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.
A simple, SIR-like but individual-based epidemic model: Application in comparison of COVID-19 in New York City and Wuhan

Xiaoping Liu

Department of Medicine, Department of Neuroscience, West Virginia University Health Science Center, Morgantown, WV 26506, United States

ARTICLE INFO

Keywords:
Epidemic model
Individual-based model
COVID-19
Latent period
Infectious period
Time-dependent transmission rate
Transient incidence rate

ABSTRACT

In this study, an individual-based epidemic model, considering latent-infectious-recovery periods, is presented. The analytic solution of the model in the form of recursive formulae with a time-dependent transmission coefficient is derived and implanted in Excel. The simulated epidemic curves from the model fit very well with the daily reported cases of COVID-19 in Wuhan, China and New York City (NYC), USA. These simulations show that the transmission rate of NYC’s COVID-19 is nearly 30% greater than the transmission rate of Wuhan’s COVID-19, and that the actual number of cumulative infected people in NYC is around 9 times the reported number of cumulative COVID-19 cases in NYC. Results from this study also provide important information about latent period, infectious period and lockdown efficiency.

Introduction

In early December 2019, a new type of pneumonia, which is named COVID-19 by the World Health Organization, was found in Wuhan, Hubei province, China [1,2], and then the disease was found in all provinces of China and almost all countries around the world. This extensive spread of infectious disease in the world was officially described as a pandemic on March 11, 2020. As of the middle of August 2020, >750,000 people are killed by COVID-19 globally. To prevent the disease of further transmission, many cities, states or provinces, and countries declared lockdown.

Since the beginning of the COVID-19 epidemic, mathematical models have played an important role in many key policy discussions about COVID-19, and a common approach is to use compartmental models, such as the susceptible-infected-recovered (SIR) or SIR-derived models to analyze the characteristic of the COVID-19 [3-11]. The SIR model was first published nearly 90 years ago [12]. Here, $S$ stands for the number of susceptible, $I$ for the number of infectious (latent period was not considered in the classical SIR model), and $R$ for the number recovered (or immune) individuals [13]. In the SIR model, changes of $S$, $I$ and $R$ with time $t$ are determined by three differential equations. For example, $\frac{dS}{dt} = \beta SI - \gamma S$, where the parameters $\beta$ and $\gamma$ are transmission rate and recovery rate, respectively [14]. Both of the two parameters are constant. These equations require that an individual in $S$, $I$ and $R$ has the same probability appearing at different locations in the whole area studied. So, the SIR model can be considered as a “well-mixed” based model [15]. However, individuals in the real world actually have their own daily moving paths. They are not well-mixed in the whole population. In the SIR model, $I$ is usually considered as the theoretical value of the daily reported cases of an infectious disease studied. However, it has been known that there were many asymptomatic and mild cases of COVID-19, and only a small portion of infected individuals was reported [16]. Furthermore, increasing capacity of viral tests and tracing/testing all close contacts of COVID-19 patients can increase daily reported cases of COVID-19, which can also change the epidemic curve. Therefore, a parameter may need to be induced in an epidemic model for linking the actual number of cases to the reported number of cases. In the SIR model, the transmission rate is considered as a constant. However, during the COVID-19 pandemic, many interventions (i.e. closing schools, closing restaurants or other businesses, staying at home, wearing masks) have been used to slow down the COVID-19’s transmission rate for flattening the epidemic curve. Therefore, a time-dependent transmission coefficient, in response to the applied interventions, may need to be considered in an epidemic model.

In this study, a new individual-based $I-i$ AIR model, which does not require a well-mixed population, is presented. This model has considered a limited constant latent period ($l$), a limited constant infectious period ($i$) and an unlimited recovery phase for each infected individual. A time-dependent transmission coefficient $k_t$, and a transient incidence rate $r$ of the infected people are induced in the model to examine the effect of social distancing (including mask wearing) and the testing capacity of the virus on the epidemic curves, respectively. Recursive
formulae for all variables ($A_n$, $I_n$ and $R_n$) in this model have been derived. Here, $A_0$, $I_0$ and $R_0$ stand for the actual active infectious individuals in the infectious period, the actual cumulative infected individuals, and the actual cumulative recovered individuals, respectively. These recursive formulae can be easily placed in Excel for calculating $A_n$, $I_n$ and $R_n$. All the three variables are functions of $k_0$, and all these variables are dependent on the parameters $l$ and $i$. The calculated $A_0$ is linked to the reported daily confirmed number ($y_0$) of COVID-19 cases through the transient incidence rate $\alpha$. In the following sections, we will introduce this $l\cdot i$ AIR model; describe how to calculate these variables in Excel and how to determine parameters $l$, $i$, and $\alpha$; and validate the $l\cdot i$ AIR model qualitatively by known characteristics of epidemic curves and quantitatively by COVID-19 data from Wuhan, China and New York City (NYC), USA.

Theory: $l\cdot i$ AIR model

If a person is infected by an infectious disease, the person will experience a latent period ($l$), an infectious period ($i$), and then recovery or death. In this model, $l$ and $i$ are two parameters; and the three letters $A$, $I$ and $R$ are variables. $A$ stands for the actual active infectious individuals or the individuals in the infectious period, $I$ for the actual cumulative infected individuals, and $R$ for the actual cumulative recovered individuals. It is assumed that (a) the length of latent period is $l$ days or $l$ time units (1 time unit can be 1 day or less); (b) the length of infectious period is $i$ days or $i$ time units; and (c) the infectious individual infects one person per day or per time unit in the infectious period. Under these assumptions, if $i = n$, it means that an infectious individual will infect 1 person per day (or per time unit) for $n$ days (or $n$ time units) in the infectious period, or will infect a total of $n$ people in the infectious period before the infectious individual recovers or dies.

Set up the $l\cdot i$ AIR model.

In the following, a special example is given for examining how an infectious person spreads the diseases in the $l\cdot i$ AIR model. At first, it is assumed that the time unit is 1 day/unit; the latent period is 2 days ($l = 2$); the infectious period is 3 days ($i = 3$); and the first infected individual has been generated on day 1. A column of 5 cells (Fig. 1A) is used to represent the infectious status of an infected individual. The transmission pattern of the epidemics is illustrated in Fig. 1B:

On day 1, the infected individual is in the latent period, so $A_1 = 0$ because there is no one in the infectious period. However $I_1 = 1$ because the first infected individual exists already. There is no one recovered, so $R_1 = 0$. Thus, the three variables ($A_1$, $I_1$, $R_1$) = (0, 1, 0).

On day 2, the first infected individual is still in the latent period, who is not able to infect others and is not recovered yet, so ($A_2$, $I_2$, $R_2$) = (0, 1, 0).

On day 3, the first infected person enters the infectious period, so its infectious status changes from 0 to 1 ($A_3 = 1$). In this infectious status, the individual can infect one person ($I_3 = I_2 + 1 = 2$) per day. The new infected person is in the latent period. No one has recovered ($R_3 = 0$). Thus, we have ($A_3$, $I_3$, $R_3$) = (1, 2, 0).

On day 4, the first infected person is in the infectious period; the second infected person is still in the latent period; so there is only one person in the infectious period ($A_4 = 1$), and they can only infect 1 more person ($I_4 = I_3 + 1 = 2 + 1 = 3$), and there is still no recovered persons ($R_4 = 0$). So, we have ($A_4$, $I_4$, $R_4$) = (1, 3, 0).

On day 5, in addition to the first infected person, the second infected person also enters the infectious period ($A_5 = A_4 + 1 + 1 = 2$). Since $A_5 = 2$, they can infect 2 more persons, or ($I_5 = I_4 + 2 = 3 + 2 = 5$). There are no recovered people ($R_5 = 0$). Thus we have ($A_5$, $I_5$, $R_5$) = (2, 5, 0).

On day 6, the first infected person has recovered, so we have $R_6 = R_5 + 1 = 1$. The number of persons in the infectious period decreases by 1, but the third infected person enters the infectious period, so we have $A_6 = A_5 - 1 + 1 = 2$. Since $A_6 = 2$, they can only infect 2 more persons. Thus $I_6 = I_5 + 2 = 5 + 2 = 7$, or ($A_6$, $I_6$, $R_6$) = (2, 7, 1).

On day 7, only the first infected person is in the recovered status ($R_7 = 1$). In addition to infected persons 2 and 3, infected persons 4 and 5 enter the infectious period ($A_7 = A_6 + 2 = 4$). Infected persons 6 and 7 are in latent period, so they can only infect 4 more persons, $I_7 = I_6 + 4 = 11$. Thus ($A_7$, $I_7$, $R_7$) = (4, 11, 1).

On day 8, in addition to the first infected person, the second infected person also entered the infectious period ($A_8 = A_7 + 1 + 1 = 2$). The number of the infectious persons needs to be subtracted by 1. However, the infected persons 6 and 7 enter the infectious period, so the number of the active infectious persons $A_8 = A_7 - 1 + 2 = 5$. Since $A_8 = 5$, they can infect 5 more persons, so $I_8 = I_7 + 5 = 16$. Thus, ($A_8$, $I_8$, $R_8$) = (5, 16, 2).

Considering all of the above steps, we obtain the following recursive formulae to calculate the nth term ($A_n$, $I_n$, $R_n$) after its previous terms are determined:

$$A_n = A_{n-1} + (I_{n-1} - I_{n-1-1}) - (I_{n-c} - I_{n-c-1}) \quad (1a)$$

$$I_n = I_{n-1} + A_n \quad (1b)$$

$$R_n = R_{n-1} + I_{n-c} - I_{n-c-1} \quad (1c)$$

In the above equations, $c = l + i$. From Eqn. (1) we can easily find the total number of infected persons in both the latent period and infectious period, ($T_n$), which is equal to the number of cumulative infected individuals ($I_n$) subtracted by the number of cumulative recovered individuals ($R_n$):

$$T_n = I_n - R_n \quad (1d)$$

In the above example, we have $l = 2$ and $i = 3$, so we can calculate the
second term from the first term assuming that \((A_1, I_1, R_1)=(0, 1, 0)\) and 
\((A_n, I_n, R_n)=(0, 0, 0)\) as \(n \leq 0\).

On day 2:
\[
A_2 = A_1 + I_0 - I_{-1} - (I_{-1} - I_{-2}) = 0 + 0 - 0 - (0 - 0) = 0
\]
\[
I_2 = I_1 + A_2 = 1 + 0 = 1
\]
\[
R_2 = R_1 + I_{-1} - I_{-2} = 0 + 0 - 0 = 0
\]
Thus, \((A_2, I_2, R_2)=(0,1,0)\).

In the same way, we can use Eqn. (1) to calculate \((A_n, I_n, R_n)\) from \(n = 3 \) to \(n = 8\). The calculated results are exactly the same as these \((A_n, I_n, R_n)\) terms derived from the transmission pattern showing Fig. 1B.

**l-i AIR model with social interventions.**

If the outbreak of an epidemic occurs in a geographic area and the area is lockdown or other social interventions are applied in the area for	

...air model with social interventions.

Methods

Equations (1) and (2) are recursive formulae for calculating the nth term of variables \(A_n, I_n, R_n\) and \(T_n\) from previous terms of these variables after the first term is given. Excel is an excellent tool for calculating the nth term of a variable from a recursive formula because Excel has implanted the ability to do this type of work easily. In this study, Excel 2010 or a higher version is used in Windows 10 on a desktop or laptop computer. The following is a brief procedure for calculating \(A_n, I_n, R_n\) in Excel at \(l = 2\) and \(l = 4\) or \(c = l + i = 6\) as \(n\) increases from 1 to any number \(N\).

**Excel basic**

After Excel is opened, a worksheet is shown on the screen. The worksheet is made up of rows, columns and cells. A row is identified by the number (1, 2, 3, ...)) on the left side of the row. A column is identified by a column header (A, B, C, ...)) on the top of the column. A cell is the intersection of a row and column. It is identified by the column header and row number such as A2 (at the intersection of column 1 and row 2), B3 (at the intersection of column 2 and row 3), and so on. Different numbers, characters and a math formula can be input into each cell. For example, we can input the number “2” in A1 and the number “4” in B1. In the cell A2, if we type “=A1+B1” and click the “Enter” button, we will see the number “6” shown in the cell A2, which is the sum of 2 and 4. Alternatively in Excel, we don’t need to type “A1+B1” at all. Instead, after typing “=” in A2, we can just click on the cell A1 by the mouse, and then type “+” followed by clicking on the cell B2.

**Write a notation on worksheet and label columns**

To calculate \(A_n, I_n\) (or \(T_n\)), and \(R_n\) on the worksheet, we can first write a notation in the cell A1 in row 1 to show the pair of parameters \((l, i)\) used in the following calculation(s). In row 2, input \(n, An, In, Rn\) into the cells A2, B2, C2, respectively. If \(n\) increases from 1 to any number \(N\), we can fill column E with the number \(n\) across a range of cells in column A until the desired cell is reached. In this way, a series of numbers (0,1,2,3,4,..) increasing from 0 to \(N\) is generated in column A. Assuming that \(kn\) is the constant 1, we can fill column E with the number “1”. Since we had “1” in E11 already, we can select E11, and then drag the fill handle down across a range of cells in column E until the desired cell is reached.
Input formulae for calculating \( A_n \), \( I_n \) and \( R_n \) at \( n = 2 \)

The formulae for calculating \( A_n \), \( I_n \) and \( R_n \) are given in Eqsns. (1a), (1b) and (1c). At \( n = 2 \), \( l = 2 \) and \( c = 4 \) or \( c = l + 1 \) and \( i = 6 \) (the procedure is similar if using a different pair of \( l \) and \( c \) or \( i \)), the formulae of \( A_2 \), \( I_2 \) and \( R_2 \) can be simplified as:

\[
A_2 = A_1 + I_0 - I_{-1} - (I_{-1} - I_{-2}) - I_1 + I_2 + R_2 = R_1 + (I_{-4} - I_{-5})
\]

Similarly, the formula of \( I_2 \) (or \( l = 9 + l_0 + 12 + l_2^2 \)) is input into the cell C12; and the formula of \( R_2 \) (or \( l = 11 + 6 + C_5 \)) is input into the cell D12. From these recursive formulae in B12, C12 and D12 in row 12, we can see that these formulae need to use data stored in C5, C6, C9 and C10 in above rows where \( n < 1 \). This is the reason why we need to leave enough blank rows above the row at which \( n = 1 \).

Calculating the values of \( A_n \), \( I_n \) (or \( T_n \)) and \( R_n \) as \( n \) increases from 3 to \( N \)

In Excel, we don’t need to repeatedly input these recursive formulae in the row at which \( n = 3 \) or greater. Instead, this can simply be done by selecting the three cells B12, C12 and D12 (which contain the formulae of \( A_2 \), \( I_2 \) and \( R_2 \) respectively) and dragging the fill handle on the bottom right corner of the selected cells down across a range of cells in columns B, C and D. After this operation is done, all these cells will be filled with their recursive formulae from \( n = 3 \) to \( n = N \); and the calculated results for \( A_n \), \( I_n \) and \( R_n \) will be shown in these cells immediately. If we hope to examine how \( k_n \) affects the variables \( A_n \), \( I_n \) and \( R_n \), we can change the value of \( k_n \) (n is between 2 and N) in any cell of column E from \( n = 1 \) to 0. The above example has been uploaded to the Mendeley Data repository [37].

Programming for varied parameters \( l \) and \( c \) (or \( i \))

In the above procedure, we assumed the parameters \( l = 2 \) and \( c = 6 \). If we change the values of the two parameters, then we need to rewrite the formulae of \( A_n \), \( I_n \) (or \( T_n \)) and \( R_n \) correspondingly in the row at which \( n = 2 \); and then select the three cells that contain the formulae of \( A_2 \), \( I_2 \) and \( R_2 \); afterwards, we should drag the fill handle down across a range of cells. If it is needed to change the parameters \( l \) and \( c \) frequently in calculations of variables \( A_n \), \( I_n \) and \( R_n \), it would be better to put the two parameters \( l \) and \( c \) in two specific cells on the top row, and then we put the location of these parameters in the formulae of \( A_n \), \( I_n \) and \( R_n \) in each cell. As a result, these formulae will calculate the values of \( A_n \), \( I_n \) and \( R_n \) in each cell as soon as the two parameters in the two special cells on the top row are revised. These programs can be written by using INDIRECT function and absolute cell references in the formulae. Interested readers can refer to our data files submitted [17].

Verification of Excel programs and validation of the \( I-I \) AIR model

In this study, all programs for calculating the variables \( A_n \), \( I_n \) (or \( T_n \)) and \( R_n \) are written in the Excel worksheet as described above. One method used for verifying these programs is the manual calculation of the first many terms of variables \( A_n \), \( I_n \) and \( R_n \) and then compare them with those terms calculated from the program. If the manual calculation and the program calculation do not match, then we can check where the problems are and correct them. Validation of the \( I-I \) AIR model is demonstrated by several examples in the following two sections qualitatively (section: Simulated Results) and quantitatively (section: Comparing COVID-19 Outbreak in Wuhan, China and in NYC, USA).

The first day (\( n = 1 \)) of the calculated variables \( A_n \), \( I_n \) and \( R_n \) is set at the day on which the first infected person was found and reported in the literature. After \( l \), \( i \) and \( \alpha \) are determined as described above, these parameters are then used to calculate \( A_n \), \( I_n \) and \( R_n \) as \( n \) increases from 1 to any \( N \). After social interventions are applied to flatten the epidemic curve, \( k_n \) will reduce from its initial value \( k_0 \) = 1 to \( k_0 \) < 1. By regulating \( k_n \) in column E in the order from small \( n \) to large \( n \), we can fit the calculated daily number \( (\alpha A_n) \) of COVID-19 cases to the reported daily number of cases \( y_n \). When \( k_n \) changes during fitting \( \alpha A_n \) to \( y_n \), all of the calculated \( A_n \), \( I_n \) and \( R_n \) will change correspondingly. In this way, we are able to calculate \( I_0 \), the actual number of cumulative infected people, by fitting \( \alpha A_n \) to \( y_n \). Since the actual number of cumulative infected people in NYC, which was detected by antibody tests, was reported on some certain dates, this reported number on a certain date can be used to validate the \( I-I \) AIR model.

Simulated results

From the following simulations, we can examine the characteristics of \( A_n \), \( I_n \) and \( R_n \) with/without social interventions, and compare the characteristics of these variables to those of known epidemic curves in the literature.

1. No epidemic outbreak at \( i = 1 \). The rapidity of outbreak of epidemics is greatly related to the time length of \( l \) and \( i \). In Fig. 2, we demonstrate how the total number of infected persons in both the latent period and infectious period (\( T_n \)) changes with time (\( n \)) for different \( l \) at \( i = 1 \). No epidemic outbreak can be seen from the figure. At \( i = 1 \), the present infectious person will only infect one new person before the former person loses his/her transmissibility by recovery. As a result, the number of infectious individuals can’t be accumulated up to \( \geq 2 \). Although the total number of infected persons in the latent and infectious periods (\( T_n \)) jumps up and down between 1 and 2 repeatedly, the number of recovered individuals (\( R_n \)) can be accumulated up with time. However, the number of infectious persons (\( A_n \)), similar to \( T_n \), jumps between 0 and 1 repeatedly (Fig. 3).

2. Propagated epidemic curves as \( l > i > 1 \). Propagated epidemic curves usually consist of a series of waves with successively larger peaks [19]. By simulations, we observed that the propagated epidemic curves are formed when \( l > i \). In Fig. 4, we illustrate three calculated curves of \( A_n \) for \( l = 4, 6 \) and 8 at \( i = 2 \). These calculated curves are very similar to the propagated epidemic curves reported in the literature.

3. Exponential epidemic outbreak as \( i \geq l \geq 1 \) except \( i = 1 \). In the early stage of epidemic outbreak, it is often seen that epidemic curves of new cases exponentially grow with time initially [20,21]. By simulation using the \( I-I \) AIR model, we can see that all curves of \( A_n \), \( I_n \) and \( R_n \) exponentially increase with time if \( l \geq l \geq 1 \) except \( i = 1 \). A special example of the calculated \( A_n \), \( I_n \) and \( R_n \) at \( l = 2 \) and \( i = 4 \) is shown in Fig. 5A. The logarithm of \( A_n \), \( I_n \) and \( R_n \) at \( l = 2 \) and \( i = 4 \) is linear with \( n \) (Fig. 5B), indicating that \( R_n \), \( I_n \) and \( T_n \) exponentially increase with \( n \).

4. Epidemic curves with lockdown intervention. During an epidemic outbreak, if the lockdown intervention of the epidemic area is performed, the forward tendency of the epidemic curve will change depending on how strict the lockdown intervention is being implemented. In an ideal situation, the infection of an infectious person to an uninfected person is completely blocked, or the transmission coefficient \( k \) in Eqsns. (2) is 0, then Eqs. (2) can be simply replaced by Eqsns. (2a). Using Eqs. (1) and Eqs. (2a), we calculated \( R_n \), \( A_n \) and \( T_n \) at different combination of \( l \) and \( i \). The calculated \( A_n \) and \( T_n \), assuming that the first infected person appears on day 1 and that the lockdown is performed on day 41, are demonstrated in Fig. 6.

In comparison, if the virus transmission is not completely blocked by the lockdown intervention or \( k_n \) is >0, then the shape of epidemic curves
Fig. 2. Changes of the total number of infected individuals in both the latent and the infectious periods $T_n$ with $n$ for different $l$ at $i = 1$. The values of $T_n$ vary between 1 and 2.

Fig. 3. Changes of $R_n$ and $A_n$ with time $(n)$ for different values of $l$ at $i = 1$. 
will depend on $k_n$. In Fig. 7, we present the simulated epidemic curves for the nonzero value of $k_n$ at $l = 3$ and $i = 5$ to show how the shape of $A_n$ curve changes with $k_n$ ($A_n$ is proportional to the number of the reported daily new cases).

When the number of the cumulative infected individuals increases to a range comparable to the initial susceptible population ($N$), the remaining susceptible people is largely reduced. In this case, the chance to contact an uninfected person is reduced, so $k_n$ will decrease even if there is no lockdown intervention. Assuming that $k_n$ is proportional to the ratio of the number of uninfected people ($N-I_n$) to the number of initial susceptible population $N$, $k_n=(N-I_n)/N$, the simulated epidemic curves of $A_n, I_n$ and $R_n$ at $l = 1, i = 2$ and $N = 1 \times 10^6$ are demonstrated in Fig. 8.

**Comparing COVID-19 outbreak in Wuhan, China and in NYC, USA**

The COVID-19 pandemic provides a lot of data to test this $l$-$i$ AIR model. In the following, we will test how well the calculated epidemic curves from this new model fit with the reported data from Wuhan and NYC, and examine whether the number of infected people calculated from this new model is close to the reported data. In the meantime, we will use this new epidemic model to analyze the characteristics of the COVID-19 transmission process in Wuhan and NYC, and examine the effect of lockdown intervention and increasing the number of viral tests on epidemic curves.

**Simulation of COVID-19 outbreak in Wuhan**

In the $l$-$i$ AIR model, $A$ stands for the number of infected individuals in the infectious period, and $A_n$ is defined as $A$ at day $n$. To compare the calculated epidemic curve with the reported data in an epidemic event, we need to have an assumption: the number of daily new COVID-19 cases at day $n$ ($y_n$) can be considered as the transient incidence of infectious disease at day $n$, which is proportional to $A_n$ or $y_n = aA_n$. This relationship between $y_n$ and $A_n$ means that if the number of infected people in the active infectious period is $A_n$ on a certain day, only a portion of the infected people, $aA_n$, can be confirmed as COVID-19 patients on that day. In Fig. 9A, we compare the calculated daily new cases with the reported daily new cases of COVID-19 in Wuhan. The red line with closed circles represents the three-day averages of the reported daily new cases in Wuhan. The daily new cases were reported by Hubei Health Commission. Between 2/12/2020 and 2/18/2020, a total of 7 days, clinical diagnosis of COVID-19 was added as a diagnostic criteria, resulting in large increase in case number on February 12 and subsequent days. This change in diagnosis of COVID-19 disease creates difficulties in using the reported cases for data analysis[8]. Fortunately, both the number of the total new cases and the number of cases diagnosed clinically were listed in the first 4 days between 2/12/2020 and 2/15/2020, so it is easy to find the number of daily new cases confirmed from the viral tests by subtracting the clinically diagnosed case number from the total new case number. The daily new cases in the last three days between 2/16/2020 and 2/18/2020 can be estimated from the reported daily new cases by using information (the ratio of the numbers measured by the two diagnosis methods) provided in the previous days. To calculate daily new cases from Eqns. (1) and (2), we set: (a) the first COVID-19 case appeared on December 8, 2019, and the lockdown day of Wuhan started on January 23, 2020; (b) $l = 1$ unit (3 days) [22,23], $i = 4$ units, and each time unit is 3 days; (c) the transmission coefficient $k_n$ is 1 before lockdown; gradually decreased to 0.02 in the first 10 days after lockdown of the city; and (d) the transient incidence rate $\alpha$ is $1/85$ or $\sim 1.2\%$. From Fig. 9A, it can be seen that the calculated number of daily new cases (green line) is consistent with the reported daily new cases (red line with closed circles) before and on the date 2/15/2020. After this date, it seems that the reported number of daily new cases makes a turn and deviates from the calculated curve. Changing the parameter $k_n$ or $\alpha$ can make the calculated curve fit the reported number of daily new cases better. However, although increasing $k_n$ can moderately improve the fitting, a bigger $k_n$ will be too large to be reasonable considering the strict lockdown that was performed during that time. This deviation is more likely caused by changes in $\alpha$ because of some other intervention used during this time, such as the large increase in the number of viral tests and the use of 16 Fangcang shelter hospitals to admit about 12,000

![Fig. 4. Propagated epidemic curves for $l = 4, 6$ and 8 at $i = 2$. Each curve consists of a series of waves with successively larger peaks.](image-url)

![Fig. 5. An example of exponential growth of $R_n, A_n$ and $T_n$ as $i > l > 1$. (A) The curves are calculated from Eqn. 1 assuming $l = 2$ and $i = 4$. (B) The logarithm of $A_n$, $T_n$ and $R_n$ linearly increase with $n$.](image-url)
patients[24]. By gradually increasing \( \alpha \) from 2/15/2020 to 3/1/2020, the calculated number of daily new cases fits with the reported number of daily new cases pretty well (Fig. 9B). R-squared (or \( R^2 \)), which is defined as \( R^2 = 1 - \frac{\text{SSE}}{\text{SST}} \) (SSE: the sum of squared errors/residuals, SST: the total sum of squares), can be used to determine how close the data are to a calculated curve[25]. For the reported data of daily new cases and the simulated epidemic curve in Fig. 9A, the R-squared value is 0.89. In contrast, for the reported data and the simulated epidemic curve in Fig. 9B, the R-squared value is 0.99, which is very close to 1. If R-squared equals 1, it means that all these reported data are exactly on the simulated line.

Simulation of COVID-19 outbreak in NYC

The first COVID-19 case in NYC was confirmed on March 1, 2020[26]. Research indicates that coronavirus started to transmit in the New York state between late January to mid-February[27]. A state of emergency in New York was declared on March 7 after 89 cases were confirmed in the New York state. All schools, bars, and restaurants in NYC were closed on March 17, 2020[28]. The statewide stay-at-home order of New York was effective on March 22, 2020[29]. To simulate the reported daily new cases of COVID-19 in NYC, a different combination of parameters \( l \) and \( i \), and different start date of the first infected person who initiated the COVID-19 transmission in NYC were tried. It was found that the simulated curve of daily new cases can fit well with the reported daily new cases if \( l = 4[30], i = 10, \) and the start day (day 1 or \( n = 1 \)) of the first infected person appeared in NYC was on February 9, 2020. In the simulation, the transmission coefficient \( k_n \) was largely dropped from 1 on March 17 (\( n = 38, k_{38} = 0.45 \)), the first day that NYC closed schools, restaurants and bars, and then gradually decreased to 0.064 by April 20, 2020. In mid May, \( k_n \) gradually increased from 0.064 to 0.075 and has been staying at 0.075 since then. The transient incidence rate \( \alpha \) is 1/85 or \( \sim 1.2\% \). The reported daily new cases in NYC and the simulated epidemic curve are demonstrated in Fig. 10A. The calculated actual number of cumulative infected people (\( I_n \)), based on Eqns. (1) and (2) using the same parameters \( l, i, k_n \) and \( \alpha \) for Fig. 10A, is shown in Fig. 10B.
Comparison of the spreading characteristic of COVID-19 in Wuhan and in New York City

The values of $l$ and $i$ of COVID-19 in NYC are different from those of the COVID-19 in Wuhan. For Wuhan’s COVID-19, $l$ is 1 time unit and $i$ is 4 time units, with the time unit being 3 days/unit. This means: (a) the latent period $l$ is 3 days; (b) the infectious period $i$ is 12 days; and (c) each infectious individual can infect one person every 3 days or infect 4 persons in 12 days. For NYC’s COVID-19, $l = 4$, $i = 10$, and time unit is 1 day/unit. This means that the latent period $l$ and infectious period $i$ are 4 days and 10 days respectively; and each infectious individual can infect 1 person per day for a total of 10 persons in 10 days. To compare the transmission characteristic of the two types of COVID-19, simulations were performed by assuming that the outbreak of COVID-19 occurs in an area having 1 million susceptible people and no intervention is used to slow down the disease outbreak. In the simulations, it was assumed that the transmission efficiency is gradually reduced while the number of uninfected people is gradually reduced with time by assuming that $k_n$ is proportional to $(N-I_n)/N$, where $N = 1,000,000$ in the simulations. The simulated epidemic curves of $A_n$, $I_n$ and $R_n$ are shown in Fig. 11. COVID-19 in Wuhan takes 72 days to reach the outbreak peak in an area with 1 million susceptible people, but COVID-19 in NYC only needs 54 days to reach the outbreak peak in the same area with the same population.

Discussion

The $l$-$i$ AIR epidemic model is presented, and its analytic solution in the form of recursive formulae is given in this paper. The procedure of
implementing the recursive formulae into Excel to find values of $A_0$, $I_0$, and $T_0$, and verifying the programs has been described in the Methods section. The model is validated qualitatively in the Simulated Results section and quantitatively in the “Comparing COVID-19 outbreak in Wuhan, China and in NYC, USA” section.

At $i = 1$, or an infected individual can only pass the virus to one another person before the infected individual is recovered or dead, simulations based on Eqn. (1) (Figs. 2 & 3) show that there will be no epidemic outbreak regardless of the length $l$ of the latent period. This is consistent with our basic knowledge in transmission of infectious diseases. Figs. 2 & 3 also show that the cumulative recovered people $(R_n)$ will be linearly accumulated up with time. However, this transmission process may be stopped at any time when a new infectious individual is not able to infect anyone else before recovery because of any interference.

For any other situations of $i > 1$, all of $A_n$, $T_n$, and $R_n$ will continuously increase with time initially. For special conditions of $i > 1 > i$, simulations (Fig. 4) show the repeated wave-shape in the epidemic curves, which have been observed in the real world, named as propagated epidemic curves, usually transmission by direct person-to-person contact[19]. From the viewpoint of the $i$-AIR model, the propagated epidemic curves can happen when $l > i > 1$.

For all other conditions of $i > 1 > i > 1$ at $i = l = 1$, all of $A_n$, $T_n$, and $R_n$ will exponentially increase with time initially (Fig. 5). This is expected because epidemics naturally exhibit exponential behavior in the early stages of an outbreak [20].

As shown in Fig. 6, if the transmission process is completely blocked by a lockdown intervention or $k_0 = 0$, then $T_n$ will stop to increase on the lockdown day and decrease to $0$ in $l + i$ days (the total length of the latent period and the infectious period). In comparison, $A_n$ will stop to increase $l$ days later because those individuals infected immediately before the lockdown starting day will get symptoms after the latent period of $l$ days (assuming that time in latent period and time in incubation period are similar to each other). However, $R_n$ will decrease to $0$ in $i$ days. As a result, both $T_n$ and $R_n$ decrease to $0$ on the same day even though $A_n$ starts to increase or reaches the peak $l$ days later.

If the lockdown process is not strict enough in such a way that the transmission process is not completely blocked, then $k_0$ will be $>0$. In this situation, the epidemic curve may decrease, go flat or even continuously increase depending on whether $k_0$ is $<1$, equals $1$, or $>1$. Here, $k_0$ is similar to the basic production number $R_0$ used in the SIR model[31]. A simulation is demonstrated in Fig. 7, assuming that $l = 3$ and $i = 5$. When $k_0 = 1$, $k_0$ will equal $1/i = 0.2$. It can been seen that the epidemic curve increases when $k_0 > 0.2$, goes flat at $k_0 = 0.2$, and decreases when $k_0 < 0.2$.

In some situations, the number $(I_n)$ of the cumulative infected individuals becomes large enough to compare with the number of the initial susceptible population $N$ so that the chance of the infected individuals to meet uninfected individuals gradually decreases. A simulation was done by assuming that the chance of infected individuals meeting uninfected individuals is proportional to $(N-I_0)/N$ (Fig. 8). It can be seen that a peak appears in the epidemic curve when $I_n$ is comparable to $N$. To make the epidemic curve decrease, $I_n$ must be large enough to meet the condition $(N-I_0)/N < 1/l$. Thus, we have $I_n > N(1-1/l)$. If $i = 2$, $I_n$ must be $>0.5N$ or $I_0/N > 0.5$, to make the epidemic curve decrease. As shown in Fig. 6, the peak of $A_n$ appears after $I_n$ is $>0.5N$. The fraction $I_0/N$ reached at the point when the epidemic curve starts to decrease is called herd immunity [21].

The above simulated results based on the $i$-AIR model are consistent with known knowledge in epidemiology. Some of these simulated results may be easily obtained from SIR model too, but the propagated epidemic curves shown in Fig. 4 cannot be generated from the classic SIR model, even though the propagated epidemic curves were observed in the real world. After the $i$-AIR model was validated qualitatively from the simulated results, we further tested the $i$-AIR model with the COVID-19 pandemic data of Wuhan, China and NYC, USA. The first example is the COVID-19 outbreak in Wuhan, where the first COVID-19 patient was seen in early December 2019[8]. After testing different combinations of $l$ and $i$, the final simulated epidemic curve (Fig. 9) fits the reported daily new cases in Wuhan pretty well by meeting all of the following requirements: (a) the epidemics started on December 8, 2019; (b) the lockdown intervention was started on January 23, 2020; (c) the latent period $(l = 1$ unit or 3 days) is within the range of reported data [22,23]; and (d) the simulated epidemic curve fits well with Wuhan’s daily new case. It can be noticed that the reported daily cases (the red line with closed circles) are deviated from the simulated epidemic curve (blue line) after February 15, 2020 (Fig. 9A). This may be related to interventions implanted during this period, such as the large increase in the number of viral tests and the use of 16 Fangcang shelter hospitals to admit about 12,000 patients between February 5 and March 10, 2020 [24]. These interventions can help to find many more patients with mild symptoms, but may generate a greater apparent transient incidence rate of epidemics than usual. By gradually increasing the incidence rate $\alpha$ from February 15 to March 1, 2020, the calculated daily new case fits the reported data very well (Fig. 9B).

The second example is COVID-19 outbreak in NYC, USA. As shown in Fig. 10, the simulated epidemic curve fits well with the reported daily new cases in NYC. The simulated curve meets the following requirements: (a) the first case started on February 9, 2020, which is within the time range estimated in the report[27]; (b) the intervention to slow down the COVID-19 transmission started on March 17, 2020 and more strict interventions were added later; (c) the latent period $(l = 4$ days) is within the range of the reported data; and (d) the simulated epidemic curve fits well with New York City’s daily new cases (Fig. 10A). Using the same parameters for Fig. 10A, the calculated number of the cumulative infected people $(I_n)$ is shown in Fig. 10B. On May 1, 2020, $I_n$ is 1.58 million, which is very close to the estimated 1.67 million people [32] infected with SARS-CoV-2 in NYC that was determined from the results of antibody tests [33], but is 9 times the reported number of total cases of COVID-19 confirmed by viral tests in NYC on May 1, 2020 [34]. Ambrosio B. et al has used time-dependent $SIR$ model to analyze data of COVID-19 in the New York state [35]. They used 4 different transmission rates in the period between March 1, 2020 and April 1, 2020. Their calculated numbers of cumulative infected people fit very well with the reported numbers, but their time-dependent $SIR$ model could not explain this nearly 10 times difference between the actual infected people (2,710,000 reported on April 23,2020[36]) detected by antibody test and the reported cumulative infected cases confirmed by viral test (272,000 on April 23, 2020[26]). However, by using the same procedure for calculating the actual numbers of the cumulative infected people as we did for analyzing epidemic data of NYC, we calculated the actual number of cumulative infected people in the New York state based on the $i$-AIR model, which is 2,700,000 on April 23, 2020. This calculated number from the $i$-AIR model, which is nearly 10 times the reported number of cumulative infected cases confirmed by viral tests, is very close to that detected by antibody tests. The calculated curves and the calculation program for analyzing COVID-19 data of the New York state can be found in the Excel file uploaded in the Mendeley Data [38].

Research has shown that the earliest cases of COVID-19 in New York were likely brought in by travelers from Europe[27]. Simulated curves in Fig. 11 show that COVID-19 in Wuhan takes 72 days to reach the outbreak peak in an area with 1 million susceptible people, while COVID-19 in NYC needs only 61 days to reach the outbreak peak in the same area with the same population. This indicates that the transmission rate of NYC’s COVID-19 is nearly 30% greater than the transmission rate of Wuhan’s COVID-19. Furthermore, the value of $i$ for NYC’s COVID-19 is 10, but it is 4 for COVID-19 in Wuhan. Because an intervention must reduce $k_0$ to below $1/i$ for making the epidemic curve decrease, it needs to reduce $k_0$ down to $1/10$ or 0.1 to make the epidemic curve decrease in New York. However, it just needs to reduce $k_0$ down to 1/4 or 0.25 in Wuhan. Therefore, it is likely that COVID-19 in NYC has stronger...
infectivity than that in Wuhan.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

Kerui Liu was involved in reviewing and editing the manuscript.

**References**

[1] Huang C, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020;395(10223):497–506.

[2] Li Q, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020.

[3] Tolles J, Luong ThaiBinh. Modeling Epidemics With Compartmental Models. JAMA 2011;305(2):137–45.

[4] Weitz JS, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. PLoS ONE 2020;15(4):e0230405.

[5] Prem K, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic and evaluation of population-wide interventions. International Journal of Infectious Diseases 2020;96:376–84.

[6] Chowell ON, Shea K, Krzywinski M, Altman N. Modeling infectious epidemics. Nat Methods 2020;17(5):455–6.

[7] Giordano G, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020;26(6):855–60.

[8] Neipel, J., Bauermann, J., Bo, S., Harmon, T. & Julicher, F. Power-law population heterogeneity governs epidemic waves. PloS one 15, e0239678 (2020).

[9] Li Q, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020;382(13):1199–207.

[10] CDC. Principles of Epidemiology in Public Health Practice, (U.S. Department of Health and Human Services, Atlanta, GA, 2006).

[11] Anastassopoulou, C., Russo, L., Tsakris, A. Modeling COVID-19 and forecasting the spread of COVID-19 in different communities. Chaos, Solitons Fractals 2020;139:110057.

[12] Brauer, F., Allen, L.J.S., Van den Driessche, P. Mathematical Epidemiology, in: Lecture Notes in Mathematics, No. 1945, Mathematical Epidemiology, in: Lecture Notes in Mathematics, No. 1945, Springer-Verlag, Berlin-Heidelberg, (2008).

[13] Huppert A, Karlst E. Mathematical modelling and prediction in infectious disease epidemiology. Clin Microbiol Infect 2013;19(11):999–1005.

[14] Gostel L, Barton DAW, Johansson A. Analytical Modelling of the Spread of Disease in Confined and Crowded Spaces. Sci Rep 2015;5(1). https://doi.org/10.1038/srep08456.

[15] Reese, H., et al. Estimated incidence of COVID-19 illness and hospitalization - United States, February-September, 2020. Clinical Infectious Diseases: an official publication of the Infectious Diseases Society of America (2020).

[16] Liu X. Excel generated epidemic curves for the paper "A Simple, SIR-like but Individual-Based Epidemic Model: Application in Comparison of COVID-19 in New York City and Wuhan. Mendeley Data 2020;V3. https://doi.org/10.17632/3yv27y9ykj.3.

[17] Cooper I, Mondal A, Antonopoulos G. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons Fractals 2020;139:110057.

[18] Wikipedia. COVID-19 pandemic in New York (state). https://en.wikipedia.org/wiki/COVID-19_pandemic_in_New_York_(state) (2020).

[19] Gonzalez-Reiche AS, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science 2020;369:297–301.

[20] Worldometer. Coronavirus (COVID-19) Mortality Rate. https://www.worldometers.info/coronavirus/coronavirus-death-rate/ (2020).

[21] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[22] Bailey NTJ. The mathematical theory of infectious diseases and its applications. London: Griffin; 1975.

[23] Worldometer. COVID-19 in different communities. Chaos, Solitons Fractals 2020;139:110057.

[24] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[25] Bailey NTJ. The mathematical theory of infectious diseases and its applications. London: Griffin; 1975.

[26] Wikipedia. COVID-19 pandemic in New York City. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_New_York_City#March (2020).

[27] Worldometer. Coronavirus (COVID-19) Mortality Rate. https://www.worldometers.info/coronavirus/coronavirus-death-rate/ (2020).

[28] Worldometer. Coronavirus (COVID-19) Mortality Rate. https://www.worldometers.info/coronavirus/coronavirus-death-rate/ (2020).

[29] Wikipedia. COVID-19 pandemic in New York City. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_New_York_City#March (2020).

[30] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[31] Bailey NTJ. The mathematical theory of infectious diseases and its applications. London: Griffin; 1975.

[32] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[33] Worldometer. Coronavirus (COVID-19) Mortality Rate. https://www.worldometers.info/coronavirus/coronavirus-death-rate/ (2020).

[34] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[35] Worldometer. Coronavirus (COVID-19) Mortality Rate. https://www.worldometers.info/coronavirus/coronavirus-death-rate/ (2020).

[36] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[37] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[38] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[39] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[40] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[41] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[42] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[43] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[44] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[45] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.

[46] Rothe C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020;382(10):970–7.