



Evaluation and prediction of the COVID-19 variations at different input population and quarantine strategies, a case study in Guangdong province, China

Zengyun Hu^{a,b,1}, Qianqian Cui^{c,1}, Junmei Han^d, Xia Wang^e, Wei E.I. Sha^f, Zhidong Teng^{g,*}

^a State Key Laboratory of desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

^b Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China

^c School of Mathematics and Statistics, Ningxia University, Yinchuan, Ningxia 750021, China

^d The third Hospital of Jincheng, Jincheng 048000, China

^e School of Mathematics and Information Science, Shaanxi Normal University, Xian 710119, China

^f College of Information Science Electronic Engineering, Zhejiang University, Hangzhou 310027, China

^g College of Mathematics and System Sciences, Xinjiang University, Urumqi, Xinjiang 830046, China

ARTICLE INFO

Article history:

Received 9 March 2020

Received in revised form 30 March 2020

Accepted 2 April 2020

Keywords:

Coronavirus disease 2019 (COVID-19)

Guangdong province

SEIRQ model

Population migration

Quarantine strategies

Scenario analysis

ABSTRACT

In this study, an epidemic model was developed to simulate and predict the disease variations of Guangdong province which was focused on the period from Jan 27 to Feb 20, 2020. To explore the impacts of the input population and quarantine strategies on the disease variations at different scenarios, four time points were assumed as Feb 6, Feb 16, Feb 24 and Mar 5 2020. The major results suggest that our model can well capture the disease variations with high accuracy. The simulated peak value of the confirmed cases is 1002 at Feb 10, 2020 which is mostly close to the reported number of 1007 at Feb 9, 2020. The disease will become extinction with peak value of 1397 at May 11, 2020. Moreover, the increased numbers of the input population can mainly shorten the disease extinction days and the increased percentages of the exposed individuals of the input population increase the number of cumulative confirmed cases at a small percentage. Increasing the input population and decreasing the quarantine strategy together around the time point of the peak value of the confirmed cases, may lead to the second outbreak.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

In the past more than fifty days, the Chinese government and all the people in China fought against the COVID-19 disease, and employed extremely and rigorously controlling measures to protect the people avoiding the infection of the COVID-19 virus, such as the lockdown of many cities in Hubei province (e.g. Wuhan city) and initiating a top-level emergency response to rein in the outbreak of the epidemic associated with COVID-19 in the other provinces of China. With these strong and effective strategic policies, the number of the daily new confirmed COVID-19 cases was significantly decreased from the largest value of 3887 at Feb 4, 2020 to the value of 648 at Feb 22, 2020 from the National Health

Commission of the People's Republic of China (<http://www.nhc.gov.cn/>)(excluding more than 140,000 cases at Feb 12, 2020).

Recently, more and more researchers have been paid large attention on the COVID-19 variations in China, such as detecting the clinical characteristics (Guan et al., 2020), estimating the spreading characteristics (Wu et al., 2020; Zhao et al., 2020a,b) and exploring the effects of the control strategies (Chinazzi et al., 2020; Huang et al., 2020; Lin et al., 2020; Tang et al., 2020a,b). The individual behavioural reaction and governmental actions played a key role in controlling the spread of the COVID-19 outbreak for the public health in the world, e.g. holiday extension, travel restriction, hospitalisation and quarantine (Chinazzi et al., 2020; Lin et al., 2020).

Until now, there are only few researches about the effects of different population migration and quarantine strategies on the COVID-19 variations in China. Guangdong province has the largest gross domestic product (GDP) than the other provinces in China. Moreover, according to the present COVID-19 variations and control strategies, the Guangdong province adjusted the

* Corresponding author.

E-mail address: zhidong@xju.edu.cn (Z. Teng).

¹ Qianqian Cui and Zengyun Hu contributed equally to this work.

emergency response level of epidemic prevention and control from the first level response to the second level at Feb 24, 2020 (<http://www.gd.gov.cn/>). More and more workers will come back to Guangdong province from other provinces. Thereby, we choose Guangdong province as a case study to explore the effects of the population migration and quarantine strategies on the COVID-19 variations. Based on the present rigorous and extreme control measures in Hubei province, input population from Hubei province are not considered.

In this study, we focus on the input population and quarantine strategies influencing on the disease variations, including the peak values of the cumulative confirmed cases, the daily new increased confirmed cases and the confirmed cases, and the corresponding times. The organization of this paper is as follows. In the next section, the establishment of SEIRQ model, data and methodology are illustrated. In “Result” section, the input population and quarantine strategies at different scenarios are investigated which are our main results. A brief discussion is provided in “Discussion” section.

SEIRQ model, data and methodology

SEIRQ model

In this study, according to the characteristics of the COVID-19 transmission, the whole population at time t is divided into seven compartments which include the susceptible individuals $S(t)$, exposed individuals $E(t)$, infectious individuals $I(t)$, removed individuals $R(t)$, quarantined susceptible individuals $S_q(t)$, quarantined exposed individuals $E_q(t)$ and quarantined infectious individuals $I_q(t)$. The COVID-19 disease is transmitted from $I(t)$ to $S(t)$ with the incidence rate of β , and from $E(t)$ to $S(t)$ with the incidence rate of $\sigma\beta$, respectively. The susceptible individuals $S(t)$ is partly quarantined with the rate of $q_1(t)$. We assume that exposed individuals $E(t)$ and quarantined exposed individuals $E_q(t)$ are transmitted to infectious individuals $I(t)$ and quarantined infectious individuals $I_q(t)$ with the same transition rate of ν . The quarantined rates of exposed individuals $E(t)$ and infectious individuals $I(t)$ are $q_2(t)$ and q_3 . The death rate induced by the COVID-19 disease is α in both infectious individuals $I(t)$ and quarantined infectious individuals $I_q(t)$ which removed to the removed individuals $R(t)$. $\gamma(t)$ is the recovery rate of quarantined infected individuals $I_q(t)$ which is the mainly part of removed individuals $R(t)$.

Moreover, based on the population migration, we assume that the input population and output population have constant numbers. Susceptible individuals $S(t)$, exposed individuals $E(t)$ and infectious individuals $I(t)$ have their respective input individuals of $p_1(t)A(t)$, $p_2(t)A(t)$ and $p_3(t)A(t)$, and the parameters $p_i(t)$, $i = 1, 2, 3$ are the rates of susceptible individuals, exposed individuals, infectious individuals in the total input number of $A(t)$

from other provinces. The output population are B_1 , B_2 and B_3 for the susceptible individuals $S(t)$, exposed individuals $E(t)$, infectious individuals $I(t)$. The COVID-19 disease transmission and population migration are demonstrated by Fig. 1 in details.

The SEIRQ epidemic model can be described by the following system of ordinary differential equations

$$\begin{cases} S' = p_1(t)A(t) - \beta SI - \sigma\beta SE - q_1(t)S - B_1, \\ E' = p_2(t)A(t) + \beta SI + \sigma\beta SE - \nu E - q_2(t)E - B_2, \\ I' = p_3(t)A(t) + \nu E - q_3 I - \alpha I - B_3, \\ S_q' = q_1(t)S \\ E_q' = q_2(t)E - \nu E_q \\ I_q' = q_3 I + \nu E_q - \gamma(t)I_q - \alpha I_q \end{cases} \quad (1)$$

where the prime (') denotes the differentiation with respect to time t . Here, parameters $0 < \beta, \nu, \gamma(t), \alpha < 1$ and the quarantined rates $0 \leq q_1(t), q_2(t), q_3 \leq 1$. All the initial values of different individual groups: $S(0), E(0), I(0), R(0), S_q(0), E_q(0), I_q(0)$ are non-negative.

Data

In this study, the COVID-19 cases of Guangdong province, Hubei province and mainland China are obtained from the Health Commission of Guangdong Province (<http://wsjkw.gd.gov.cn/>), the Health Commission of Hubei Province (<http://wjw.hubei.gov.cn/>), and the National Health Commission of the People's Republic of China (<http://www.nhc.gov.cn/>), respectively. The data are from Jan 20, 2020 to present which include the number of the cumulative confirmed cases, the number of the confirmed cases, the number of the cumulative cured cases and the number of cumulative death cases. The numbers of the total population of Guangdong Province, Hubei Province and mainland China are employed at the end of 2018 from the National Bureau of Statistics (<http://www.stats.gov.cn/>).

The numbers of the input and output population from Hubei province and the other provinces of mainland China to Guangdong province are from the Baidu migration (<http://qianxi.baidu.com/>). These data are covering the period of Jan 1, 2020 to Feb 20, 2020 which are employed to display the population migration variations from other provinces to Guangdong province. Because the input population from Hubei province to Guangdong province is significantly decreased from 26.86% of the total input population at Jan 26, 2020 to the 6.84% at Jan 27, 2020, for the Guangdong province, the starting date of the COVID-19 disease data is from Jan 27, 2020.

Methodology

In this study, for the COVID-19 variations, we focus on the cumulative confirmed cases and confirmed cases. The largest value of the cumulative confirmed cases means the total number of the

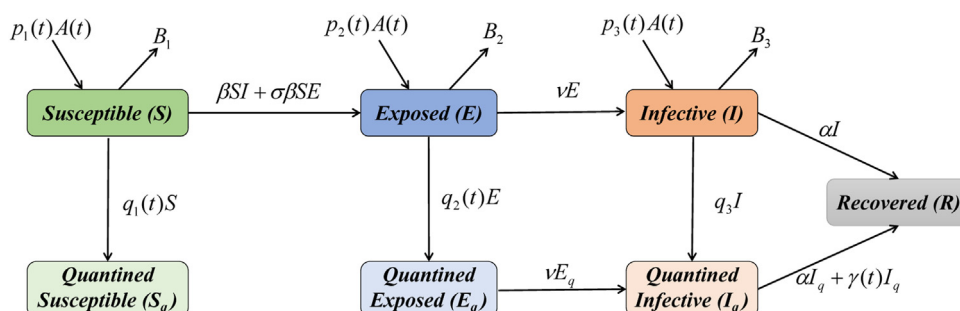


Figure 1. Flowchart of COVID-19 SEIRQ epidemic model.

population infected by the COVID-19 disease. The disease extinction time is defined as the day with no confirmed case which is the time of $I_q(t)=0$.

The initial values and parameters can be obtained from the Text methodology of the supplementary information. The baseline parameters noted as $(A, B, p_1, p_2, p_3, q_1, q_2, q_3, \alpha, \beta, \nu, \sigma, \gamma) = (A^*, B^*, p_1^*, p_2^*, p_3^*, q_1^*, q_2^*, q_3^*, \alpha^*, \beta^*, \nu^*, \sigma^*, \gamma^*)$ is obtained from the simulation result of the cumulative confirmed cases, the daily new confirmed cases, the confirmed cases and the recovered cases.

To compare with the baseline results, three aspects from the perspectives of the input population and quarantine strategies on the COVID-19 variations are analyzed: (1) aspect 1, effects of the input population at different scenarios; (2) aspect 2, effects of quarantine rates at different scenarios and (3) aspect 3, effects of both input population and quarantine rates at different scenarios. To evaluate the accuracy of our model, five statistical indices are applied, including the absolute error (AE), relative error (RE), mean absolute percentage error (MAPE), determinant coefficient R^{*2} which is the square of correlation coefficient R^* and distance between indices of simulation and observation (DISO) (Hu et al., 2016, 2019). The details are displayed in Text methodology of the supplementary information.

Result

Simulation and prediction of the COVID-19 disease variations

In this section, the variations of the COVID-19 in Guangdong province are simulated and predicted based on our SEIRQ model only considering the input population from the other provinces of China (excluding Hubei province). The simulated period are from Jan 27, 2019 to Feb 19, 2020. The parameter values and the initial values of our simulation and prediction are provided in Table 1. The performance is evaluated by the data from Feb 20, 2020 to Feb 23, 2020, and R^{*2} , AE, RE, RMSE, MPAE and DISO are employed to quantify the accuracy. The simulation and prediction results are displayed in Table 2 and Fig. 2.

Our model has the ability to simulate and to predict the COVID-19 variations with the very high accuracy (Table 2 and Fig. 2). Particularly, the determinant coefficients R^* of the cumulative confirmed cases, confirmed cases and recovered cases are highly to

0.9973, 0.9898 and 0.9934, respectively (Table 2). Very small estimations are obtained with the AE values of -5.33 , -2.63 and -3.38 for the cumulative cases, confirmed cases and recovered cases. The comprehensive accuracies of our model are quantitatively measured by the DISO with the values of 0.06, 0.11 and 0.17 for the cumulative cases, confirmed cases and recovered cases. For the validation at Feb 20, Feb 21, Feb 22 and Feb 23, 2020, the very small RE values of the cumulative confirmed cases, confirmed cases and recovered cases indicate that our model also has very high accuracies and it can be employed to predict the future variations of the COVID-19 disease (Table 2).

Moreover, the largest number of cumulative confirmed cases is 1397 at May 7, 2020 which indicates that the COVID-19 disease will become extinction after 102 days in Guangdong province (Fig. 2A, STable 1). The peak value time of daily new confirmed cases is Feb 1, 2020 which is highly agreement with the reported time at Jan 31, 2020 (Fig. 2B). For the confirmed cases, the peak value and the corresponding time are both obtained by our model with the simulated values of 1002 at Feb 10, 2020 and reported values of 1007 at Feb 9, 2020 (Fig. 2C). The number of the recovered cases will reach about 1400 which is consist with the future changes of the cumulative confirmed cases (Fig. 2D).

In order to further explore the forecasting accuracy of our model, we have been compared the forecasting result with the observed data prolonged 11 days from Feb 24, 2020 to Mar 4, 2020. The absolute values of RE (relative error) of the cumulative confirmed cases are smaller than 1% (Table 3). The corresponding figures also display that our model can capture the temporal variations in a relative longer period (see SFigure 1 in the supplementary information).

Effects of input population at different scenarios

The input population variations include the percentage changes p_2 of the exposed individuals and the number changes A of the input population which impact the disease on the peak value of the cumulative confirmed cases and the disease extinction time (Figs. 3 and 4). For the first time point $t_1 = 10$ (i.e. Feb 6, 2020), the days of disease extinction (DDE) are shortened to 78 days (i.e. Apr 13, 2020) and 69 days (i.e. Apr 4, 2020) at Sce 1: $(p_2, A) = (p_2^*, 1.5A^*)$ and Sce 2: $(p_2, A) = (p_2^*, 2A^*)$, and the maximum values of

Table 1
Parameter estimates for COVID-19 in Guangdong province.

Parameter	Definitions	Estimated value	Source
β	Transmission incidence rate	2.45×10^{-8}	Estimated
σ	The fraction of transmission incidence rate for exposed individuals	0.63	Estimated
α	Disease-induced death rate	0.00375	Estimated
ν	Transmission rate of exposed individuals to the infected class	0.183	Zhao et al., 2020a,b
$\gamma(t)$	Recovery rate	$0.008 + \frac{0.19}{(1+e^{5.0126-0.1846t})}$	Estimated
$q_1(t)$	Quarantined rate of susceptible individuals	0.28	Estimated
$q_2(t)$	Quarantined rate of exposed individuals	0.76	Estimated
q_3	Quarantined rate of infected individuals	0.89	Estimated
$A(t)$	Input number	86926	data
B_1	Output number	21356	data
p_1	The fraction of input population into susceptible class	0.9999927	Computed
p_2	The fraction of input population into exposed class	0.0000073	Computed
p_3	The fraction of input population into infected class	0	Assumed
Initial values	Definitions	Estimated value	Source
$N(0)$	Initial total population	113460000	GSY
$S(0)$	Initial susceptible population	113346174	Estimated
$E(0)$	Initial exposed population	31	Estimated
$I(0)$	Initial infected population	19	Estimated
$S_q(0)$	Initial quarantined susceptible population	113460	Estimated
$E_q(0)$	Initial quarantined exposed population	128	data
$I_q(0)$	Initial quarantined infected population	184	data
$R(0)$	Initial recovered population	4	data

Note: GSY: Guangdong Statistical Yearbook, 2019.

Table 2

Evaluation results of the simulation and prediction in Guangdong province.

Different cases	Simulation				Prediction			
	R^2	AE	MAPE (%)	DISO	20/2 RE (%)	21/2 RE (%)	22/2 RE (%)	23/2 RE (%)
Cumulative confirmed cases	0.9973	−5.33	2.54	0.06	−0.38	−0.45	−0.37	−0.37
Confirmed cases	0.9898	−2.63	3.86	0.11	2.68	1.51	0.81	7.07
Recovered cases	0.9934	−3.38	43.32	0.17	−2.09	−1.38	−3.75	−10.41

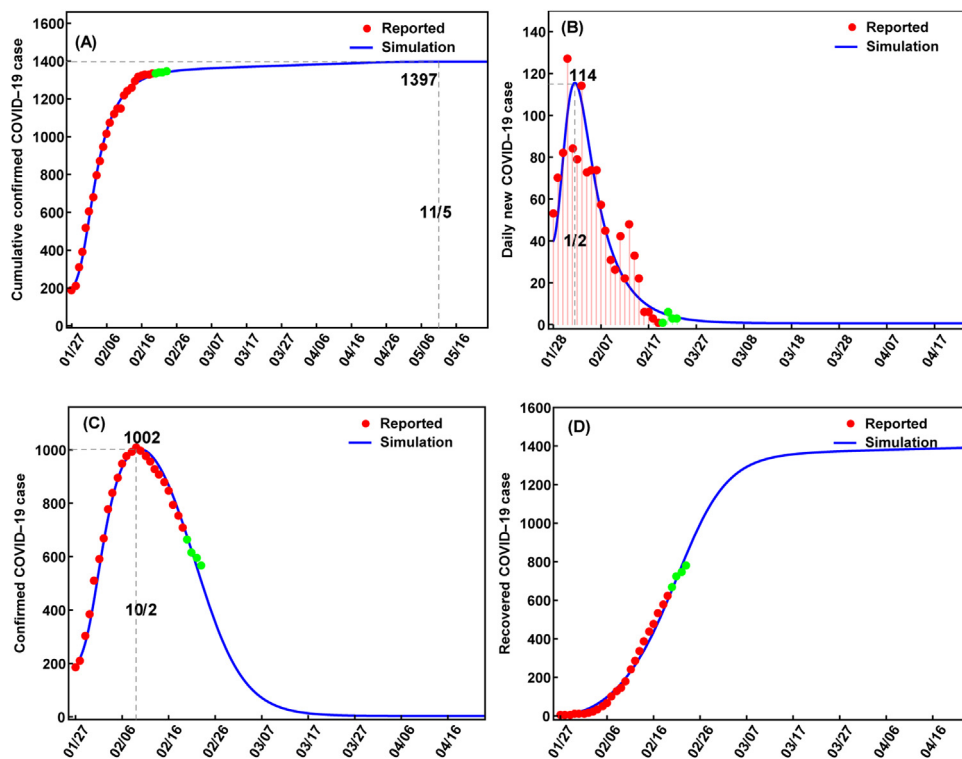


Figure 2. Simulation and prediction of the COVID-19 in Guangdong province. (A) cumulative confirmed cases; (B) daily new confirmed cases and (C) difference of increased confirmed cases. The initial values and parameters are $S(0) = 113346174$, $E(0) = 31$, $I(0) = 19$, $R(0) = 4$, $S_d(0) = 113460$, $E_d(0) = 128$, $I_d(0) = 184$, $A = 86926$, $B = 21356$, $p_1 = 0.9999927$, $p_2 = 0.0000073$, $p_3 = 0$, $q_1 = 0.28$, $q_2 = 0.76$, $q_3 = 0.89$, $\alpha = 0.00375$, $\beta = 2.45 \times 10^{-8}$, $\nu = 0.183$, $\sigma = 0.63$, $\gamma(t) = 0.008 + 0.19/(1 + e^{5.0126 - 0.1846t})$.

Table 3

Evaluation results of the prediction in Guangdong province.

RE (%)	24/2	25/2	26/2	27/2	28/2	29/2	1/3	2/3	3/3	4/3
Cumulative confirmed cases	−2.30	−0.41	0.12	0.20	0.25	0.37	0.40	0.49	0.58	0.66
Confirmed cases	−14.98	−19.21	−24.22	−26.74	−27.64	−30.81	−36.19	−35.94	−33.52	−34.68
Recovered cases	9.60	11.35	13.09	12.57	10.88	10.67	11.35	9.35	7.08	6.31

the cumulative confirmed cases (MVCCC) have the numbers of 1396 and 1397 [Fig. 3A, Supplementary table 1 (STable 1)]. For the confirmed cases, the peak values are nearly close to the baseline value with the number of 1003, and the corresponding times are same as the baseline value (STable 1). Moreover, the confirmed cases of Sce 1 and Sce 2 have the same variations as the baseline result with their early disease extinction that are consist with the variations of the cumulative confirmed cases (Fig. 2A and 3 A). For Sce 4, Sce 5, Sce 7 and Sce 8, compared with the baseline results, the DDE of these scenarios are 81 days (i.e. Apr 16, 2020), 59 days (i.e. Mar 25, 2020), 83 days (i.e. Apr 18, 2020) and 73 days (i.e. Apr 8, 2020), respectively which indicate the early extinction of COVID-19 (STable 1). The MVCCC of the four scenarios are larger than the

baseline result with the largest value (1448) in Sce: 8 (Fig. 3A, STable 1). For the confirmed cases, these scenarios are similar as these of the baseline results (Fig. 4A, STable 1).

For Sce 3: $(p_2, A) = (1.5p_2^*, A^*)$ and Sce 6: $(p_2, A) = (2p_2^*, A^*)$, the increased percentage of the exposed individuals only impacted the number of the cumulative confirmed cases with the values of 1422 and 1447, and the corresponding DDE have only small changes with 105 days for Sce 3 and 107 days for Sce 6 (Fig. 3A, STable 1). For the confirmed cases, they have the very similar variations as the baseline result in the peak value and the peak value time (Fig. 4A, STable 1). For the other three time points $t_1 = 20$, $t_1 = 28$ and $t_1 = 38$, the differences of the scenarios results are similar as the these of $t_1 = 10$. Moreover, for each scenario, the changes in the input

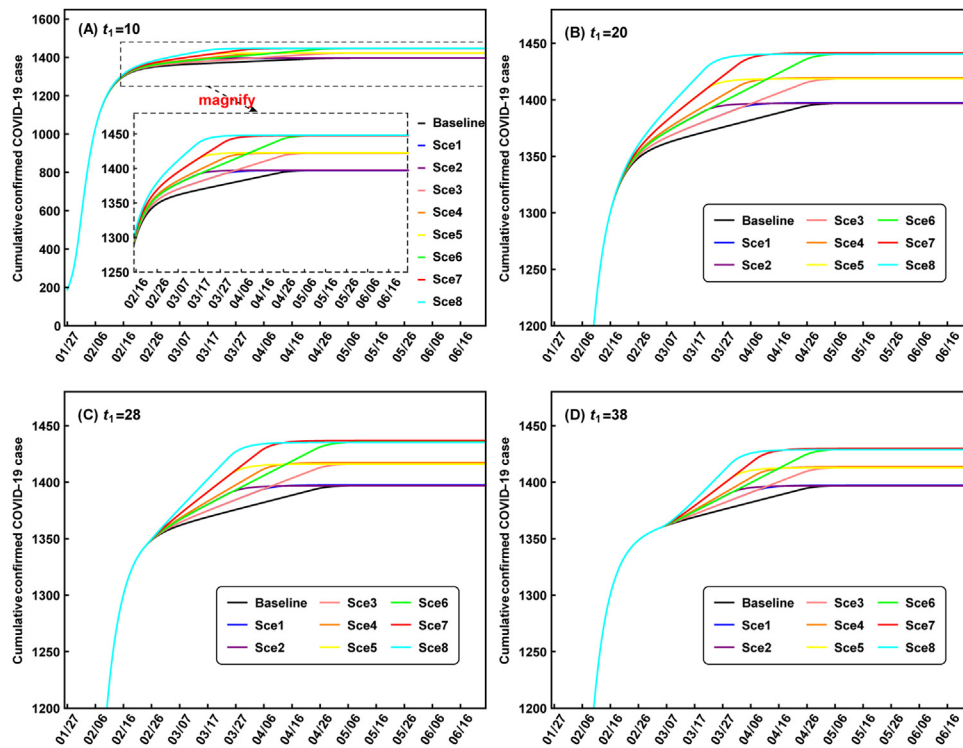


Figure 3. Scenarios results of input population impacting on the cumulative confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

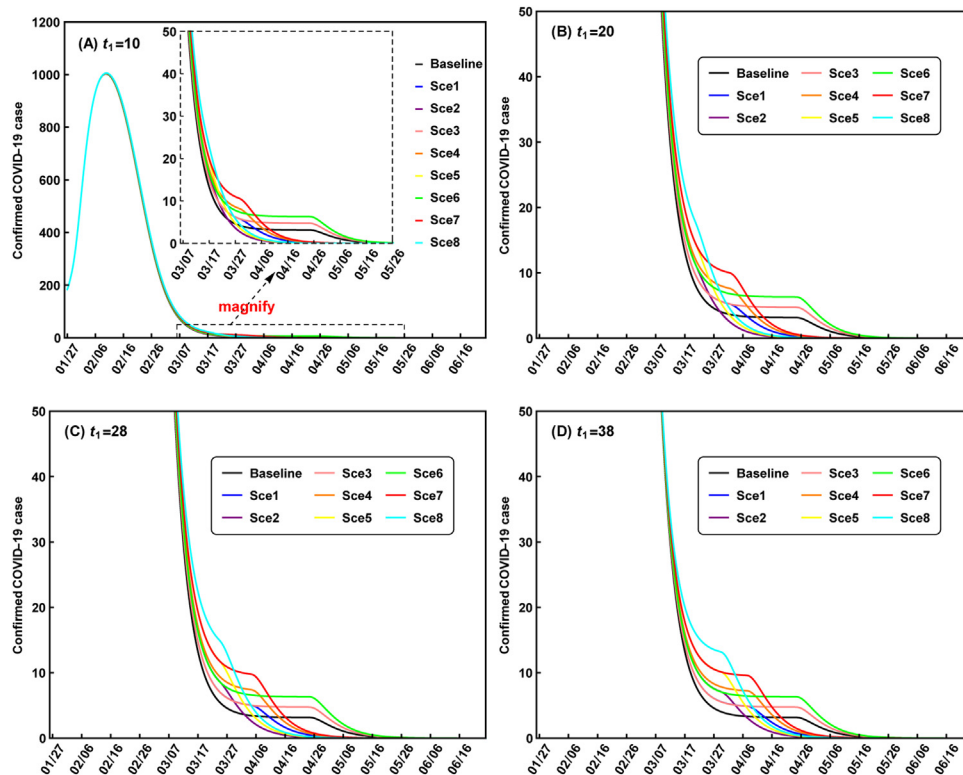


Figure 4. Scenarios results of input population impacting on the confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

population have the nearly same impacts on the disease variations among the four time points which display that the same input population strategies at different time points have no significant difference on the disease.

From the above analysis, it can be concluded that the increased numbers of the input population can mainly shorten the disease extinction days and the increased percentages of the exposed individuals of the input population increase the number of cumulative confirmed cases at a small percentage. Both the increased input population and the increased exposed individuals have no impacts on the peak values and peak value times of the confirmed cases.

Effects of quarantine rates at different scenarios

In this section, the effects of quarantine rates at six scenarios on the COVID-19 variations are displayed in Figs. 5 and 6. For the first time point $t_1 = 10$, Feb 6, 2020, Sce 1 (q_1, q_2) = ($0q_1^*$, $0q_2^*$) has significantly negative impacts on the COVID-19 variations with the disease outbreak again which suggest the very high risks appear at the quarantine strategy of Sce 1 (Figs. 5 and 6A). Specifically, the confirmed cases reaches its first peak value as the baseline result at Feb 10, 2020, and then the number is decreased close to 97 at Mar 14, 2020. A sharp increase is detected to the second peak value of the confirmed cases with the number of 1016704 at 165 days (Fig. 6A). The disease will become extinction after 361 days with the MVCCC dramatically reaching to more than 9 million (Figs. 5A and STable 2). Sce 2: (q_1, q_2) = ($0q_1^*$, $0.5q_2^*$) and Sce 3: (q_1, q_2) = ($0q_1^*$, q_2^*) have the similar impacts on the disease variations with the largest cumulative confirmed values of 1444 at 110 days (i.e. May 15, 2020), and 1416 at 105 days (i.e. May 10, 2020). The DDE and MVCCC of Sce 4: (q_1, q_2) = ($0.5q_1^*$, $0.5q_2^*$), Sce 5: (q_1, q_2) = ($0.5q_1^*$, q_2^*) and Sce 6: (q_1, q_2) = (q_1^* , $0.5q_2^*$) are agreement with the baseline results (STable 2). These three

scenarios have very weak influences on the confirmed case variations compared with the baseline result (Fig. 6A, STable 2).

For the other three time points, Sce 1: (q_1, q_2) = ($0q_1^*$, $0q_2^*$) increased the MVCCC and prolonged the DDE with the values of 1430 at 123 days (i.e. May 28, 2020), 1416 at 115 days (i.e. May 20, 2020) and 1409 at 112 days (i.e. May 17, 2020) (STable 2). The disease variations of the other scenarios are agreement with the baseline results which indicates the weak impacts of these scenarios (Fig. 5A, STable 2).

Moreover, we also explored that the second outbreak of the disease appears when both the values of q_1 and q_2 are nearly close to zero, such as (q_1, q_2) = ($0.01q_1^*$, $0.01q_2^*$), ($0q_1^*$, $0.05q_2^*$) at $t_1 = 10$, and (q_1, q_2) = ($0q_1^*$, $0q_2^*$) at $t_1 = 11$ (Fig. 7, STable 3). This suggests that no quarantine or very weak quarantine on the susceptible individuals and exposed individuals before the days of the peak values of the confirmed cases may lead to the disease outbreak again.

Effects of both input population and quarantine rates at different scenarios

The impact results of both the input population and quarantine rates on the COVID-19 disease are displayed in Fig. 8, 9 and STable 3. According to the results in "Effects of input population at different scenarios" and "Effects of quarantine rates at different scenarios" sections, the second outbreak of the disease are obtained in the scenarios with no or very weak quarantine strategy. Therefore, Figs. 8 and 9 only provide the COVID-19 disease variations of the scenarios with second outbreak, and the disease variations in other scenarios are not provided. STable 4 provides the results of all the scenarios.

For time point $t_1 = 10$, Sce 1: (p_2, A, q_1, q_2) = ($1.5p_2^*$, $1.5A^*$, $0q_1^*$, $0q_2^*$), Sce 2: (p_2, A, q_1, q_2) = ($1.5p_2^*$, $2A^*$, $0q_1^*$, $0q_2^*$), Sce 7: (p_2, A, q_1, q_2) = ($2p_2^*$, $1.5A^*$, $0q_1^*$, $0q_2^*$) and Sce 8: (p_2, A, q_1, q_2) = ($2p_2^*$,

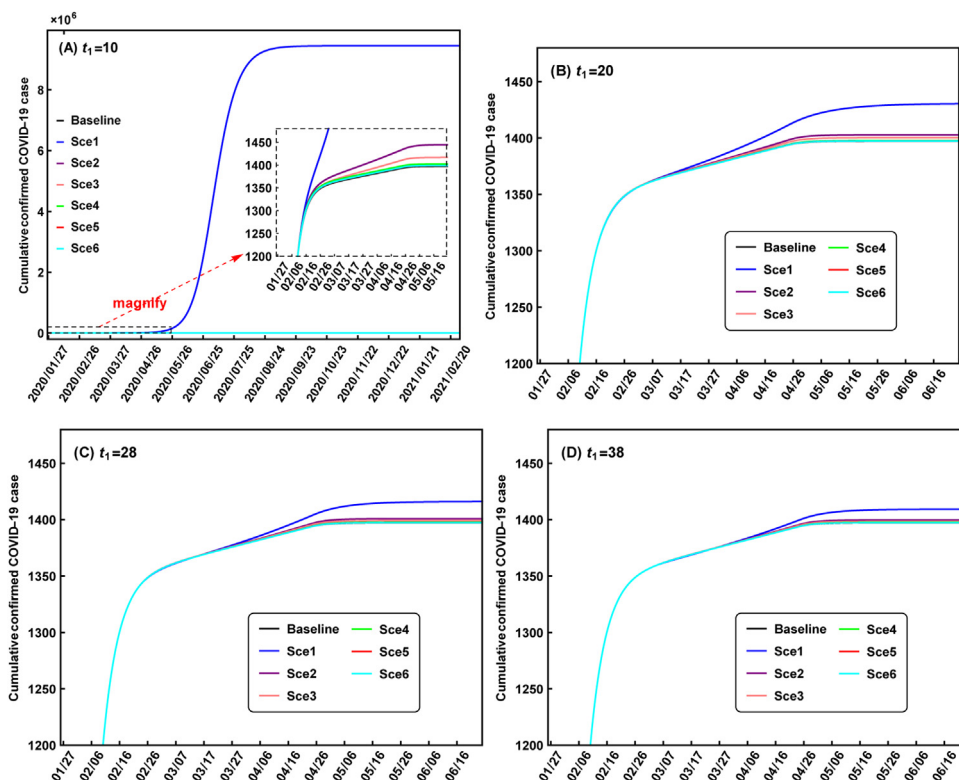


Figure 5. Scenarios results of quarantine rates impacting on the cumulative confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

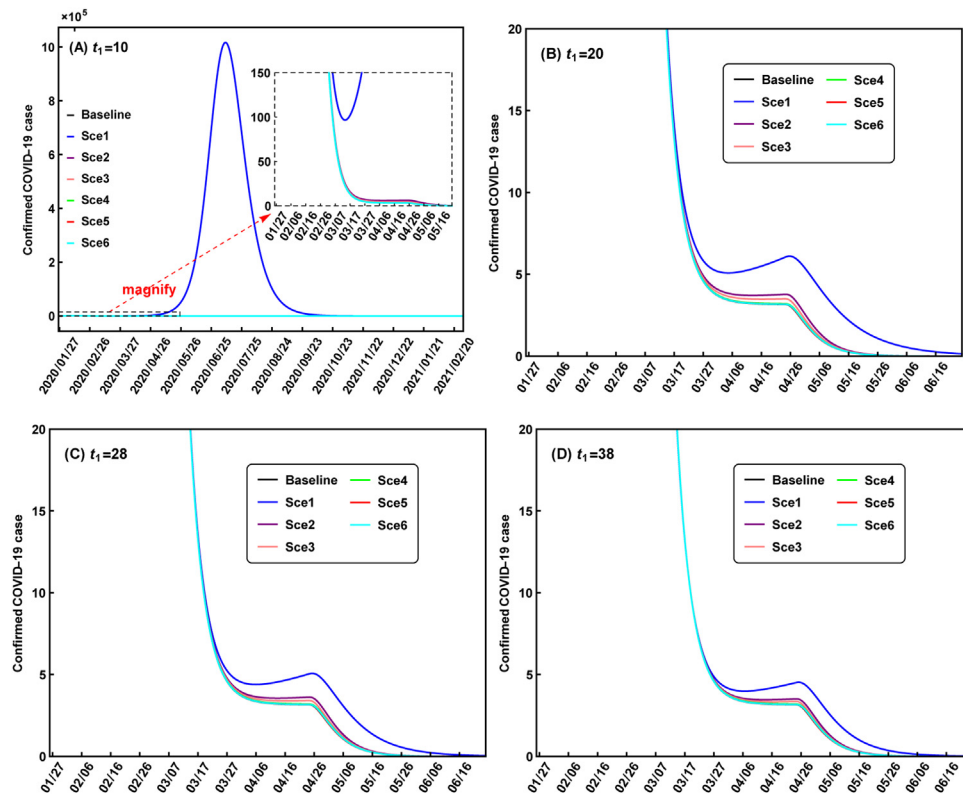


Figure 6. Scenarios results of quarantine rates impacting on the confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

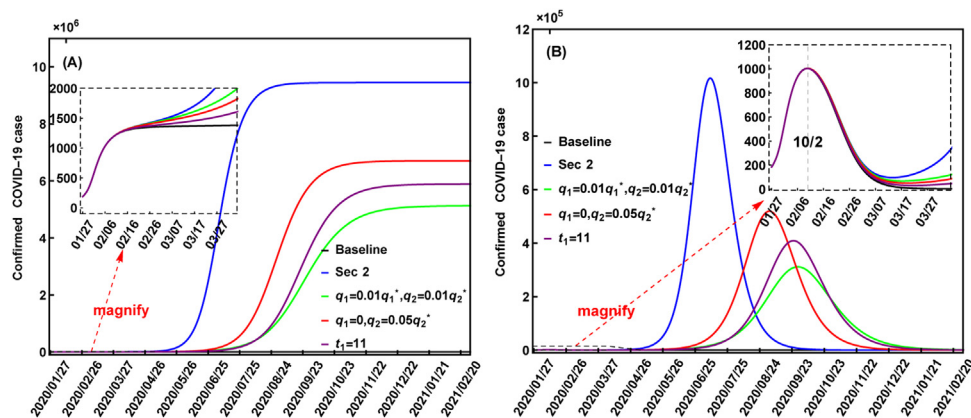


Figure 7. Cumulative confirmed COVID-19 cases (A) and confirmed COVID-19 cases (B) at the scenarios of aspect 2 with $(q_1, q_2) = (0.01q_1^*, 0.01q_2^*)$, $(0q_1^*, 0.05q_2^*)$ at $t_1 = 10$, and $(q_1, q_2) = (0q_1^*, 0q_2^*)$ at $t_1 = 11$, and the other parameters as the baseline values.

$2A^*$, $0q_1^*$, $0q_2^*$) have the MVCCC larger than 10 million at 328, 313, 327 and 312 days (Fig. 8A, STable 3). In fact, they have the two outbreaks of the disease with the confirmed cases having the first peak value as the baseline result at Feb 10, 2020 and the second peak values larger than 1 million at 142 days, 132 days, 141 days and 130 days for Sce1, Sce 2, Sce 7 and Sce 8, respectively (Fig. 9A, STable 3). The magnified figure in the period of Jan 27, 2020–Apr 26, 2020 clearly displays the second outbreak of this disease (Fig. 9A). Moreover, the weak changes of the four scenarios in the quarantine rates or around the time point $t_1 = 10$, the second outbreak also resulted in the second outbreak of the disease.

If the control measures employed as the four scenarios after the other three time points $t_1 = 20$, $t_1 = 28$, and $t_1 = 38$, the MVCCC are rapidly decreased with still larger than the baseline results, and the

DDE are prolonged except the Sce 2 and Sce 8 of $t_1 = 28$, and $t_1 = 38$ (STable 4).

For the other scenarios: Sce 3–Sce 6 and Sce 9–Sce 12 of the four time points, the DDE become smaller than the baseline result due to the larger input population and more exposed individuals. Moreover, the weaker quarantine rates together with the more input population resulted in the more infected individuals and increased the MVCCC (STable 4).

Discussion

Since the COVID-19 disease reported in Wuhan city, Hubei province of China, the Chinese government and all the people have been fighting against the disease for more than two months. Now,

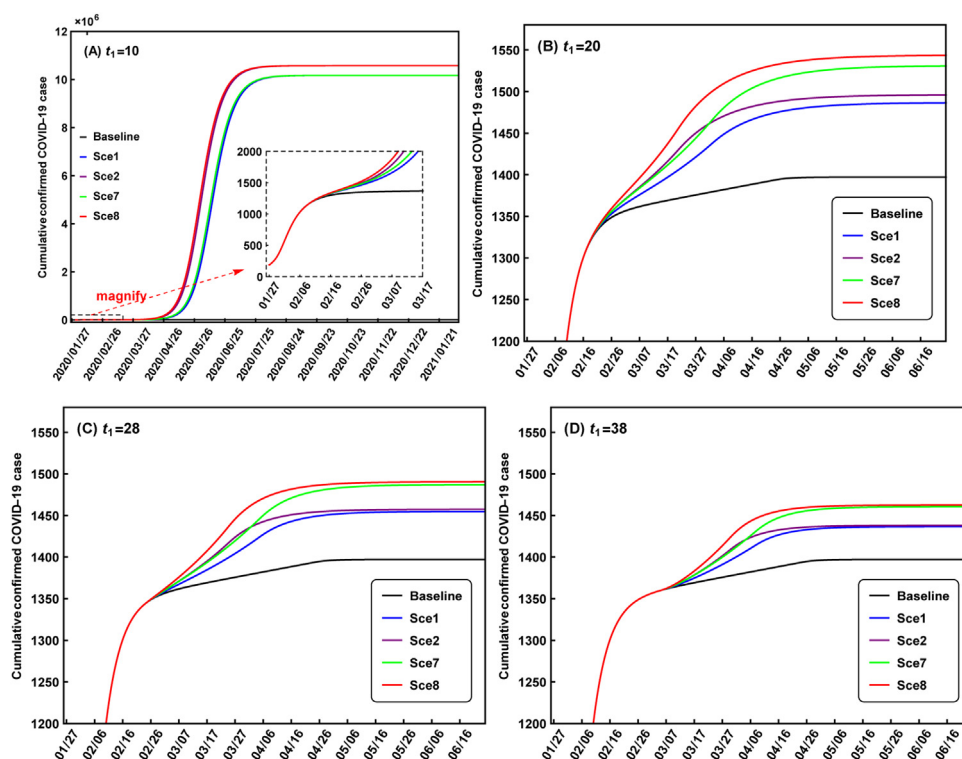


Figure 8. Scenarios results of both input population and quarantine rates impacting on the cumulative confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

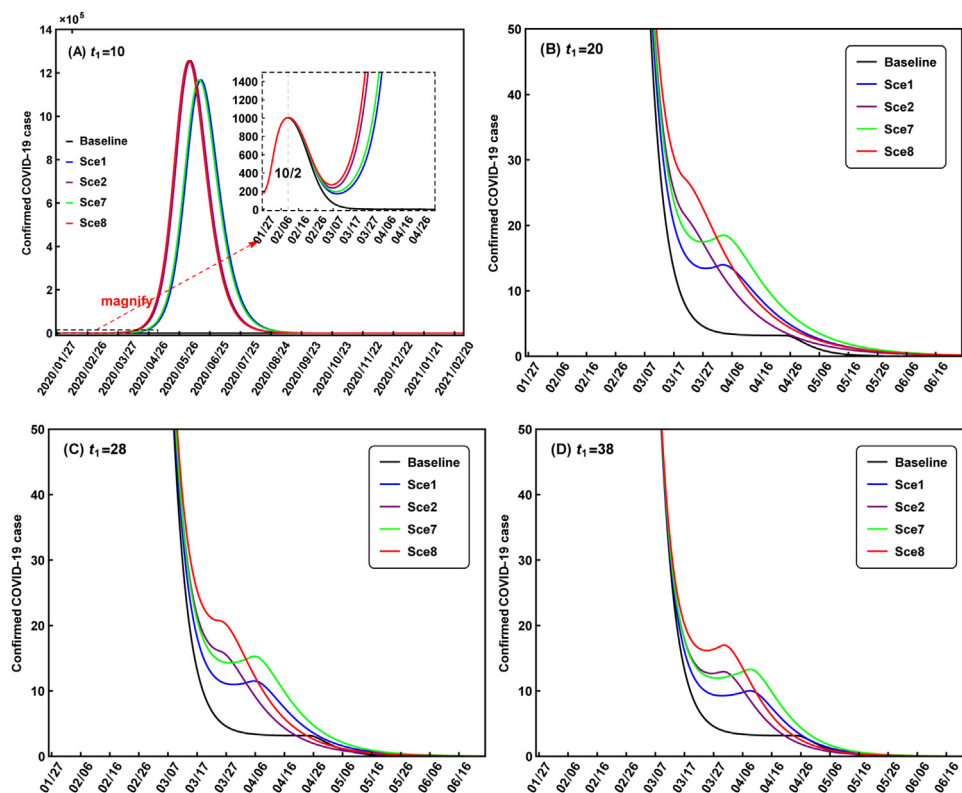


Figure 9. Scenarios results of both input population and quarantine rates impacting on the cumulative confirmed COVID-19 cases at four time points: (A) $t_1 = 10$, (B) $t_1 = 20$, (C) $t_1 = 28$ and (D) $t_1 = 38$ corresponding to Feb 6, 2020, Feb 16, 2020, Feb 24, 2020 and Mar 5, 2020.

the daily new confirmed cases have been continuously decreasing, and the latest value is 427 at Feb 28, 2020 from the National Health Commission of the People's Republic of China (<http://www.nhc.gov.cn/>).

According to the present COVID-19 disease situation, some provinces have been adjusted the emergency response level of epidemic prevention and control from the first level response to the second level, such as Guangdong province. More and more workers are coming back to Guangdong province from other provinces. To address the effects of the input population on the disease variations, taking Guangdong province as a case study, the impacts of the input population and quarantine strategies are explored using a dynamical epidemic model at three aspects. They include aspect 1: effects of the input population at different scenarios; aspect 2: effects of quarantine rates at different scenarios and the last aspect (i.e. aspect 3): effects of both input population and quarantine rates at different scenarios.

For the population flow, recent study (Tang et al., 2020a,b) considered the data from the Baidu migration website in a stochastic discrete transmission dynamic model. Both our study and Tang et al. (2020a,b) obtained the risk of the secondary outbreak when the population flow are changed at a serious input population flow. In Tang et al. (2020a,b), with more data from the Health Commission of Shanxi Province, they estimated the daily new increased confirmed cases, and the daily new increased infectious individuals from the population flow by the Poisson distribution. In our study, constrained by the data policy of the Health Commission of Guangdong Province, the input population is defined as the deterministic and continuous input. Moreover, the ratio of the exposed individuals accounting for the input population is defined as the percentages of the exposed individuals in the total population of China excluding Guangdong and Hubei provinces which is derived from the daily new increased confirmed cases according to the 3–7 days latent periods.

In the development of the COVID-19 model, Tang et al. (2020a,b) considered the quarantined susceptible individuals returned back to susceptible individuals after 14 days quarantine. While this condition is not included in our study the major reasons are displayed as follows. Under the present quarantine strategies in China, the susceptible individuals are quarantined in the forms of home quarantine, community quarantine. Although the quarantined susceptible individuals can be returned to susceptible individuals after 14 days, they will certainly employ very strict other controlling strategies against the COVID-19 virus, such as wearing the medical masks and washing their hands frequently, and which result in only very small part of the quarantined susceptible individuals back to the truth susceptible individuals.

For the simulation and prediction abilities of our model, it displayed that our model can well capture the COVID-19 variations with high accuracy. In general, it is very hard to capture the disease variations with high accuracy by the dynamical models. We have been compared our forecasting with the observed data prolonged 11 days from Feb 24, 2020 to Mar 4, 2020. The absolute values of RE (relative error) of the cumulative confirmed cases are smaller than 1% (Table 2). The corresponding figures also display that our model can capture the temporal variations in a relative longer period (see SFigure 1 in the supplementary information).

The weaker forecasting capabilities from Feb 24, 2020 to Mar 4, 2020 than these from Feb 20, 2020 to Feb 23, 2020 are resulted by the parameter estimation period of Jan 19, 2020 to Feb 19, 2020. At the same time, it inspired that if we want to obtain a high accuracy in a relative longer period the dataset used to estimate the parameters should be changed or prolonged with the time.

Our result indicated that the increased numbers of the input population can mainly shorten the disease extinction days and the increased percentages of the exposed individuals of the input population increase the number of cumulative confirmed cases at a small percentage. Both the increased input population and the increased exposed individuals have no impacts on the peak values and peak value times of the confirmed cases.

For the impacts of aspect 2, no quarantine or very weak quarantine on the susceptible individuals and exposed individuals before the days of the peak values of the confirmed cases may lead to the disease outbreak again. This proves the significant role of the quarantine strategy on the disease control.

If we increase the input population and decrease the quarantine strategy together around the time point of the peak value of the confirmed cases, there will appear second outbreak of the disease exponentially. Moreover, the weaker quarantine rates together with the more input population resulted in the more infected individuals and increased the number of the cumulative confirmed cases.

More information about our simulation and quarantine situation can be explored if more data can be obtained. In this study, to address the quarantine situation in Guangdong province, 108 scenarios are listed from the input population and quarantine strategies which may include the present quarantine strategies in Guangdong province. The other further analysis of the COVID-19 variations, such as the daily number of people under medical observation, will be explored when more new data are obtained in future.

Based the above analysis, we have the major conclusions as follows.

- (1) The COVID-19 disease variations can be simulated by our models with very high accuracy, including the cumulative confirmed cases and confirmed cases.
- (2) Under the present daily input population and quarantine strategy, the COVID-19 disease will become extinction in May 11, 2020, with the cumulative confirmed cases number of 1397.
- (3) In Guangdong province, the adjustment of the emergency response level of epidemic prevention and control from the first level response to the second level at Feb 24, 2020 is reasonable which is also predicted by our model.
- (4) The disease will have a second outbreak risk when the input population is remarkably increased and the present quarantine strategy rapidly decreases to the values around zero.

Ethics approval and consent to participate

Because no individual patient's data was employed, the ethical approval or individual consent was not applicable.

Availability of data and materials

All data are publicly available.

Funding

This research was supported by National Natural Science Foundation of P.R. China [11771373].

Disclaimer

The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflict of interests

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work.

Authors' contributions

Study design: Zengyun Hu, Qianqian Cui, Junmei Han and Zhidong Teng; Conceptualization: Zengyun Hu, Qianqian Cui; Data collection: Junmei Han, Zengyun Hu and Qianqian Cui; Data analysis: Zengyun Hu, Qianqian Cui; Visualization: Qianqian Cui, Junmei Han; Writing: Zengyun Hu; Review and editing: Zhidong Teng, Zengyun Hu. In the revised processes, Dr. Wei E. I. Sha from Zhejiang University provided important suggestions to address the quarantine strategy and improved the manuscript in English grammar. Dr. Xia Wang from Shaanxi Normal University addressed the comments on the differences of population flow between our model and Tang et al. (2020a,b).

Acknowledgements

The authors would like to thank Prof. Zhipeng Qiu from Nanjing University of Science and Technology, Prof. Tailei Zhang from Changan University, Dr. Jiao Huang from Huazhong University of Science and Technology and Mr Zhiming Jiang from University of Chinese Academy of Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ijid.2020.04.010>.

References

- Chinazzi M, Davis J, Ajelli M, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020];, doi:<http://dx.doi.org/10.1126/science.aba9757>.
- Guan W, Ni Z, Hu Y, et al. Clinical characteristics of 2019 novel coronavirus infection in China. *N Engl J Med* 2020];, doi:<http://dx.doi.org/10.1056/NEJMoa2002032>.
- Huang S, Peng Z, Jin Z. Studies of the strategies for controlling the COVID-19 epidemic in China: estimation of control efficacy and suggestions for policy makers. *Sci Sin Math* 2020];50;. doi:<http://dx.doi.org/10.1360/SSM-2020-0043>.
- Hu Z, Chen X, Zhou Q, et al. DISO: a rethink of Taylor diagram. *Int J Climatol* 2019];39:2825–32.
- Hu Z, Hu Q, Zhang C, et al. Evaluation of reanalysis, spatially interpolated and satellite remotely sensed precipitation data sets in central Asia. *J Geophys Res Atmos* 2016];121:5648–63.
- Lin Q, Zhao S, Gao D, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis* 2020];93:211–6.
- Tang B, Wang X, Li Q. Estimation of the transmission risk of 2019-nCoV and its implication for public health interventions. *J Clin Med* 2020a]; Available at SSRN: <https://ssrn.com/abstract=3525558>.
- Tang S, Tang B, Bragazzi N, et al. Analysis of COVID-19 epidemic traced data and stochastic discrete transmission dynamic model. *Sci Sin Math* 2020b];50, doi: <http://dx.doi.org/10.1360/SSM-2020-0053>.
- Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020];, doi:[http://dx.doi.org/10.1016/S0140-6736\(20\)30260-9](http://dx.doi.org/10.1016/S0140-6736(20)30260-9).
- 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* 2020a];92:214–7.
- Zhao S, Musa S, Lin Q, 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* 2020b];9:388, doi:<http://dx.doi.org/10.3390/jcm9020388>.
- Guangdong Statistical Yearbook 2019. 2019. http://stats.gd.gov.cn/gdtjnj/content/post_2639622.html.