Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
COVID-19: Is it safe now? Study of asymptomatic infection spread and quantity risk based on SAIR model

Liu Ying, Tang Xiaoping

School of Science, Shaoyang University, Shaoyang 422000, China
School of Statistics & Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai 201620, China

ARTICLE INFO

Article history:
Received 22 May 2020
Revised 11 November 2020
Accepted 14 May 2021
Available online 24 May 2021

Keywords:
COVID-19
SAIR model
Asymptomatic case
Quantity risk

ABSTRACT

Based on the characteristic of the COVID-19 asymptomatic infection, and due to the shortage of traditional mathematical models of transmission dynamics of infectious diseases, we propose a new SAIR model. This SAIR model fully considers the infectious characteristics of asymptomatic cases and the transformation characteristics between the four kinds case. According to the data released by the National Health Commission of P.R.C, the model parameters are calculated, and the transmission process of the COVID-19 is simulated dynamically. It is found that the SAIR model data are in good agreement with the actual data, and the time characteristics of the infection rate are particularly accurate, proving the accuracy and effectiveness of the model. Then, on the basis of the differences between the model data and the real data, the standard deviation of the error is calculated. From the standard deviation, the functional intervals of the confirmed infection rate and the asymptomatic infection rate, the interval of the total number of cases in the model, and the interval of the number of asymptomatic cases in the society are also calculated. The number of asymptomatic cases in society is of important and realistic significance for the assessment of risk and subsequent control measures. Then, according to the dynamic simulation data of the model with changed value of parameters, the remarkable effects of strict quarantines are discussed. Finally, the possible direction of further study is given.

© 2021 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Since the COVID-19 case was discovered in Wuhan in late December, 2019, the virus has been spreading rapidly in China and beyond. As of April 3, 2020, within only four months, there have been more than 80,000 confirmed cases and over 3000 deaths in China; and almost all of the world’s more than 100 countries and regions have been hit by the virus, with a total of 1 million confirmed cases and more than 50,000 deaths. Obviously, the number of confirmed cases and deaths of COVID-19 greatly surpass that of SARS in 2003. Now, the COVID-19 has caused panic among global people, and also raised concern of governments and the World Health Organization. It is a major public health emergency in the world. The prevention and treatment of COVID-19 is in dire need for an appropriate mathematical model for the development of the virus, so as to predict and control in follow-ups, and properly arrange the distribution of materials and personnel, treatment and implementation of control measures.

In particular, the COVID-19 has created a new case, that is, asymptomatic cases (suspected patients), who are virus carriers without showing symptoms (or only mild symptoms). However, this kind of case is also infectious. A few cases are as infectious as confirmed cases, while most of them are less infectious. There are three possibilities for the later development of asymptomatic cases. One, after a period of time, it turns into confirmed cases with more severe symptoms; two, the patients experience self-cure and recover in the future, with virus antibodies remaining in their body; three, the patients remain as asymptomatic virus carriers for a long time.

Due to the high cost of COVID-19 testing which needs to be done several times (at present, testing kits are not that efficient and people need to be tested at least twice for confirmation). Therefore, it is generally impossible to carry out testing of the population in the epidemic area on a large scale. Since asymptomatic cases do not show obvious symptoms but are infectious, they unknowingly spread the virus among people. It is fair to say that they are invisible “bombs” in the crowd, highly risky for public safety.
It is very important to quarantine these asymptomatic cases and trace the close contacts. Therefore, it is urgent and meaningful to establish a proper mathematical model and estimate the number of asymptomatic cases.

Famous traditional mathematical models of infectious diseases transmission include Exponential, SIS, SIR and SEIR model. However, each of these models has its own shortcomings. For the Exponential model, SIS model and SIR model, the asymptomatic case is not taken into consideration at all. According to the data released by the National Health Commission, there are a large number of asymptomatic cases. The incubation period of asymptomatic cases is not long, with an average incubation period of 5.2 days. To be on the safe side, the community generally requires a quarantine of 14 days. Therefore, the above mathematical models are inconsistent with the actual situations of COVID-19.

There are four types of case for SEIR model: S (susceptible), I (infection), R (recovery) and E (exposed) which refers to people who have been exposed to the infected but are unable to infect others. However, the asymptomatic cases of COVID-19 are infectious within the incubation period, so it is inconsistent with the E (exposed) in SEIR model. Moreover, SEIR is only applicable to infectious diseases with a long incubation period (at least more than 20 days), but the incubation period of COVID-19 is not long. Therefore, SEIR model is not suitable for the transmission dynamics of COVID-19. In view of the new characteristics of COVID-19 asymptomatic cases, a new infection dynamics SAIR model is proposed in this paper. Our SAIR model fully considers the infectious characteristics of asymptomatic cases and its three transformation possibilities, and the relationship characteristics between the four kinds of case. At present, there are many papers [1] study COVID-19, and most of COVID-19 research papers use SEIR model or its derivative model [2, 3], so they have theoretical defects.

Alfan et al. have done a good research on the nonlinear dynamic model [4], its novel hybrid forecasting model is very enlightening to this paper.

2. SAIR model

Model assumptions

(1) The total number of people in the area remains unchanged during the spread of the epidemic, regardless of births, deaths and migration. The resident population of Hubei is 59.17 million, with 11.081 million in Wuhan city. Also after the government locked down the city and province on January 23 and 27 in 2020 respectively, people were not allowed to travel in or out. That is to say, in the short term, compared with such a large population base, it can be approximately assumed that the total number of people remains unchanged.

(2) The population is divided into four groups: S (Susceptible people), A (Asymptomatic case), I (infection) and R (recovery).

(3) Suppose that the total population in the area is N, t represents a certain time of t, S (t) is the number of susceptible (i.e. healthy) persons at time t, A (t) is the number of asymptomatic cases at time t, I (t) is the number of infected cases (i.e. patients) at time t, R (t) is the number of recovered persons at time t.

\[
\text{SoS}(t) + A(t) + I(t) + R(t) = N.
\]

Their ratio is that as follows:

\[
\frac{S(t)}{N} = s(t), \quad \frac{A(t)}{N} = a(t), \quad \frac{I(t)}{N} = i(t), \quad \frac{R(t)}{N} = r(t)
\]

And

\[
s(t) + a(t) + i(t) + r(t) = 1.
\]

The relationship between the four kinds of people is shown in the following Fig. 1.

As the derivative is the growth rate, so the dynamic equations for each variable are as follows:

\[
\begin{align*}
\frac{dS(t)}{dt} &= -\beta_1 \cdot s(t) \cdot i(t) - \beta_2 \cdot s(t) \cdot a(t) \\
\frac{dA(t)}{dt} &= \beta_2 \cdot s(t) \cdot a(t) - \alpha_1 \cdot a(t) \cdot i(t) - \alpha_2 \cdot a(t) \cdot r(t) \\
\frac{dI(t)}{dt} &= \beta_1 \cdot s(t) \cdot i(t) - \gamma \cdot i(t) + \alpha_1 \cdot a(t) \cdot i(t) \\
\frac{dR(t)}{dt} &= \gamma \cdot i(t) + \alpha_2 \cdot a(t) \cdot r(t)
\end{align*}
\]

Where the parameter \(\beta_1\) is the susceptible-to-confirmed cases infection rate, and \(\beta_2\) is the susceptible-to-asymptomatic infection rate; \(\alpha_1\) is the conversion rate of the asymptomatic cases to the confirmed cases, \(\alpha_2\) is the conversion rate of the asymptomatic cases to the recovered; \(\gamma\) is the recovery rate of the confirmed cases.

Suppose that the average number of effective contacts per patient per day is a constant R0 (Note: where R0 is not the number of the recovered, but a medical term), which is called daily average contact rate; when the patient effectively contacts with a healthy person, the infection rate of turning the healthy person into an infected case is \(\beta_1\), and \(R_0 = \frac{\beta_1 \cdot S(t) \cdot O(t)}{\gamma} = \frac{\beta_1}{\gamma} = T\), where \(T\) is the average number of days of treatment to cure a patient, and \(\gamma\) is the recovery rate. With these parameters and formula, the model can be calculated. The data presented in the text are based on the facts that some data are being updated, that there are some missing medical-specific data, and that infectious disease control measures vary from place to place over time. Also, as treatment measures and methods continue to advance, data conclusions vary. In short, the data vary from one time period to another, and the conclusions of the calculations vary from one institution to another. For each parameter, we try to adopt the conclusions of the more prudent and reputable institutions and our own fitted data, which are calculated using the least square method.

3. COVID-19 data and the simulation of SAIR model

3.1. Data and their selection

On December 30, 2019, Dr. Li Wenliang sent a "whistle blowing" warning, and some hospitals in Wuhan city officially admitted the first COVID-19 cases on January 7 and 8, 2020. On January 23, 2020, the main expressways in Wuhan city were closed, and on January 27, the main expressways in Hubei province were also closed. However, before the lockdown of the city, news reports said, 5 million people left Wuhan city and moved across the country. Before the lockdown of Hubei province, the development and treatment of COVID-19 and its transmission in Wuhan city was relatively stable, and there were no significant changes. However, from January 7 to January 23, some data are missing. Around February 10, 19 provinces and cities across the country began to support Wuhan city and Hubei province on a large scale for COVID-19 treatment. Strict grid management was carried out in all districts of Hubei province and Wuhan city, and the treatment situation was
improved, making the parameters after that time presumably different from those before. Therefore, for the calculation of parameter value, this paper mainly selects the complete data in the previous period from January 23 to February 10. All the following data are from the National Health Commission [5].

The data is shown in Fig. 2.

Fig. 3 below shows the growth rate of each factor.

3.2. Determination of parameters

According to the latest evaluation data of WHO [6,7], the R0 of COVID-19 is 3.77, and the estimated interval is [2.23,4.82], which is large, so it is highly infectious. Since the average cure time of patients is more than 9 days, so here \( T = 10 \), thus \( \gamma = 0.1 \), \( \beta_1 = 0.377 \). The other parameters are obtained by model data fitting using the least square method. The results are shown in Table 1 below.

According to the parameter values, the SAIR model is used to simulate the development of the COVID-19, as shown in Fig. 4.

From the simulation data from Fig. 4, it can be clearly seen that the infection rate of the COVID-19 will reach its climax after about 20 days, and then gradually decrease. The result is basically consistent with the real situation. In fact, the first COVID-19 cases were received on January 7. Twenty days later on January 27, the real infection growth rate (confirmed cases) exactly reached a peak, and then gradually declined (as shown in Fig. 2). In addition, the growth rate of asymptomatic patients (suspected cases) is at the highest after about 19 days, that is, on January 26, which is also consistent with the real data. In this paper, the initial value of simulation is from January 7, 2020.

The above simulation process, especially the time characteristics of these data, fully demonstrates that the SAIR model is accurate and the simulation of virus development process is successful.

3.3. Number of confirmed cases and asymptomatic cases

According to the simulation process data of the mathematical model SAIR, comparison is made between the simulated infection data and the real day-by-day confirmed infection data, as is shown in the left part of Fig. 5 below (since the magnitude difference between the model infection rate and the real day-by-day confirmed case number is too large, the number of confirmed cases is divided by 100, and the simulated infection rate is multiplied by 100 to make the figure clear.). The standard deviation of the error between the simulated data and the real day-by-day data can be easily calculated: \( \sigma_1 = 0.0565 \). Note that on February 12, 2020, leadership personnel changes in Hubei province and changes in the case confirmation criteria, resulted in dramatic changes in the data on February 12 and 13. In this paper, the data from both two days are smoothed, and different treatments may lead to different conclusions. In addition, the symbol “△” in the mimic diagram represents the real data, and the symbol “- - -” represents the simulated data of the SAIR model.
Similarly, the real asymptomatic infection data is compared with the data simulated by the model, as shown in the right part of Fig. 5. The processing method is to divide the real asymptomatic data by 100, and multiply the model data by 100. The standard deviation of the error between the simulated data and the daily real data can be calculated as $\sigma^2 = 0.0286$. It can be found that the real data is highly consistent with the simulated data of the model, which proves the accuracy of the mathematical model in this paper.

The infection rate function area taking into consideration the error is shown in Fig. 6 below.

The curve area of asymptomatic infection rate considering the errors is shown in Fig. 7.

The interval of final cumulative number of confirmed cases (integer form) is calculated, as follows Table 2.

As of April 3, the number of confirmed cases in China was 81,639, which is in the case number interval in the SAIR model and fewer than the average of 101,793. This proves that the result is reliable. It shows that the control measures of "city lock-down" and community closure in China to reduce the flow of people and the infection rate have played a significant role, thus reducing the number of cases by about 20,000, or 24.5% of the total number. In fact, the infection rate has been falling since it peaked around January 27, 2020. If strict control measures were adopted earlier than the "city lock-down" on January 23, for example, by 5 days, the effect would be even more significant. This will be discussed later in this article.

Similarly, this model was used to calculate the number of asymptomatic cases remaining in the society (as of 3rd April 2020), as shown in Table 3.

Table 3 reveals that, although the epidemic in China has been controlled, with almost zero domestic cases and new cases mainly imported ones. However, the country should not lose guard in its control measures, because there are still at least 637 and up to 13,267 asymptomatic cases. Compared with confirmed cases, these people are 39.7% likely to transmit the virus. Therefore, social gatherings, especially schools, nursing homes, prisons, cinemas and theatres, large-scale concerts and sports events are susceptible to large-scale virus spread once transmission happens, wasting all control efforts made in the previous two months and triggering a second wave of epidemic will break out.

### 3.4. Containment of the COVID-19

Since the virus has not mutated and antibodies have yet formed in a short term, some parameters of the model are temporarily unchanged, but some parameters can be changed. We try to change the parameter value that can be changed, so as to reduce the damage of virus transmission. As $R_0 = N \cdot Contact\_prob$, and $N$ repre-
Table 3
Number of asymptomatic cases.

| Parameter | Value |
|-----------|-------|
| Lower bound | 637 |
| Upper bound | 13,267 |
| Average value | 6952 |
| Standard deviation | 0.0286 |

Table 4
Parameter value (changed).

| Parameter | Value |
|-----------|-------|
| \( \gamma \) | 0.12 |
| \( \beta_1 \) | 0.30 |
| \( \beta_2 \) | 0.36 |
| \( \alpha_1 \) | 0.10 |
| \( \alpha_2 \) | 0.31 |

Table 5
Number of confirmed cases (changed parameters).

| Parameter | Value |
|-----------|-------|
| Lower bound | 42,220 |
| Upper bound | 68,369 |
| Average value | 55,295 |
| Standard deviation | 0.0549 |

Similarly, the number of asymptomatic cases after the parameter was changed (on April 3rd) can be calculated, as shown in Table 6 below.

Fig. 8 clearly shows that the peak of infection comes about 27 days after the initial value, 7 days later than in the previous case (the previous parameter’s peak came 20 days after the initial value), thus gaining an additional 7 days of valuable response time. Also, the maximum infection rate is reduced (the previous parameter’s confirmed infection rate maximum was up to 26%, now the peak is only 14.2%). In addition, the peak of asymptomatic cases (suspected cases) was also delayed by about 2 days.

As can be seen from Table 5, the total number of infection cases decreased considerably after the control measures were taken, i.e., after adjusting the parameters, and the average value of the interval was 55,295, which is 54.4% of the average value of 101,793 before the parameters were changed. This result confirms the effectiveness of the control measures taken, i.e., the increased quarantine of the infected, suspected and healthy persons, the implementation of grid management to reduce the movement of people, the requirement that people wear masks, and the increased treatment of the infected and suspected persons. All these will reduce the total number of cases and will greatly accelerate the final control or elimination of the virus and the victory in beating over the epidemic.

As can be seen from Tables 6 and 5, when strict control measures were adopted to restrict the free movement of people, RO was changed from 3.77 to 2.5. Although population N did not change, the effective contact rate of patients was reduced by about 1/3. The average total number of confirmed cases will then fall from 101,793 to 55,295. Subsequently, the asymptomatic infections rate dropped by 84%, from 6952 to 1117. Since asymptomatic infections tend to be concentrated in the first part of the period and drops rapidly in the second part, the numbers are very small.

It should be noted that in this paper, as the simulation process software solves the differential equation, the MATLAB software essentially uses the numerical solution obtained from difference equation over non-equal time periods, and the daily infection rate was replaced by value of the closest point. If we use the integral of the whole time, the estimation result should be more accurate.

4. Conclusion and further discussion

In this paper, a new SAIR model is proposed based on the SIR model, aiming at the characteristics of asymptomatic infected cases with short incubation period and infectious of asymptomatic cases. We believe that this model has good theoretical innovation, at the same time, it has the characteristics of practical pertinences, it can better quantify the number of asymptomatic cases in the society, and the value of asymptomatic cases infection rate. It provides an important quantitative basis for risk assessment. Furthermore,
the number of asymptomatic cases in China, as well as its infection rate, is obviously of great significance for other countries in the world to evaluate the number of asymptomatic cases and the asymptomatic infections rate in their own countries. From the previous simulation process and data, we can see that the SAIR model dynamics theory is reasonable, and it is successful in describing the development process of virus transmission. In particular, it is accurate in describing time characteristics.

Quarantine and other control measures are proved to be effective. The model data reveal that the number of asymptomatic cases is still large, who are highly infectious. A total relaxation in control measures is likely to cause a second wave of outbreaks, which will wipe out the previous quarantine results.

We believe that the success of this paper is mainly based on the following reasons: 1. The values of main parameters are based on the data of authoritative institutions (for example, R0=3.77); 2. The values of other parameters are obtained by the least square method of error fitting; 3 For the abnormal data (for example, the data of 12 February and 12 February 2020), we regard it as missing data and use the third order spline interpolation method to get it, because the parameter value of the model is very sensitive to the error.

Of course, the model theory of this paper has its own limitations. For example, once the parameters are set, they cannot be changed. In reality, however, parameters do change after new measures are adopted.

The model can be improved in following aspects: (1) The parameters (8–17) in the process can be changed. (2) When the total population in the region is not very large (for example, on the “Diamond Princess”), the number of deaths must be taken into account instead of assuming the total number unchanged. (3) In a large community, infection should be related to the social network of residents, since residents seldom interact with their neighbors, the infection will be a social network graph model. The differential equation model, nevertheless, assumes that the social personnel are fully "mixed" with "homogeneous" distribution. Therefore, it is unreasonable. It is hoped that someone interested in the model can further improve it making it more universal and accurate.

Credit author statement

Tang Xiaoping: Theoretical study on infectious diseases, Model, Data collection, Computer simulation
Liu Ying: Conceptualization, Assist in data collection

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

References

[1] Simone AD, Piangretelli M. A bayesian approach for monitoring epidemics in presence of undetected cases. Chaos Solitons Fractals 2020 accepted. doi:10.1016/j.chaos.2020.110367.
[2] Tian BC, Yuan R. Traveling waves for a diffusive SEIR epidemic model with standard incidences. Sci China Math 2017;60(5):813–32.
[3] Chen DX, Yang YF, Zhang YF, Yu WW. Prediction of COVID-19 spread by sliding mSEIR observer. Sci China Inf Sci. 2020 accepted. doi:10.1007/s11432-020-3034-y.
[4] Altan A, Karasu S, Bekiros S. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos Solitons Fractals 2019;126:325–36.
[5] National Health Commission. Latest COVID-19 pneumonia situation at 24 February 2020 [EB/OL]. 2020-02-29[http://www.nhc.gov.cn/xcs/yqtb/202002/ 4ef0b2271bda47469da3145ac376ed9.shtml].
[6] Yang Y., et al., Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. Medrxiv, Posted February 11, 2020.
[7] World Health Organization. Coronavirus disease (COVID-19) situation reports-28 [EB/OL]. [2020-02-18]. https://www.who.int/emergencies/diseases/ novel-coronavirus-2019/situation-reports
[8] Tang X Q, Liu N Z, Wang H X, Bai Y Q, A class of new two-variable chaotic polynomials. J Lanzhou Univ (Natural Sciences) 2012;48(2):105–12.
[9] Tang X Q, Liu N Z, Wang H X, Bai Y Q, The two-variable chaotic polynomial of regular tree study. Acta Mathematicae Applicatae Sinica 2013;36(4):761–8.
[10] Tang X Q, New expected value expansions of rooted graphs. Acta Mathematicae Applicatae Sinica, Eng Ser 2015;31(1):81–8.
[11] Tang XQ, Wang HX. Stepwise confidence interval method for finding drug’s minimum effective dose in trial. Drug Eval Res 2010;33(1):58–62.
[12] Liu NZ, Tang XQ, Wang HX. Maximum safe dose of drug based on continuous reassessment method. Drug Eval Res 2011;34(6):425–7.
[13] Liu Y, Tang X Q, Gong Y F. Assessment of postgraduate’s innovative ability based on renew parameters probability model of grey clustering. Fuzzy Syst Math 2010;33:118–27.
[14] Tang XQ, Liu Y, Liu NZ. Reliability and stability study of rooted net. J Chongqing Normal Univ (Nat Sci) 2017;34:65–71.
[15] Liu Y, Tang X Q. Comparison of some classes of two-variable chaotic polynomials. J Natl Sci Hunan Normal Univ 2014;37:67–72.
[16] Liu Y, Tang X Q, Wang S C. Uniform form of some graphs’ two probability reliability based on characteristic equation. J Southwest China Normal Univ(Natural Science Edition), 2018; 43: 18–21.
[17] Bamparis GD, Tsironis GP. Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. Chaos Solitons Fractals 2020 accepted.