Epidemiological Model for COVID-19 in China

Shanshan Wu¹,²†, Panpan Sun¹†, Ruiling Li², Yanli Wang¹, Lifang Jiang¹, Jinbo Deng¹*

¹National Health Commission Key Laboratory of Birth Defects Prevention, Henan Key Laboratory of Population Defects Prevention, Henan Institute of Reproduction Health Science and Technology, Zhengzhou, Henan Province, China
²School of Nursing and Health, Henan University, Kaifeng, Henan Province, China
†These authors contributed equally to this work.

*Correspondence Author: Jinbo Deng, Email: jinbo_deng2017@126.com, Tel: +86-371-88966141, Fax: +86-371-65961650

Received: July 31, 2020; Accepted: November 9, 2020; Published: December 11, 2020 DOI: 10.36922/itps.v3i2.938

Copyright: © 2020 Wu, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International 4.0 (CC BY-NC 4.0), which permits all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract:

Background. The epidemic of coronavirus disease 2019 (COVID-19) caused by the SARS-CoV-2 first broke out in Wuhan, Hubei Province in China, and then spread quickly worldwide.

Objective. This study aimed to dissect the spread and end of the epidemic in China with a precise mathematical model.

Methods. Various data were obtained from the official websites of the Chinese National Health from January 20 to July 8, 2020. The Chinese study participants were divided into three groups, namely, Hubei (including Wuhan), nationwide without Hubei, and Henan. The basic reproduction number ($R_0$), effective reproduction number ($R_t$), and gender and age ratio of COVID-19 were calculated, and the epidemic’s predicted curves or fitting curves with peak time and end time were plotted with SIR model. These predicted curves were compared with actual scatter plots.

Results. The fitting curve of the Hubei group showed a parabola with a peak on February 18, 2020, with 51,673 cases and the gradual decrease of infected patients, which culminates with a downhill after May 2020. During early outbreak, the highest recorded $R_0$ was 6.13, which declined gradually forming a S-type curve, and it approached zero in early May. Similar to Hubei group, the fitting curve of the nationwide without Hubei group also showed a parabola, recording a peak of 9145 cases on February 10, 2020. At first, its $R_0$ was as high as 2.35 but declined to zero in early April. The epidemic in the Henan group also reached its peak on February 10, 2020, and ended in early April as well.

Conclusion. The epidemic development of COVID-19 in China followed the shape of parabolic curves. This model provides insights into how to strategize for epidemic control.

Keywords: Basic reproduction number, Coronavirus disease 2019, Epidemic forecast, Epidemiology, Mathematical model

1 Introduction

Novel coronavirus pneumonia or coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 has become a rampant epidemic in Mainland China. The confirmed cases of COVID-19 demonstrated severe decrease in systemic immunity, and serious respiratory symptoms, and high mortality rate [1]. First recorded in Wuhan, Hubei Province of China, the SARS-CoV-2 infections transmitted in multiple routes and manifested high level of contagiousness, presenting difficulties in the prevention and control efforts which were further encumbered by the presence of asymptomatic cases [2,3]. Therefore, the World Health Organization (WHO) officially listed COVID-19 as a “public health emergency of international concern” [4]. By July 8, 2020, there were 85,366 confirmed cases in China. The tremendous and rapid spread of the infections was attributed to the travel rush (known as “Spring Movement” in China) happened during
Chinese Spring Festival. About 5 million people left Wuhan before a lockdown was imposed that stranded 9 million people in the city starting January 23, 2020, and the outbreak quickly escalated into an epidemic in Wuhan [5]. The people who left Wuhan at that point became new source of infections in other areas of China. While the number of infections was growing in Wuhan and other areas of China, it also expanded to the other countries [6,7]. The basic reproduction number ($R_0$), a measure of an infection’s contagiousness, of COVID-19 was as high as 4.6 from January 11 to 18 and 3.2 on February 29 [8].

Predicting the development of epidemic of an infectious disease is a crucial part of effective prevention and control. Epidemiologists can estimate the epidemiological development based on parameters such as disease onset, peak and culmination, as well as predict pathogenicity, onset region and season according to their understanding about an infectious disease. However, mathematical tools are regarded as scientific and accurate tools to forecast the epidemic’s development. Some of the mathematical models include SI model, SIS model, SIR model, SEIR model, SLIAR model, and SEQIJR model. However, classic SIR (susceptible-infected-recovered) model is popular in the study of some infectious diseases, especially those that develop acquired immunity after infection, such as influenza [9], malaria [10], Ebola [11], dengue disease [12], SARS [13], and COVID-19 [14,15].

SIR model has its advantages over other models for its scientific advance, practicality, and simplicity. While using a SIR model for predicting the disease development, a population is divided into three categories: Susceptible crowd, infectious crowd, and recovered crowd. The crowds are changeable from one to another, depending on the disease transmission ($\beta$) and recovery rate ($\gamma$). SIR model analyzes the disease’s transmission and computes the epidemiological parameters that describe the disease development. The comparison of the resultant fitting curve with actual scatter plots might reveal some useful and insightful information about the disease. Therefore, in the present study, SIR model was selected to study the transmission of COVID-19. In addition, the $R_0$ and dynamic values of effective reproduction number ($R_e$) were described. The actual scatter plots and fitting curves of COVID-19 in different groups were compared.

2 Methods

2.1 Data source and grouping

The data of confirmed COVID-19 cases with other information were obtained from the official websites of National Centers for Disease Control (CDC) and National Health Committee (NHC) of the People’s Republic of China from January 20 to July 8, 2020. The confirmed cases were diagnosed according to the diagnostic criteria for COVID-19, including exposure history, clinical symptoms (especially computed tomography imaging), and nucleic acid test results, which were suggested by the National Health Committee of China [16,17].

The cases were classified into three groups according to their geographical regions: (i) Hubei (including Wuhan city), since the disease first broke out in the Hubei Province and the region was the first place put under lockdown; (ii) nationwide except Hubei Province (abbreviated as nationwide thereafter), since the people who escape the city before lockdown were sources of infections; and (iii) Henan, since Henan Province, with a population of more than 100 million, is the most populous province in China and has frequent population exchanges with Wuhan.

2.2 Parameters

Some epidemiological indicators, such as morbidity and mortality, are adopted in the epidemiological study. Since the gender and age are closely related to COVID-19, the gender and age ratio was analyzed. The data of Chinese population of various ages were derived from the 2015 census obtained from the National Bureau of Statistics, People’s Republic of China [18]. Several relevant formulas are as follows:

\[
\text{Attack rate} (\%) = \frac{\text{Number of confirmed cases}}{\text{Number of persons in a certain period}} \times 100
\]

\[
\text{Case fatality rate (\%)} = \frac{\text{Number of death attributed to COVID-19}}{\text{Number of confirmed cases}} \times 100
\]
Gender ratio = \frac{\text{Number of male cases}}{\text{Number of female cases}}

Age–related attack or crude death rate = \frac{\text{Number of cases of various ages}}{\text{Total number of people of same age in the population}}

\( R_0 \), which refers to how many persons can be infected by a patient [19,20], was also calculated. If \( R_0 > 1 \), the disease will spread out quickly with a radial pattern, whereas \( R_0 < 1 \) means that the infection is controllable. Thus, 1 is regarded as an important threshold to distinguish whether the disease will spread [20]. \( R_0 \) can be calculated with the formula, \( R_0 = \frac{S_0 \times \beta}{\gamma} \). The dynamic reproduction number \( (R_t) \) was also calculated. During the spread of an infectious disease, \( R_t \) is subject to change. In fact, one of the strategies of epidemiological control is to find a way to reduce \( R_t \) to below 1 as much as possible.

The above-mentioned indicators, which are only descriptive parameters, are incapable of providing a clear insight into the transmission of COVID-19; therefore, more works should be done on the SIR model. The susceptible cases (S), infected cases (I), recovered cases (R), transmission rate (\( \beta \)), recovery rate (\( \gamma \)), and \( R_0 \) are necessary for building the SIR model. With three-order Runge-Kutta method, least square state estimation and serial differential equations \([N=S(t)+I(t)+R(t), dS(t)d(t)=-\beta S(t)I(t), dI(t)d(t)=\beta S(t)I(t)-\gamma I(t), dR(t)d(t)=\gamma I(t)]\), the important information about COVID-19, such as onset, peak, and ending time, can be determined.

2.3 Statistical analysis

Microsoft Excel was used to record data and calculate some simple parameters. The SIR model was built with MATLAB software, and the fitting curves or forecast curves were created according to original data. In this study, the fitting curve was plotted in a way to imitate the epidemic in reality as much as possible, portraying the trend, and peak and end of the epidemic. In addition, the curves of \( R_t \) were drawn as well. Finally, the results were compared among the three groups, such as Hubei, nationwide, and Henan. The comparison with \( P < 0.05 \) is considered statistically significant.

3 Results

3.1 Actual scatter plots and disease infection

The actual scatter plots were created based on the original data. During the early stage of COVID-19 outbreak, either patient number or incidence increased with exponential curves in all three groups (Table 1 and Figure 1). For instance, from January 20 to February 14, the cases increased quickly from 270 to 54,406 in Hubei group, and correspondingly, the attack rates rose from \( 0.46 \times 10^{-5} \) to \( 91.95 \times 10^{-5} \). From January 20 to February 14, the cases increased from 0.00 to 12,086 in Henan group, and the attack rates rose from 0.00 to \( 0.90 \times 10^{-5} \). From January 20 to February 11, the attack rates of the nationwide group rose from 0.00 to \( 0.84 \times 10^{-5} \). Compared to Hubei group, the disease transmission in the nationwide group had been maintained at low level, but epidemiological curve shows exponential increase which is similar to Hubei group (Figure 2A-C). Following the peaks, the number of infected cases decreased in all three groups until the end of the epidemic.

The cumulative confirmed cases, attack, and fatality rates of the three groups are shown in Figure 1A-C. The attack rate was higher in Hubei Province and increased quickly as compared to other areas of China. In addition, we also analyzed the attack rate of COVID-19 related to gender and age. Our results showed that the attack rates were different between males and females. For instance, male-to-female ratio was 1.01/1 (44,627 cases for the statistical analysis), and the gender ratio for mortality was about 1.67/1.00 (male/female). The correlation between infection and age was also analyzed for their attack and mortality rates (Figure 1D, \( P < 0.05 \)). For instance, the infection and mortality rates of those aged >45 years increased, and the attack and death rates of those aged >80 years were 0.35 % and 14.77 %, respectively, (Figure 1D).

3.2 Mathematical model for the epidemic

With the original data and a series of formulae, the SIR model was used to analyze the epidemic, and the fitting curves were plotted for the three groups (Figure 2A-C). The fitting curves were plotted to imitate the epidemic in reality to forecast the epidemic spread. In the current study, the predicted
## Table 1. Epidemiological data among the Hubei, nationwide, and Henan groups

<table>
<thead>
<tr>
<th>Date*</th>
<th>Hubei Cases</th>
<th>Attack rates (×1/10^5)</th>
<th>Case fatality rates (%)</th>
<th>Nationwide Cases</th>
<th>Attack rates (×1/10^5)</th>
<th>Case fatality rates (%)</th>
<th>Henan Cases</th>
<th>Attack rates (×1/10^5)</th>
<th>Case fatality rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/20</td>
<td>270</td>
<td>0.46</td>
<td>2.20</td>
<td>21</td>
<td>0.00</td>
<td>--</td>
<td>0</td>
<td>0.00</td>
<td>--</td>
</tr>
<tr>
<td>1/25</td>
<td>1052</td>
<td>1.78</td>
<td>4.94</td>
<td>923</td>
<td>0.07</td>
<td>0.43</td>
<td>83</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>2/4</td>
<td>16678</td>
<td>28.19</td>
<td>4.94</td>
<td>7646</td>
<td>0.57</td>
<td>0.17</td>
<td>764</td>
<td>0.80</td>
<td>0.26</td>
</tr>
<tr>
<td>2/9</td>
<td>29631</td>
<td>50.08</td>
<td>7.65</td>
<td>10540</td>
<td>0.79</td>
<td>0.35</td>
<td>1073</td>
<td>1.12</td>
<td>0.56</td>
</tr>
<tr>
<td>2/14</td>
<td>54406</td>
<td>91.95</td>
<td>2.87</td>
<td>12086</td>
<td>0.90</td>
<td>0.55</td>
<td>1212</td>
<td>1.26</td>
<td>1.07</td>
</tr>
<tr>
<td>2/19</td>
<td>62457</td>
<td>105.56</td>
<td>3.25</td>
<td>12545</td>
<td>0.94</td>
<td>0.71</td>
<td>1265</td>
<td>1.32</td>
<td>1.50</td>
</tr>
<tr>
<td>2/24</td>
<td>64786</td>
<td>119.11</td>
<td>3.96</td>
<td>12872</td>
<td>0.96</td>
<td>0.78</td>
<td>1271</td>
<td>1.32</td>
<td>1.49</td>
</tr>
<tr>
<td>2/29</td>
<td>66907</td>
<td>114.23</td>
<td>4.13</td>
<td>12917</td>
<td>0.96</td>
<td>0.84</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>3/5</td>
<td>67592</td>
<td>114.23</td>
<td>4.34</td>
<td>12960</td>
<td>0.97</td>
<td>0.86</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>3/10</td>
<td>67773</td>
<td>114.54</td>
<td>4.49</td>
<td>13005</td>
<td>0.97</td>
<td>0.86</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>3/15</td>
<td>67798</td>
<td>114.59</td>
<td>4.57</td>
<td>13062</td>
<td>0.97</td>
<td>0.87</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>3/25</td>
<td>67801</td>
<td>114.59</td>
<td>4.67</td>
<td>13484</td>
<td>1.01</td>
<td>0.88</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>4/4</td>
<td>67803</td>
<td>114.59</td>
<td>4.73</td>
<td>13866</td>
<td>1.03</td>
<td>0.86</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>4/24</td>
<td>68128</td>
<td>115.14</td>
<td>6.62</td>
<td>14688</td>
<td>1.10</td>
<td>0.82</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>5/14</td>
<td>68128</td>
<td>115.14</td>
<td>6.62</td>
<td>14805</td>
<td>1.10</td>
<td>0.82</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>6/3</td>
<td>68128</td>
<td>115.14</td>
<td>6.62</td>
<td>14894</td>
<td>1.11</td>
<td>0.82</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>6/23</td>
<td>68128</td>
<td>115.14</td>
<td>6.62</td>
<td>15302</td>
<td>1.14</td>
<td>0.80</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
<tr>
<td>7/3</td>
<td>68128</td>
<td>115.14</td>
<td>6.62</td>
<td>15417</td>
<td>1.15</td>
<td>0.79</td>
<td>1272</td>
<td>1.32</td>
<td>1.73</td>
</tr>
</tbody>
</table>

Before March 25, the original data were released once every 5 days. Afterward, data were released once every 20 days. For statistical analyses, day-by-day data were used for all analyses. *Dates in 2020 (month/day)

results recapitulated the actual situations with negligible differences either in developmental trend or the number of infected cases ($P < 0.05$). In Figure 2A, the simulated curve of Hubei group showed that the number of infected patients increased and reached its peak on February 18 with a total number of 51,673 cases at the beginning, and then the curve went down. Finally, the number of infected patients fell to the lowest level after early May, heralding the end of the epidemic. The forecast curves were also fitted in the nationwide and Henan groups (Figure 2B and C, $P < 0.05$). The number of cases of the nationwide and Henan groups similarly reached their peaks with 9145 cases and 925 cases, respectively, on February 10. After the peaks, the cases gradually decreased and then bottomed out in the early of April 2020, and the epidemic was predicted to end subsequently. According to Figure 2, the trend of fitting curves and actual curves resemble each other, suggesting that our model is useful and practical for the forecast and surveillance of COVID-19 epidemic.

**Figure 2D** shows that $R_t$ changes in Hubei group with a S-type curve. It started from 6.13, decreased gradually to 1 around February 18, and approached 0 in early May ($P < 0.05$). Both nationwide and Henan groups also demonstrated dynamic changes in $R_t$, which started at 2.35 and 2.40, respectively, and decreased to <1 after February 10. Finally, they approached zero in early April. Both fitting curves and $R_t$ values indicate that the epidemic would end in early April and early May in the nationwide and Hubei groups, respectively.

### 4 Discussion

Since COVID-19 is a severe and highly transmissible illness, understanding its epidemiological development would, therefore, be an invaluable approach for forecasting its trend. With the use of mathematical models, the scientists and epidemiologists can compute some epidemiological parameters, including peak time and end time, but these still require more investigations for ascertainment.

SIR was chosen in our study since it is more suitable for acquired immune diseases, such as COVID-19, whose transmission shows a typical
process. Our results show that the development of COVID-19 in China also followed a typical course of development trend, encompassing phases such as onset, peaking, flattening, and ending. According to our study, the exponential increase of confirmed cases was accompanied by its peaks approximately on February 10 in various parts of Mainland China except Hubei Province, and on February 18 in Hubei Province. Subsequently, the number of COVID-19 patients decreased gradually, and the epidemic disappeared in early April (for the nationwide group) and early May (for the Hubei group). Our forecast mirrored the actual situation, even though our findings or forecasts were somehow different from Wu et al. [21]. For instance, the peak value and cumulative cases in the present study are lesser, and the predicted end time happened earlier. Such differences may be caused by the different sets of data and time. Since the data used, in the present study, were derived from official publications, it is probable that our mathematical model could simulate the actual situations to a greater extent.

\( R_0 \) is used to evaluate the contagiousness of infectious disease. According to the WHO, the typical range of \( R_0 \) for disease transmission is 1.4 – 2.5 [22]. Several studies reported that in the early outbreak of COVID-19 in Wuhan, its \( R_0 \) was found to be in a range from 2.24 (95% CI: 1.96 – 2.55) to 3.58 (95% CI: 2.89 – 4.39) [23] or from 2 to 5 [24-27], consistent with \( R_0 \) of SARS-CoV (\( R_0 \): 2 – 5) [28,29] and Middle East respiratory syndrome (\( R_0 \): 2.0 – 6.7) [30]. Our study showed that the \( R_0 \) ranged from 2.44 to 6.13, which is higher than the above-mentioned studies, suggesting the high contagiousness of COVID-19. Following April 2020, \( R_0 \) reduced to <1, suggesting that its transmission became weaker. Then, \( R_t \) also approached 0 after May in Hubei Province. In line with the model prediction as shown on the

**Figure 1.** The cumulative confirmed cases, attack, and fatality rates of COVID-19 in Hubei, nationwide, and Henan groups. (A) The number of cumulative confirmed cases of the Hubei group was higher than that of other groups. (B) The attack rate was higher in Hubei group comparing to other areas of China. Since the morbidity in other groups is very low, their curves almost overlap with each other. (C) The fatality rate of Hubei group was higher than the others. (D) The age-related attack rate (red) and death rate (blue) rose rapidly with age.
epidemic curves and Rt curves, the transmission of COVID-19 was successfully put under control in early May 2020; therefore, the lockdown order was lifted after May, allowing a return to normal life. However, since the viral infection is still very rampant in other countries, stringent surveillance was put in place as a preliminary effort to keep foreign import of cases at bay [31]. Nevertheless, our study showed that the SIR model may be able to predict the development trend of COVID-19 in China, thereby providing some important insights into COVID-19 elimination efforts.

Figure 2. Epidemic trends of COVID-19 with SIR mathematical model in Hubei, nationwide, and Henan groups. (A) The actual curve (red) and fitting curve (blue) of Hubei group peaked on February, and the epidemic was predicted to end in May. The actual situation bore a resemblance to the predicted results with little difference. (B) The actual curve (red) and fitting curve (blue) of the nationwide group peaked on February 10, and the epidemic was forecasted to end in early April. The actual situation bore a resemblance to the predicted results with little difference. (C) The actual curve (red) and fitting curve (blue) of Henan group peaked on February 10, and the epidemic was predicted to end at the end of March. (D) The dynamic Rₜ presented S-type curves in Hubei, nationwide, and Henan groups. In the Hubei group (green), at first, Rₜ was as high as 6.13 but decreased gradually, nearing 0 in early May. The Rₜ of the nationwide group (blue) approached 0 in early April. Interestingly, the curve of Henan group (red) almost overlaps with that of the nationwide group.

SARS-CoV-2 has sequences that harbor 79.5% and 51.8% similarity to SARS-CoV and MERS-CoV, respectively [32], indicating the high contagiousness of this novel coronavirus [33]. The sudden outbreak and spread of COVID-19 are more severe and rapid than those of SARS or MERS-CoV. In terms of contagion, the R₀ value in Hubei was 6.13 which was higher than 2.35 in other areas of China. With the SIR model, we forecasted that the epidemic would end in early May in Hubei, which was 1 month later than other parts of China.

In general, the success of outbreak control in China benefited from the efficient medical and administrative interventions, such as city lockdown and quarantine of infected and suspected patients. For instance, closed-off measures for towns and communities were also carried out in many areas throughout the country, traffics in city were blocked, and works and studies were temporarily suspended to abrogate potential transmission of infection. In addition, medical isolation in hospital
and solitary confinement at home are also part of the approach to battling the COVID-19 epidemic. To prevent further spread of the infection, the citizens were demanded to stay at home to break the chain of infection [34]. Furthermore, dispatching more front-liners and medical supplies, such as diagnostic kits and medicines, to COVID-19 hotspots is instrumental in controlling the epidemic.

5 Conclusion

In summary, with the use of classic SIR models, the fitting curves are able to simulate the spread of COVID-19 in China, and \( R_0 \) curves are able to evaluate the contagiousness of the disease. Based on our study, the development of the epidemic can be represented as a parabolic trend, and the SIR model may be a powerful and practical mathematical tool for the forecast and surveillance of COVID-19 epidemic development.

Acknowledgments

This study was supported by the research funds from the Medical Science and Technology Project of Henan Provincial Health Commission (LHGJ20190828), Scientific and Technical Project of Henan Science and Technology Department (192102310134), Open Project of National Health Commission Key Laboratory of Birth Defect Prevention (ZD201903) and Medical Foundation of Health and Family Planning Commission of Henan Province (No. 2018020589).

Conflict of interest

The authors declare that they have no conflicts of interest.

Author contributions

S.W. and P.S. wrote the initial draft. J.D. designed the project, provided financial support and edited subsequent revisions. All authors provided critical feedback and approved the final draft of the manuscript.

Editorial disclosure

A preprint version of this article is available at https://doi.org/10.1101/2020.02.21.20026229.

References


