

http://www.aimspress.com/journal/MBE

MBE, 17(4): 3052–3061. DOI: 10.3934/mbe.2020173 Received: 11 February 2020 Accepted: 16 March 2020 Published: 08 April 2020

Research article

Current trends and future prediction of novel coronavirus disease (COVID-19) epidemic in China: a dynamical modeling analysis

Kai Wang^{1,†}, Zhenzhen Lu^{2,†}, Xiaomeng Wang³, Hui Li⁴, Huling Li⁵, Dandan Lin⁵, Yongli Cai³, Xing Feng⁵, Yateng Song⁵, Zhiwei Feng⁵, Weidong Ji⁵, Xiaoyan Wang⁵, Yi Yin¹, Lei Wang^{1,*} and Zhihang Peng^{2,6,*}

- ¹ College of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830011, China
- ² Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, 211166, China
- ³ School of Mathematics and Statistics, Huaiyin Normal University, Huaian, 223300, China
- ⁴ Central Laboratory of Xinjiang Medical University, Urumqi, 830011, China
- ⁵ College of Public Health, Xinjiang Medical University, Urumqi, 830011, China
- ⁶ Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, 211166, China
- [†] These authors contributed equally.
- * Correspondence: Email: wlei81@126.com (L. Wang), zhihangpeng@njmu.edu.cn (Z. Peng).

Abstract: The novel coronavirus disease 2019 (COVID-19) infection broke out in December 2019 in Wuhan, and rapidly overspread 31 provinces in mainland China on 31 January 2020. In the face of the increasing number of daily confirmed infected cases, it has become a common concern and worthy of pondering when the infection will appear the turning points, what is the final size and when the infection would be ultimately controlled. Based on the current control measures, we proposed a dynamical transmission model with contact trace and quarantine and predicted the peak time and final size for daily confirmed infected cases by employing Markov Chain Monte Carlo algorithm. We estimate the basic reproductive number of COVID-19 is 5.78 (95%CI: 5.71–5.89). Under the current intervention before 31 January, the number of daily confirmed infected cases is expected to peak on around 11 February 2020 with the size of 4066 (95%CI: 3898–4472). The infection of COVID-19 might be controlled approximately after 18 May 2020. Reducing contact and increasing trace about the risk population are likely to be the present effective measures.

Keywords: novel coronavirus; dynamical model; Markov Chain Monte Carlo; peak time; basic reproduction number

1. Introduction

On 31 December 2019, the novel coronavirus disease 2019 (COVID-19) infection broke out in Wuhan, China. As shown in Figure 1, the reported number of COVID-19 cases gradually increases every day, and the cumulative number of confirmed cases and death cases has rapidly multiplied since 18 January 2020. From the point of view for the spread of COVID-19, the epidemic situation is severe and spreads to all over the country (the mainland China) with the arrival of Chunyun (see the Figure 2). Up to 20 January 2020, the confirmed infected cases have been found in 31 provinces of mainland China. By the end of 4 February 2020, there have been 24,363 confirmed cases of COVID-19 infections in mainland China, including 491 deaths and 892 hospital released cases, and 23,260 suspected cases under follow-up [1]. So far, in addition to China, the United States, Thailand, Japan, South Korea and other countries have confirmed COVID-19 cases.



Figure 1. The data about cumulative number of confirmed cases and deaths of COVID-19 from Jan 10 to Feb 4 2020, in mainland China. The source of data is from Chinese Center for Disease Control and Prevention [1].

Since the 21st century, the global epidemics of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) have been caused by human coronavirus, with alarming morbidity and mortality [2]. The novel coronavirus, occurring in Wuhan, is a novel coronavirus strain that has never been found in humans [3], which is defined as the seventh member of the family of coronaviruses that infect humans [4]. Within a few weeks, the infection of COVID-19 has caused a global awareness of the public health emergency [5].

Recently, there have been some relative studies about the estimation for the infection of COVID-19, the risk of transmission *et al.* For example, Li *et al.* [6] reached a conclusion that human-to-human transmission has occurred among close contacts since the middle of December 2019, and they estimated the epidemic doubling time and the basic reproductive number. Their findings [7] were consistent with person-to-person transmission of COVID-19 in hospital and family settings, and the reports

of infected travelers in other geographical regions. Nishiura *et al.* [8] used a spatial back-calculation method and analyzed thirteen exported cases to estimate the cumulative incidence of COVID-19 cases in China in real time. Wu *et al.* [9] adopted a metapopulation SEIR model to simulate the epidemic across the many affected cities in mainland China. Zhao *et al.* [10] considered the under-reporting was likely to have resulted in 469 unreported cases during the first half of January 2020 and estimated the basic reproduction number of COVID-19 at 2.56.



Figure 2. The geographic information map about cumulative number of confirmed cases of COVID-19 from 20 January to 31 January 2020, in mainland China. The source of data is from National Health Commission of the People's Republic of China [11].

2. Objectives and methods

2.1. Mathematical model

In this paper, by taking into account the contact trace and then quarantine measure, we propose a compartmental model to simulate the data on the number of cumulative confirmed cases from 10 January 2020 to 4 February 2020 (in mainland China) and predict the trend of the infection.

We stratify the total population as the susceptible (not traced) denoted by S, the latent (not traced) denoted by E, the infectious (not traced) denoted by I, the uninfected but traced (these individuals stay at home and take self isolation) denoted by S_T , the infected and traced denoted by E_T , the infectious and quarantined (traced) denoted by I_q , the recovered (through self-isolation or treatment in hospital

settings) denoted by R and the dead D.

10

Because in the whole country, most population is impossible to contact the COVID-19 except (for) the population from Hubei province, i.e., a certain fraction of the population is protected from the COVID-19. In this paper, we assumed that only a small proportion of the whole population is susceptible. The date, 10 January 2020, is the starting date for the dynamical model (i.e., t = 0). We do not consider the birth and death for the total population because the time of the study is short (that is, the total population $N(t) = S(t) + E(t) + I(t) + S_T(t) + E_T(t) + I_a(t) + R(t) + D(t) \equiv S_0$. Let β is the contact rate, ϕ is the probability of infection and the trace rate is ρ . Because the latent individuals have the strongly infectious ability [6], however, no stronger than the ability of infectious individuals. So, we use ε to denote the proportion of infectious ability in latent individuals relative to that in infectious individuals. Then the variation of the susceptible S consists of three parts: the uninfected but traced $\rho(1-\phi)\beta S(I+\varepsilon E)$, the infected and traced $\rho\phi\beta S(I+\varepsilon E)$ and the infected but not traced $(1 - \rho)\phi\beta S(I + \varepsilon E)$. The infected but not traced move to the latent compartment E, while the infected and traced move to E_a , and the uninfected but traced move to S_T . The latent individuals no matter whether they were quarantined are becoming the infectious with the rate α . The infectious but not quarantined individuals are sent to some hospital settings (i.e., they move to I_a) with the rate $\theta \gamma$. The infectious but not quarantined individuals are dead with the rate $(1 - \theta)\eta\gamma$. The infectious and quarantined individuals are dead with the rate $\eta_q \gamma_q$. The infectious but not quarantined individuals recover with the rate $(1 - \theta)(1 - \eta)\gamma$. The infectious and quarantined individuals recover with the rate $(1 - \theta_a)\gamma$. Given the model structure (Figure 3), we obtain that the following dynamical model described by eight differential equations.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{-\rho(1-\phi)\beta S(I+\varepsilon E)}{\mathrm{uninfected}} \frac{-\rho\phi\beta S(I+\varepsilon E)}{\mathrm{infected}} \frac{-(1-\rho)\phi\beta S(I+\varepsilon E)}{\mathrm{infected}}$$
release of those traced but uninfected
$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{(1-\rho)\phi\beta S(I+\varepsilon E)}{\mathrm{uninfected}} - \frac{\alpha E}{\mathrm{d}E},$$
infected but not traced becoming infectious are not quarantined
$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\alpha E}{\alpha E} - \frac{(1-\theta)(1-\eta)\gamma I}{\theta \gamma I}$$
becoming infectious are not quarantined and recover
$$- \frac{(1-\theta)\eta\gamma I}{\theta \gamma I} - \frac{\theta\gamma I}{\theta \gamma I},$$
not quarantined and recover quarantined and recover
$$\frac{\mathrm{d}S_T}{\mathrm{d}t} = \frac{\rho(1-\phi)\beta S(I+\varepsilon E)}{\theta \beta S(I+\varepsilon E)} - \frac{\omega S_T}{\theta \gamma I},$$
infected but traced release of those traced but uninfected
$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\rho(1-\phi)\beta S(I+\varepsilon E)}{\theta \beta S(I+\varepsilon E)} - \frac{\omega S_T}{\theta \gamma I},$$
infected but traced release of those traced but uninfected
$$\frac{\mathrm{d}E_T}{\mathrm{d}t} = \frac{\rho \phi \beta S(I+\varepsilon E)}{\theta \beta S(I+\varepsilon E)} - \frac{\omega E_T}{\theta \gamma I},$$
infected and traced traced contacts becoming infectious and quarantined
$$\frac{\mathrm{d}I_q}{\mathrm{d}t} = \frac{\alpha E_T}{\theta \gamma I} + \frac{\theta \gamma I}{\theta \gamma I},$$
infected and traced traced contacts becoming infectious and quarantined
$$\frac{\mathrm{d}I_q}{\mathrm{d}t} = \frac{\alpha E_T}{(1-\eta)\gamma I} + \frac{\eta \gamma \eta \eta}{\eta \gamma I},$$
reconsidered and recover quarantined and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\eta)\gamma I} + \frac{\eta \gamma \eta \eta}{\eta \gamma I},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\eta)\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \gamma I},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \gamma I},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \gamma I},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \gamma \eta \gamma I},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \gamma \eta \eta \eta},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta\gamma I} + \frac{\eta \eta \gamma \eta}{\eta \eta \eta \eta \eta},$$
reconsidered and dead
$$\frac{\mathrm{d}D}{\mathrm{d}t} = \frac{(1-\theta)\eta\gamma I}{(1-\theta)\eta \eta \eta \eta \eta \eta \eta \eta}$$

Mathematical Biosciences and Engineering

Volume 17, Issue 4, 3052-3061.

the interpretations of all parameters are listed in Table 1.

The basic reproduction number for model (2.1) is defined by $R_0 = (1 - \rho)\phi\beta(\frac{\varepsilon}{\alpha} + \frac{1}{\gamma})S_0$ according to the concepts of next generation matrix and reproduction number presented in [12, 13].



Figure 3. Flow diagram of model (2.1).

2.2. Data sources and parameter estimation

We collected the data on the cumulative number for the confirmed infected cases and death cases from 10 January 2020 (as the starting date for the dynamical model, i.e., t = 0) to 4 February 2020 from the National Health Commission of the People's Republic of China [1].

The mean incubation period for COVID-19 was 5.2 days [6]. The mean latent period for COVID-19 was not clear, so we take $1/\alpha = 5.2$. The mean hospital stay was 12.5 days $(1/\gamma_q = 12.5)$ [14]. Suspected cases can be released after 14 days of follow-up [15], so we take $\omega = 1/14$. The remaining parameters in the model need to be estimated. We model cumulative cases as a Poisson-distributed random variable because the Poisson distribution describes the number of observed events in an interval of time. We calibrate the model by sampling from the posterior distribution of parameter vector $\theta | \mathbf{y} = \{\beta, \phi, \rho, \varepsilon, \theta, \eta, \eta_q, \gamma\} | \mathbf{y}$, where vector \mathbf{y} is derived from $\frac{d}{dt} Y(t) = \alpha E_T + \theta \gamma I$ and Y(t) denotes the reported cumulative cases. We conduct sampling via Markov Chain Monte Carlo using a Metropolis-Hastings acceptance rule. The posterior density is

$$f_{\Theta|\mathbf{y}}(\theta|\mathbf{y}) = \prod_{T} \mathcal{L}(Y(t)|\theta) f_{\Theta}(\theta).$$

The likelihood $\mathcal{L}(Y(t)|\theta)$ describes the probability of the data with a Poisson distribution. The prior density $f_{\Theta}(\theta)$ is the joint probability of eight univariate priors. We consider that β , ϕ , ρ , ε , θ , η , η_q and γ are distributed according to $\mathcal{U}(0, 1)$. The program was implemented in R version 3.6.0 *deBInfer* package [16]. We sampled from 100,000 MCMC iterations and discarded the first 90,000 samples as a burn-in period. On the basis of these 10,000 samples, the point estimates and 95% confidence intervals for those parameters were calculated. The initial conditions can be obtained from the actual data as R(0) = 2, $I_q(0) = 41$, D(0) = 1. We left S(0), E(0), I(0), $S_T(0)$ and $E_T(0)$ as estimated parameters. The

data from 10 January to 31 January were used for modeling and the one from 1 February to 4 February were used for validation. The results are shown in Table 1.

Parameters	Value (95% CI)	Dimension	Interpretation	Source
β	$2.6 \times 10^{-6} (2.4 \times 10^{-6}, 2.9 \times 10^{-6})$	dimensionless	Contact rate	MCMC
ϕ	0.0923 (0.0917,0.0963)	dimensionless	Probability of transmission per contact	MCMC
ρ	0.0171 (0.0151,0.0198)	dimensionless	Traced rate	MCMC
ε	0.914 (0.860,0.939)	dimensionless	Infectious ability of the latent	MCMC
$\frac{1}{\alpha}$	5.2	day	Mean incubation period	[6]
θ	0.0952 (0.0890,0.0977)	day ⁻¹	Quarantined rate of the infectious	MCMC
η_q	0.00161 (0.0145,0.0177)	day ⁻¹	Disease induced death rate for the infectious and quarantined	MCMC
η	0.00392 (0.00352,0.00427)	day ⁻¹	Disease induced death rate for the infectious not quarantined	MCMC
γ	0.071 (0.063,0.078)	day ⁻¹	Remove rate for the infected not quarantined	MCMC
γ_q	1/12.5	day ⁻¹	Remove rate for the infected and quarantined	[14]
ω	1/14	day ⁻¹	Release rate for the uninfected but traced	[15]
S(0)	1326000(1324000,1328000)	Humans	Initial susceptible population	MCMC
E(0)	3349 (2764,4327)	Humans	Initial latent population	MCMC
<i>I</i> (0)	28 (0,102)	Humans	Initial infectious population	MCMC
$S_T(0)$	509 (344,674)	Humans	Initial uninfected but traced population	MCMC
$E_T(0)$	80 (25,97)	Humans	Initial infected and traced population	MCMC

 Table 1. Parameter values and interpretations.

3. Results and discussion

Based on the fitted model, we estimate the based reproductive number as $R_0 \approx 5.78$ (95%CI: 5.71– 5.89), which means each infected person might infect an average of 5.78 people during the period of infection. The daily confirmed data from 1 February to 4 February were used to verify the model, which showed good prediction effect. Using the estimated parameter values, we can predict the trend of COVID-19 infection (see Figure 4). Under the current intervention (before 31 January)– intensive intensive contact tracing followed by quarantine and isolation, the number of daily confirmed infected individuals is expected to peak on around 11 February with a peak size of 4066 (95%CI: 3898–4472) daily confirmed infected individuals. Starting from 18 May, there will be no new confirmed cases, the cumulative number of cases will not increase, and the outbreak will be over.

We plot the cumulative infected cases and daily infected cases with varying contact rate β and traced rate ρ to examine the possible impact of enhanced interventions on COVID-19 infection (see Figure 5). As shown in Figure 5(a), reducing contact rate persistently decreases the peak value and may delay the peak. To be specific, it shows that reducing contact β by 0.5 times or 0.25 times will delay the peak by 28 days or 113 days, and lead to reduction of peak value in the number of daily infected cases by 54% or by 92%. Meanwhile, increasing the trace proportion of susceptible people decreases the peak value but may not delay the peak, as shown in Figure 5(b). This result indicates that interventions to reduce contact rate come first in control of the outbreak.



Figure 4. Estimation of the cumulative infected cases is shown in (a) and that of daily infected cases data is shown in (b).



Figure 5. Sensitivity analyses with respect to the contact rate β (a) and the traced rate ρ (b) on the new daily infected cases.

We estimated that the basic reproductive number for COVID-19 was 5.78 (95%CI: 5.71–5.89), which is higher than the basic reproductive number in previous studies, (for instance, 2.68 (95%CI: 2.47–2.86) in [9], 2.56 (95% CI: 2.49–2.63) in [10].) This may be due to the fact that there are some under-reported new cases in the early outbreak. We did not considered this factor and directly fitted the

data without adjustment, then the basic reproductive number may be overestimated, which is in line with previous estimated [17] (6.47 (95%CI: 5.71–7.23)).

Based on our predictions, around 11 February, the number of daily confirmed infected individuals may reach a peak, and the outbreak will appear an inflection point. Although the inflection point may occur in the near future, it is important to keep it under control to avoid a new wave or outbreak, as secondary outbreaks tend to be more severe.

Reducing contact rate persistently decreases the peak value and may delay the peak, which is in line with the government's call for a reduction in unnecessary gathering. The delay in peak could buy time for new drug and vaccine development. Though increasing the trace proportion of susceptible people may not delay the peak, it is still a necessity to reduce the peak value. The resumption of work after the Spring Festival may bring potential risks to the control of the epidemic. It is suggested to pay close attention to the changes of the epidemic and choose the time of resumption carefully.

Because the daily data could contain more noise during the early stage, we use the cumulative incidence data rather than the daily incidence to make the simulations. Cumulative cases eliminate data fluctuations and will lead to a narrow 95% confidence interval. This is a limitation of the estimation in our paper. In our future work, we would use more detailed and accurate daily data to improve the estimation. A model is always an abstraction of reality and never a mirror of reality. This paper is only a mathematical results for reference. As more data on outbreaks follow, more accurate mathematical models will emerge.

4. Conclusions

In this paper, it is demonstrated that the current trends and future prediction of COVID-19 can be investigated by an appropriate mathematical model. Our model predicts that the daily confirmed COVID-19 cases will continuously increase in the next several days and reach the peak around 11 February. The outbreak is expected to end around 18 May. The basic reproductive number is estimated to be 5.78 (95%CI: 5.71–5.89) which means this outbreak is severe. At present, the most effective control is to reduce the contact rate which can not only reduce the peak value but also delay the peak.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China (11961071, 82041026, 81673275, 91846302), the National S&T Major Project Foundation of China (2018ZX10715002, 2018ZX10713001).

Conflict of interest

The authors declare no conflict of interest.

References

1. *Chinese Center for Disease Control and Prevention*, Report about 2019-nCoV, 2020. Available from: http://www.chinacdc.cn/yyrdgz/202001/P020200128523354919292.pdf.

- 2. C. I. Paules, H. D. Marston, A. S. Fauci, Coronavirus infections-more than just the common cold, *JAMA*., **323** (2020), 707–708.
- 3. J. F. Chan, K. H. Kok, Z. Zhu, H. CHu, K. K. To, S. Yuan, et al., Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, *Emerg. Microbes Infec.*, **9** (2020), 221–236.
- 4. N. Zhu, D. Y. Zhang, W. L. Wang, X. W. Li, B. Yang, J. Song, et al., A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.*, **382** (2020), 727–733.
- 5. C. Wang, P. W. Horby, F. G. Hayden, G. Gao, A novel coronavirus outbreak of global health concern, *Lancet.*, **395** (2020), 470–473.
- Q. Li, X. H. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus infected pneumonia, *N. Engl. J. Med.*, 382 (2020), 1199– 1207.
- 7. J. F. Chan, S. F. Yuan, K. H. Kok, K. K. To, H. Chu, J. Yang, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet.*, **395** (2020), 514–523.
- 8. H. Nishiura, S. M. Jung, N. M. Linton, R. Kinoshita, Y. Yang, K. Hayashi, et al., The Extent of Transmission of Novel Coronavirus in Wuhan, *J. Clin. Med.*, **9** (2020), 330.
- J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *Lancet.*, 395 (2020), 689–697.
- S. Zhao, S. S. Musa, Q. Lin, J. Ran, G. Yang, W. Wang, et al., Estimating the Unreported Number of Novel Coronavirus 2019-nCoV Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak, J. Clin. Med., 9 (2020), 388.
- 11. Situation report of the pneumonia cases caused by the novel coronavirus, released on January 31, 2020 by the National Health Commission of the People's Republic of China (in Chinese), 2020. Available from: http://www.nhc.gov.cn/xcs/yqtb/202002/84faf71e096446fdb1ae44939ba5c528.shtml.
- O. Diekmann, J. A. P. Heesterbeek, J. A. J. Metz, On the definition and the computation of the basic reproduction ratio R0, in models for infectious diseases in heterogeneous populations, *J. Math. Biol.*, 28 (1990), 365–382.
- 13. P. Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Math. Biosci.*, **180** (2002), 29–48.
- 14. Central People's Government of the People's Republic of China, Available from: http://www.gov.cn/xinwen/2020-02/01/content\$_\$5473742.htm.
- 15. National Health Commission of the People's Republic of China, Available from: http://www.nhc.gov.cn/xcs/kpzs/202001/a4499dd30c484457b1c05e3590fae794.shtml.
- 16. P. H. Boersch-Supan, S. J. Ryan, L. R. Johnson, deBInfer: Bayesian inference for dynamical models of biological systems in R, *Methods Ecol. Evol.*, **8** (2017), 511–518.

 B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. xiao, et al., Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions, *J. Clin. Med.*, 9 (2020), 462.



© 2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)