



Investigating time, strength, and duration of measures in controlling the spread of COVID-19 using a networked meta-population model

Jiang Zhang · Lei Dong · Yanbo Zhang · Xinyue Chen · Guiqing Yao · Zhangang Han

Received: 16 April 2020 / Accepted: 14 June 2020 / Published online: 27 June 2020
© Springer Nature B.V. 2020

Abstract Policy makers around the world are facing unprecedented challenges in making decisions on when and what degrees of measures should be implemented to tackle the COVID-19 pandemic. Here, using a nationwide mobile phone dataset, we developed a networked meta-population model to simulate the impact of intervention in controlling the spread of the virus in China by varying the effectiveness of transmission reduction and the timing of intervention start and relaxation. We estimated basic

reproduction number and transition probabilities between health states based on reported cases. Our model demonstrates that both the time of initiating an intervention and its effectiveness had a very large impact on controlling the epidemic, and the current Chinese intense social distancing intervention has reduced the impact substantially but would have been even more effective had it started earlier. The optimal duration of the control measures to avoid resurgence was estimated to be 2 months, although would need to be longer under less effective controls.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11071-020-05769-2>) contains supplementary material, which is available to authorized users.

Keywords COVID-19 · Meta-population epidemic model · Network dynamics · Intervention

J. Zhang (✉) · Z. Han
School of Systems Science, Beijing Normal University,
Beijing 100875, China
e-mail: zhangjiang@bnu.edu.cn

X. Chen
School of Geography and Planning, Sun Yat-Sen
University, Guangzhou 510275, China

L. Dong
School of Earth and Space Sciences, Institute of Remote
Sensing and Geographical Information Systems, Peking
University, Beijing 100871, China

G. Yao
Department of Health Sciences, University of Leicester,
Leicester LE1 7RH, UK

L. Dong
Senseable City Lab, Massachusetts Institute of
Technology, Cambridge, MA 02139, USA

Y. Zhang
School of Earth and Space Exploration, Arizona State
University, Tempe, AZ 85281, USA

1 Introduction

The outbreak of coronavirus disease 2019 (COVID-19) led to a global epidemic [1]. The World Health Organization (WHO) declared the outbreak as a Public Health Emergency of International Concern and subsequently a pandemic [2]. Several countries introduced tough measures including lockdown of borders and major cities, social distancing, and isolation of the infected population. All these efforts were aimed at controlling the spread of this highly infectious virus and thereby to reduce severe illness in humans. However, it is still unclear what role these intervention policies have played in controlling the spread of the epidemic.

On March 27, 2020, at the time of writing this paper, 4 months after the “first” case was reported (December 8, 2019) [3], 81,394 infected cases had been confirmed in mainland China with 3295 reported deaths. Of these, 61.4% of the infections and 77.1% of the deaths had occurred in Wuhan [4], the initiating center of this outbreak. Globally, the USA, Italy, Spain, Germany, France, Iran, UK, and Switzerland now each have more than 10,000 confirmed cases [5].

Recent studies have shown that human mobility plays a key role in the spread of epidemics [6–12]. Specifically, epidemic spread in urban networks can be regarded as a reaction–diffusion process in which local reaction occurs within cities, and city-wide diffusion is driven by the human mobility across those cities [13]. However, most conventional epidemic spread models (e.g., susceptible–infected–recovered, SIR) [14] cannot fully account for the diffusion dynamics caused by mobility, which are extremely important in highly connected urban networks. The situation in Wuhan also highlighted the role of human mobility in spreading and controlling COVID-19. As a central transportation hub, Wuhan has direct links to its surrounding cities and most of the major cities in other parts of China (Fig. 1a), which made the epidemic quickly spread to China through the urban transportation networks. In order to control the spread of the epidemic, Wuhan has been “locked down” since January 23, 2020 [15], and strictly controls population movement (Fig. 1b). In terms of infected cases, the lockdown of the city has achieved remarkable results, with fewer than 100 new cases reported daily in Wuhan since March 6, 2020 (the population of Wuhan is over 10 million) [16].

Right now, policy makers in all countries are facing unprecedented challenges in making decisions on when

and what degrees of intervention measures should be implemented to tackle the pandemic. Meanwhile, a large number of publications provide medical and policy support from different perspectives [17]. The biological, clinical [18–21], and spreading characteristics of the virus [3, 22, 23] have been extensively discussed, and several modeling-based studies have addressed the spread of the virus through human migration [24–26], and different measure policies. Literatures [27–30] review the modeling frameworks on general disease dynamics and COVID-19 virus transmission. However, few have modeled the effect of intervention timing and its strength on controlling the spread of the epidemic [31–34].

Here, by using a nationwide human movement dataset, we developed a networked meta-population model to estimate the spread of COVID-19 across 371 prefectural-level cities in mainland China under different intervention scenarios. We found that both the time of initiating an intervention and its effectiveness had a very large impact on controlling the epidemic. Under the least effective scenario, i.e., doing nothing, we estimated that the virus would have infected over one billion people, approximately 85% of the Chinese population. By implementing a tough intervention as imposed in Wuhan (city-wide quarantine, strictest travel ban, minimized social contact, widespread testing and isolation, etc.), the spread was reduced substantially, and the cumulative infected population in China was modeled to be 0.28 million (95% CI 0.178–0.604) in 7 months, 0.02% of 1.18 billion with no intervention. If the time of the intervention had been brought forward by ten days, the estimated number of infected cases for the same period would be 65,200 (95% CI 4100–13,900), which was a quarter of the tough intervention case. For a less effective intervention, the model estimated a tenfold increase in infected cases and deaths, but even here if it was implemented earlier, the number of the cases would reduce substantially.

2 Methods

2.1 Data

2.1.1 Population migration dataset

The city-level mobility network is constructed from the Baidu Migration Project [35]. This dataset is

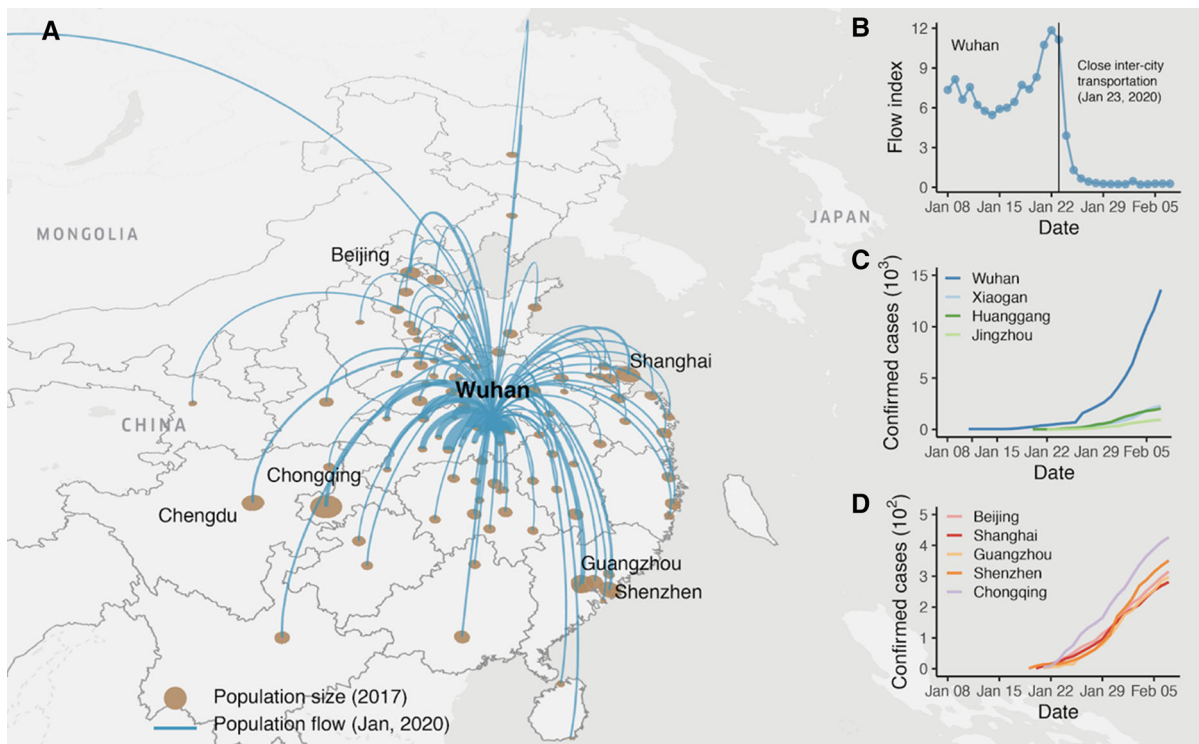


Fig. 1 Population flow and the cumulative number of confirmed cases. **a** Population flow from Wuhan to other cities of China as of January, 2020. **b** Outflow index of Wuhan from January 1, 2020, to February 7, 2020. The quarantine policy was implemented on January 23, 2020. The outflow dropped

calculated based on the movement of mobile phones using Baidu location-based services. It includes two parts: (1) a daily origin–destination matrix recording the migration percentage P_{ij}^t of the top 100 inflow and outflow cities for 371 prefecture-level cities in China. The total population of these cities is 1.39 billion, accounting for over 99% of the national population. P_{ij}^t denotes the proportion of migration from the city j to i , to the total population that moved out from the city j in day t ; and (2) a flux index F_j^t of the daily inflow and outflow for each city j in day t . This index is proportional to the number of mobile phone users who migrate into or out of a certain city (Fig. 1a, b). Both parts cover movements between January 1, 2020, and February 7, 2020. To alleviate the fluctuation of P_{ij}^t , an “average” mobility matrix P over the whole period is derived, where $P_{ij} = \Sigma_t \left(P_{ij}^t F_j^t \right) / \Sigma_t F_j^t$.

dramatically after that day. The dataset to calculate population flow is detailed in Methods. **c, d** The number of cumulative confirmed cases of four cities from R dataset in Hubei and five main cities (see Sect. 2)

2.1.2 Case dataset

We collected city-level reported case data from the R package *nCov2019* [36], which includes the cumulative number of confirmed, dead, and recovered individuals each day in each city (Fig. 1c, d). To estimate the parameters of our model, such as incubation time, we also used individual case data derived from the Wolfram database [37]. The original sources of these case data are from the National Health Commission, local health commissions, and Chinese Centers for Disease Control and Prevention [38]. We supplemented the cumulative confirmed case data before January 10, 2020 (the start date of the case dataset provided by the R package) with cases reported by Li et al. [3].

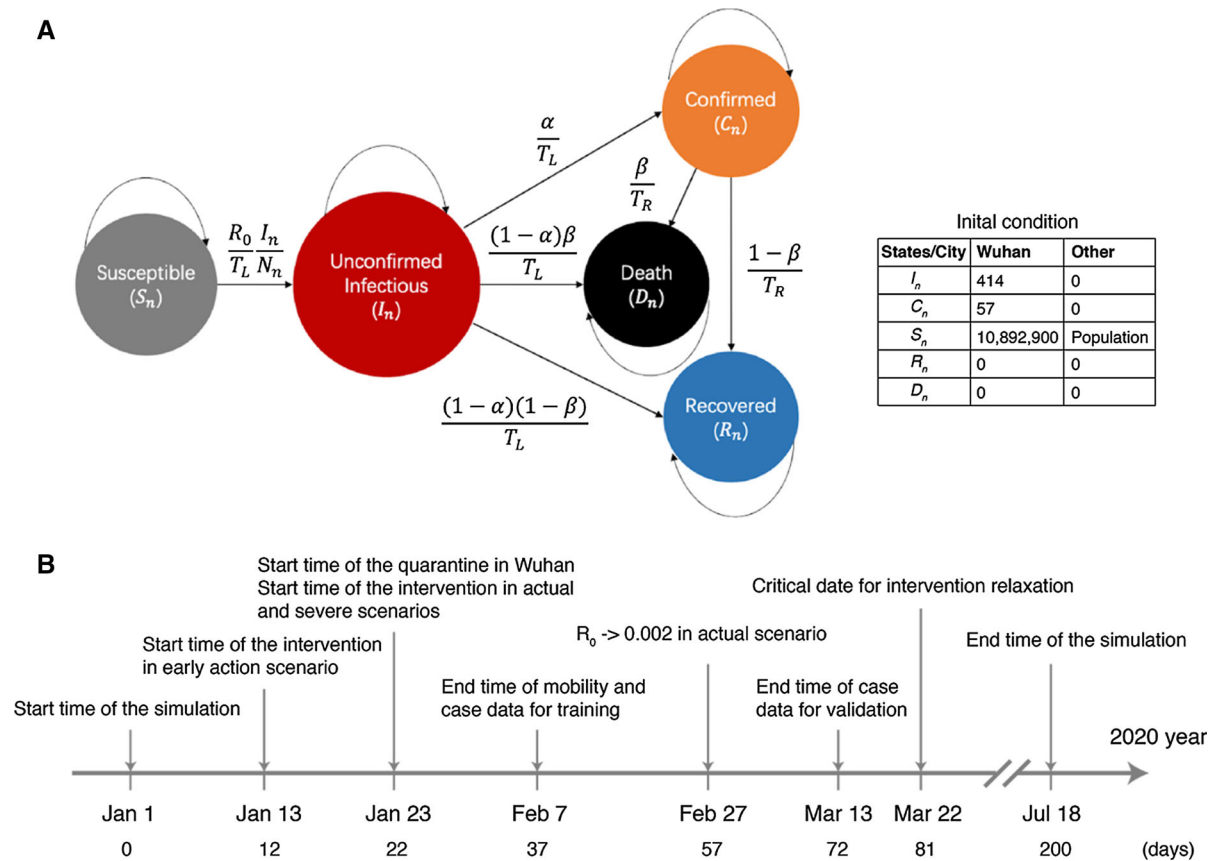


Fig. 2 Schematic of the SICRD model, the initial conditions, and timeline of the simulation. **a** Model schematic and initial conditions. S_n , I_n , C_n , D_n , and R_n are the populations of the states, and s_n , i_n , c_n , d_n , and r_n are the relative fractions of the states, respectively, which can be easily calculated by dividing S , I , C , D , and R with the city’s population. The explanation of model parameters is listed in Table 1. As the initial condition for

Wuhan, the value of the I state is estimated by fitting the case data (see SI Section 2). The number of confirmed cases, $C = 57$, is based on Li et al. [3]. S , the initial susceptible number, is assumed to be the city population size, which is collected from city population data (<https://www.citypopulation.de/>). **b** The timeline of the simulation

2.2 Model

2.2.1 The networked meta-population model

We extended the classic SIR model [39] by introducing two new health states, i.e., unconfirmed cases and death, as well as taking into account inter-city human mobility [9]. In our model, health states were defined by susceptible, unconfirmed infectious, confirmed, recovered, and dead (SICRD model). The setting of the five-state model reflects the real-life situation in which an infectious person may recover or die before receiving a formal diagnosis (unconfirmed cases). The model structure and the transition among the five states are illustrated in Fig. 2a.

Specifically, health states were defined as follows:

- Susceptible state: healthy individual without COVID-19; a person in this state can be infected if they contact an infectious person.
- Unconfirmed infectious state: persons in this state can be symptomatic or asymptomatic but they are infectious. A person in this state can stay in the same state, or move to a confirmed state, recovered state, or dead state.
- Confirmed state: those with viral confirmation. We have assumed that a person in this state will be isolated; therefore, he/she will have no chance to infect other people. A person in a confirmed state can stay in the same state or move to recovered or dead states.

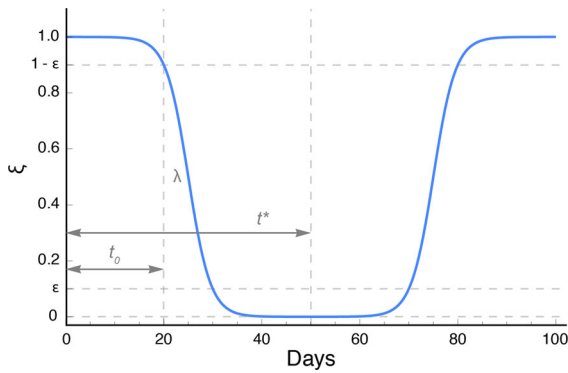


Fig. 3 The functional form of the intervention and relaxation term when $t^* < \infty$. The two terms in Eq. (2) are symmetric along the axis $t = t^*$. That is, we assume that the relaxation process reverses the intervention process. When $t < t^*$, ζ decays as an S-shaped function of t because any policy needs time to implement, while ζ grows in an S-shaped curve when $t > t^*$, and this simulates the slow relaxation of intervention. The detailed illustration of Eq. (2) can be read in SI Section 3

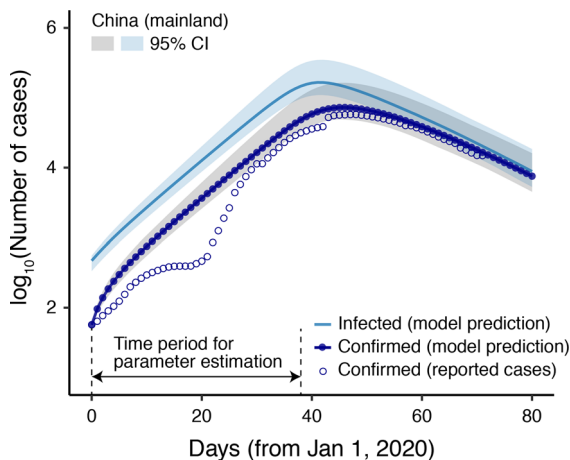


Fig. 4 Model validation. The x -axis is the time starting from January 1, 2020, in days. The confirmed cases (lines with dots) predicted by the model fit the reported case data (circles) well, indicating the effectiveness of our model and parameters

- Recovered state: those who are cured and immune. We have assumed that a person in this state would not be infectious or be infected again.
- Death state: the deaths due to COVID-19

To be noticed that, our model is different from the traditional SEIR model (susceptible, exposed, infectious, and recovered). The people in the unconfirmed infectious state contain the exposed state in the conventional SEIR model. While in our model, we

assume that the infected cases are infectious immediately.

Mathematically, the model is defined by the following equations [see Supplementary Information (SI) Section 1 for the derivation]:

$$\begin{aligned} \frac{\partial s_n}{\partial t} &= -\frac{R_0}{T_L} s_n i_n + \omega \sum_{m \neq n} P_{mn} (s_m - s_n) \\ \frac{\partial i_n}{\partial t} &= \frac{R_0}{T_L} s_n i_n - \frac{\alpha}{T_L} i_n - \frac{1-\alpha}{T_L} i_n + \omega \sum_{m \neq n} P_{mn} (i_m - i_n) \\ \frac{\partial c_n}{\partial t} &= \frac{\alpha}{T_L} i_n - \frac{1}{T_R} c_n \\ \frac{\partial d_n}{\partial t} &= \frac{\beta(1-\alpha)}{T_L} i_n + \frac{\beta}{T_R} c_n \\ \frac{\partial r_n}{\partial t} &= \frac{(1-\beta)(1-\alpha)}{T_L} i_n + \frac{1-\beta}{T_R} c_n + \omega \sum_{m \neq n} P_{mn} (r_m - r_n) \end{aligned} \tag{1}$$

where $s_n = S_n/N_n, i_n = I_n/N_n, c_n = C_n/N_n, r_n = R_n/N_n, d_n = D_n/N_n$ are the fractions of susceptible, unconfirmed infectious, confirmed, recovered, and dead individuals in city n , respectively, and N_n is the population size of city n . P is the mobility matrix calculated based on the migration data (see Sect. 2), and $0 \leq P_{mn} \leq 1$ denotes the proportion of migration from city n to the city m , to the total population who left city n . Infectious people could move to other cities to induce epidemic outbreaks in other cities. The initial states are listed in Fig. 2a.

Detailed descriptions and estimation of parameters $R_0, T_L, T_R, \alpha, \beta, \omega$, and $I_{\text{wuhan}}(0)$ are shown in Table 1 and SI Section 2. T_L, T_R , and β can be estimated from patient clinical records, and ω can be estimated based on the Baidu migration dataset, which is calculated based on the movement of mobile phone users (see Sect. 2). The remaining three free parameters R_0, α , and $I_{\text{wuhan}}(0)$ need to be estimated with reported city-level case data by fitting the number of confirmed cases for all Chinese cities before February 7, 2020, in order to minimize the loss function: $\min_{R_0, I_{\text{wuhan}}(0), \alpha} L =$

$$\sum_n \sum_t [\log_{10}(c_n(t)N_n) - \log_{10}(C_n^*(t))]^2 \quad (\text{see Methods and SI section 2}).$$

To solve this optimization problem, we use the differentiable ODE solver implemented by the framework of PyTorch [41]. The model and the estimated parameters are robust to changes in other parameters (SI Section 2). All the following results and the

Table 1 Parameters of the SICRD model. We report the median values of R_0 , α , and $I_{\text{wuhan}}(0)$ and mean values of T_L , T_R , and ω with 95% CIs

Parameter	Definition	Value [95% CI]	Description
R_0	Basic reproductive number	2.22 [2.15–2.33]	The average number of cases directly infected by one infectious case in a meta-population. It is estimated by fitting the number of confirmed cases (SI Section 2)
T_L	Time spent in days from being infectious (when the case can infect other people) to be confirmed, recovered, or dead	8.3 [7.5–9.1]	This is estimated based on patient medical records (Fig. S3) [37]
T_R	Time span from confirmation to recovery or death (no longer infectious, i.e., immune)	9.2 [7.4–11.0]	By assuming the death and recovery events as independent events, we can estimate T_R based on patient medical records [37]
α	The proportion of confirmed cases among all cases transferred from the unconfirmed infectious state	0.93 [0.65–0.99]	This is the probability of an unconfirmed case being confirmed after T_L days. It is estimated by fitting the confirmed case data
β	The fatality rate	0.023	The average ratio of deaths to the total number of infectious people for a certain period of time. The value is from Ref [40]
ω	The average migration rate	0.03 [0.014–0.046]	The total traffic flux (based on the migration dataset) between cities divided by the total population in China, see SI Section 2
$I_{\text{wuhan}}(0)$	The number of unconfirmed infectious cases in Wuhan on January 1, 2020	414 [277–497]	The initial condition for unconfirmed infected people in Wuhan. It is estimated by fitting the confirmed case data, see SI Section 2

confidence intervals (CIs) are obtained after 360 repeated experiments (Table 1).

2.2.2 Intervention and relaxation

Further, we introduce an intervention term to simulate the intervention strategy that the Chinese government implemented to reduce contact or interaction among the population. By leaving all the equations and parameters unchanged in Eq. (1), we altered the model by multiplying ξ to the infectious term, i.e., $s_n i_n$ in the first and the second lines of Eq. (1), such that:

$$\xi(t, t_0, t^*, \lambda) = 1 / \left[1 + e^{\lambda(t-t_0) - \ln(1/\epsilon - 1)} \right] + 1 / \left[1 + e^{\lambda(2t^* - t - t_0) - \ln(1/\epsilon - 1)} \right]. \quad (2)$$

There are three key parameters to describe the intervention term ξ : (1) The initiating time t_0 captures the time when the government started to apply intervention policies (e.g., traffic control, social distancing, and quarantine patients). (2) λ models the rate at which the intervention can take effect to reduce the

reproduction number from R_0 to ϵR_0 ; a larger value means the spread can be reduced in a shorter time. ϵ is the ratio of an acceptable reproduction number that can control the spread of the virus. (3) The timing t^* when we can relax the intervention. Overall, the term ξ models the temporal and effective dimensions of an intervention that governments use to prevent the epidemic spreading. The functional form of Eq. (2) is shown in Fig. 3. The values of t_0 , λ , t^* , and ϵ are presented in Table 2.

3 Results

3.1 Model validation

We validated the model by comparing the results estimated from the model under the actual scenario (the definition of different scenarios is detailed below) with the observed cases reported in China. Figure 4 compares the model-predicted confirmed cases with the reported confirmed cases at the national level. The

Table 2 Parameters of the intervention and relaxation terms in SICRD model

Parameter	Definition	Value [95% CI]	Description
t_0	The time span from January 1, 2020, to the implementation of the intervention policy	22 or 12	Wuhan reduced the inter-city traffic on January 23, 22 days after January 1, the initial reports of COVID-19 cases [15]
λ	The speed of the intervention implementation	> 0	How fast the intervention can take effect to reduce the spread of the virus, i.e., to reduce R_0 to ϵR_0
ϵ	The ratio of an acceptable reproduction number to R_0 that can control the spread	0.001	If the reproduction number is reduced to ϵR_0 , then intervention could stop
t^*	The date when the intervention begins to relax	$> t_0$	The critical value of t^* such that the number of infectious cases will never rise again

model predicted well beyond the time period for parameter estimation, which demonstrates the validity of our model. However, we notice that since February 7, the model and actual data have deviated in individual cities (Table S4). This is likely to be due to the fact that Wuhan has changed the diagnosis criteria to include all clinical confirmed cases, while the rest of the cities still used the same diagnoses tool as before. The details of error analysis can be referred to SI Section 3.3.

Our model can also predict the number of “real” infected people (confirmed + unconfirmed infectious), as shown by the straight lines in Fig. 4. The results show that the real situation was likely to be much more serious than the reported data for all cities, especially those in Hubei Province. For example, our model predicted that the total number of infectious (confirmed + unconfirmed infectious) cases in Wuhan was more than three times than the confirmed number at most time points, and seven times higher at the beginning of the epidemic [155 reported cases and 1140 (95% CI 760–1680) model-predicted cases, as of January 7, 2020]. The large gap between the numbers of confirmed and infected cases in Wuhan was documented in a recent study [42]. This reflects the fact that a large number of infectious people could not obtain any appropriate treatment due to the limited medical resources in Wuhan, and/or they had only mild symptoms or were even asymptomatic.

3.2 Scenario analysis

We ran the model for 200 days (from January 1, 2020, to July 18, 2020). We will discuss the simulation

results under five scenarios: (1) base case scenario: without intervention; (2) actual scenario: with intervention started from January 23, 2020, and reduced the reproduction number to $\epsilon R_0 = 0.002$ within five weeks; (3) early action scenario: same as the actual scenario but the intervention was implemented ten days earlier; (4) less effective scenario: same as the actual scenario but weaker intervention strength; and (5) early action but less effective scenario: same as the early action scenario but weaker intervention strength.

3.3 Base case scenario (without intervention)

According to our estimation, 1640 (95% CI 1220–1910) infectious cases were exported from Wuhan to all over the country before January 23, 2020, when the city-wide quarantine intervention policy was implemented.

The dynamics of some representative cities are shown in Fig. 5. There are two key findings. First, there would be two peaks in the spread of the disease (Fig. 5). The first peak would be the cities in Hubei Province, e.g., Wuhan, Xiaogan, and Huanggang. The second peak would be large cities such as Guangzhou. In those cities, this peak would be delayed by approximately three to four weeks compared with Wuhan. Second, the city with the most infections would be Chongqing (not Wuhan) because Chongqing has the largest population in China. Several other large cities (e.g., Beijing, Shanghai, and Guangzhou) would have more infections than Wuhan. The cumulative number of infectious cases for mainland China would be 1.18 billion (95% CI 1.16–1.22), accounting for 85% of the population of China.

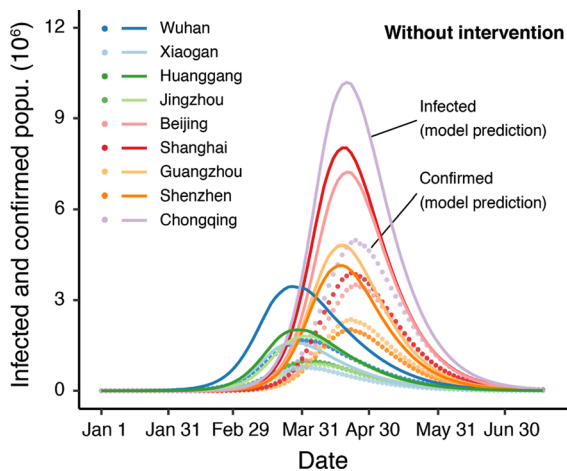


Fig. 5 Basic case scenario without any intervention. The y-axis is the existing confirmed and infected populations

3.4 Actual scenario ($t_0 = 22$, $\lambda = 0.4$, $t^* \rightarrow \infty$)

The Chinese government implemented strong intervention policies to prevent the spread of the epidemic. For example, most cities in Hubei Province closed their inter-city transportation and implemented strict control over intra-city traffic. Schools and factories were closed, and only few daily necessities stores were allowed to open. Therefore, to derive better simulation results, we incorporated these interventions into the model as shown in Eq. (2).

Our model shows that if interventions could significantly reduce the reproduction number to $\epsilon R_0 = 0.002$ in a short time window (~ 35 days, $\lambda = 0.4$) after the outbreak, which was close to the real situation in China, the peaks of the infected individuals of Wuhan, Huanggang, Chongqing, and Shanghai would reach 48,000 (95% CI 22,300–176,100), 12,600 (10,600–14,800), 1770 (1430–220), and 1770 (1430–2250), respectively (Fig. 6a, b) and the cumulative infected population for the whole simulation period would be approximately 284,000 (178,000–604,000) for the entire country, approximately 0.02% of the base case scenario.

The predicted trend of the epidemic fits well with the latest data, with the predicted curves peaking in mid-February (Fig. 6a, b). However, cities may have different behaviors due to the variation in intervention strength, which cannot be reflected in our model because all the modeled parameters were shared among all cities.

The model predicted that a total of 6500 (95% CI 4100–13,900) people would die as a result of COVID-19 in China out of 284,000 infected cases. In fact, as treatment levels continue to improve, the fatality rate may decrease, and the corresponding total number of deaths may also fall below this predicted value.

3.5 Early action scenario

$$(t_0 = 12, \lambda = 0.4, t^* \rightarrow \infty)$$

In this scenario, we assumed that the government implemented a strict intervention policy ten days before the actual date, i.e., January 13, 2020 ($t_0 = 12$), while the control intensity is the same as in the previous scenario ($\lambda = 0.4$). The simulation shows that the infected population would still reach a large number of people (14,000, 95% CI 7100–40,400) in Wuhan, while the number would be less than one-third of the number in the actual scenario and the peaks would arrive ten days earlier (Fig. 6c). The early action scenario would have a larger reduction in the size of the affected population across the entire country, and the cumulative infected population would be 65,200 (95% CI 42,000–120,900) and the total number of deaths would have been 1500 (95% CI 970–2780) or only a quarter of the predicted actual scenario. This shows the importance of early intervention. According to our simulation, the infected population and the number of deaths would decay exponentially with the advance of intervention time, at a rate of 0.212 (95% CI 0.208–0.219) in a log2 base, indicating that if the intervention time was advanced by 5 days, the total number of infected people would be reduced to half of the present total. This conclusion is similar as in [32].

3.6 Less effective scenario

$$(t_0 = 22, \lambda = 0.2, t^* \rightarrow \infty).$$

We also consider a less effective scenario in which the intervention strength is weaker than the actual situation which means the reproduction number would be reduced to $\epsilon R_0 = 0.002$ in about 70 days ($\lambda = 0.2$), the infected individuals in Wuhan would peak at 0.27 million (95% CI 0.12–1.15) in early March (Fig. 6e). The total number of deaths in mainland China would increase to 63,000 (95% CI 38,700–149,800).

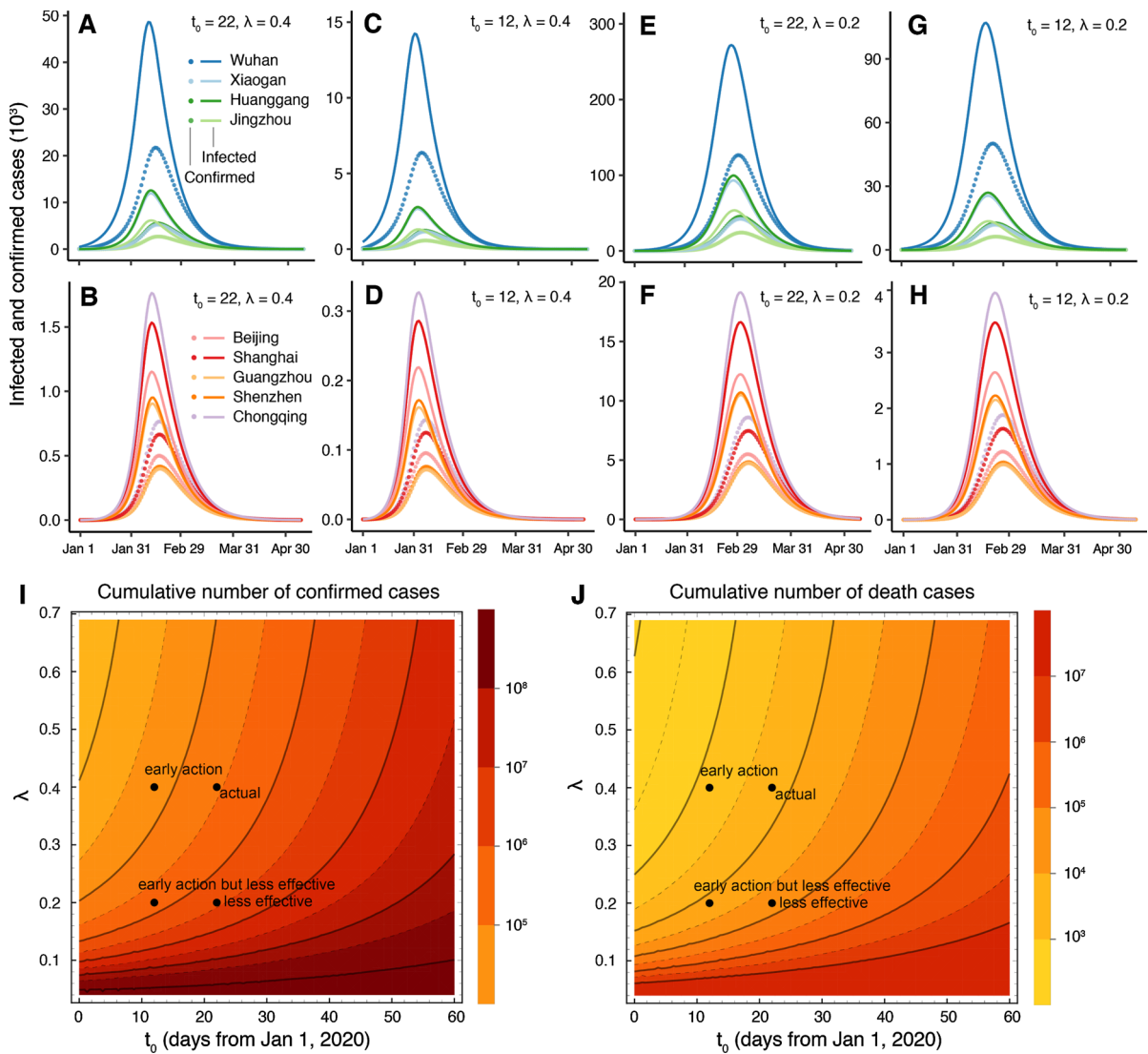


Fig. 6 Epidemic predictions under different scenarios of policy intervention. The actual (a, b), early action (c, d), less effective (e, f), and early but less effective scenarios (g, h) for four cities in Hubei Province and five main cities of China. The y-axis is the

existing median confirmed or infected populations. **i, j** The phase diagram of the cumulative number of confirmed (i) and death (j) cases under different intervention starting dates t_0 and strengths λ

3.7 Early action but less effective scenario ($t_0 = 12, \lambda = 0.2, t^* \rightarrow \infty$).

This scenario corresponds to the policy that most of European countries have implemented, that is implementing the intervention earlier ($t_0 = 12$) but with less strict enforcement ($\lambda = 0.2$). The predicted number of cumulative infectious cases and death cases would be 652,000 (95% CI 403,000–1,490,000) and 15,000 (9270–34,200).

We summarize the predicted cumulative infected and confirmed cases after the epidemic in different scenarios in Fig. 6.

3.8 Intervention relaxation ($t_0 = 22, \lambda = 0.4, 0 < t^* < \infty$)

Under the actual scenario ($\lambda = 0.4, t_0 = 22$), the earliest date t^* across all cities to relax the intervention to avoid resurgence is around the 81st day (i.e., March

22, 2020), that means the tough intervention should be maintained for about eight weeks. In the early action but less effective scenario, the measure would have to be maintained for about 12 weeks, see SI Section 3 for details.

4 Discussion

In this study, we proposed a five-state model and estimated the spreading dynamics of COVID-19 within China by analyzing the latest national population mobility data. The model can capture the early exponential growth of the outbreak at the city level, as well as perform scenario analyses based on different strengths of prevention interventions and time of start and relaxation.

Both the time of starting an intervention and its strength play important roles in controlling the outbreak size. Without any population intervention, the virus would have spread extensively through the population with over one billion cases ($\sim 85\%$ of population) in China and with nearly 27 million deaths. The current Chinese city policy has reduced the impact substantially but would have been even more effective had it started earlier.

Our model shows that the intervention needs to be sustained for about eight weeks to prevent a recurrent epidemic. Moreover, if it had been less effective in preventing population mixing, which may be the case in many other countries, then the number of cases and deaths would increase significantly, and by tenfold under the assumptions of our less effective scenario.

Our estimation of the basic reproduction number R_0 is slightly lower than some previous literature: 2.47–2.86 [23], 2.28–3.68 [22], and 2.39–4.13 [26], as we estimated this using the actual data to February 7, which includes two weeks after effective measures to reduce population mixing began in China. Comparing our findings with others then for the number of infected cases, our prediction for Wuhan is smaller than some recent works: 37,304–130,330 as of January 25 [24] and 11,090–33,490 as of January 22 [31] because they may have overestimated R_0 or did not take the strong intervention into account. In other major cities, our result is broadly consistent with others [24].

Our work demonstrates the value of data sharing and information disclosure [43]. All the data used in

this article came from public sources, including data released by large companies (migration data), as well as by start-ups such as DXY.cn, a Chinese physician's platform [44]. We notice that many volunteers are also working on collecting data via crowdsourcing, and several applications have been developed immediately [45].

However, despite adopting a state-of-the-art modeling framework, we must acknowledge that our model has several major limitations. First, we assume that different cities share the same model parameters. However, the geographical location, demographic structure, and level of city governance of different cities are different, and thus, the actual disease transmission parameters are likely to be different. Second, the case data and mobile phone data used in training the model are as of February 7, 2020, and the fatality data are also based on the results of early patient studies [40]. If updated and more recent data were added, the model's predictive performance could be improved (SI Section 3). Third, our model did not consider imported or exported cases between China and other countries. Given global mobility network data and reported case data, further analysis can be extended to simulate infection in different countries, which could be of crucial importance for public health planning, transportation intervention, and logistics support at the current stage, when large-scale outbreaks of COVID-19 are happening in Europe, Iran, and other regions.

The intense social distancing intervention in China has been effective in limiting the scale of the epidemic but needs to be sustained. The model suggests that delay in start and a less effective intervention would increase the number of cases and deaths substantially and prolong the period of social distancing and other control measures.

Acknowledgements We thank Professor Paul Roderick, Keith Abrams, Paul Little, and Mr. Martin Williams for their valuable comments on the manuscript. J.Z. was supported by the Fundamental Research Funds for the Central Universities (No. 2020KJZX004) and National Natural Science Foundation of China (NSFC) (No. 61673070), L.D. was supported by the NSFC (No. 41801299), and Z.H. was supported by NSFC (No. 61374165). G.Y. was supported by the Newton Fund through a UK-China AMR Partnership Hub award (No: MR/S013717/1) and the National Institute for Health Research (NIHR) Health Technology Assessment Programme (No. 13/34/64).

Author contributions JZ, LD, and GY designed the research. JZ and GY designed the model structure and pathway. JZ developed the model. JZ, LD, YZ, and XC collected and analyzed data. GY contributed substantially to the interpretation of results and critique revision of the manuscript. GY and ZH supervised the research. All authors contributed to the writing of the manuscript.

Funding This study is supported by Fundamental Research Funds for the Central Universities (Grant Reference No: 2020KJZX004), National Natural Science Foundation of China (NSFC) (Grant Reference Numbers: 61673070, 41801299, 61374165), the Newton Fund through a UK-China AMR Partnership Hub award (Grant Reference No: MR/S013717/1), and the National Institute for Health Research (NIHR) Health Technology Assessment Programme (Reference Number 13/34/64).

Data and code availability The data and code are available via: https://github.com/jakezj/SICRD_model_COVID19_in_China

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F.: A novel coronavirus outbreak of global health concern. *Lancet* **395**(10223), 470–473 (2020)
- World Health Organization: Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) (2020). <https://www.who.int/>. Accessed 16 Feb 2020
- Li, Q., Guan, X., Wu, P., et al.: Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* (2020). <https://doi.org/10.1056/NEJMoa2001316>
- National Health Commission of the People's Republic of China: Daily briefing on novel coronavirus cases in China (2020). https://en.nhc.gov.cn/2020-03/28/c_78396.htm/. Accessed 28 Mar 2020
- World Health Organization: Coronavirus disease 2019 (COVID-19) Situation Report—67 (2020). <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200327-sitrep-67-covid-19.pdf>. Accessed 28 Mar 2020
- Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A.: The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**(7), 2015–2020 (2006)
- Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* **106**(51), 21484–21489 (2009)
- Belik, V., Geisel, T., Brockmann, D.: Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**(1), 011001 (2011)
- Brockmann, D., Helbing, D.: The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**(6164), 1337–1342 (2013)
- Pei, S., Kandula, S., Yang, W., Shaman, J.: Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci.* **115**(11), 2752–2757 (2018)
- Soriano-Paños, D., Lotero, L., Arenas, A., Gómez-Gardeñes, J.: Spreading processes in multiplex metapopulations containing different mobility networks. *Phys. Rev. X* **8**(3), 031039 (2018)
- Jia, J.S., Lu, X., Yuan, Y., et al.: Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **582**, 389–394 (2020)
- Gómez-Gardenes, J., Soriano-Panos, D., Arenas, A.: Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nat. Phys.* **14**(4), 391–395 (2018)
- Hethcote, H.W.: The mathematics of infectious diseases. *SIAM Rev.* **42**(4), 599–653 (2000)
- China Central Television: Wuhan cut the inter-city traffic, airport, and railway station (2020). <https://m.news.cctv.com/2020/01/23/ARTIW8nDZOFyhQQQoAoMKqIR200123.shtml/>. Accessed 16 Feb 2020 (**in Chinese**)
- People's Daily: National Health Commission: Wuhan newly diagnosed cases dropped to double digits (2020). <https://health.people.com.cn/n1/2020/0307/c14739-31621626.html/>. Accessed 7 Apr 2020 (**in Chinese**)
- Cui, M., Ma, T.H., Li, X.-E.: Spatial behavior of an epidemic model with migration. *Nonlinear Dyn.* **64**, 331–338 (2011)
- Huang, C., Wang, Y., Li, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan China. *Lancet* **395**(10223), 497–506 (2020)
- Holshue, M.L., DeBolt, C., Lindquist, S., et al.: First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* **382**, 929–936 (2020)
- Zhou, P., Yang, X.-L., Wang, X.-G., et al.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020)
- Guan, W.-j, Ni, Z.-y, Hu, Y., et al.: Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* (2020). <https://doi.org/10.1056/NEJMoa2002032>
- Zhao, S., Lin, Q., Ran, J., 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. *Int. J. Infect. Dis.* **92**, 214–217 (2020)
- Liu T, Hu J, Kang M, et al.: Transmission dynamics of 2019 novel coronavirus (2019-nCoV) (2020). Available at SSRN: <https://ssrn.com/abstract=3526307>
- 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* **395**(10225), 689–697 (2020)
- Lai, S., Bogoch, I.I., Watts, A., Khan, K., Li, Z., Tatem, A.: Preliminary risk analysis of 2019 novel coronavirus spread

- within and beyond China. World pop (2020). <https://www.worldpop.org/events/china/>. Accessed 16 Feb 2020
26. Chinazzi, M., Davis, J.T., Ajelli, M., et al.: The effect of travel restrictions on the spread of the 2019 novel coronavirus (2019-nCoV) outbreak. medRxiv (2020). <https://doi.org/10.1101/2020.02.09.20021261>
 27. Tian, H., Liu, Y., et al.: An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **2020**, eabb6105 (2020)
 28. Ferretti, L., Wymant, C., et al.: Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **2020**, eabb6936 (2020)
 29. Cobey, S.: Modeling infectious disease dynamics. *Science* **368**(6492), 713–714 (2020)
 30. Vespignani, A., Tian, H., Dye, C., et al.: Modelling COVID-19. *Nat. Rev. Phys.* **2**, 279–281 (2020). <https://doi.org/10.1038/s42254-020-0178-4>
 31. 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). <https://doi.org/10.1101/2020.01.23.20018549>
 32. Lai, S., Ruktanonchai, N.W., Zhou, L., et al.: Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak: an observational and modelling study. medRxiv (2020). <https://doi.org/10.1101/2020.03.03.20029843>
 33. Lv, W., Ke, Q., Li, K.: Dynamical analysis and control strategies of an SIVS epidemic model with imperfect vaccination on scale-free networks. *Nonlinear Dyn.* **99**, 1507–1523 (2020)
 34. Tripathi, J.P., Abbas, S.: Global dynamics of autonomous and nonautonomous SI epidemic models with nonlinear incidence rate and feedback controls. *Nonlinear Dyn.* **86**, 337–351 (2016)
 35. Baidu: Baidu migration project (2020). <https://qianxi.baidu.com/>. Accessed 16 Feb 2020
 36. Wu, T., Ge, X., Yu, G.: An R package and a website with real-time data on the COVID-19 coronavirus outbreak. medRxiv (2020). <https://doi.org/10.1101/2020.02.25.20027433>
 37. Wolfram Research: Patient medical data for novel coronavirus COVID-19 (2020). <https://doi.org/10.24097/wolfram.11224.data/>. Accessed 16 Feb 2020
 38. China Centers for Disease Control and Prevention: Distribution of new coronavirus pneumonia (2020). <https://2019ncov.chinacdc.cn/2019-nCoV/>. Accessed 16 Feb 2020 (in Chinese)
 39. Anderson, R.M., Anderson, B., May, R.M.: Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford (1992)
 40. Novel, C.P.E.R.E.: The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**(2), 145–151 (2020)
 41. Chen, T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. *Adv. Neural Inf. Process. Syst.* **2018**, 6571–6583 (2018)
 42. Li, R., Pei, S., Chen, B., et al.: Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (COVID-19). *Science* **2020**, eabb3221 (2020)
 43. Heymann, D.L.: Data sharing and outbreaks: best practice exemplified. *Lancet* **395**(10223), 469–470 (2020)
 44. DXY: Daily data report on the COVID-19 (2020). https://ncov.dxy.cn/ncovh5/view/pneumonia_peopleapp/. Accessed Feb 16 Feb 2020 (in Chinese)
 45. Reuters: Chinese citizens turn to virus tracker apps to avoid infected neighborhoods (2020). <https://www.reuters.com/article/us-china-health-apps/chinese-citizens-turn-to-virus-tracker-apps-to-avoid-infected-neighborhoods-idUSKBN1ZX2IH/>. Accessed 16 Feb 2020