

Phase- and epidemic region-adjusted estimation of the number of coronavirus disease 2019 cases in China

Ruijie Chang^{1,*}, Huwen Wang^{1,*}, Shuxian Zhang^{1,*}, Zezhou Wang^{2,*}, Yinqiao Dong^{3,*}, Lhakpa Tsamlag^{1,*}, Xiaoyue Yu^{1,*}, Chen Xu^{1,*}, Yuelin Yu^{1,*}, Rusi Long^{1,*}, Ning-Ning Liu^{1,*}, Qiao Chu^{1,*}, Ying Wang¹, Gang Xu¹, Tian Shen¹, Suping Wang¹, Xiaobei Deng¹, Jinyan Huang⁴, Xinxin Zhang (✉)⁵, Hui Wang (✉)¹, Yong Cai (✉)¹

¹School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; ²Department of Cancer Prevention, Shanghai Cancer Center, Fudan University; Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200025, China; ³Department of Environmental and Occupational Health, School of Public Health, China Medical University, Shenyang 110122, China; ⁴State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, National Research Center for Translational Medicine (Shanghai), Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; ⁵Research Laboratory of Clinical Virology, National Research Center for Translational Medicine (Shanghai), Ruijin Hospital and Ruijin Hospital North Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract The outbreak of the coronavirus disease 2019 in China was first reported in Wuhan in December 2019 and gradually spread to other areas in China. After implementation of prevention and control measures, the estimation of the epidemic trend is needed. A phase- and region-adjusted SEIR model was applied for modeling and predicting the number of cases in Wuhan, Hubei Province and regions outside Hubei Province in China. The estimated number of infections could reach its peak in late February 2020 in Wuhan and Hubei Province, which is 55 303–84 520 and 83 944–129 312, respectively, while the epidemic peaks in regions outside Hubei Province in China could appear on February 13, 2020 with the estimated 13 035–19 108 cases. According to the estimation, the outbreak would abate in March and April all over China. Current estimation provided evidence for planned work resumption under stringent prevention and control in China to further support the fight against the epidemic. Nevertheless, there is still possibility of the second outbreak brought by the work resumption and population migration, especially from Hubei Province and high intensity cities outside Hubei Province. Strict prevention and control measures still need to be considered in the regions with high intensity of epidemic and densely-populated cities.

Keywords SEIR model; COVID-19; estimate; China

Introduction

The outbreak of the coronavirus disease 2019 (COVID-19) in China was first reported in Wuhan in December 2019 and gradually spread to other areas in China [1]. On January 30, 2020, World Health Organization (WHO) announced that it had listed the COVID-19 outbreak as a Public Health Emergency of International Concern [2]. As

of March 11, 2020, the cumulative number of confirmed cases in China (including Hong Kong, Macao, and Taiwan) reached 80 793; 67 781 of which were reported from Hubei Province (49 986 in Wuhan) [3]. Based on the data reported by National Health Commission of the People's Republic of China (NHC), we found that the epidemic in Hubei Province (including Wuhan) was significantly different from that outside. Most studies published or preprinted online [4–7] focused mainly on the epidemic in Wuhan before prevention and control measures were taken. The trend of the epidemic needs to be further investigated and discussed, especially in the whole Hubei Province and regions outside Hubei Province in China after the public health measures putting into action.

Received March 12, 2020; accepted March 14, 2020

Correspondence: Yong Cai, caiyong202028@hotmail.com;

Hui Wang, huiwang@shsmu.edu.cn;

Xinxin Zhang, zhangx@shsmu.edu.cn

*These authors contributed equally to this manuscript.

In the early stage of the disease outbreak, measures for disease prevention and control were relatively weak due to insufficient understanding of its infectivity. Meanwhile, studies focusing on the speed of disease transmission started to predict the number of infections. Forecasts of the basic reproduction number, R_0 , which represents the average number of new infections generated by each infected person [4], were carried out by many institutions. The value of R_0 was estimated to reach 2.2 [8], 2.6 (1.9–3.1) [4], 3.8 [5], or even 6.47 [9], respectively. All these results suggested that the 2019 novel coronavirus (2019-nCoV) had a strong ability of human-to-human transmission.

R_0 plays a crucial role in predicting the prevalence of infectious diseases. Generally, $R_0 > 1$ indicates that effective prevention and control measures have not been taken, thus the prevalence of diseases would be self-sustaining. While $R_0 < 1$ implies that the number of new cases would decrease with the time passing by, thus the prevalence of diseases would gradually subside [4]. The corresponding prevention and control measures may effectively control the spread of the disease despite the lack of vaccines and effective treatment. Taking SARS as an example, the effective reproduction number, R_e (i.e., the average number of secondary cases under different control scenarios) [10], was decreased to less than 1 after strict quarantine [11].

After the COVID-19 outbreak, many unprecedented public health measures were implemented in China. The Huanan Seafood Wholesale Market was closed on January 1, 2020 once an epidemiological link to the market of the initial cluster was detected [12,13]. Furthermore, the public transportation to and from Wuhan, the epidemic focus of the outbreak, was halted since January 23, 2020 [14]. By January 26, 2020, 30 provinces in China had initiated the first-level public health emergency response and conducted medical examinations on each individual from Hubei [15]. On the same day, the traditional Chinese New Year festival was extended and the spring semester for students was postponed to delay the large scale population migration and gathering [16]. Transportation restriction and quarantine were gradually established, with all close contacts and suspected cases receiving medical observation and quarantine for 14 days [17]. As the number of infections increased in early February, China took on more stringent prevention and control measures since early February. Two specialized hospitals and several mobile cabin hospitals were established in Wuhan [18,19]. Community-oriented public health management stated goal of “Never miss a single household, never miss a single individual,” was carried out in many cities of Hubei Province [20]. Since February 21, 2020, the diagnosis has been mainly based on contact history, clinical manifestation, and imaging evidence of pulmonary lesion suggestive of pneumonia, while viral detection with RT-PCR is still being performed

in a part of patients, to detect possible infections more timely [21].

All these actions may work to reverse the epidemic trend of COVID-19. Additionally, it might neglect many comprehensive factors that occurred as the epidemic progresses by using a single R_0 to predict the number of infections. Previously, we have forecasted the phase-adjusted epidemic trend of COVID-19 in Wuhan with control measures taken into consideration [7], which was proved to match the actual trend well according to the recent data reported by NHC. However, considering the rapid progression of the epidemic and a great difference between Wuhan and Hubei Province and regions outside Hubei Province in China, updated estimation in Wuhan and region-adjusted estimation in the whole China is further needed. This study aimed at (1) estimating the epidemic trend in Wuhan, Hubei Province and regions outside Hubei Province in China, respectively using phase- and region-adjusted R_e with different levels of prevention and control measures taken into consideration, (2) predicting the possible infection scale and epidemic characteristics of COVID-19, (3) providing potential evaluation criteria for the effectiveness of current prevention and control measures, (4) finally offering theoretical support for current decision-making on infectious disease prevention.

Materials and methods

Model

We employed an infectious disease dynamics model (Susceptible, Exposed, Infectious, and Removed model (SEIR model)) for modeling and predicting the number of COVID-19 cases in Wuhan, Hubei Province and regions outside Hubei Province in China. The model is classically applied to analyze the infectious disease which has a definite latent period, and has been proved to be predictive for a variety of acute infectious diseases in the past such as Ebola and SARS [11,22–27]. And utilization of this mathematical model is significantly instructive in assessing the impact of isolation of symptomatic cases as well as medical observation of asymptomatic contact cases and promoting evidence-based decisions and policies.

The model was established based on the following assumptions: no new transmissions from animals, no differences in individual immunity, no super-spreaders, the time-scale of the epidemic is much faster than characteristic times for demographic processes (natural birth and death), and no differences in natural births and deaths. In this model, individuals are classified into four types: susceptible (S; at risk of contracting the disease), exposed (E; infected but not yet infectious), infectious (I; capable of transmitting the disease), and removed (R; those who

recover or die from the disease). The total population size (N) is given by $N = S + E + I + R$. It is assumed that susceptible individuals who have been infected first enter a latent (exposed) stage, during which they may have a low level of infectivity. The differential equations of the SEIR model are given as [28,29]:

$$dS/dt = -\beta SI/N$$

$$dE/dt = \beta SI/N - \sigma E$$

$$dI/dt = \sigma E - \gamma I$$

$$dR/dt = \gamma I$$

$$\beta = R_0\gamma$$

where β is the transmission rate; σ is the infection rate calculated by the inverse of the mean latent period; γ is the recovery rate calculated by the inverse of the infectious period.

R software (Version 3.6.2) was applied for all the calculations and estimates in the current study.

Data collection and parameter values

Estimation of the epidemic trend in both Hubei Province and Wuhan

We first estimated the epidemic trend in Hubei Province and Wuhan, respectively, with phase adjusted R_e to keep in line with the actual levels of prevention and control measures in different phases. At baseline, S was assumed to be the population of Hubei Province (59.17 million) and Wuhan (11 million), respectively [8,30]. The initial assumed numbers of cases caused by zoonotic exposure were 40 (I) in Wuhan according to Imai *et al.*'s estimation [4] and 60 (I) in Hubei Province calculated by the proportion of Wuhan infectious cases in Hubei Province from daily reported data released by NHC. We assumed E at 20 times of I in accordance with Read *et al.* [5]. Absence of fever in COVID-19 cases is more frequent than in SARS-CoV and MERS-CoV infection [31]. Such patients may be missed for diagnosis since the current surveillance case definition focused mainly on fever detection. Accordingly, the possibility of E at 30 times of I cannot be excluded. Therefore, we also explored the model with E at 30 times of I to provide a possible range of infections. R was set as 0. σ was set as $1/5.2$ according to the latest article by Li *et al.* [8], which calculated the mean incubation period of COVID-19 to be 5.2 days. Chen *et al.* [32] calculated the average hospitalization period of 31 discharged patients to be 12.39 ± 4.77 days. Yang *et al.* calculated the median time from disease onset to diagnosis among confirmed patients to be 5 days [33]. γ was accordingly set as $1/18$ (ceiling of $12.39 + 5$ is 18) [7].

R_e was chosen by various phases. The first phase ranges from December 1, 2019 to January 23, 2020 and can be regarded as the early phase of the epidemic when a few prevention and control measures were implemented. R_0

was set as 3.1 consistent with Imai *et al.*'s estimation of high transmission level [4]. On January 23, 2020, airplanes, trains, and other public transportation within the city were restricted and other prevention and control measures such as quarantine and isolation were gradually established in Wuhan and Hubei Province [14]. So the second phase began on January 24, 2020 and R_e was set as 2.6 consistent with Imai *et al.*'s estimation of moderate transmission level [4]. February 2, 2020 was the last day of the extended Spring Festival holiday and more medical resources were mobilized to support Wuhan and Hubei Province ever since [34]. Two specialized hospitals which came into service to admit more infections, and mobile cabin hospitals, which were originally gymnasium, were in operation to admit cases with mild to moderate symptoms [18,19]. In addition, more and more medical teams arrived in Wuhan and Hubei Province. Therefore, the third phase began on February 3, 2020 and R_e was set as 1.9 according to Imai *et al.*'s estimation of low transmission level [4]. It might take the longest incubation period (about 14 days) for all these measures to take effect. Finally, the last phase began on February 16, 2020 and R_e was set as 0.9, 0.5, and 0.25, respectively, assuming the prevention and control measures are sufficient and effective to depict three different levels of preventive measures in reducing transmission probability [7].

Estimation of the epidemic trend in regions outside Hubei Province in China

Next, we provided an estimation of the epidemic trend in the regions outside Hubei Province in China, trying to fill the research gaps and depict the disease holistically. Considering the large population flow between cities and provinces during the Spring Festival, we chose the median date (January 12, 2020) of the travel rush period as the starting date of estimation. The set of S , R , σ , and γ was similar to the first section. The initial values of I , 541 and 793, was set based on the previous estimation of numbers of infections in Wuhan with $E = 20I$ and $30I$, respectively [7], and two assumptions. The first assumption indicates that the outflow rate of the population from Hubei Province was constant. In the second assumption, the proportion of infected people in the outflow population was the same as the proportion of infected people in Wuhan. We assumed that exposed cases in regions outside the epidemic focus are relatively less and calculated based on data reported by NHC that E was approximately at 10 times of I in regions outside Hubei Province in China.

R_e was also selected by phases but different from those in Wuhan and Hubei Province, considering the relatively low cumulative number of COVID-19 cases and prompt prevention and control measures in regions outside Hubei Province in China. The first phase ranged from January 12

to 23, 2020. This is the peak period of Spring Festival travel rush. A few measures were taken during this time. Transportation from Wuhan was not yet stopped. Regions outside Hubei Province in China were assumed to be confronted with the same transmission risk as Hubei Province. Therefore, R_0 was set as 3.1 consistent with Imai *et al.*'s estimation of high transmission level [4]. By January 26, 2020, 30 provinces in China had initiated the first-level public health emergency response and conducted medical examinations on each individual coming from Hubei [15]. The epidemic entered into a new phase with stringent measures implemented. The second phase began on January 24, 2020 and R_e was set as 1.9, consistent with Imai *et al.*'s low estimation of moderate transmission level [4]. After a long incubation period, February 6, 2020 was regarded as a new time point when various measures began to take effect. The number of newly-confirmed cases decreased continuously for 14 days in regions outside Hubei Province in China [35]. For the third phase, it began on February 7, 2020 and R_e was set as 0.9. Since February 14, 2020, the number of confirmed cases was significantly reduced and the number of cases discharged from hospital exceeded that of the newly-confirmed cases in several provinces [3]. Therefore, the last phase began on February 14, 2020 and R_e was set as 0.5, 0.3, and 0.15, respectively, to predict the epidemic situations after different levels of actual intervention taking effect.

Regional distribution of COVID-19 cases in China

Due to the limited daily data source in different regions in China every day, we were unable to conduct estimation for

each region outside Hubei Province, but regarded them as a whole continuous region. To supplement this, we described the regional distribution of COVID-19 cases in China to show the regional disparity of the current epidemic. The epidemic maps of COVID-19 in China and Hubei Province were drawn using ArcGIS 10.2 (Environmental Systems Research Institute, Inc., California, USA) based on data from National Earth System Science Data Center and National Bureau of Statistics of China.

Results

Estimation of the epidemic trend in Wuhan

COVID-19 infections in Wuhan would reach 17 656–25 875 by the end of the first phase (December 1, 2019–January 23, 2020) with R_0 as 3.1. At the end of the second phase (January 24, 2020–February 2, 2020), it would be 32 061–46 905 with R_e decreased to 2.6. During February 3 to 15, 2020 (the third phase), the number of infections would reach 53 070–77 390 in the end when R_e reduced sequentially to 1.9. Finally, if R_e could gradually drop to 0.9, 0.5 or 0.25 during the fourth phase as we estimated (February 16, 2020–April 30, 2020), the peaks could occur on February 23, February 19 or February 18, 2020 in Wuhan. The number of infections could be 58 077–84 520, 55 869–81 393 or 55 303–80 588 with $R_e = 0.9, 0.5$ or 0.25, respectively, and then might continue to decline in March and April until the epidemic disappears. The actual confirmed cases reported by NHC peaked on February 18, 2020, in line with our estimated epidemic trend with $R_e = 0.25$ (Fig. 1 and Fig. 2).

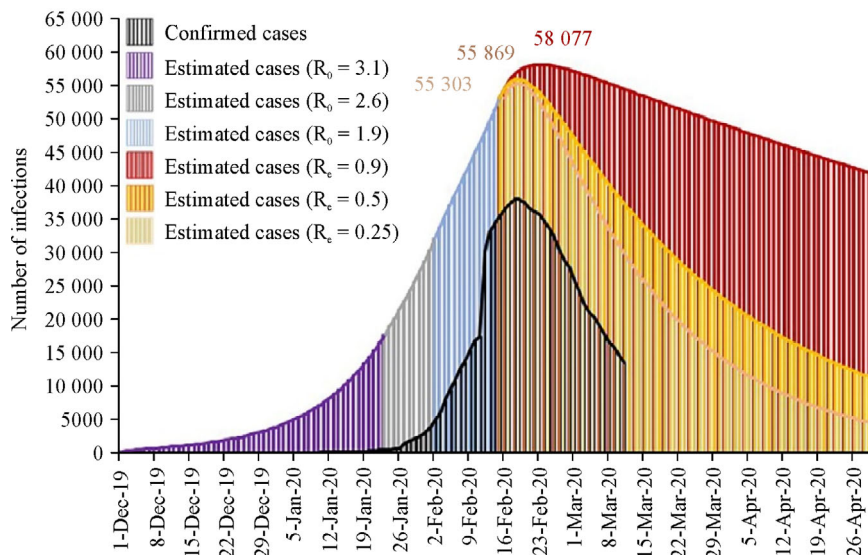


Fig. 1 Phase-adjusted estimation of the number of COVID-19 cases in Wuhan, China (December 1, 2019–April 30, 2020, $E = 20I$).

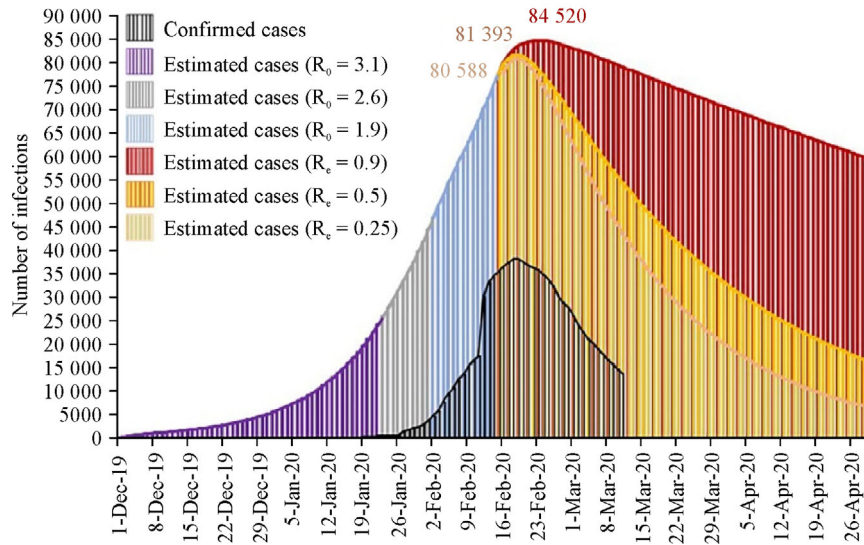


Fig. 2 Phase-adjusted estimation of the number of COVID-19 cases in Wuhan, China (December 1, 2019–April 30, 2020, $E = 30I$).

Estimation of the epidemic trend in Hubei Province

As shown in Fig. 3 and Fig. 4, the division of epidemic phases and the determination of R values in Hubei Province were the same as those in Wuhan. The number of infections in Hubei Province would reach 26 559–38 967 by the end of the first phase (December 1, 2019–January 23, 2020) with R_0 as 3.1 and would increase to 48 351–70 906 by the end of the second phase (January 24, 2020–February 2, 2020) as R_e decreased to 2.6. When R_e reduced to 1.9, the number of infections would reach 80 441–117 856 at the end of the third phase (February 3,

2020–February 15, 2020). In the fourth phase (February 16, 2020–April 30, 2020) when R_e gradually dropped to 0.9, 0.5 or 0.25 after effective prevention and control measures taking effect, the epidemic peaks were expected to occur on February 23, February 19 or February 18, 2020, respectively. The number of infections could eventually be 88 310–129 312, 84 819–124 231 or 83 944–122 946 with $R_e = 0.9, 0.5$ or 0.25, respectively, and might keep dropping for next two months. The actual confirmed cases reported by NHC peaked on February 18, 2020, which was exactly the same day as we estimated with $R_e = 0.25$ (Fig. 3 and Fig. 4).

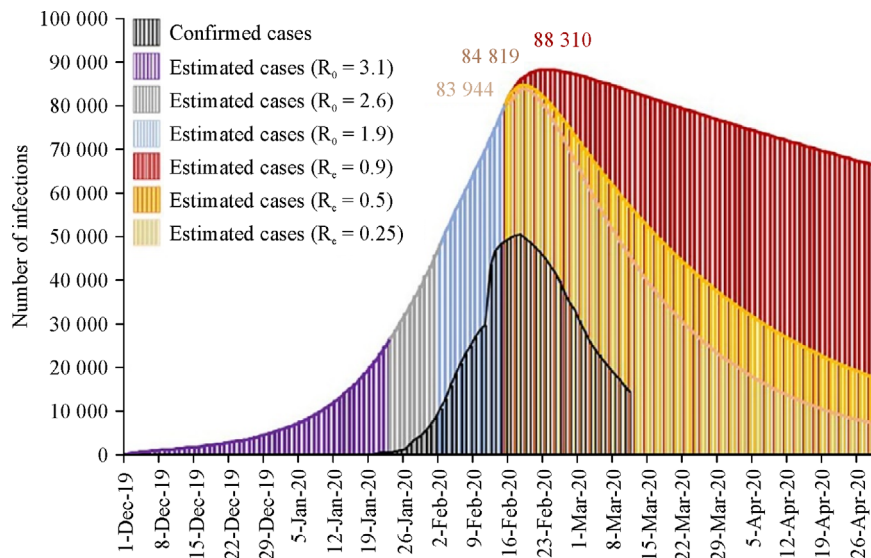


Fig. 3 Phase-adjusted estimation of the number of COVID-19 cases in Hubei Province, China (December 1, 2019–April 30, 2020, $E = 20I$).

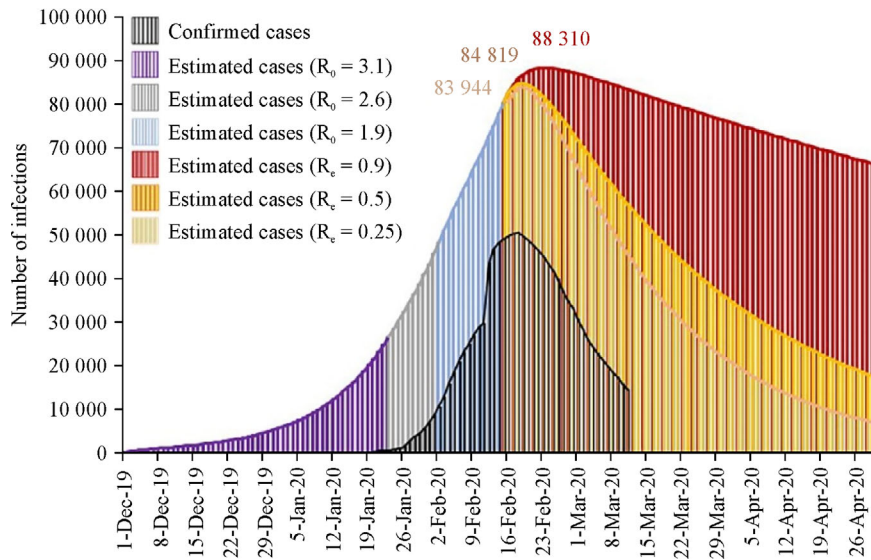


Fig. 4 Phase-adjusted estimation of the number of COVID-19 cases in Hubei Province, China (December 1, 2019–April 30, 2020, $E = 30I$).

Estimation of the epidemic trend in regions outside Hubei Province in China

According to our prediction model, the number of infections in regions outside Hubei Province in China would reach 6531–9573 by the end of the first phase (January 12, 2020–January 23, 2020) with R_0 as 3.1 when the number of initial infections were 541 and 793 on January 12, 2020, respectively. However, it would increase almost twofold to 11 870–17 400 by the end of the second phase (January 24, 2020–February 6, 2020) as R_e decreased to 1.9. When R_e was reduced sequentially to 0.9, the number of infections would decline correspond-

ingly to 13 035–19 108 by the end of the third phase (February 7, 2020–February 13, 2020). The epidemic peaks would appear on February 13, 2020 and the total number of infections would keep falling after that. In the last phase (February 14, 2020–April 30, 2020), R_e was expected to drop gradually to 0.5, 0.3, and 0.15 with effective prevention and control measures in effect. By then the number of infections would be 2346–3439, 1051–1541, and 540–790 with $R_e = 0.5, 0.3,$ and $0.15,$ respectively, on April 30, 2020. The actual confirmed cases in regions outside Hubei Province peaked on February 11, 2020, two days ahead of our estimated peak day (Fig. 5 and Fig. 6).

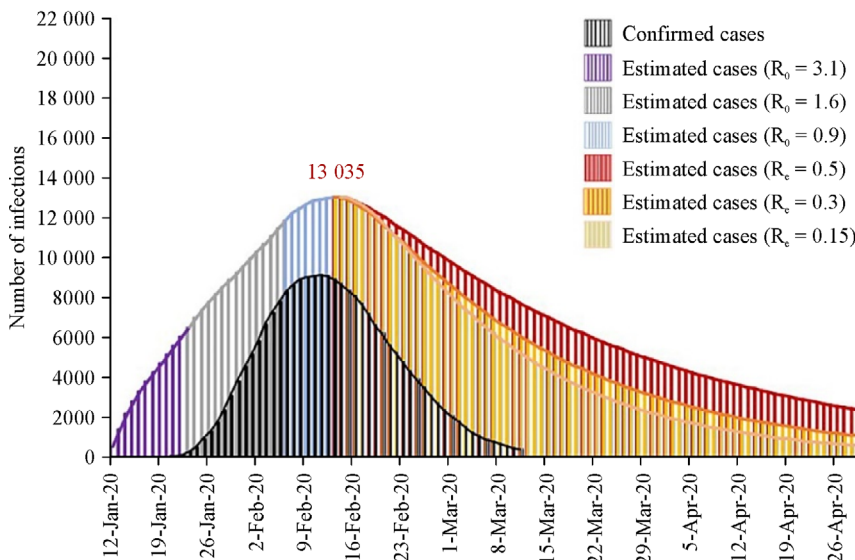


Fig. 5 Phase-adjusted estimation of the number of COVID-19 cases in regions outside Hubei Province, China (January 12, 2020–April 30, 2020, $E = 10I, I = 541$).

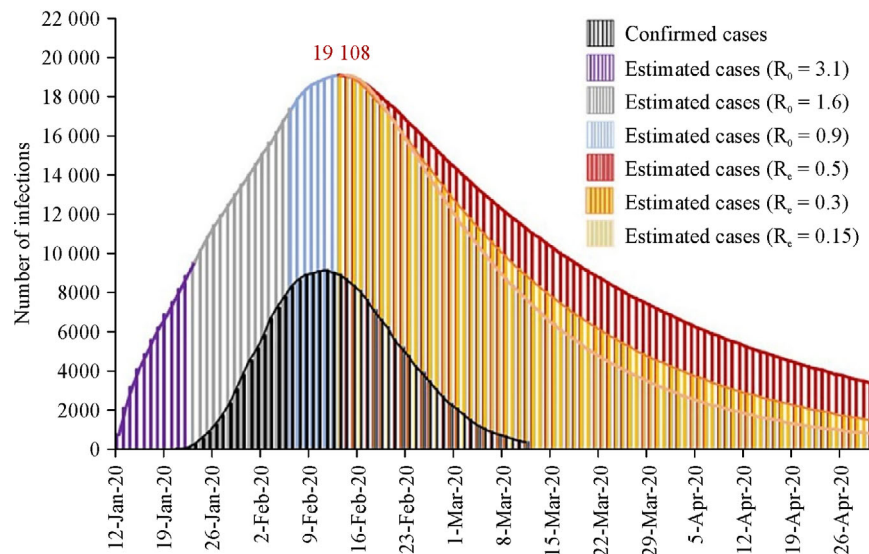


Fig. 6 Phase-adjusted estimation of the number of COVID-19 cases in regions outside Hubei Province, China (January 12, 2020– April 30, 2020, $E = 10I$, $I = 793$).

The number of deaths of COVID-19 cases might range from 2686 to 6207 in Hubei Province (including 2212–5071 in Wuhan). It is estimated that there would be about 91–191 deaths due to COVID-19 in regions outside Hubei Province (Table 1).

Table 1 Estimating the number of deaths of COVID-19 cases in China

	Number of deaths	Number of cases
Wuhan	2212–5071 ^a	55 303–84 520
Hubei Province	2686–6207 ^b	83 944–129 312
Regions outside Hubei Province	91–191 ^c	13 035–19 108

^a The estimated percent of deaths is about 4%–6%.

^b The estimated percent of deaths is about 3.2%–4.8%.

^c The estimated percent of deaths is about 0.7%–1%.

Regional distribution of COVID-19 cases in China

We compared the cumulative number of confirmed COVID-19 infections and infection rate (per million population) on January 20, 2020 with that on March 11, 2020 in China, particularly in Hubei Province.

By January 20, 2020, there were 270 cases in Hubei Province and 21 imported cases in regions outside Hubei including 14 in Guangdong Province, 5 in Beijing, and 2 in Shanghai. As of March 11, 2020, the cumulative number of confirmed cases in China (including Hong Kong, Macao, and Taiwan) reached 80 980. In regions outside Hubei Province, there is a more severe epidemic in major financial regions (Beijing, Guangdong, and Shanghai) and regions neighboring Hubei Province (Anhui, Chong-

qing, Henan, Hunan, Jiangxi, and Jiangsu), where more than 10 per million population were infected. Inner Mongolia, Liaoning, Qinghai, Taiwan, and Tibet are among the few regions where the number of confirmed diagnoses per million population was less than 3 in China.

Within Hubei Province, the cumulative number of confirmed cases reached 67 781, surging nearly 250 times from January 20 to March 11, 2020. Wuhan, the epidemic focus of COVID-19, accounted for 73.7% of the total confirmed number in Hubei Province (more than 4500 per million population). There are differences in the prevalence of different distances from Wuhan. In the areas adjacent to Wuhan (Ezhou, Xiaogan, Xiantao, Huanggang, and Xianning), the average infection rate was 330 to 1214 per million population. Conversely, the infection rate of three relatively distant provincial cities (Shiyan, Enshi, and Shennongjia Forestry District) was 75 to 196 per million population.

Discussion

The epidemic trend and transmission risk of COVID-19 called for timely estimation to evaluate the effectiveness and sufficiency of the current control and treatment measures. Data reported by NHC indicate a great difference between the epidemic focus regions Wuhan and Hubei Province and regions outside Hubei Province in China, necessitating the region-adjusted estimation of the epidemic trend. In early phase of the epidemic, few control measures resulted in the rapid growth of the outbreak. Based on previous studies, more than 100 000 people would be infected in China without effective prevention

and control measures ($R_e > 3$) [5,6]. On January 23, 2020, public transportation to and from Wuhan were restricted, and more and more prevention and control measures such as quarantine and isolation were gradually established in China. Since then, R_e was expected to gradually step down with strict measures coming into effect. Therefore, in the current study, a phase- and region-adjusted SEIR model was performed to estimate the number of infections in China.

Hubei Province and its capital city Wuhan are the epidemic focus regions where the COVID-19 outbreak emerged. The trend of the estimated cases in Hubei Province including Wuhan is in accordance with the trend of currently confirmed cases. The relatively big difference in number may be due to the possible existence of a large number of mild and asymptomatic cases and the limitations of current diagnostic measures. Therefore, the epidemic peaks are highly possible to occur in Wuhan and Hubei in late February 2020 as estimated in the current study. The estimated peak number of infections would be 55 303–84 520 and 83 944–129 312, respectively, and would occur around February 20, 2020. Our estimated epidemic trend was in line with the actual trend reported by NHC, increasing the credibility of the estimation model.

At the early stage of the outbreak, for lack of awareness of the disease and sufficient prevention and control measures, a huge number of latent infections flowed out from Hubei Province across the country during the Spring Festival. Regions outside Hubei Province might be confronted with similar risk of COVID-19 transmission as Hubei. However, in later phases, R_e values in regions outside Hubei Province in China may decline faster than those of Hubei Province and Wuhan because of the following reasons. Compared with the epidemic focus regions, there were fewer infections in other regions in the early stage. The epidemic intensity of the disease was not so high as in Wuhan and Hubei. And the medical resources were relatively more sufficient. The daily number of newly-confirmed cases in regions outside Hubei Province in China has decreased since February 4, 2020. The number of newly-added recovered cases exceeded that of newly-confirmed cases while the number of existing confirmed cases has begun to decline since February 12, 2020. Our estimation findings support the assertion that the epidemic peaks may have already emerged [3]. It's further proved by our estimation. In regions outside Hubei Province in China, the turning point of the model appeared on February 13, 2020. The peak number of infections was 19 108 as estimated. It is reasonable to speculate that the outbreak would be halted soon with maintaining the effective prevention and control measures all over the country.

At present, the situation of epidemic prevention and control in China is showing a positive trend, the work of enterprise epidemic prevention and control still needs to

implement. The deployment of strengthening epidemic control and resumption of work issued by the Communist Party of China Central Committee and the State Council is used to guide all regions to further promote the effective implementation of enterprise resumption of work and epidemic prevention and control [36]. This announcement together with our estimation provide evidence for planned work resumption under stringent prevention and control in regions outside Hubei Province in China to further support the fight against the epidemic. Prevention and control measures for returning population migration should be paid great attention, especially in the regions with high intensity of epidemic such as Hunan, Guangdong, Zhejiang provinces as well as densely-populated cities such as Shanghai and Beijing. A 14-day medical observation period and protective measures in the workplace or other public spaces are still of great importance to prevent the reversal of the epidemic trend, even when the peak of the confirmed cases seemed to have emerged.

It is hard to estimate the number of deaths of COVID-19 cases during the outbreak stage. Early research has calculated the case fatality rate (CFR) of 11% [32], however with the application of potentially effective new therapies such as convalescent plasma from the recovered patients and cytokine storm monoclonal antibody drugs [37], the cure rate of severe and critically ill patients is expected to increase. We calculated the CFR of COVID-19 to be 4%–6% in Wuhan and 3.2%–4.8% in Hubei Province according to NHC report [3]. The CFR outside Hubei Province was calculated to be 0.7%, but considering that there are still some severe and critically ill patients in the hospital, the CFR was extended to 0.7%–1.0%. At present, great importance must be attached to further strengthening the treatment of severe patients in order to effectively reduce the CFR.

The present study has some limitations. Firstly, the SEIR model was based on a few essential assumptions (e.g., no multiple zoonotic sources of the virus, no infectivity during incubation period, and no super-spreaders), which lacked supportive evidence currently. Specifically, due to the inherent limitation of the SEIR model [25], which assumes diseases spread evenly across homogeneous population, the number of cases in the current study might be underestimated constrained by the possible existence of super-spreaders and asymptomatic infectors. We could do essential modification to the estimation model in future research. Secondly, the accuracy of the estimation model depends largely on the accuracy of the parameters it used (S , E , I , R , β , σ , and γ). Data lags might exist since the parameters (such as recovery period) were based on published literatures rather than real-time data. There are inevitable deviations between the prediction and reported data. The predicted data are relatively stable, while the reported data may be affected by many real-time human factors. For example, the currently not stringent enough

standard for recovery (proven by the occurrence of those discharged but turned out to be not fully recovered) may make the reported peak falsely advanced. Despite the deviations, the epidemic trend estimated by SEIR model could provide theoretical reference for disease prevention and control. With more precise parameters obtained as the epidemic progresses and more published researches, our estimation model will also be more precise. We will further revise the model in the future research. Thirdly, since R_e is a dynamic parameter, it would be hard to avoid bias in the model when using a fixed R_e value in every single phase. Furthermore, the model involved so many parameters that the multi-sensitivity analysis would be very complex. For that reason, we only considered the relationship between number of people exposed to disease (E) and number of people infected (I), which we thought would impact the model mostly, and made two possible assumptions (i.e., $E = 20I$ and $E = 30I$). SEIR model was also estimated with two initial data in regions outside Hubei Province. The number of infected people leaving Wuhan was calculated as 541 and 793 based on our predictive number in Wuhan under two different assumptions. They could be considered as equivalents to sensitivity analyses in some extent. Also, the model fit information should be supplemented once the data corresponding to actual situations are available in the future. With practical restriction, we regarded regions outside Hubei Province in China as a whole mainly because the big differences among regions outside Hubei Province are not as significant as those between Hubei Province and other regions, and it is uncertain to predict the epidemic trend for every province and city separately with uncertain essential parameters. Lastly, the SEIR model in the current study is usually used to estimate the epidemic in a closed crowd. The relatively large population flowing in regions outside Hubei Province in China may make the application of the model not as reasonable as in Wuhan and Hubei Province. Due to the inconstant domestic population migration and the increasing potential cases imported from abroad in the coming weeks, it is hard to make accurate epidemic forecast in the SEIR model, especially for the possible second outbreak.

Despite the limitations mentioned above, this study was the first phase- and region-adjusted estimation of the epidemic trend in strict-controlled Hubei Province and regions outside Hubei in China based on the latest data. The peak of the number of infections would probably occur by the end of February in Hubei Province and Wuhan, and may have occurred in regions outside Hubei Province. Nevertheless, there is still possibility of the second outbreak brought by the work resumption and population migration, especially from Hubei Province and high intensity cities outside Hubei Province. Therefore, strict prevention and control measures still need to be considered, especially in the regions with high intensity of

epidemic such as Hunan, Guangdong, Zhejiang provinces as well as densely-populated cities such as Shanghai and Beijing.

Acknowledgements

This work is funded by Medicine and Engineering Interdisciplinary Research Fund of Shanghai Jiao Tong University (No. YG2020YQ06), the National Key Research and Development Project (Nos. 2018YFC1705100, 2018YFC1705103, and 2018YFC2000700) and the National Natural Science Foundation of China (Nos. 71673187 and 81630086), the Key Research Program (No. ZDRW-ZS-2017-1) of the Chinese Academy of Sciences, Innovative research team of high-level local universities in Shanghai. We acknowledge all health-care workers involved in the diagnosis and treatment of patients all around China. We thank National Health Commission of the People's Republic of China for coordinating data collection for patients with COVID-19.

Compliance with ethics guideline

Ruijie Chang, Huwen Wang, Shuxian Zhang, Zezhou Wang, Yinqiao Dong, Lhakpa Tsamlag, Xiaoyue Yu, Chen Xu, Yuelin Yu, Rusi Long, Ning-Ning Liu, Qiao Chu, Ying Wang, Gang Xu, Tian Shen, Suping Wang, Xiaobei Deng, Jinyan Huang, Xinxin Zhang, Hui Wang, and Yong Cai declare no competing interests. The data sets analyzed for this study can be found in the reports of the National Health Commission of the People's Republic of China (NHC) (http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml). This manuscript does not involve a research protocol requiring approval by the relevant institutional review board or ethics committee.

References

1. WHO. Pneumonia of unknown cause—China. January 5, 2020. <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/> (accessed March 12, 2020)
2. WHO. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). January 30, 2020. [https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)) (accessed March 12, 2020)
3. National Health Commission of the People's Republic of China. Update on the incidence of pneumonia with novel coronavirus infection. March 12, 2020. http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml (in Chinese) (accessed March 12, 2020)
4. Imai N, *et al.* Report 3: Transmissibility of 2019-nCoV. March 11, 2020. <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-3-transmissibility-of-covid-19/> (accessed March 12, 2020)
5. Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological

- parameters and epidemic predictions. *MedRxiv* 2020; doi: 10.1101/2020.01.23.20018549
6. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang MH. Preliminary estimation of the basic reproduction number of novel coronavirus. *Int J Infect Dis* 2020; 92: 214–217
 7. Wang H, Wang Z, Dong Y, Chang R, Xu C, Yu X, Zhang S, Tsamslag L, Shang M, Huang J, Wang Y, Xu G, Shen T, Zhang X, Cai Y. Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell Discov* 2020; 6(1): 10
 8. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Li M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JTK, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020 Jan 29. [Epub ahead of print] doi: 10.1056/NEJMoa2001316
 9. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, Wu JH. Estimation of the transmission risk of 2019-nCoV and its implication for public health interventions. *SSRN* 2020 Jan 24. doi: 10.2139/ssrn.3525558
 10. Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. *J Theor Biol* 2015; 384: 33–49
 11. Chen SC, Chang CF, Liao CM. Predictive models of control strategies involved in containing indoor airborne infections. *Indoor Air* 2006; 16(6): 469–481
 12. WHO. Novel coronavirus—China. January 12, 2020. <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/> (accessed March 12, 2020)
 13. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A, Petersen E. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 2020; 91: 264–266
 14. The Central People’s Government of the People’s Republic of China. Notice on Wuhan novel coronavirus infection of pneumonia epidemic prevention and control. 2020. http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm/ (in Chinese) (accessed March 12, 2020)
 15. National Health Commission of the People’s Republic of China. Full mobilization to contain the epidemic. January 27, 2020. <http://www.nhc.gov.cn/xcs/xwbd/202001/5e03b7c4bf3046ef99d46d57ce64aff6.shtml> (in Chinese) (accessed March 12, 2020)
 16. The Central People’s Government of the People’s Republic of China. Notice on prolonging the Spring Festival Holiday. 2020. http://www.gov.cn/zhengce/content/2020-01/27/content_5472352.htm (in Chinese) (accessed March 12, 2020)
 17. National Health Commission of the People’s Republic of China. Prevention and Control Plan for New Coronavirus-Infected Pneumonia (version 3). January 28, 2020. <http://www.nhc.gov.cn/jkj/s7923/202001/470b128513fe46f086d79667db9f76a5.shtml> (in Chinese) (accessed March 12, 2020)
 18. The People’s Government of Wuhan. Wuhan Huoshenshan Hospital was completed today. February 2, 2020. http://www.wuhan.gov.cn/sy/whyw/202003/t20200316_960407.shtml (in Chinese) (accessed March 12, 2020)
 19. The People’s Government of Hubei Province. Wuhan Leishenshan Hospital welcomed its first confirmed 2019-nCoV patients. February 9, 2020. http://www.hubei.gov.cn/zhuant/2020/gzxxgzbd/qfqk/202002/t20200209_2022182.shtml (in Chinese) (accessed March 12, 2020)
 20. National Health Commission of the People’s Republic of China. Do the best to prevent and control the epidemic from the source. February 6, 2020. <http://www.nhc.gov.cn/wjw/xwdt/202002/174ec162cd2e4b61a7036d32ff67f50c.shtml> (in Chinese) (accessed March 12, 2020)
 21. Health Commission of Hubei Province. Update on 2019-nCoV cases in Hubei Province. February 13, 2020. http://wjw.hubei.gov.cn/bmdt/ztl/fkxxgzbdgrfyq/xxfb/202002/t20200213_2025580.shtml (in Chinese) (accessed March 12, 2020)
 22. Fang H, Chen J, Hu J. Modelling the SARS epidemic by a lattice-based Monte-Carlo simulation. *Conf Proc IEEE Eng Med Biol Soc* 2005; 2005(7): 7470–7473
 23. Saito MM, Imoto S, Yamaguchi R, Sato H, Nakada H, Kami M, Miyano S, Higuchi T. Extension and verification of the SEIR model on the 2009 influenza A (H1N1) pandemic in Japan. *Math Biosci* 2013; 246(1): 47–54
 24. Smirnova A, deCamp L, Chowell G. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model. *Bull Math Biol* 2019; 81(11):4343–4365
 25. Getz WM2, Salter R, Mgbara W. Adequacy of SEIR models when epidemics have spatial structure: Ebola in Sierra Leone. *Philos Trans R Soc Lond B Biol Sci* 2019; 374(1775): 20180282
 26. Feng Z, Yang Y, Xu D, Zhang P, McCauley MM, Glasser JW. Timely identification of optimal control strategies for emerging infectious diseases. *J Theor Biol* 2009; 259(1): 165–171
 27. Peak CM, Childs LM, Grad YH, Buckee CO. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc Natl Acad Sci USA* 2017; 114(15): 4023–4028
 28. Chowell G, Nishiura H, Bettencourt L. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J R Soc Interface* 2007; 4(12): 155–166
 29. Chowell G, Hengartner NW, Castillo-Chavez C, Fenimore PW, Hyman JM. The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *J Theor Biol* 2004; 229(1): 119–126
 30. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, Li Q. Notes from the field: an outbreak of NCIP (2019-nCoV) infection in China — Wuhan, Hubei Province, 2019–2020. *China CDC Weekly* 2020; 2(5):79–80 <http://weekly.chinacdc.cn/en/article/id/e3c63ca9-dedb-4fb6-9c1c-d057adb77b57> (accessed March 12, 2020)
 31. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, Liu L, Shan H, Lei C, Hui DSC, Du B, Li L, Zeng G, Yuen K, Chen R, Tang C, Wang T, Chen P, Xiang J, Li S, Wang J, Liang Z, Peng Y, Wei L, Liu Y, Hu Y, Peng P, Wang J, Liu J, Chen Z, Li G, Zheng Z, Qiu S, Luo J, Ye C, Zhu S, Zhong N. Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv* 2020; doi: 2020.02.06.20020974
 32. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia Ja, Yu T, Zhang X, Zhang L. Epidemiological and

- clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; 395 (10223):507–513
33. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, Dean N, Longini I, Halloran ME, Xu B, Zhang X, Wang L, Liu W, Fang L. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *medRxiv* 2020; doi: 2020.02.10.20021675
 34. The People's Government of Hubei Province. Pooling high-quality medical resources all over the country to win the battle of preventing and controlling epidemics. February 11, 2020. http://www.hubei.gov.cn/zwgk/hbyw/hbywqb/202002/t20200211_2023527.shtml (in Chinese) (accessed March 12, 2020)
 35. The China News Online Service. The number of newly confirmed cases fell for 14 consecutive days in regions outside Hubei Province in China. February 18, 2020. <http://ent.chinanews.com/gn/2020/02-18/9095950.shtml> (in Chinese) (accessed March 12, 2020)
 36. National Health Commission of the People's Republic of China. Circular of the General Office of the National Health Commission on strengthening the guidance of epidemic prevention and control during the resumption of work. March 7, 2020. <http://www.nhc.gov.cn/xcs/zhengcwj/202003/cc7158449091457b9a88a4c5b88e075d.shtml> (in Chinese) (accessed March 12, 2020)
 37. National Health Commission of the People's Republic of China. Diagnosis and treatment scheme for COVID-19 (the sixth edition). February 18, 2020. <http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2/files/b218cfeb1b-c54639af227f922bf6b817.pdf> (in Chinese) (accessed March 12, 2020)