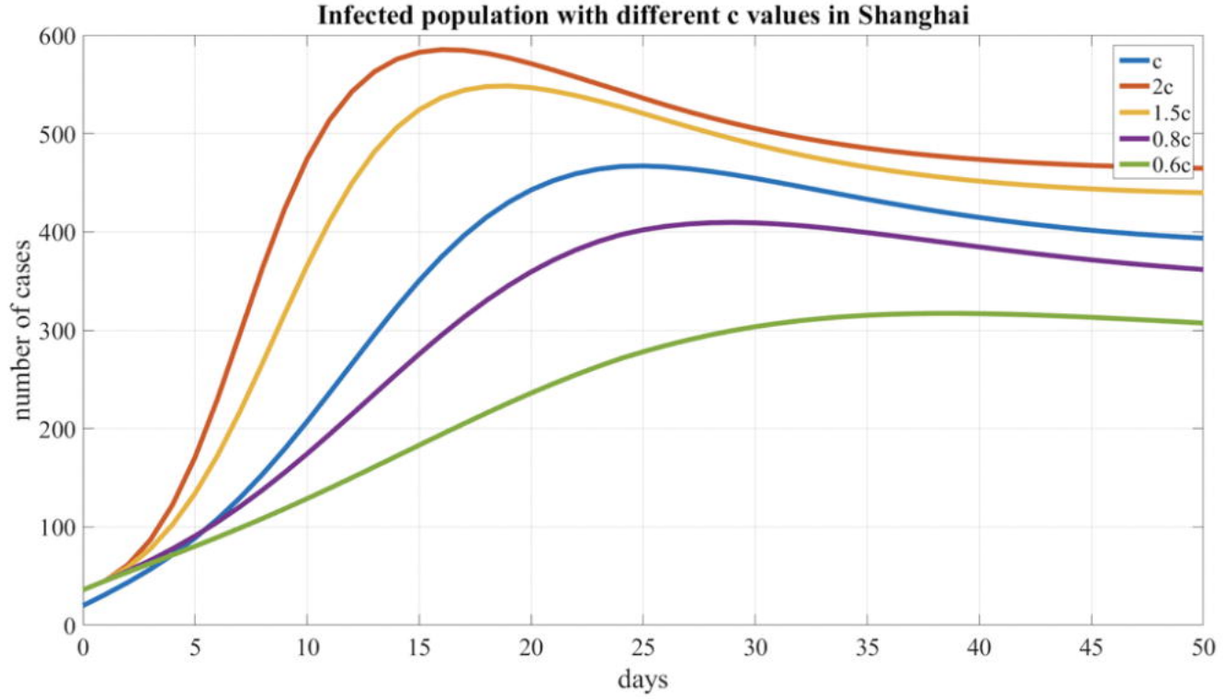

Prediction results of Beijing
Infected
Recovered
Reported data of infected
Reported data of recovered
number of cases
days

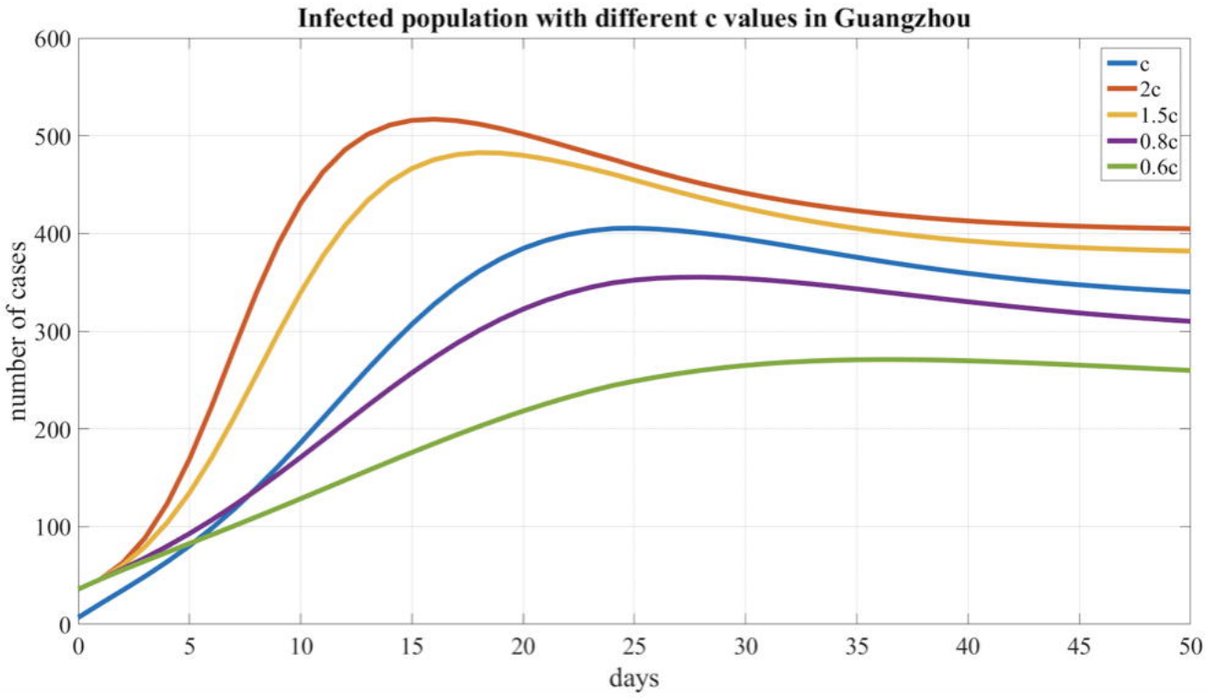

Prediction results of Shanghai
Infected
Recovered
Reported data of infected
Reported data of recovered
number of cases
days

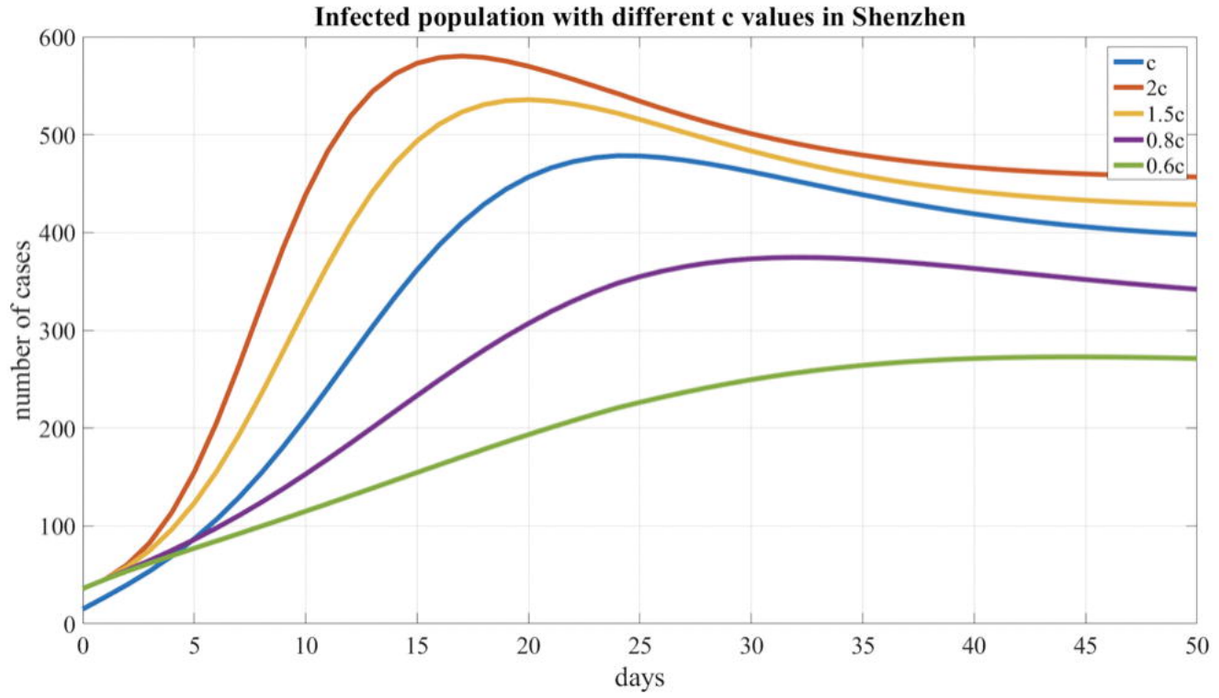

Prediction results of Guangzhou
Infected
Recovered
Reported data of infected
Reported data of recovered
number of cases
days

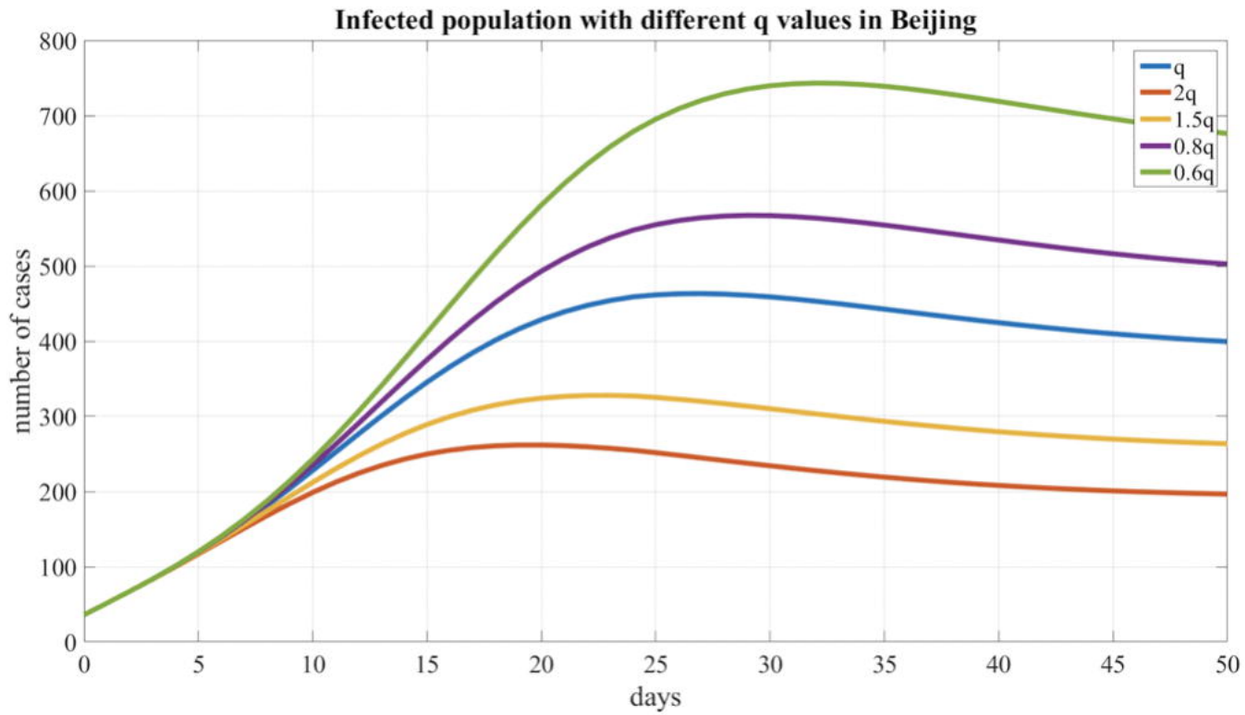

Prediction results of Shenzhen
Infected
Recovered
Reported data of infected
Reported data of recovered
number of cases
days

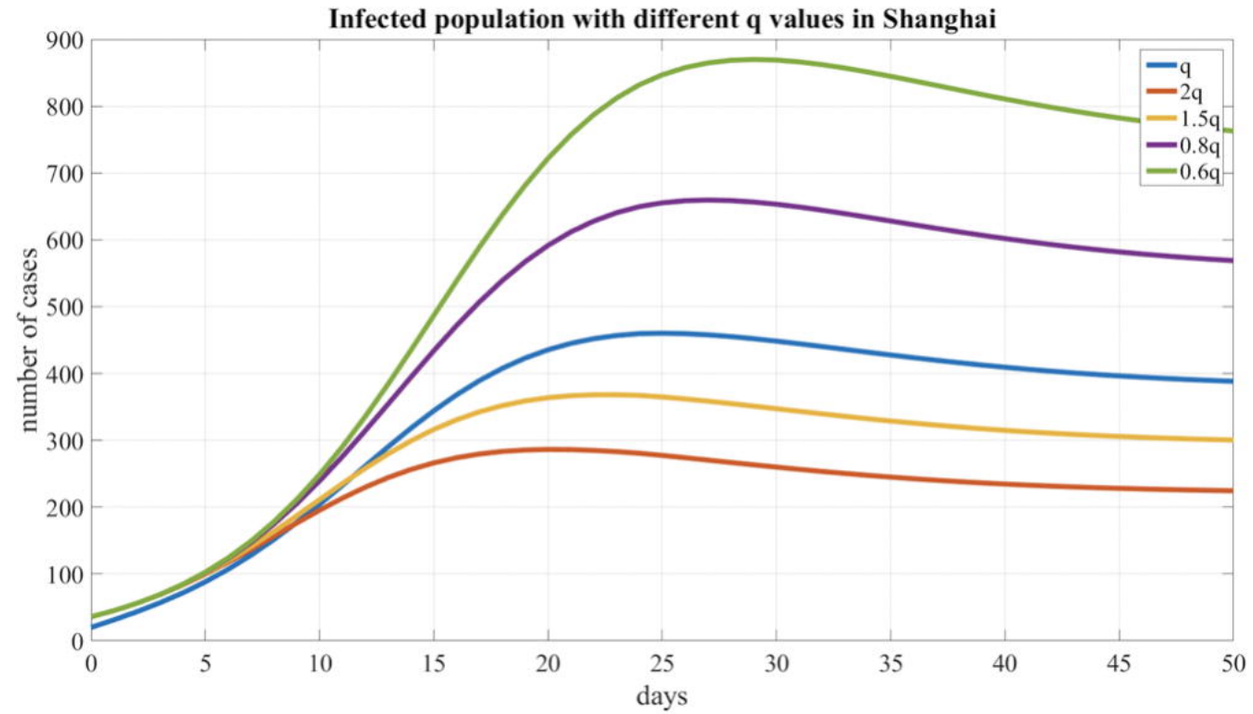

Infected population with different c values in Beijing
number of cases
days
c
2c
1.5c
0.8c
0.6c

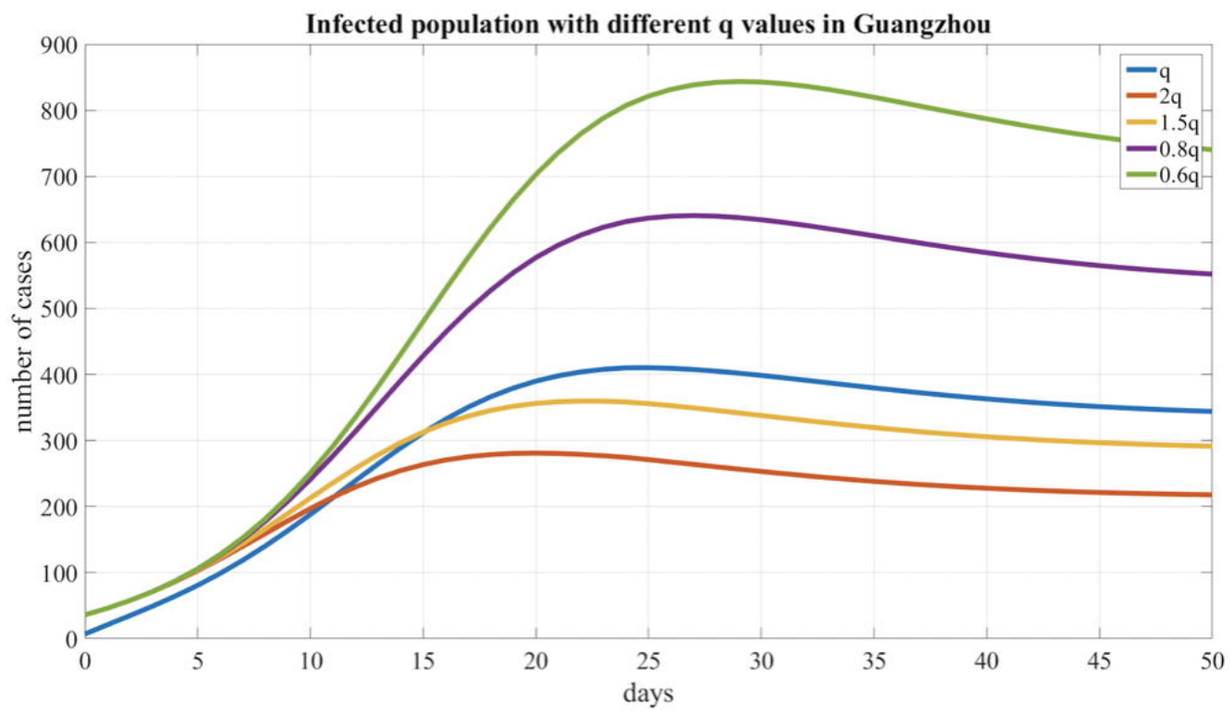

Infected population with different c values in Shanghai
number of cases
days
c
2c
1.5c
0.8c
0.6c

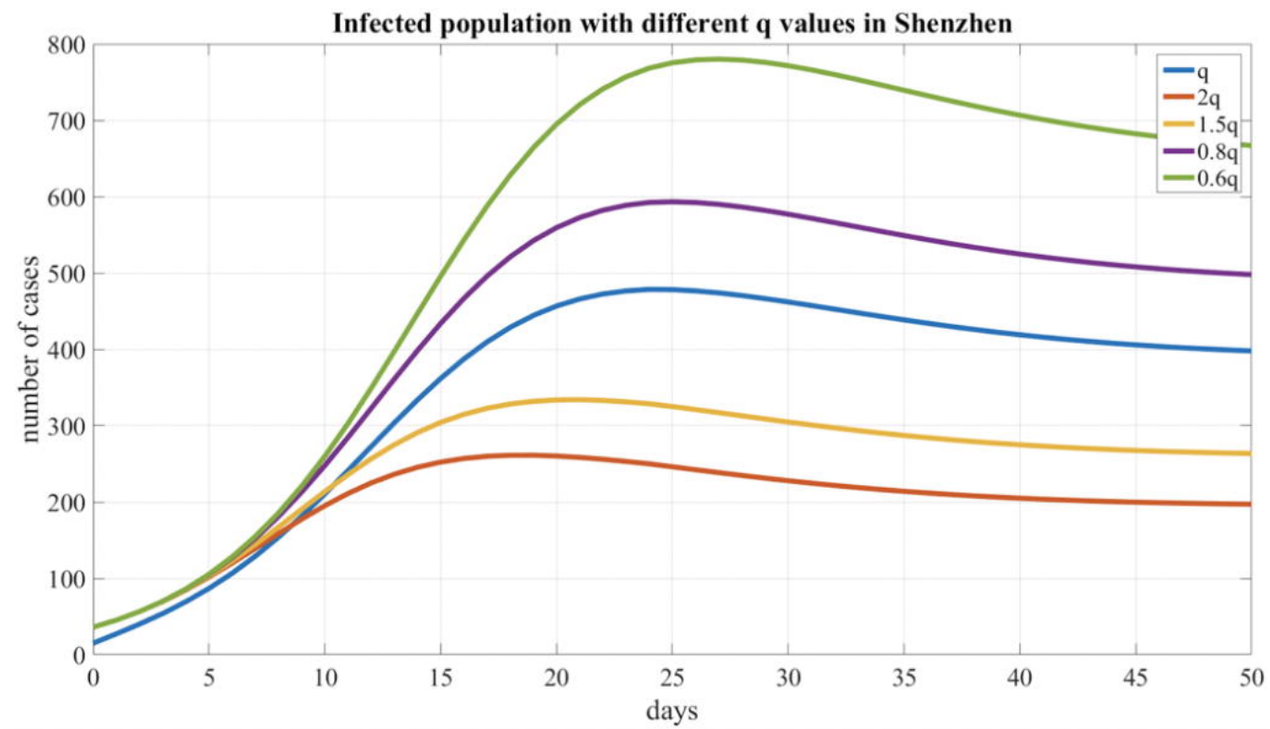

Infected population with different c values in Guangzhou
number of cases
days
c
2c
1.5c
0.8c
0.6c

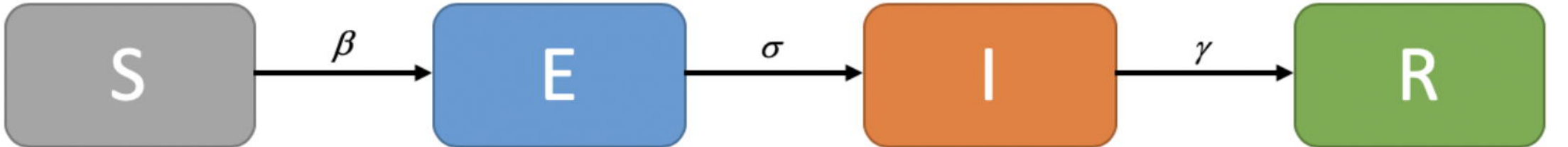

Infected population with different c values in Shenzhen
number of cases
days
c
2c
1.5c
0.8c
0.6c

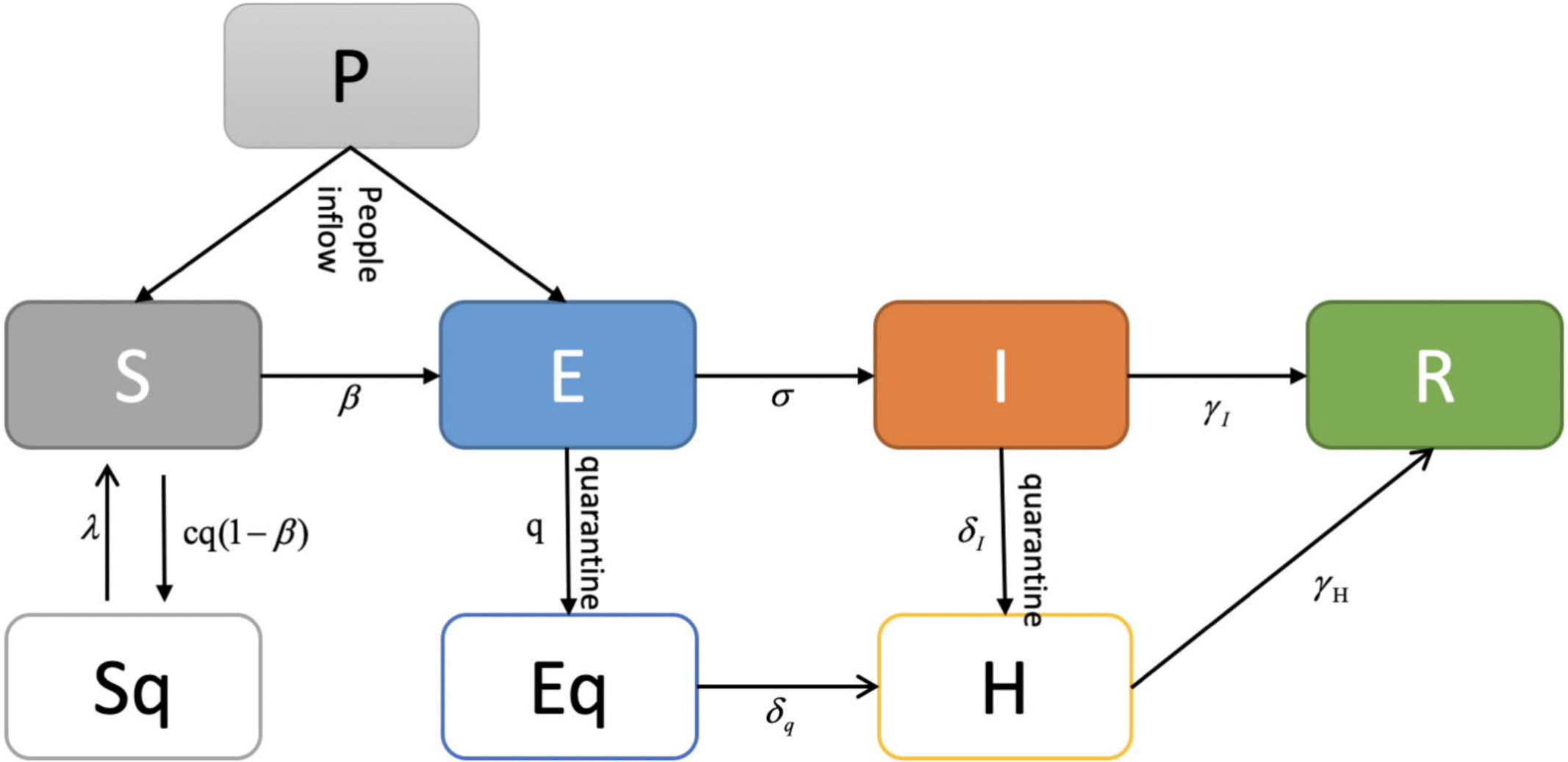

Infected population with different q values in Beijing
number of cases
days
q
2q
1.5q
0.8q
0.6q


Infected population with different q values in Shanghai
number of cases
days
q
2q
1.5q
0.8q
0.6q


Infected population with different q values in Guangzhou
number of cases
days
q
2q
1.5q
0.8q
0.6q


Infected population with different q values in Shenzhen
number of cases
days
q
2q
1.5q
0.8q
0.6q


S
β
E
σ
I
γ
R


P
People inflow
S
E
I
R
β
σ
γ_I
λ
$cq(1-\beta)$
q
quarantine
δ_I
quarantine
γ_H
Sq
Eq
H
δ_q