Assessing the Global Tendency of COVID-19 Outbreak

Qinghe Liu[1], Zhicheng Liu[1], Junkai Zhu[1], Yuhao Zhu[2], Deqiang Li[1],

Zefei Gao[1], Liuling Zhou[1], Yuanbo Tang[1], Xiang Zhang[1], Junyan Yang[3], Qiao Wang[1,4]

[1] School of Information Science and Engineering, Southeast University, Nanjing, China

[2] Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands

[3] School of Architecture, Southeast University, Nanjing, China

[4] Shing-Tung Yau Center, Southeast University, Nanjing, China

*Correspondence to Q. Wang (Email: qiaowang@seu.edu.cn)

Version 3.0,   April 1th, 2020

ABSTRACT

COVID-19 is now widely spreading around the world as a global pandemic. In this report, we estimate the global tendency of COVID-19 and analyze the associated global epidemic risk, given that the status quo is continued without further measures being taken.

The results show that the global R_0, excluding China, is estimated to be 2.49 (95% CI: 2.15 – 2.92). The United States, Germany, Italy and Spain have peak values over 100,000. According to dynamical model and cluster analysis, we category the globe into four type regional epicenters of the outbreak: Southeast Asia extending southward to Oceania, the Middle East, Western Europe and North America. Among them, Western Europe will become the major center of the outbreak. The peak values in Germany, Italy and Spain are estimated to be 105,903, 127,283 and 152,539, respectively. The United States is the country with the most serious outbreak trend. Based on the current control measures by Mar. 27, 2020, the peak value in the United States will reach 400,892. Above all, if the current control measures are maintained, the cumulative number of patients worldwide will be 1,442,523 (95% CI: 1,052,577 – 8,981,440). We also estimated the diagnosis rate, recovery rate and infection degree of each country or region, and use clustering algorithm to retrieve countries or regions with similar epidemic characteristics. Different suggestions are proposed for countries or regions in different clusters.

Keywords: COVID-19; SARS-Cov-2; SEIR model; Global epidemic; t-SNE; BIRCH algorithm.

1. INTRODUCTION

In Dec. 8, 2019, the first case of the novel coronavirus pneumonia (NCP) was confirmed in Wuhan, Hubei Province, China, caused by a new type of coronavirus named '2019 new coronavirus [1] (SARS-CoV-19)' [2]. The disease spread across China during the traditional Chinese Spring Festival. As of 24:00 on March 14, China had passed the peak of the epidemic and started to recover gradually, with a total of 80,844 confirmed cases reported [3]. However, new cases of disease began to appear in other parts of the world and increases rapidly. Because of the rapid spread of the disease, the World Heath Organization (WHO) has assessed COVID-19 as a pandemic [4], which is the first pandemic caused by a coronavirus. As of March 14, 2020, the cases of COVID-19 have been reported in 135 countries and regions worldwide, with a total of 142,539 confirmed cases (61,518 outside of China) and 5,393 deaths (2,199 outside of China) [5]. Among them, a total of 17,660 cases were confirmed in Italy, 11,364 in Iran, 8,086 in Republic of Korea, 1,678 in The United States, furthermore, Europe became the epicenter of the pandemic, with more reported cases and deaths than the rest of the world combined, apart from China [6].

With the outbreak of COVID-19, countries or regions have been taking different measures to cope with the spread of the pandemic, but the number of infected people is still increasing. Meanwhile, the spread of the pandemic has also caused a huge impact on the trade flows and economic affairs of the world. These situations raised many urgent problems. How will the epidemics spread in countries or regions around the world? When will the spread of epidemic arrive the peak or turn to stabilize? How many people will be at risk of infection? Therefore, it is of great significance to analyze the trend of COVID-19 and predict the arrival of peaks for the prevention and control of the pandemic all over the world .

Based on the global data [7], we perform an analysis on epidemic status in 96 countries, which had the number of confirmed cases over 100 in Mar. 27, 2020. The Section 2 of this paper briefly discusses the most up-to-date literature related to COVID-19 epidemic status. Section 3 describes the transmission dynamics model in details. Section 4 presents the simulation and analysis of the estimated trend. Section 5 analyzes the four epidemic transmission clusters with different characteristics. Finally, conclusion and discussion will be respectively presented in section 6 and 7.

2. RELATED WORK

On March 1, 2020, Li Y *et al* [8], used the outside-China diagnosis number released by WHO and built a mathematical model to capture the global trend of epidemics outside China, and they found that 34 founder patient outside China were not found, the worldwide epidemic trend is approximately exponential, and may grow 10 folds every 19 days.

On March 2, 2020, Zhuang Z *et al* [9], used a stochastic model to simulate the transmission process of South Korea and Italy under two corresponding assumptions of exponential growth periods, the results indicated that the reproductive number of the Republic of Korea and Italy are 2.6 (95% CI: 2.3-2.9) and 3.3 (95% CI: 3.0-3.6), respectively. In addition, the estimation of dispersion term were conducted, which indicated that there are fewer super-spreading events.

On March 8, 2020, Zhang Z *et al* [10], applied a data-driven coding method for the prediction of the global spreading epidemic of COVID-19. Based on the historical epidemiological data and the sets of parameters of augmented SEIR model in 367 cities of China, they selected the best fit profiles to predict the trend of epidemic with any population by comparing the given an early epidemic and the historical profiles. In their study, the peaks of infectious cases in South Korea, Italy, and Iran are expected to occur at the end of March, and the percentages of population infectious will less than 0.01%, 0.05% and 0.02%, respectively.

On March 10, 2020, Zhang Z *et al* [11], used the state transition matrix model to predict the epidemic trend of South Korea, Italy and Iran. By matching and fitting different scenarios, the inflection point arrival time will be March 6-12 for South Korea, March 10-24 for Italy and March 10-24 for Iran, and the cumulative number of cases will reach 20k in South Korea, 209k in Italy and 226k in Iran, respectively.

On March 14, 2020, Li L *et al* [12], made a propagation analysis and the wordwide prediction of COVID-19, and they also realized the backward inference of the epidemic starting day. The main results show that the epidemic in South Korea will be basically under control at the end of March, and the inflection day is January 7, before the control the reproductive number is 4.2 and 0.1 after the control. Epidemic size in Italy and Iran will reach 200,000 and 20,000, respectively at the end of March, and the inflection starting day are both on January 13, the reproductive number are in decline from 4.2, 4.0 to 0.1, 0.2. And their work could be a reference to our prediction results.

On March 17, Darwin R *et al* [13], used the progression of the epidemic curve and the defined frame work to estimate the instantaneous reproductive number combined with the whole genome sequencing, wich can provide the genomic evolution and variation in the context of the outbreak dynamics. They divided the epidemic into different serial interval scenarios, and estimated the R values. The R value in Japan, Germany, Spain, Kuwait and France are over 2, in Italy, Iran and South Korea are even over 10, however, the R will be low after the social distancing intervention.

In our previous work [14], we predicted the time and value of inflection point and peak point in both Hubei and outside Hubei of China. Inspired by it, we conducted a worldwide prediction of value of the peak point and analyzed the main characteristics and appearances of COVID-19 in the world. The predicted results are based on the current medical situation, if there is specific medicine for COVID-19 in the future, the time and value of peak point of all countries will be advanced and shrinking compared with none-medicine situation.

3. COVID-19 TRANSMISSION MODEL: TRANSMISSION DYNAMICS

3.1. Transmission Dynamics Model

We constructed a transmission dynamics model to infer the epidemiological characteristics and the peak size and trend of COVID-19 based on the existing infectious data and recovered data of various countries or regions. The population of this paper are divided into four main categories based on the SEIR model, i.e. S, E, I and R referring to [15]: the Susceptibles, Exposed, Infectious, and Recovered, respectively [16][17]. In the study of China's epidemic situation, many factors such as Spring Festival, control policies and so on are taken into account, but in contrast, this paper will analyze the heterogeneity of each country or region on the basis of the common spread of infectious diseases. Although these countries have different policies and customs, their basic spread pattern of the epidemic still meets the principle of dynamics [18]. *S*, the susceptibles, is equal to the total population (*N* in equation (3.1)) in the research of COVID-19 due to the general susceptibility of the population. The main assumptions of this model are as follows:

(1) Because the difference of infectivity between the exposed and the infectious population is unknown, both groups of *E* and *I* are set to follow the same coefficient β to represent the average infection level of COVID-19.
(2) In the predicted time scale, the existing influences of policies or culture of the target object will be unchanged.
(3) During the transmission of virus, the probability of infection in contact with each person is equal in the same target group.

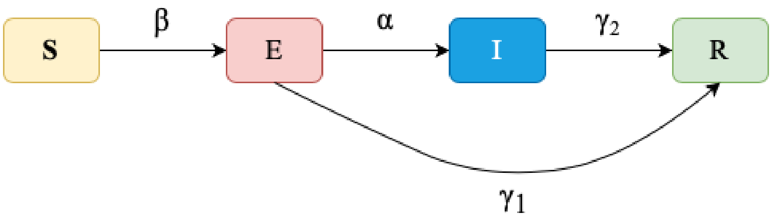

Figure 1. SEIR system state transition diagram.

Table 1. List of Symbols

Constant	N	Total population of object of study
State variable	S	Number of susceptible people
	E	Number of exposed people
	I	Number of infectious people
	R	Number of recovered people
Transition variables	α	Rate of those exposed and being infectious.
	β	Rate of those susceptible and being infected.
	γ_1	Transmission rate to the recovered from the exposed.
	γ_2	Transmission rate to the recovered from the infectious.

$$\begin{cases} \frac{dS}{dt} = -\beta(E+I)\frac{S}{N} \\ \frac{dE}{dt} = \beta(E+I)\frac{S}{N} - (\alpha+\gamma_1)E \\ \frac{dI}{dt} = \alpha E - \gamma_2 I \\ \frac{dR}{dt} = \gamma_1 E + \gamma_2 I \end{cases} \tag{3.1}$$

where *t* means time, and *S*, *E*, *I*, *R* represent four different state variables, respectively. Class *S* represents a healthy population without virus. Once transferred to class *E*, it means the population has been infected until transferred to class *R*, recovered. The difference between group E and group I is that people in class E are in the incubation period of transmission, while people in class I have been diagnosed due to symptom detection. Formula (3.1) can be represented by the state transition diagram in Figure 1. All parameters and physical interpretations of the model are shown in Table 1.

3.2 Practical Consideration

We divided the spread of the COVID-19 into three phases according to the evolution of the size of the confirmed cases, and the schematic diagram of each stage for the number of patients is shown in Figure 2.

Phase one: the time range of this phase starts from the beginning of SARS-CoV-2 transmission and ends when the number of patients begins to increase significantly. COVID-19 spreads during incubation period at this phase, but the confirmed cases officially counted remains at a low level, and the public's awareness of defense is light.

Phase two: the main feature is that a large number of patients have been diagnosed, and the statistical data of confirmed cases increase significantly. But there still has no policy intervention and large-scale quarantine measures.

Phase three: as the intensity of policy intervention increases, the spread of COVID-19 is effectively suppressed, and growth rate of confirmed cases is gradually controlled until the end of the epidemic. The main feature of this phase is a strong human intervention.

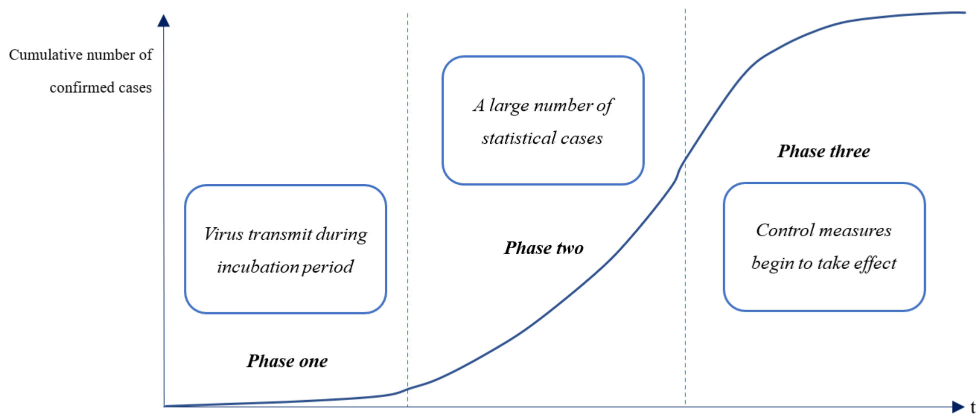

Figure 2. Phase diagram of confirmed cases changing with time in a certain region.

It is crucial to get as much information as possible from the data in three stages. Because this paper aims at the analysis of the peak size and its arrival time in each country or region, two latter transmission phases will be our focus in this paper. In phase two, SEIR will be used to analyze the trend of the epidemic, while in phase three, the simulation of policy intervention will be carried out in combination with the prior information of China studied in [14].

In phase two, SARS-CoV-2 spread rapidly. The infection rate remains at a high level, and the recovery level is low because the government had not carried out obviously effective policy control. From phase three, the infection rate decreases and the recovery rate begins to increase. Through the study of the historical data of COVID-19 in China, we dicided to use the power law to simulate the trend of infection rate and recovery rate in phase three, as is shown in equation (3.2). Since the coefficient c plays a role in inhibiting the growth of the number of confirmed cases, we define c as the incremental inhibition ratio.

$$\begin{cases} \beta(t) = \beta_0(1-c)^t \\ \gamma(t) = \gamma_0(1+c)^t \end{cases} \tag{3.2}$$

where β_0 and γ_0 are the initial values of β and γ and t is time from the beginning of phase three.

3.3. Estimation of R_0

The basic reproduction number, R_0, refers to the average value of how many people an infected person can transmit the virus to through natural transmission without external intervention [19]. From the second generator approach in [20], the equation (3.3) is derived as follows:

$$R_0 = \beta\left(\frac{1}{\alpha+\gamma_1}\right) \tag{3.3}$$

where α, β, γ_1 are transfer variables of SEIR model. Considering that the calculation of R_0 is in the case of natural transmission, so we use the parameters of the starting time of epidemic spread in each country or region to estimate R_0 to avoid the impact of human intervention.

3.4. Data Source

In this study, we use open dataset from Johns Hopkins University (https://bluehub.jrc.ec.europa.eu/migration/app/). The details of data source are shown in Appendix 1.

Johns Hopkins University. Johns Hopkins University has shared their data on GitHub for academic and scientific research. We have compiled their open data into a chronology of confirmed data for prediction and visualization use.

4. MAIN RESULTS OF SEIR MODEL

4.1. Estimate of R_0

While estimating the R_0 of the epidemic globally, we ignore the countries whose ultimate cumulative number of confirmed cases are less than 100 by Mar. 28, 2020 and focus on the remaining 96 countries (please refer to appendix). The average basic reproduction number of these countries is 2.49 (95% CI: 2.15 – 2.92). As we can notice, COVID-19 is lashing a great part of the world, and the sufferings are not likely to end soon. By estimating the average value of R_0 , we obtain the objective information of current situations as well as the importance of epidemic prevention.

Among these 96 mainly researched countries, the R_0 of Franch, Iran, Japan, Singapore and Turkey are higher than 3, suggesting the great potential of a destructive outburst. France is one of these countries with a basic reproduction number of 3.61, much higher than the global average(See Appendix 2 for more details). In addition to France, the basic reproduction number of Germany, the United Kingdom, Italy and Spain are all above the global average, reflecting the severe epidemic status in European countries and confirming that they have become the epicenter of COVID-19 outbreaks.

The number of the United States is 2.81, which is also well above the global average, revealing the urgency of the current outbreak in the United States. The lowest value of R_0 is 1.68 in Kuwait, still greater than 1.

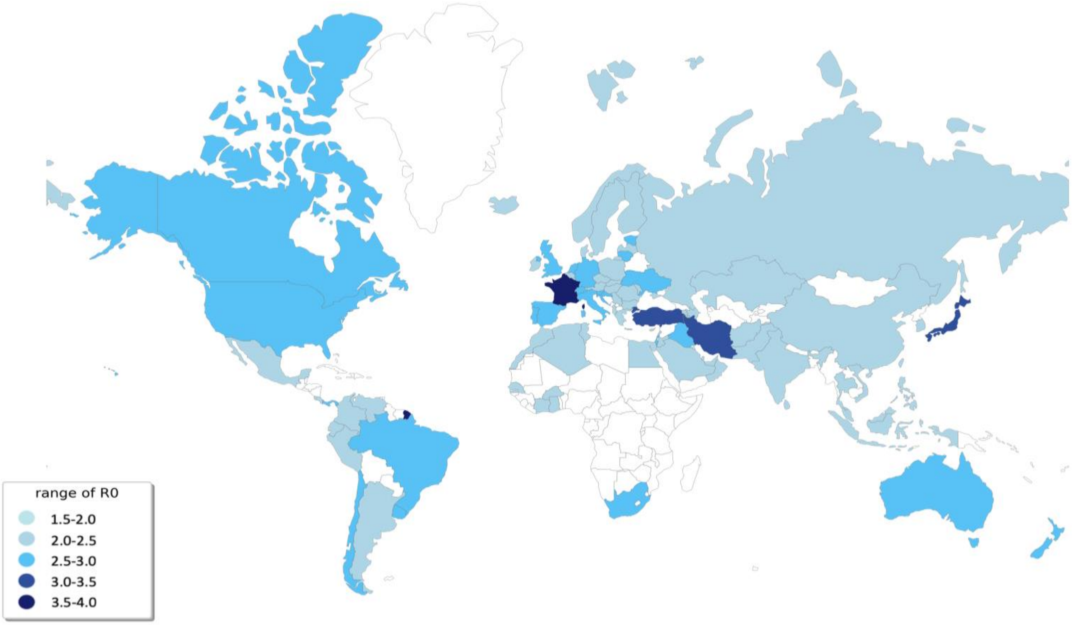

Figure 3. R_0 in 96 countries or regions over the world, which has the number of confirmed cases exceeding 100 by March 28, 2020. Blue means the higher the value of R_0, while white means the lower the the value of R_0.

4.2. Estimate of the Peak of Confirmed Cases in Country-Level

The peak value is defined as the cumulative number of confirmed cases. After this peak point, the temporary number of cases gradually decrease or slightly fluctuate, and the epidemic subsides [14]. The estimated trend of 96 countries are all listed in Appendix 2 for details.

The United States, Spain, Italy and Germany have peak values higher than 100,000 based on simulations. 16 countries including France, Iran and the United Kingdom have peak values ranging from 10,000 to 100,000. 44 countries have peak values ranging from 1000 to 10,000. Jordan and other 31 countries have peak values ranging from 100 and 1,000.

Furthermore, 32 countries will not reach the peak value of 1000, which accounts for 30% of all the estimated 96 countries. We also compared peak arrival time in countries whose peak value are higher than 1000, and find that most of the countries will reach their peak value in May.

The United States is the country with the highest exposure to the risk of the coronavirus. According to our estimation, the peak number of real-time confirmed cases of the United States is predicted to reach 400,892 after July 21, 2020, provided the present coping strategy is not changed. It is the only country in our prediction that the number of confirmed cases will exceed 200,000.

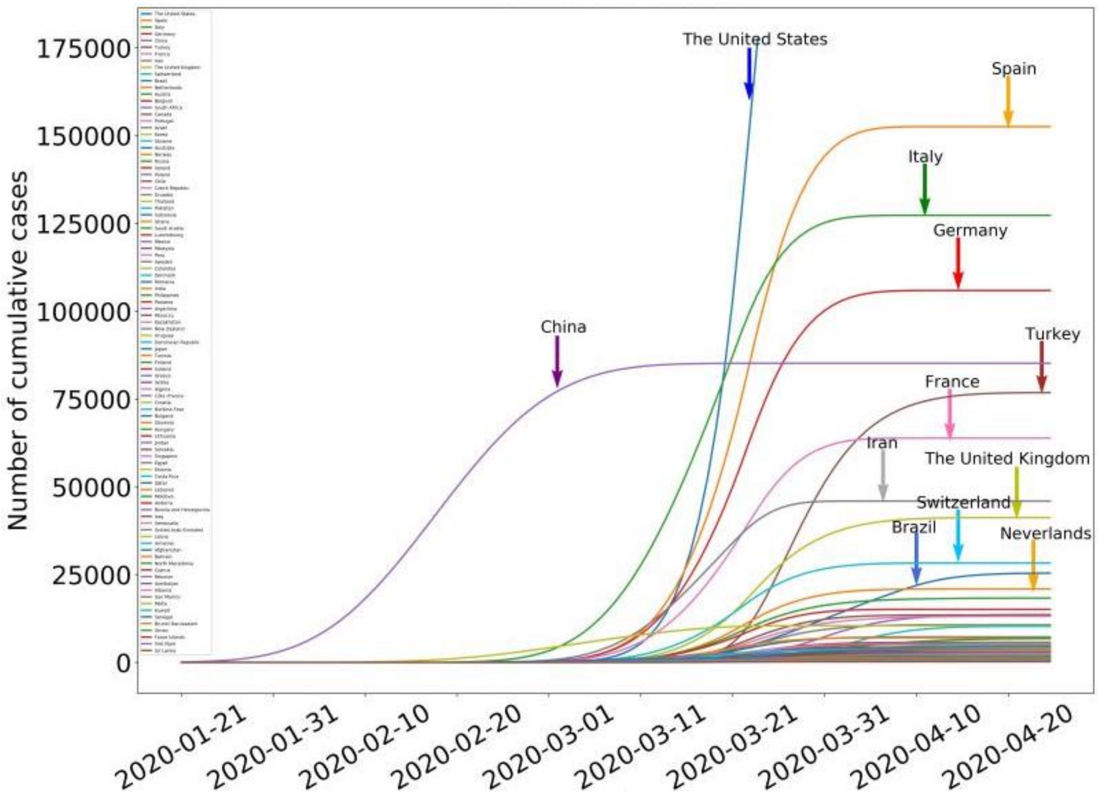

Figure 4. Simulated trends of cumulative number of confirmed cases in 96 countries or regions, which has the number of confirmed cases exceeding 100 by March 28, 2020. The horizontal axis is the date, and the vertical axis is the number of confirmed cases.

Table 2. Sum of peak values of different continents

Peak interval	Countries or regions
>100,000	The United States, Spain, Italy, Germany
[10,000, 100,000]	China, Turkey, France, Iran, The United, Kingdom, Switzerland, Brazil, Netherlands, Austria, Belgium, South Africa, Canada, Portugal, Israel, Korea, Ukraine
[3,000, 10,000]	Australia, Norway, Russia, Ireland, Poland, Chile, Czech Republic, Ecuador, Thailand, Pakistan, Indonesia, Ghana, Saudi Arabia, Luxembourg, Mexico, Malaysia, Peru, Sweden, Colombia, Denmark, Romania, India
[1,000, 3,000]	Philippines, Panama, Argentina, Morocco, Kazakhstan, New Zealand, Uruguay, Dominican Republic, Japan, Tunisia, Finland, Iceland, Greece, Serbia, Algeria, Côte d'Ivoire, Croatia, Burkina Faso, Bulgaria, Slovenia, Hungary, Lithuania
<1,000	Jordan, Slovakia, Singapore, Egypt, Estonia, Costa Rica, Qatar, Lebanon, Moldova, Andorra, Bosnia and Herzegovina, Iraq, Venezuela, United Arab Emirates, Latvia, Armenia, Afghanistan, Bahrain, North Macedonia, Cyprus, Réunion, Azerbaijan, Albania, San Marino, Malta, Kuwait, Senegal, Brunei Darussalam, Oman, Faroe Islands, Viet Nam, Sri Lanka

Spain, Italy and Germany monopolize one category, out of all 96 countries, due to its high risk of the coronavirus. The peak number of real-time confirmed cases of Italy is predicted to reach 127,283.The trend of COVID-19 in Italy is expanding, though efforts had been taken by the government of Italy. Other 16 countries whose peak values are estimated to reach over 10,000, such as the Turkey, France, Iran, the United Kingdom and Canada, are all in severe situation at present. Australia, Norway, Russia and several other countries are the next worst-hit areas with their peak value approaching 10,000.

4.3. Estimate of the Peak of Confirmed Cases in Continent-Level

Table 3. Sum of peak values of different continents

Continents	Peak value
Europe	749,911 (95% CI: 561,598 – 2,814,499)
Asia	184,722 (95% CI: 178,414 – 206,180)
North America	420,685 (95% CI: 239,546 – 5,788,917)
South America	52,686 (95% CI: 44,867 – 104,299)
Oceania	7,247 (95% CI: 5,273 – 17,720)
Africa	27,272 (95% CI: 22,879 – 49,825)

We collected the peak estimate results of six continents in Table 3. Oceania has the minimum peak value and there are only two countries in Oceania for our study, the sparse population makes it less severe than other continents. It is estimated that Asia's peak value would reach 184,722 (95% CI: 178,414 – 206,180). As of March 28, 2020, the daily increment of confirmed cases only in The United States is 19,821. This also led to an outbreak trend in North America, which is estimated to reach a peak of 420,685 (95% CI: 239,546 – 5,788,917) under the current policy intervention. Europe has the highest peak value indicating the severe situation in Europe. South America has a much lower peak value than North America.

4.4. Geographical Distribution Analysis of COVID-19

The estimated final epidemic scale is visualized on the world map to show the geographical characteristics of the predicted epidemic peak of confirmed cases. The cumulative number of global patients might finally attain 1,442,523 (95% CI: 1,052,577 – 8,981,440). According to [21], geographic proximity transmission is very prominent in epidemic transmission, which means that the regional transmission is popular. This opinion can also be confirmed on the map shown in Figure 5. The outbreak shows multiple epicenters. It can be found that North Korea is geographically close to South Korea, Japan and China, which has the worst situation.

The choropleth map indicates that the spread of COVID-19 will form four major regional clusters. The first epicenter ranges from East Asia to Oceania, including China, South Korea, Japan, Indonesia, the Philippines, Australia, etc. South Korea, Japan, Indonesia, Malaysia, and the Philippines will have peak values range from 1,000 to 10,000.

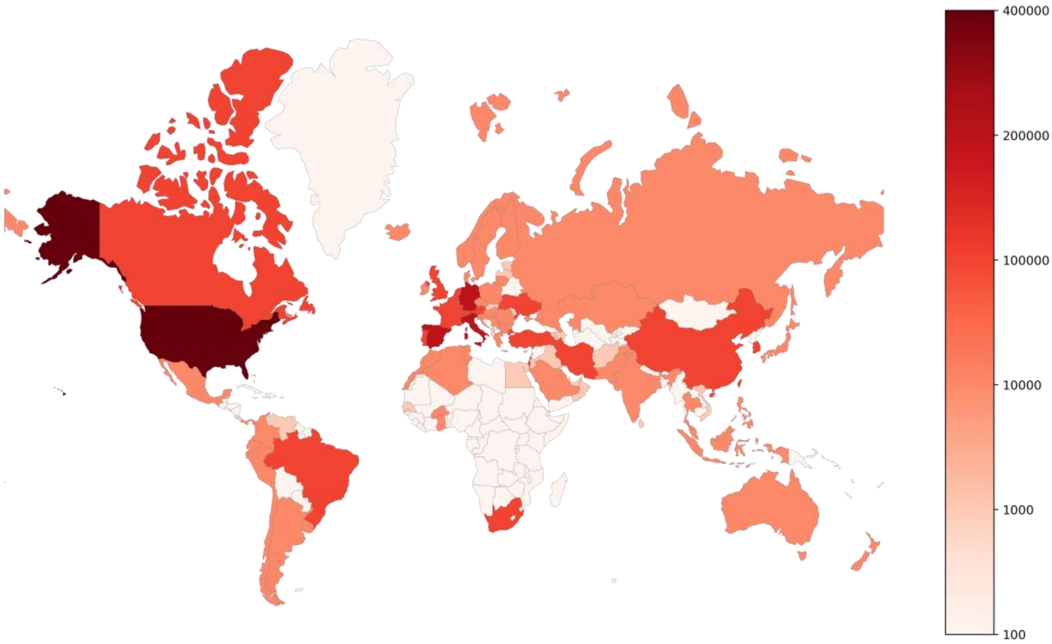

Figure 5. Peak value of confirmed cases in 96 countries or regions over the world, which has the number of confirmed cases exceeding 100 by March 28, 2020. Red means the higher the diagnosis cases peak, while white means the lower the diagnosis cases peak.

The second outbreak epicenter occurs in Western Europe, including Spain, Italy, France, Germany, and many other neighboring countries, as well as Turkey. Germany, Italy and Spain are estimated to have 105,903, 127,283 and 152,539 confirmed cases, respectively, given that the current policies are not changed. Turky, France, The United Kingdom, Switzerland, Netherlands are all expected to have peak values of over 20,000. Considering the number and density of European countries, there would be a serious cross-border phenomenon. Under the current transmission environment and control measures as of Mar. 27, 2020, the estimated number of patients is likely to be 749,911, with 740 million population in Europe. Based on the above data, the per capita prevalence rate in Europe will be at least 17.2 times higher than that in China.

The third epicenter is in the Middle East. According to the current trend, the peak number of confirmed cases in Iran will reach to 45,932. According to the current trend, the epidemic situation in Iran has gradually slowed down. In addition, the peak values of Saudi Arabia is estimated to be over 4,000. Estimated by the current situation, the peak of diagnosis in the Middle East will come in early May this year.

The last epicenter occurred in North America. Given that the epidemic has spread in The United States (140,886 cases in total, as of Mar. 30, 2020), if the current situation is maintained, the peak number of cases in The United States will reach 400,892, which may have a huge impact on American people's livelihood and social economy. If The United States follows the existing policy as of Mar. 27, 2020, the peak arrival time of COVID-19 in The United States will be after Jun. 1, 2020. In addition, Canada is expected to have a peak value of 13,468, making it the second most affected country in the region.

5. CLUSTER ANALYSIS OF EPIDEMIC CHRACTERISTICS

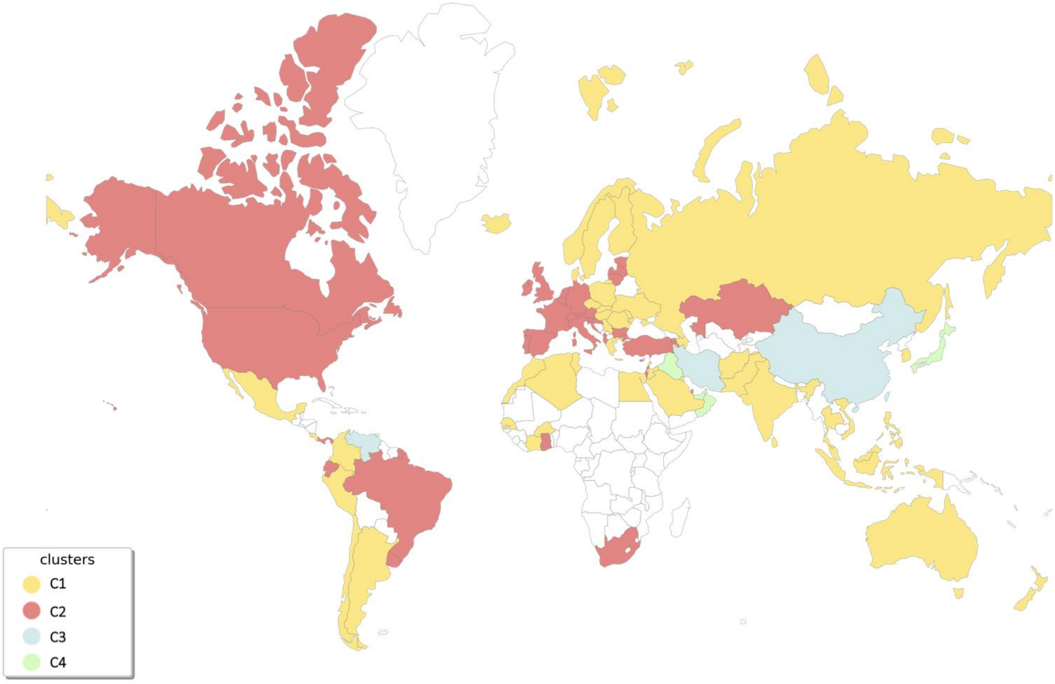

Figure 6. Clustering of the 96 countries or regions with the five SEIR parameters: α, β, γ_1, γ_2, c. In this world map, white represents the countries not participating in the clustering.

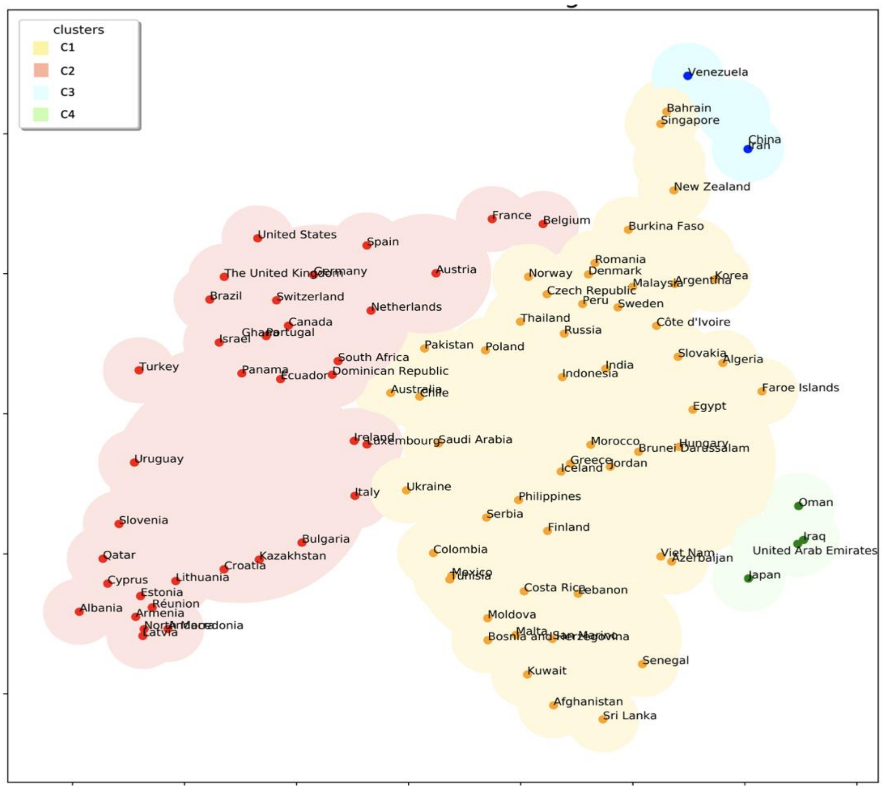

Figure 7. Illustration of SEIR parameter-proximity, i.e. α, β, γ_1, γ_2, c, of the 96 countries using t-SNE dimension reduction. Owing to the proximity preservation property of t-SNE, the proximity relationship in 5-dimensional parameter space can be visualized in the above 2-dimensional plannar. Colors indicate the clusters resulted from clustering algorithm.

To further analyze the characteristics of epidemic transmission among different countries or regions, we took the five parameters, i.e. α, β, γ_1, γ_2, c, of the SEIR model obtained in the simulation as feature vectors. Then the BIRCH (Balanced Iterative Reduction and Clustering Using Hierarchies) algorithm was used to perform clustering based on these feature vectors to retrieve countries or regions with similar epidemic patterns, see Figure 7 and Appendix 3. The average of parameter in each cluster is shown in Table 4. To visualize the proximity relationship of epidemic characteristics of these countries or regions, t-SNE (t-distributed stochastic neighbor embedding) algorithm is used to reduce the dimension of data to a 2-dimensional plannar, while preserving sample proximity in the 5-dimentional parameter space, see Figure 7.

The number of clustering countries or regions indicates the universality or particularity of the epidemic pattern reflected by the cluster. Cluster one accounts for 54% of the total number of studies, and has the parameters most close to the overall average. This includes Australia, South Korea, New Zealand, etc. Although the transmission degree is on average, in order to prevent the further spread of COVID-19, the policy of intervention and medical treatment for the confirmed cases still needs to be improved.

Cluster two is represented by The United States, Italy, Spain, Germany, etc. Their $\bar{\beta}$ are of the highest levels, 1.80. However, $\overline{\gamma_1}$ and $\overline{\gamma_2}$ were the lowest, 0.19% and 0.16% respectively, which shows that the epidemic situation in such countries or regions is the most serious, but has not yet got the effective recovery level. It is necessary to remind them to take timely prevention and control measures for the flow of people and invest in medical treatment, otherwise the epidemic will be likely to further spread.

Cluster three contains only three countries: China, Iran and Venezuela. Because of the high $\bar{\beta}$, their epidemic spread quickly. But the three countries are excellent at diagnosing and treating patients. $\bar{\alpha}$, $\overline{\gamma_1}$ and $\overline{\gamma_2}$ are much higher than the average, which makes their epidemic situation be better suppressed. For example, intervention measures such as traffic blockade and quarantine taken in time at the beginning of the outbreak in China are very effective and have successfully restrained the further growth of confirmed cases. This is also recommended for other countries to learn from.

Cluster four has the lowest risk among the four clusters. $\bar{\beta}$ = 1.20 indicates the lowest level of virus transmission, while $\overline{\gamma_1}$ = 0.65% and $\overline{\gamma_2}$ = 0.51% are 62.5% and 50.0% higher than the average respectively. Their epidemic spread is very light, and is more easily suppressed. The diagnosis rate, $\bar{\alpha}$=0.38, is the lowest, which may have the main reason that the demand for diagnosis level is not high.

It should be noted that clustering is only used to help us find out similar epedemic patterns. It is not that any country or region is fixed as the description of the above four clusters. Figure 7 shows that some countries are on the edge of clustering. The cluster assignments of these countries might be sensitive to the estimated SEIR parameters. We have reason to believe that Singapore and Bahrain located at the edge of Cluster one and Cluster Three also have good epidemic control status.

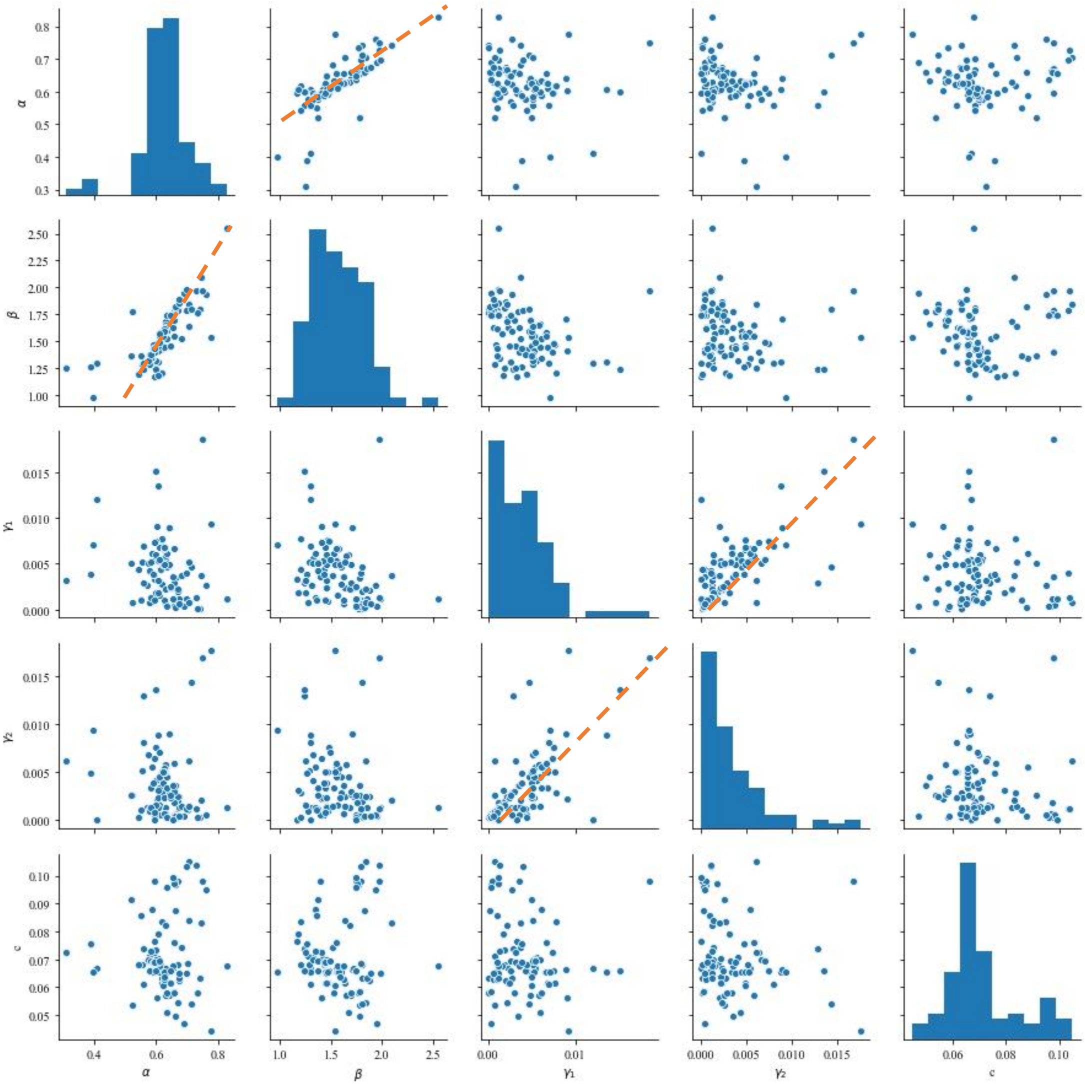

Figure 8. Frequency histograms and scatters of α, β, γ_1, γ_2, c, which show that there is a linear relationship between α and β, and there is a linear relationship between γ_1 and γ_2. This shows that countries with rapid COVID-19 transmission are also more efficient in diagnosis, so as to reduce the burden of increasing patients. Among the 96 countries or regions, China, Kazakhstan, Bulgaria, etc. not only keep the diagnosis rate at a high level, but also effectively reduce the spreading speed of the epidemic. Most of them have adopted a more timely travel restriction and contact precaution in face with COVID-19.

Table 4. Attributes list of cluster centers

Cluster	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\gamma_1}$	$\bar{\gamma_2}$	$\bar{c}$	#Countries	Countries or Regions
One	0.60	1.42	0.0050	0.0038	0.069	52	Australia, Korea, etc.
Two	0.67	1.80	0.0019	0.0016	0.074	37	The United States, Italy, Spain, Turkey, etc.
Three	0.75	1.77	0.0108	0.0163	0.065	3	China, Iran, etc.
Four	0.38	1.20	0.0065	0.0051	0.070	4	Japan, Oman, Iraq, United Arab Emirates.
Total	0.62	1.57	0.0040	0.0034	0.07	96	-

Australia, Pakistan, Norway, Poland and other countries in cluster one are very close to the second cluster, which indicates the possibility of further outbreaks of COVID-19.

The high risk areas of the outbreak are concentrated in the Americas, Europe and the Middle East (cluster two). The severity of the epidemic in South Africa is very prominent on the geographical map. On the one hand, we once again call for timely treatment and isolation measures, on the other hand, it is necessary to remind Southern African countries of the prevention of COVID-19 geographic proximity transmission.

6. CONCLUSION

We performed simulation analysis of COVID-19 using SEIR model assuming that current policies by Mar. 27, 2020 remains. From analysis, we draw the following conclusions:

(1) We collected data before March 28, 2020 to simulate the transmission of the SARS-CoV-2 in 96 countries. The average basic reproductive number R_0 of these countries was estimated to be 2.49 (95% CI: 2.15 – 2.92).

(2) Four countries will have a peak value over 100,000, including the United States, Spain, Italy and Germany. And 16 of the 96 countries including France, Turkey, The United Kingdom, Switzerland will have peak values between 10,000 and 100,000. Furthermore, 32 countries will not reach the peak value of 1000.

(3) Among the six continents, Europe has the highest peak value of 749911 (95% CI: 561,598 – 2,814,499). The lowest peak of 7,247 (95% CI: 5,273 – 17,720) appears in Oceania. Asia has a peak value of 184,722 (95% CI: 178,414 – 206,180. (see Table 2 for more details.)

(4) Western Europe, Southeast Asia to Oceania, the Middle East, and North America will be the four epicenter with the most severe situation. The epidemic situation in North America and Western Europe are estimated to far exceed that in China. In addition, North America will be the worst-hit epicenter.

(5) The 96 countries or regions were divided into four clusters by clustering algorithm based on epidemiology related parameters α, β, γ_1, γ_2, c. The discovered four clusters represent low risk, medium risk, high risk and effective control respectively. Based on the similarity of epidemic characteristics, we gave early warning to many countries, including Australia, Ukraine, etc.

7. DISCUSSION

In this paper, we mainly estimated and analyzed the transmission trend of COVID-19 based on the daily updated data of confirmed cases and recovered cases. In fact, the trend of COVID-19 is related to multiple factors. China's prior data were used in the third phase of the epidemic (see Figure 2), however, the testing policies and standards in each country and region are different, which is highly related to the governance system and safety consciousness. In addition, in the second phase of epidemic transmission, community transmission plays an important role. The intensity of people flow directly affects the infection degree of the virus. Some factors, such as subway traffic, catering facilities, customs or habits of each region remains to be studied.

All the analysis above is based on the assumption that the official statistics of confirmed cases and of recovered cases reflect the real status quo of epedemics. As a matter of fact, there may be deviations in the norms of statistical data in various countries, which may also lead to bias of our results.

Abbreviations

COVID-19: Coronavirus Disease 2019; SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2; CI: Confidence Interval; SEIR: Susceptible-Exposed-Infectious-Recovered.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The datasets used and analyzed during the current study is available from open resources.

Competing Interests

The authors declare that they have no conflict of interest.

Funding

This work was not supported by any funding.

Authors' Contributions

Conceived and designed the experiments: Qinghe Liu, Junkai Zhu, Junyan Yang, Qiao Wang.
Performed the mathematical modelling: Qinghe Liu, Junkai Zhu, Qiao Wang.
Analyzed the data: Qinghe Liu, Junkai Zhu, Zhicheng Liu, Yuhao Zhu.
Collect the data: Zefei Gao, Deqiang Li.
Performed the computations: Yuanbo Tang, Xiang Zhang.
Wrote the paper: Junkai Zhu, Qinghe Liu, Deqiang Li, Liuling Zhou, Zefei Gao, Zhicheng Liu, Yuhao Zhu.
All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

Reference

[1] 'Definition of suspected cases of unexplained pneumonia', the National Health Commission of the People's Republic of China (in Chinese) /[http://www.nhc.gov.cn/]

[2] Li Q, Guan X, Wu P, *et al*. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia[J]. The New England Journal of Medicine, NEJM.

[3] 'Update on COVID-19 as of 24:00 on March 14' (in Chinese)/[http://www.nhc.gov.cn/]

[4] 'WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020'/[https://www.who.int/]

[5] WHO, Novel coronavirus (2019-nCoV) situation report 54. World Health Organization, 2020. /[https://www.who.int/]

[6] 'WHO Director-General's opening remarks at the media briefing on COVID-19 - 13 March 2020'/[https://www.who.int/]

[7] WHO, Novel coronavirus(2019-nCoV) situation report. World Health Organization, 2020. /[https://www.who.int/]

[8] Li Y, Liang M, Yin X, *et al*. COVID-19 Epidemic Outside China: 34 Founders and Exponential Growth. medRxiv 2020:2020.03.01.20029819.

[9] Zhuang Z, Zhao S, Lin Q, *et al*. Preliminary estimating the reproduction number of the coronavirus disease (COVID-19) outbreak in Republic of Korea and Italy by 5 March 2020. medRxiv 2020:2020.03.02.20030312.

[10] Zhan C, Tse CK, Lai Z, *et al*. Prediction of COVID-19 Spreading Profiles in South Korea, Italy and Iran by Data-Driven Coding. medRxiv 2020:2020.03.08.20032847.

[11] Zheng Z, Wu K, Yao Z, *et al*. The Prediction for Development of COVID-19 in Global Major Epidemic Areas Through Empirical Trends in China by Utilizing State Transition Matrix Model. medRxiv 2020:2020.03.10.20033670.

[12] Li L, Yang Z, Dang Z, et al. Propagation analysis and prediction of the COVID-19. medRxiv 2020:2020.03.14.20036202.

[13] Bandoy DJDR, Weimer BC. Pandemic dynamics of COVID-19 using epidemic stage, instantaneous reproductive number and pathogen genome identity (GENI) score: modeling molecular epidemiology. medRxiv 2020:2020.03.17.20037481.

[14] Qinghe Liu, Zhicheng Liu, Deqiang Li, Zefei Gao, Junkai Zhu, Junyan Yang, Qiao Wang, Assessing the Tendency of 2019-nCoV (COVID-19) Outbreak in China, medRxiv, 2020.02.09.20021444.

[15] Aron JL, Schwartz IB. Seasonality and period-doubling bifurcations in an epidemic model[J].Journal of theoretical biology, 1984, 110(4): 665-679.

[16] Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology[J]. Mathematical biosciences, 1995, 125(2): 155-164.

[17] Li MY, Smith HL, Wang L. Global dynamics of an SEIR epidemic model with vertical transmission[J]. SIAM Journal on Applied Mathematics,2001,62(1):58-69.

[18] Hethcote H W . The Mathematics of Infectious Diseases[J]. SIAM Review, 2000, 42(4):599-653.

[19] Dietz, K. The estimation of the basic reproduction number for infectious diseases[J]. Statistical Methods in Medical Research, 1993, 2(1):23-41.

[20] G. Chowell, P.W. Fenimore, M.A. Castillo-Garsow, C. Castillo-Chavez. SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism [J]. Journal of Theoretical Biology, 2003, 224(1): 1–8.

[21] Riley S , Fraser C , Ca. D , *et al*. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions[J]. Science, 2003, 300(5627):p. 1961-1966.

Appendix 1 Data Source (All data used in paper is public.)

National-Level Migration Data:
https://bluehub.jrc.ec.europa.eu/migration/app/

Appendix 2 Simulation Results in 96 Main Countries or Regions

Data used as of March 27, 2020. All countries or regions are in alphabetical order.

	Country	Peak value	Peak time	Inflection point	Estimated R_0
1	Afghanistan	537	2020-05-15	2020-04-03	2.37
2	Albania	400	2020-04-17	2020-03-28	2.51
3	Algeria	1389	2020-05-04	2020-04-01	2.36
4	Andorra	690	2020-05-02	2020-03-29	2.72
5	Argentina	2715	2020-05-04	2020-04-03	2.42
6	Armenia	599	2020-04-17	2020-03-28	2.60
7	Australia	7247	2020-04-23	2020-03-28	2.65
8	Austria	18671	2020-07-25	2020-03-28	2.59
9	Azerbaijan	428	2020-04-28	2020-03-31	2.23
10	Bahrain	524	2020-04-09	2020-03-19	2.01
11	Belgium	15097	2020-04-18	2020-03-29	2.55
12	Bosnia and Herzegovina	678	2020-04-28	2020-04-01	2.14
13	Brazil	25424	2020-05-08	2020-04-08	2.70
14	Brunei Darussalam	249	2020-04-16	2020-03-27	2.48
15	Bulgaria	1199	2020-04-24	2020-04-04	2.25
16	Burkina Faso	1218	2020-05-06	2020-04-05	2.30
17	Canada	13468	2020-04-23	2020-03-29	2.67
18	Chile	5258	2020-04-23	2020-03-30	2.62
19	China	85166	2020-04-03	2020-02-21	2.28
20	Colombia	3744	2020-05-21	2020-04-06	2.41
21	Costa Rica	821	2020-04-23	2020-04-03	2.27
22	Côte d'Ivoire	1319	2020-05-10	2020-04-07	2.55
23	Croatia	1273	2020-04-20	2020-03-29	2.67
24	Cyprus	438	2020-04-18	2020-03-29	2.29
25	Czech Republic	5218	2020-04-18	2020-03-28	2.45
26	Denmark	3555	2020-04-15	2020-03-28	2.41
27	Dominican Republic	2017	2020-04-20	2020-03-30	2.54
28	Ecuador	5191	2020-04-26	2020-03-29	2.53
29	Egypt	844	2020-04-24	2020-03-24	2.30
30	Estonia	838	2020-06-24	2020-03-23	2.68
31	Faroe Islands	226	2020-04-09	2020-03-26	2.07
32	Finland	1702	2020-04-26	2020-03-29	2.20

33	France	63851	2020-04-19	2020-03-29	3.61
34	Germany	105903	2020-04-21	2020-03-29	2.72
35	Ghana	4488	2020-05-05	2020-04-16	2.53
36	Greece	1496	2020-04-19	2020-03-26	2.27
37	Hungary	1084	2020-04-27	2020-04-03	2.27
38	Iceland	1524	2020-04-16	2020-03-28	2.15
39	India	3326	2020-05-09	2020-04-02	2.41
40	Indonesia	4638	2020-05-13	2020-04-03	2.43
41	Iran	45932	2020-04-11	2020-03-25	3.55
42	Iraq	668	2020-04-22	2020-03-24	2.86
43	Ireland	5707	2020-04-29	2020-03-29	2.59
44	Israel	10740	2020-04-25	2020-03-30	2.78
45	Italy	127283	2020-04-18	2020-03-28	2.67
46	Japan	1737	2020-05-24	2020-03-13	3.22
47	Jordan	949	2020-05-10	2020-04-03	2.28
48	Kazakhstan	2329	2020-05-04	2020-04-11	2.33
49	Korea	10451	2020-04-08	2020-03-09	2.31
50	Kuwait	295	2020-04-20	2020-03-18	1.68
51	Latvia	600	2020-04-28	2020-03-28	2.23
52	Lebanon	786	2020-04-25	2020-03-28	2.20
53	Lithuania	1069	2020-06-10	2020-03-29	2.73
54	Luxembourg	4316	2020-04-20	2020-03-29	2.44
55	Malaysia	4296	2020-04-17	2020-03-29	2.39
56	Malta	310	2020-05-26	2020-03-27	2.22
57	Mexico	4308	2020-06-17	2020-04-07	2.36
58	Moldova	775	2020-04-23	2020-04-03	2.28
59	Morocco	2397	2020-05-15	2020-04-07	2.41
60	Netherlands	20893	2020-04-25	2020-03-28	2.61
61	New Zealand	2102	2020-04-20	2020-04-01	2.62
62	North Macedonia	486	2020-05-17	2020-03-28	2.67
63	Norway	7002	2020-04-17	2020-03-28	2.34
64	Oman	230	2020-04-18	2020-03-26	2.21
65	Pakistan	4939	2020-05-13	2020-04-02	2.35
66	Panama	2854	2020-04-24	2020-03-30	2.64
67	Peru	3992	2020-05-01	2020-04-06	2.29
68	Philippines	2972	2020-05-09	2020-04-02	2.35
69	Poland	5586	2020-04-24	2020-04-06	2.51
70	Portugal	13108	2020-05-05	2020-03-30	2.66
71	Qatar	819	2020-04-17	2020-03-21	2.46
72	Réunion	429	2020-04-27	2020-03-29	2.65
73	Romania	3413	2020-04-18	2020-04-01	2.52
74	Russia	6774	2020-05-14	2020-04-06	2.52
75	San Marino	311	2020-06-04	2020-03-21	2.14

76	Saudi Arabia	4335	2020-04-27	2020-04-04	2.47
77	Senegal	280	2020-04-29	2020-03-28	2.34
78	Serbia	1481	2020-04-21	2020-04-03	2.25
79	Singapore	847	2020-04-16	2020-03-20	3.15
80	Slovakia	879	2020-04-25	2020-04-03	2.27
81	Slovenia	1168	2020-04-25	2020-03-27	2.76
82	South Africa	13614	2020-05-10	2020-04-07	2.70
83	Spain	152539	2020-04-22	2020-03-29	2.66
84	Sri Lanka	200	2020-04-23	2020-03-23	2.28
85	Sweden	3974	2020-07-23	2020-03-23	2.21
86	Switzerland	28290	2020-04-24	2020-03-28	2.64
87	Thailand	5112	2020-05-15	2020-04-03	2.32
88	The United Kingdom	41228	2020-05-02	2020-03-30	2.69
89	The United States	400892	2020-04-28	2020-03-30	2.82
90	Tunisia	1723	2020-05-06	2020-04-08	2.38
91	Turkey	76824	2020-05-03	2020-04-03	3.11
92	Ukraine	10285	2020-04-23	2020-04-10	2.69
93	United Arab Emirates	643	2020-05-07	2020-03-26	2.21
94	Uruguay	2030	2020-04-26	2020-04-01	2.64
95	Venezuela	657	2020-04-21	2020-04-02	2.42
96	Viet Nam	211	2020-04-17	2020-03-20	2.31

Appendix 3 Clustering Results and the Estimated SEIR Parameters in 96 Main Countries or Regions

Data used as of March 27, 2020. All countries or regions are in alphabetical order.

	Country	α	β	γ_1	γ_2	c	Cluster
1	Afghanistan	0.55	1.36	0.00099	0.00099	0.086	1
2	Albania	0.70	1.84	0.00082	0.00614	0.105	2
3	Algeria	0.56	1.30	0.00697	0.00805	0.061	1
4	Andorra	0.64	1.75	0.00035	0.00028	0.096	2
5	Argentina	0.60	1.48	0.00753	0.00753	0.066	1
6	Armenia	0.66	1.75	0.00120	0.00180	0.097	2
7	Australia	0.63	1.63	0.00238	0.00238	0.062	1
8	Austria	0.64	1.69	0.00488	0.00264	0.057	2
9	Azerbaijan	0.59	1.34	0.00616	0.00552	0.088	1
10	Bahrain	0.60	1.24	0.01511	0.01358	0.066	1
11	Belgium	0.63	1.65	0.00581	0.00581	0.057	2
12	Bosnia and Herzegovina	0.58	1.28	0.00077	0.00082	0.071	1
13	Brazil	0.74	1.80	0.00013	0.00018	0.063	2

14	Brunei Darussalam	0.60	1.32	0.00556	0.00412	0.071	1
15	Bulgaria	0.68	1.52	0.00378	0.00139	0.074	2
16	Burkina Faso	0.66	1.55	0.00670	0.00591	0.076	1
17	Canada	0.67	1.87	0.00211	0.00239	0.062	2
18	Chile	0.62	1.60	0.00281	0.00281	0.063	1
19	China	0.78	1.54	0.00928	0.01764	0.044	3
20	Colombia	0.59	1.46	0.00090	0.00091	0.066	1
21	Costa Rica	0.57	1.31	0.00260	0.00130	0.070	1
22	Côte d'Ivoire	0.61	1.49	0.00543	0.00710	0.069	1
23	Croatia	0.63	1.69	0.00223	0.00124	0.082	2
24	Cyprus	0.70	1.79	0.00395	0.00111	0.103	2
25	Czech Republic	0.62	1.54	0.00498	0.00498	0.061	1
26	Denmark	0.63	1.57	0.00654	0.00498	0.065	1
27	Dominican Republic	0.65	1.72	0.00238	0.00238	0.067	2
28	Ecuador	0.67	1.81	0.00191	0.00191	0.068	2
29	Egypt	0.58	1.32	0.00512	0.00673	0.070	1
30	Estonia	0.67	1.79	0.00067	0.00026	0.098	2
31	Faroe Islands	0.56	1.24	0.00295	0.01295	0.074	1
32	Finland	0.58	1.37	0.00427	0.00111	0.066	1
33	France	0.63	1.67	0.00598	0.00443	0.051	2
34	Germany	0.67	1.84	0.00224	0.00238	0.054	2
35	Ghana	0.68	1.89	0.00069	0.00042	0.065	2
36	Greece	0.61	1.41	0.00535	0.00229	0.070	1
37	Hungary	0.56	1.28	0.00516	0.00516	0.068	1
38	Iceland	0.60	1.45	0.00598	0.00136	0.070	1
39	India	0.61	1.45	0.00682	0.00337	0.065	1
40	Indonesia	0.62	1.46	0.00493	0.00385	0.065	1
41	Iran	0.71	1.80	0.00471	0.01441	0.054	3
42	Iraq	0.31	1.25	0.00317	0.00617	0.072	4
43	Ireland	0.63	1.62	0.00066	0.00066	0.063	2
44	Israel	0.70	1.98	0.00123	0.00130	0.065	2
45	Italy	0.52	1.78	0.00084	0.00261	0.054	2
46	Japan	0.41	1.30	0.01199	0.00001	0.067	4
47	Jordan	0.58	1.37	0.00516	0.00329	0.070	1
48	Kazakhstan	0.71	1.64	0.00520	0.00139	0.084	2
49	Korea	0.60	1.41	0.00905	0.00211	0.056	1
50	Kuwait	0.61	1.18	0.00179	0.00311	0.079	1
51	Latvia	0.65	1.74	0.00115	0.00000	0.099	2
52	Lebanon	0.54	1.20	0.00442	0.00019	0.068	1
53	Lithuania	0.66	1.82	0.00027	0.00031	0.088	2
54	Luxembourg	0.62	1.62	0.00111	0.00066	0.063	2
55	Malaysia	0.62	1.47	0.00760	0.00498	0.062	1
56	Malta	0.60	1.25	0.00189	0.00147	0.073	1

57	Mexico	0.59	1.44	0.00104	0.00104	0.068	1
58	Moldova	0.58	1.35	0.00130	0.00130	0.073	1
59	Morocco	0.59	1.43	0.00390	0.00390	0.069	1
60	Netherlands	0.66	1.69	0.00296	0.00296	0.058	2
61	New Zealand	0.64	1.71	0.00896	0.00896	0.066	1
62	North Macedonia	0.66	1.75	0.00055	0.00055	0.098	2
63	Norway	0.63	1.53	0.00607	0.00337	0.058	1
64	Oman	0.40	0.98	0.00712	0.00939	0.066	4
65	Pakistan	0.66	1.62	0.00356	0.00356	0.068	1
66	Panama	0.70	1.79	0.00105	0.00108	0.068	2
67	Peru	0.63	1.55	0.00545	0.00545	0.067	1
68	Philippines	0.61	1.43	0.00441	0.00050	0.068	1
69	Poland	0.62	1.59	0.00498	0.00229	0.064	1
70	Portugal	0.67	1.85	0.00124	0.00124	0.063	2
71	Qatar	0.73	1.97	0.00135	0.00118	0.104	2
72	Réunion	0.66	1.76	0.00041	0.00065	0.097	2
73	Romania	0.63	1.58	0.00662	0.00533	0.065	1
74	Russia	0.61	1.55	0.00529	0.00453	0.066	1
75	San Marino	0.59	1.17	0.00327	0.00000	0.076	1
76	Saudi Arabia	0.65	1.45	0.00241	0.00229	0.061	1
77	Senegal	0.52	1.37	0.00506	0.00262	0.091	1
78	Serbia	0.61	1.45	0.00313	0.00127	0.069	1
79	Singapore	0.61	1.30	0.01355	0.00890	0.066	1
80	Slovakia	0.59	1.40	0.00735	0.00645	0.073	1
81	Slovenia	0.76	1.93	0.00268	0.00044	0.095	2
82	South Africa	0.65	1.74	0.00255	0.00255	0.066	2
83	Spain	0.66	1.77	0.00346	0.00352	0.050	2
84	Sri Lanka	0.59	1.40	0.00361	0.00049	0.098	1
85	Sweden	0.62	1.45	0.00685	0.00498	0.061	1
86	Switzerland	0.66	1.78	0.00033	0.00033	0.059	2
87	Thailand	0.63	1.59	0.00465	0.00452	0.066	1
88	The United Kingdom	0.73	1.76	0.00013	0.00020	0.058	2
89	Tunisia	0.59	1.44	0.00103	0.00073	0.069	1
90	Turkey	0.83	2.55	0.00114	0.00131	0.068	2
91	Ukraine	0.61	1.57	0.00156	0.00156	0.068	1
92	United Arab Emirates	0.39	1.26	0.00382	0.00481	0.076	4
93	United States	0.69	1.95	0.00037	0.00043	0.047	2
94	Uruguay	0.74	2.09	0.00373	0.00208	0.083	2
95	Venezuela	0.75	1.97	0.01853	0.01686	0.098	3
96	Viet Nam	0.62	1.21	0.00775	0.00274	0.083	1