Prediction of Daily New COVID-19 Cases - Difficulties and Possible Solutions

Xiaoping Liu
Department of Medicine, Department of Neuroscience, Rockefeller Neuroscience Institute,
West Virginia University Health Science Center,
Morgantown, West Virginia, United States of America

Email: Xiaoping.liu@hsc.wvu.edu

Abstract

Epidemiological compartmental models, such as $SEIR$ (Susceptible, Exposed, Infectious, and Recovered) models, have been generally used in analyzing epidemiological data and forecasting the trajectory of transmission of infectious diseases such as COVID-19. Experience shows that accurately forecasting the trajectory of COVID-19 transmission curve is a big challenge to researchers in the field of epidemiological modeling. Multiple factors (such as social distancing, vaccinations, public health interventions, and new COVID-19 variants) can affect the trajectory of COVID-19 transmission. In the past years, we used a new compartmental model, l-i $SEIR$ model, to analyze the COVID-19 transmission trend in the United States. The letters l and i are two parameters in the model representing the average time length of the latent period and the average time length of infectious period. The l-i $SEIR$ model takes into account of the temporal heterogeneity of infected individuals and thus improves the accuracy in forecasting the trajectory of transmission of infectious diseases. This paper describes how these multiple factors mentioned above could significantly change COVID-19 transmission trends, why accurately forecasting COVID-19 transmission trend is difficult, what the strategies we have used to improve the forecast outcome, and some of successful examples that we have obtained.

Introduction

The World Health Organization (WHO) has declared an end to the COVID-19 global health emergency on May 5, 2023[1]. The COVID-19 pandemic has made great health damages to the peoples in the world. In the past three and half years, there were nearly 7 million people who died from COVID-19 globally[2]. Throughout the pandemic, researchers used mathematical models to analyze COVID-19 data for better understanding transmission patterns, monitoring disease severity, anticipating future epidemic outcomes[3], and justifying the adoption of intervention measures[4]. Among these mathematical models, compartmental models describing the disease as a sequence of different stages encountered upon infection to recovery, such as *SEIR* (*S*usceptible-*E*xposed-*I*nfectious-*R*ecovered) model, have been generally adopted to forecast or simulate future transmission trajectories[5, 6]. These compartmental models provide a parsimonious (i.e., using few parameters) approach to understanding important behaviors of epidemic pathways. Experience has shown that such models generate robust results that strengthen their usefulness[7]. However, it has been recognized that forecasting COVID-19 transmission trajectories is still a big challenge to the mathematical modelers[3, 5, 7, 8]. Multiple factors, such as interventions on social distancing, face masks, vaccination, emerging of new variants of COVID-19 with more contagious, may affect the accuracy of prediction from compartmental models[3, 5, 7]. Understanding how these factors affect forecast results is important to improving forecast accuracy. Furthermore, the compartmental models assume that infected individuals in each related compartment have no temporal heterogeneity, so all infected individuals in a compartment have the same probability to transfer to their next compartment. However, the realty is different: individuals are usually infected in differential days with a chronological order, so on average, individuals infected earlier in one compartment will be transferred to their next compartment at an earlier time. We recently developed a new compartmental model[9], the *l-i SEIR* Model. The letters l and i are two parameters in the model representing the average time length of the latent period and the average time length of infectious period. The *l-i SEIR* model takes into account of temporal heterogeneity or the chronological order of infected individuals. It was demonstrated that, when calculating the transfer rate of infected individuals from one compartment of the *SEIR* model to the next compartment, the temporal homogeneity approximation in the conventional *SEIR* model leads to calculation errors that increase linearly with the rate of change in the number of infectious individuals. Despite the improvement in calculation accuracy of the *SEIR* model after taking into account of the chronological order of infected individuals in the model, other factors mentioned above are still able to largely affect prediction values from the model. In this paper, we will describe the difficulties that we have encountered in predicting transmission trajectories of COVID-19 from the *l-i SEIR* model and some strategies that we have used to overcome these difficulties. Our data analysis focuses on the data of COVID-19 transmission caused by COVID-19 omicron variants in the United States.

Methods

The *l-i SEIR* (*S*usceptible-*E*xposed-*I*nfectious-*R*ecovered) epidemic model is described by the following recursive equations:

$$S_n - S_{n-1} = -\beta_n I_n \frac{S_{n-1}}{N} \quad (1a)$$

$$E_n - E_{n-1} = (S_{n-1} - S_n) - (S_{n-l-1} - S_{n-l}) \quad (1b)$$

$$I_n - I_{n-1} = (S_{n-l-1} - S_{n-l}) - (S_{n-c-1} - S_{n-c}) \quad (1c)$$

$$R_n - R_{n-1} = (S_{n-c-1} - S_{n-c}) \quad (1d)$$

The solution of the *l-i SEIR* model in closed-form is given as below[9, 10]:

$$S_n = S_0 \prod_{m=1}^{n} (1 - \frac{\beta_m (S_{m-c} - S_{m-l})}{N}) \quad (2a)$$

$$E_n = S_{n-l} - S_n \quad (2b)$$

$$I_n = S_{n-c} - S_{n-l} \quad (2c)$$

$$R_n = N - S_{n-c} \quad (2d)$$

To connect the calculated model variable with the daily new COVID-19 cases, we assume:

$$y_n = \alpha I_n \approx \bar{y}_n \quad (2e)$$

In the above equations, S_n is the number of remaining susceptible individuals who are able to contract the disease on day n; E_n is the number of exposed individuals who are in the latent period before becoming infectious; I_n is the number of infectious individuals who are in the infectious period and are capable of transmitting the disease; R_n is the number of recovered individuals who are becoming immune; β_n represents the transmission rate coefficient on day n; y_n represents the calculated number of the daily new cases; the coefficient α is a fraction between 0 and 1; αI_n means that only a fraction of individuals in the infectious period becomes the confirmed daily new cases; and $\bar{y}_n$ represents the reported number (7-day average) of the daily new COVID-19 cases in the United States. The parameter c represents the sum ($c = l+i$) of the average time length of latent period (l) and the average time length of infectious period (i). For the initial condition of Eqns. (2a)-(2d), we assume: (a) $S_n = N$ and $E_n = I_n = R_n = 0$ as $n < 0$; and (b) $S_0 = N-1$, $E_0 = 1$, $I_0 = R_0 = 0$. Here, N is the number of susceptible people right before the infectious disease spreads out. If all people in the population are susceptible to the infectious agents before the infectious disease spreads out, N equals to the number of population P. However, if a portion of people has immunity to the infectious disease before the infectious disease spreads out, N is smaller than P. Eqns (2a) – (2d) were derived based on the following assumption: Change in S_n is proportional to I_n and proportional to S_n/N. This assumption implies that any person infected cannot be re-infected by the infectious agents because they have obtained immunity to the infectious agents.

From Eqn. (2a), we can easily calculate all S_n if parameters l and c, the coefficient β_n and the initial conditions regarding S_n ($n \leq 0$) and N are known or given. From the calculated S_n, we can further find E_n, I_n, R_n from Eqns. (2b)-(2d), and find y_n from Eqn. (1e) if α is known. We have demonstrated how to obtain the values of l, c and α from the reported daily new COVID-19 cases in the early period of the outbreak of infectious diseases[11, 12]. Thus, theoretically, we can use the *l-i SEIR* model eqns. (2a)-(2e) to predict the trajectory of COVID-19 daily new cases if β_n and N (or S_0) are known and the underlying assumption (no re-infection) to the model equations remains true.

Results and Discussion

In the past two or three years, we used the *l-i SEIR* model to simulate and predict the daily new COVID-19 cases. The first important thing being recognized was that the rate coefficient β_n of COVID-19 transmission varies with time. Why β_n of COVID-19 transmission is not a constant but varies with time?

The coefficient β_n can be considered as a weight factor reflecting the efficiency of interaction between I_n (the number of infectious people) and S_n/N (the probability that the susceptible people is going to be infected) in the rate equation (1a), which determines the transmission rate of infectious diseases. Compared to the common flu, COVID-19 is highly contagious and has a relatively high mortality rate. In the early stages of the COVID-19 outbreak, when most people were not immune to the coronavirus, the death toll from COVID-19 would be very high if interventions of social distance and face mask policies were not applied to slow the spread of COVID-19 from person to person. Therefore, some interventions including social distancing and face masking were generally used to reduce COVID-19 transmission rate via decreasing the transmission rate coefficient β_n. However, the quantitative relationship between these interventions and the value of β_n are unknown while the spread rate of COVID-19 in the early outbreak period is highly sensitive to β_n. As a result, it is difficult to accurately predict at which day the peak of daily new cases will be reached and the height of the peak in the early outbreak period of COVID-19. Furthermore, COVID-19 vaccines and/or COVID-19 variants can also greatly affect the number of COVID-19 daily new cases by changing the number of remaining susceptible people S_n and/or the number of total susceptible people N, the coefficient of transmission rate (β_n), and the ratio of reinfection as a percentage of total infections. In the following, we will describe how these problems were addressed for predicting the trajectory of COVID-19 transmission in the United States.

Simulations and predictions of spread of COVID-19 omicron variants in the United States.

After mid-December 2020, COVID-19 vaccines were given to people in the US. Since then, COVID-19 vaccines gradually became an important factor to affect the trajectory of COVID-19 transmission. In 2021, the COVID-19 alpha variant caused a transmission wave peak in mid-April and the delta variant caused a transmission wave peak in early September[13]. In this situation, multiple factors including vaccination (affect S_0), breakthrough infection[14] (affect S_n), reinfection[15] (affect rate equations), and intervention measures (affect transmission rate coefficient β_n) were able to affect the trajectory of y_n, making simulations/predictions of y_n trajectory more complicated because the coexistence of these factors made it difficult to identify who were susceptible and who were immune in the US. This complicated situation changed when the Omicron variant began to spread. The Omicron variant had the strongest breakthrough infectivity and re-infectivity compared to the other previous COVID-19 variants[14, 16, 17]. Vaccine effectiveness to omicron, comparing to Delta variants, dropped from 0.52 to 0.38 for those who had had their second dose 180 days earlier or more[17]. Considering that many people in the US have only received one dose or even have not received vaccines, the actual number of people with immunity to omicron variant of coronavirus may be less than 38% (0.38) of the US population (P=330,000,000). Our simulations show that the transmission of omicron variant in the US can be treated as the transmission of a new infectious disease from the beginning by assuming that only a fraction ~0.25 of the US population has immunity to the Omicron original variant in the early period of Omicron-induced COVID-19 outbreak in the US. This indicates that N is ~75% of the population P ($N \approx 0.75P$ =250000000). To simulate/predict the transmission process of Omicron variants, we first estimated the value of β_n from the reported number of daily new COVID-19 cases before Omicron started to spread out in the US by using the method described previously[9, 12]. This

Table 1. The determined time-dependent β_n N=250000000

Dates	β_n
Dec. 25, 2021	0.4
Dec. 27, 2021	0.3
Dec. 29, 2021	0.25
Jan. 2, 2022	0.22
Jan. 4, 2022	0.2
Jan. 9, 2022	0.17

estimated value of β_n was used as the initial value of β_n for simulating Omicron-induced daily new COVID-19 cases.

Prediction of peak height and peak day of reported daily COVID-19 cases ($\bar{y}_n$) induced by Omicron variants. Usually, in the period of the COVID-19 outbreak, the number of new COVID-19 cases rapidly increases, and thus different interventions are applied for slowing down COVID-19 transmission rate, resulting that β_n gradually decreases before seeing the peak number of reported daily new COVID-19 cases ($\bar{y}_n$). Because β_n is not a constant but decreases with time before the y_n peak passes, it is difficult to know the exact value of β_n in the future and thus it is difficult to predict the exact height and

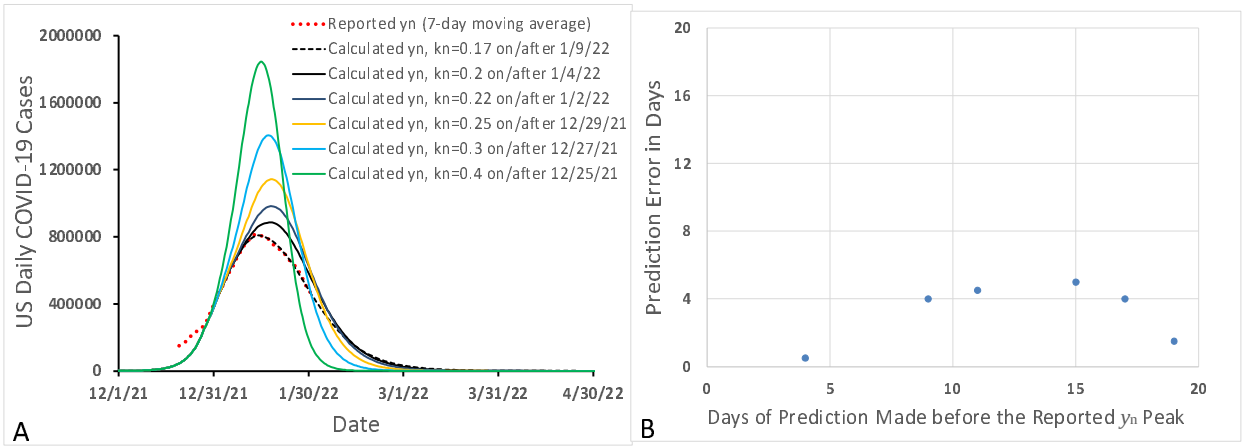

Figure 1. Prediction errors in forecasting peak height (A) and peak day (B) of Omicron-induced daily new COVID-19 cases in the United States.

date of the y_n peak. To address this problem, we examined the magnitude of the prediction error when using the latest available number of daily COVID-19 cases ($\bar{y}_n$) during the early rising phase of the COVID-19 outbreak to predict the height and date of the upcoming $\bar{y}_n$ peak. In the simulations, we assumed that l=4, i=10[9, 11, 12] and N=250,000,000. Table 1 lists these values of β_n determined before the peak of $\bar{y}_n$ from December 25, 2021 to January 9, 2022 and the corresponding dates when these β_n values were determined. These values of β_n were determined by fitting the calculated y_n to the reported daily COVID-19 cases ($\bar{y}_n$) as described previously[9, 11, 12]. The β_n values determined before December 25, 2021 were not listed. The first row of the data in Table 1 shows that β_n=0.4 on December 25, 2021. Supposing that we were on December 26, 2021, knowing all values of β_n up to December 25, 2021, and assuming that β_n remains constant 0.4 after December 25, 2021, we could calculate a y_n curve (the green line, Fig 1) from l-i $SEIR$ model. From the calculated y_n curve, we could know the height and the date of the y_n peak. In this way, we forecasted the height and date of $\bar{y}_n$ peak based on available COVID-19 data as of December 25, 2021. In the same way, we forecasted the $\bar{y}_n$ peaks and heights on other dates (Dec. 27, 29, Jan. 2, 4 and 9) listed in Table 1. The forecasted number of daily new COVID-19 cases (y_n) and the reported number of daily new COVID-19 cases 7-day average ($\bar{y}_n$) are shown in Figure 1A. The y_n peak predicted on December 25, 2021 is 1.84 million, which has the largest error comparing to the value of the reported peak $\bar{y}_n$ (0.81 million) on January 13 & 14, 2022. As the prediction day approaches to January 13, 2022, the predicted value of peak y_n approaches to the actual reported peak $\bar{y}_n$ (Figure 1A). This result implies that simulations performed on earlier dates to predict the height of peak y_n may be significantly greater than the height of $\bar{y}_n$ peak, whereas simulations performed on earlier dates to predict the day of the y_n peak may not produce a larger error. As shown in Figure 1B, the largest prediction error for the date of $\bar{y}_n$ peak occurs on the day that is 15 days before the date of the reported $\bar{y}_n$ peak. This result implies that an earlier prediction day may not cause a larger prediction error for the

date of $\bar{y}_n$ peak. Thus, by using the β_n determined from the latest reported $\bar{y}_n$ data, one can predict the date of $\bar{y}_n$ peak within a limited prediction error; and one can also predict an upper limit for the height of the $\bar{y}_n$ peak.

Prediction of lower and upper bounds of the $\bar{y}_n$ curve after $\bar{y}_n$ peak. After the number of daily new COVID-19 cases, $\bar{y}_n$, passes its peak, β_n may remain the same or even decrease a bit until it is confirmed that the peak has completely passed. Then β_n will increase because the interventions for reducing social distance and wearing face masks will be gradually lifted. Furthermore, the new sub-Omicron variants with greater infectivity may spread any day after $\bar{y}_n$ peak to increase $\bar{y}_n$ again. These unknown/undetermined factors make it almost impossible to make long-term prediction of the exact trajectory of $\bar{y}_n$. However, since β_n most likely reaches its minimum value around the $\bar{y}_n$ peak, if we use this minimum value of β_n to predict changes in $\bar{y}_n$ in the near future, the simulated y_n curve will be likely lower than the reported $\bar{y}_n$ curve. This enables us to predict the lower bound of $\bar{y}_n$ curve in the near future after the $\bar{y}_n$ peak. The lower bound of $\bar{y}_n$ curves (dashed line) in Figure 2 was obtained by assuming that β_n =0.16/day after January 22, 2022. In addition to calculating the lower bound of y_n curve, we can also calculate an upper bound of the $\bar{y}_n$ curve by assuming that β_n rapidly increases to 1 or a greater number in a short period of time (solid line in Figure 2). This period is chosen to be significantly shorter than the actual period needed to increase β_n in the real world. In the calculations, we assumed that no new variants or sub-variants appear in this time period to affect $\bar{y}_n$ largely. As shown in Figure 2, the reported daily COVID-19 cases ($\bar{y}_n$) were within the predicted lower and upper bounds for 3 months, until April 22, 2023.

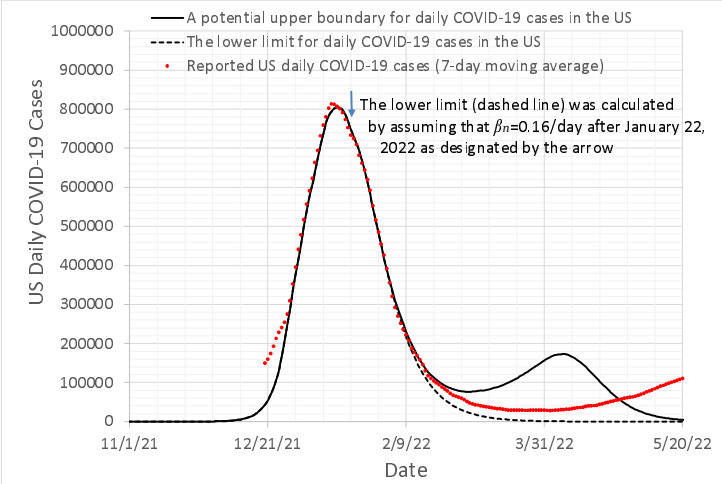

Figure 2. The simulated lower limit (dashed line) and upper limit (solid line) of the number of daily new COVID-19 cases (y_n) induced by Omicron variants in the US after the y_n peak. The reported y_n (red dotted line) is on or between the lower and upper limits for the simulated y_n.

Simulation/prediction of Omicron sub-variants induced increases in $\bar{y}_n$. After the Omicron-induced large peak of COVID-19 transmission in the US, the number of daily new COVID-19 cases, y_n, gradually decreased with time until early April 2022. Then, y_n started to increase again because of the increased β_n and transmission of multiple new sub-variants of Omicron with higher contagiousness comparing to the original Omicron variant. The major Omicron sub-variants that had significant contributions to COVID-19 transmission in the United States include (B.1.1.529 & BA.1.1), BA.2, BA.2.12.1, BA.2.75, BA.4, BA.4.6, BA.5, and (BQ.1 & BQ.1.1) [18, 19]. Among these sub-variants, B.1.1.529 and BA.1.1 were dominated in the big

peak of $\bar{y}_n$ as of mid-February 2022[20], and then other sub-variants followed separately. Considering that the later Omicron sub-variants had larger infectivity, we assume that each new sub-variant mentioned above can affect COVID-19 transmission by enlarging the number of susceptible people N. To simulate the Omicron-induced changes in y_n after mid-February 2022, we allowed N to increase on some selected dates (from N=250,000,000 to N=332,400,000) between the end of 2021 and early October 2022, while β_n gradually increases to 1 as of mid-September 2022. In this way, the simulated y_n can fit the reported $\bar{y}_n$ very well (Figure 3) as of Oct 23, 2022. It needs to be noted that, when N increases to 332400000, almost all of the population in the US has become susceptible to the highly infectious Omicron sub-variants. To predict y_n after October 23, 2022, we let β_n continuously increase to 3.5 before the end of November 2022, and remain at 3.5 after November 2022. The predicted y_n (solid line) forms a plateau from late October 2022 to the end of November 2022, and then y_n significantly decreases after early December, 2022, and y_n drops to nearly 1000 cases/day by the end of January 2023. This predicted result was uploaded to Twitter in late October, 2022[21]. The reported daily COVID-19 cases met the predicted results well until early December 2022[22]. Our simulation/prediction showed that after August 2022, especially after the y_n plateau in early December 2022, increasing β_n or emergence of more contagious Omicron variants would not push y_n up. This implies that the herd immunity to omicron has been reached in the United Sates base on the *l-i SEIR* model. In the above *l-i SEIR* model, it was assumed that any individual infected by an Omicron sub-variant would not be re-infected by any other Omicron sub-variants and any new COVID-19 variants. However, actually, Omicron-infected individuals still have a chance to be re-infected by an Omicron variant, even though the reinfection chance is very low. Therefore, the infected people are not able to form a perfect herd immunity. As we have seen, the reported daily new COVID-19 cases after late October 2022 (blue dots) formed a plateau between late October 2022 and late November 2022, which agreed with the predicted curve very well. However, the reported daily new COVID-19 cases slightly increased in the period between December 2022 and January 2023 because of the gatherings in the holiday seasons (Christmas and New Year) and a more contagious Omicron variant XBB.1.5 also appeared in this period[23]. This deviation from the predicted curve based on *l-i SEIR* model implies that a small ratio of Omicron-infected people

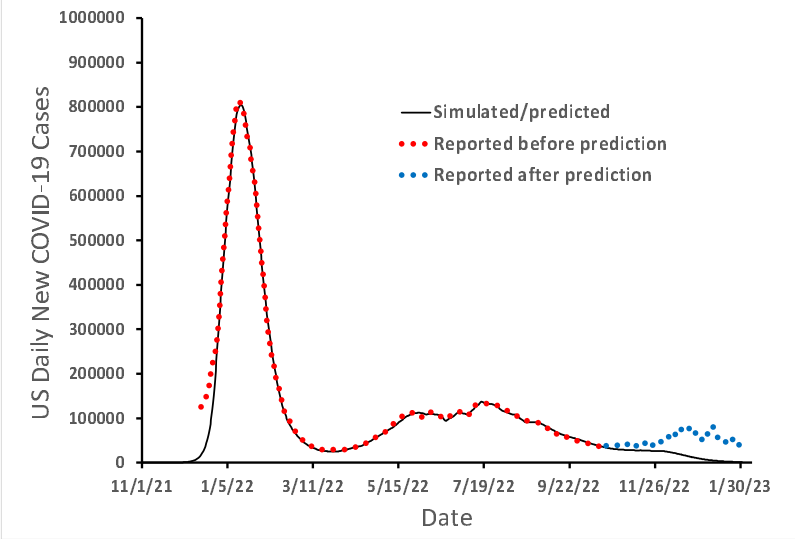

Figure 3. Simulated and reported number of Omicron-induced daily new COVID-19 cases in the United States without considering Omicron reinfection in the model.

can be re-infected by Omicron sub-variants and that the reinfection needs to be considered in the modelling.

Simulation/prediction of trajectory of $\bar{y}_n$ in the presence of reinfection of Omicron infections.

In the above *1-i SEIR* model, the number of susceptible people S_n varies between 0 and N, or $0 \le S_n \le N$. If most of susceptible people have been infected, then S_n is far smaller than N and the ratio S_n/N is near zero. Therefore, the number of daily new exposed people, $(S_n - S_{n-1})$ in Eqn. (1a), is also near 0. However, if the rate of reinfection of Omicron infected people is non-negligible, $(S_n - S_{n-1})$ must not be near zero even if all susceptible people have been infected. Thus, we suggest that in the presence of non-negligible rate of reinfection, the ratio S_{n-1}/N in Eqn. (3a) should be replaced by $[S_{n-1}(\ge 0)/N + a_n(N - S_{n-1}(\ge 0))/N]$. The second term in the square brackets consists of a reinfection coefficient a_n $(0 \le \alpha_n \le 1)$ and a weight factor $(N - S_{n-1}(\ge 0))/N$. When S_n is close to N, the weight factor is close to 0 and the first term in the square brackets plays the main role. However, when S_n is close to 0, the weight factor is close to 1 and the second term in the square brackets plays the main role. If a_n=0, it means no reinfection. In this situation, S_n can vary from N (no one is infected) to 0 (all susceptible people are infected). If reinfection rate is non-negligible, then a_n is greater than 0.

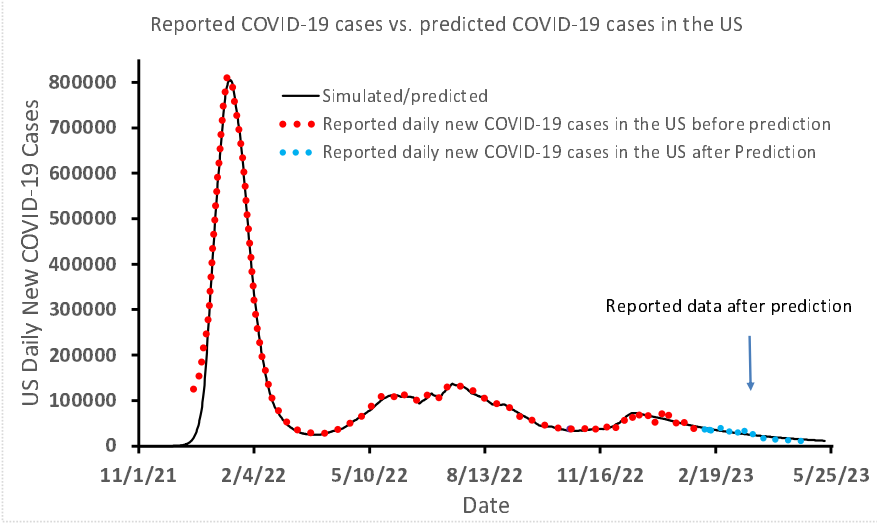

Figure 4. Simulated/predicted and reported number of Omicron-induced daily new COVID-19 cases in the United States after considering Omicron reinfection in the model.

In this situation, S_n can vary from N (no one is infected) to a negative number. The negative number means that not only all susceptible people are infected, but also a portion of them are re-infected. Thus, in the presence of reinfection, Eqn. (1a) is replaced by the following equation:

$$S_n - S_{n-1} = -k_n I_n \left(\frac{S_{n-1}(\ge 0)}{N} + a_n \frac{N - S_{n-1}(\ge 0)}{N}\right) \qquad (1a)'$$

In Eqn. (1a)', $S_n(\ge 0) = S_n$ if $S_n > 0$, and $S_n(\ge 0) = 0$ if $S_n \le 0$. Based on Eqns. (1a)' and (1b)-(1d), we simulated/predicted daily new COVID-19 cases on February 10, 2023[24], and compared them with reported data until May 5, 2023 (Figure 4)[25] when data of daily COVID-19 cases in the US were not updated anymore on websites. The red dots in Fig 4 represent the number of daily COVID-19 cases ($\bar{y}_n$) reported before the simulated/predicted trajectory of y_n (solid black line) was generated. The blue dots in Fig 4 represent $\bar{y}_n$ reported after the simulated/predicted

trajectory of y_n (solid black line) was generated. The result in Fig 4 shows that the reported $\bar{y}_n$ matches very well with the predicted trajectory of y_n.

Summary

Based on *l-i SEIR* model, the author described difficulties and discussed possible solutions in forecasting the peak day and the peak height of daily new COVID-19 cases ($\bar{y}_n$), the trajectory of $\bar{y}_n$ after the $\bar{y}_n$ peak, and the trajectory of $\bar{y}_n$ after the herd immunity was reached in the presence/absence of reinfection. Our simulations show that by using the β_n determined from the latest reported $\bar{y}_n$ data, one may predict the date of $\bar{y}_n$ peak within a limited prediction error, and also predict an upper limit for the height of the $\bar{y}_n$ peak. It is possible to accurately predict the trajectory of y_n after the $\bar{y}_n$ peak for a few of weeks (up to 4 weeks from 1/22/2022-2/19/2022 as shown in Figure 2) with a constant β_n. However, by calculating a lower limit and an upper limit of the y_n curve, one may successfully predict the trace of $\bar{y}_n$ within the range between the lower limit and upper limit of the y_n curve for more than 3 months (from 1/22/2022 to 4/28/2022 in Figure 2). The *l-i SEIR* model without considering Omicron reinfection could not explain the remaining non-negligible number of daily new COVID-19 cases after the herd immunity was reached ($S_n/N\approx 0$), suggesting that the Omicron reinfection should be taken into account in the model. The simulated y_n curve based on the *l-i SEIR* model considering Omicron reinfection can fit very well with the numbers of reported COVID-19 cases after the herd immunity has been reached, and the predicted y_n curve is in good agreement with the number of daily new COVID-19 cases reported to date May 10, 2023, twelve weeks after the prediction of y_n curve was made on February 10, 2023.

Calculation Methods and Programs.

Writing codes in Excel for calculating S_n, E_n, I_n, R_n and y_n.

We have described in detail how to put *l-i SEIR* model equations into Excel[26] to calculate model variables S_n, E_n, I_n, R_n, and y_n assuming that β_n=1 and the parameters or coefficients l, i (or c) and α are known. Below is the procedure for using Eqns. (2a)-(2e) to calculate these model variables in Excel.

1. In cells B1, C1, D1, E1, F1 (row 1 of the Excel sheet shown in Table 2), write S_0, N, l, c and α, respectively.
2. In cells B2, C2, D2, E2, F2 (row 2), input 329999999, 330000000, 4, 14 and 0.0145349, respectively. With these arrangements, we can easily know that S_0=329999999, N=330000000, l=4, c=14 and α=0.0145349. The values of these parameters were used for analyzing COVID-19 data reported from the United States[11, 12].
3. Write variable names in row 3. Especially, column A is used for date (A3: Date); column B for the number of susceptible individuals (B3: S_n Susceptible individuals), column C for the exposed individuals (C3: E_n Exposed individuals), column D for the infectious individuals (D3: I_n Infectious individuals), column E for the recovered individuals (E3: R_n Recovered individuals), column F for the calculated US daily new COVID-19 cases (F3: y_n the calculated US daily new cases), and column G for the transmission rate coefficient (G3: β_n in US). In this Excel calculation

program, we assumed that the first infection appeared on 2/2/2020[12], which is in row 28. We let n=0 on 2/2/2020.

4. Input dates in Excel in chronological order. First, we input 1/9/2020 in cell A4, and then click, hold, and drag the fill handle (a small black box on the bottom right) down until all of the cells we want to fill are selected. Release the mouse to fill the selected cells with the dates in chronological order as shown in column A of the Excel sheet (Table 2).

Table 2. Writing parameters, naming variables and setting initial conditions in Excel.

	A	B	C	D	E	F	G
1		S0	N	l	c	α	
2		329999999	330000000	4	14	0.0145349	
3	Date	S_n Susceptible individuals	E_n Exposed individuals	I_n *infectious individuals*	R_n Recorved individuals	y_n, the calculated US daily cases	k_n in US
4	1/9/2020	330000000	0	0	0	0	1
5	1/10/2020	330000000	0	0	0	0	1
6	1/11/2020	330000000	0	0	0	0	1
7	1/12/2020	330000000	0	0	0	0	1
8	1/13/2020	330000000	0	0	0	0	1
9	1/14/2020	330000000	0	0	0	0	1
10	1/15/2020	330000000	0	0	0	0	1
11	1/16/2020	330000000	0	0	0	0	1
12	1/17/2020	330000000	0	0	0	0	1
13	1/18/2020	330000000	0	0	0	0	1
14	1/19/2020	330000000	0	0	0	0	1
15	1/20/2020	330000000	0	0	0	0	1
16	1/21/2020	330000000	0	0	0	0	1
17	1/22/2020	330000000	0	0	0	0	1
18	1/23/2020	330000000	0	0	0	0	1
19	1/24/2020	330000000	0	0	0	0	1
20	1/25/2020	330000000	0	0	0	0	1
21	1/26/2020	330000000	0	0	0	0	1
22	1/27/2020	330000000	0	0	0	0	1
23	1/28/2020	330000000	0	0	0	0	1
24	1/29/2020	330000000	0	0	0	0	1
25	1/30/2020	330000000	0	0	0	0	1
26	1/31/2020	330000000	0	0	0	0	1
27	2/1/2020	330000000	0	0	0	0	1
28	2/2/2020	329999999	1	0	0	0	1

5. Set initial conditions for n≤0. Input 330000000 (S_n=N for n<0) or type =$C$2 into cells B4-B27 (corresponding to dates from 1/9/2020 to 2/1/2020 (A4-A27) because our simulations suggested that the first infection in the US appeared on 2/2/2020)[12]. This operation is easy to perform if the fill handle is used. Here $C$2 is the absolute cell reference to cell C2. When the absolute cell reference is copied from one cell to another cell, the absolute cell reference will remain unchanged. Because the value of the number in C2 is 330000000, we will see 330000000 in all cells B4-B27 if B4-B27 contain =$C$2. In this case, if we change the value in cell C2, the value in all cells B4-B27 will change accordingly. Input 0 to cells C4-C27 (E_n for n<0), cells D4-D27 (I_n for n<0), E4-E27 (R_n for n<0), and F4-F27 (calculated US daily new cases αI_n for n<0). Input 1 (β_n) to column G starting from G4 to all other cells in column G, corresponding to all dates available in column A. In row 28, n=0 and the date is 2/2/2020. The first exposed individual appeared on this day[12], so the number of susceptible individuals at n=0, S_0, is 329999999. Thus, we input 329999999 or =$B$2 in cell B28. The number of exposed individuals at n=0, E_0, is 1, so we input 1 in cell C28. However, I_0, R_0 and αI_0 are still 0 at n=0, so we input 0 in cell D28, E28 and F28. So far, we have set or input all parameters and initial values required for calculating S_n, E_n, I_n, R_n and αI_n. In the following, we will input Eqns. (2a)-(2e) into row 29 in the Excel sheet to calculate S_n, E_n, I_n, R_n and αI_n for n>0.

6. From Eqn. (2a) we know that the equation for S_1 can be simplified as below:

$$S_1 = S_0(1 - \frac{\beta_1(S_{1-c} - S_{1-l})}{N})$$

Similarly, S_2 and S_n can be written as:

$$S_2 = S_1(1 - \frac{\beta_2(S_{2-c} - S_{2-l})}{N})$$

$$S_n = S_{n-1}(1 - \frac{\beta_n(S_{n-c} - S_{n-l})}{N})$$

In our Excel program sheet, we let n=0 in row 28. Thus, row 29 will be n=1 and row 30 will be n=2 and so on. The value of S_0 is placed in cell B28, so S_1 will be placed in cell B29, and β_1 placed in G29. Furthermore, we have known that c=14, and l=4[11, 12], therefore $S_{1\text{-}c}$ or $S_{\text{-}13}$ will be placed in cell B15, and $S_{1\text{-}l}$ or $S_{\text{-}3}$ will be placed in B25. Thus, to input the above equation for S_1 in cell B29, we need to type the following codes (Excel codes) and click the Enter key:

=B28*(1-G29*(B15-B25)/330000000)

or

=B28*(1-G29*(B15-B25)/$C$2)

7. From Eqn. (2b) we know that the equation for E_1 can be simplified as below:

$$E_1 = S_{-3} - S_1$$

We have known in step 6 that $S_{\text{-}3}$ is placed in cell B25 and S_1 is placed in cell B29, so we should type the following codes in cell C29:

=B25-B29

8. From Eqn. (2c) we know that the equation for I_1 can be simplified as below:

$$I_1 = S_{1-c} - S_{1-l} = S_{-13} - S_{-3}$$

The value of I_0 is placed in cell D28, so I_1 is placed in cell D29. We have known in step 6 that $S_{\text{-}13}$ is placed in cell B15, and $S_{\text{-}3}$ is placed in B25, so we should type the following codes in D29:

=B15-B25

9. From Eqn. (2d) we can get the following equation for calculating the number of recovered individuals at n=1:

$$R_1 = N - S_{-13}$$

Because $S_{\text{-}13}$ is placed in cell B15, so we should type the following codes in E29:

=330000000-B15

or

=$C$2-B15

10. From Eqn. (2e) we can get the following equation for calculating the US daily new COVID-19 cases at n=1:

$$y_1 = \alpha I_1$$

The calculated US daily new COVID-19 cases y_1 is placed in cell F29. Thus, we need to type either line of the following codes in cell F29:

=0.0145349*D29

or

=$F$2*D29

The above Excel codes mean that the value of I_1 is in cell D29 and the value of α is in cell F2. We can directly input the value of α, 0.0145349, into F29 or input the absolute cell reference of α, $F$2, into F29.

11. So far, we have finished the Excel program for calculating S_1, E_1, I_1, R_1 and y_1. To calculate any other S_n, E_n, I_n, R_n and y_n for n>1, we just need to copy cells B29, C29, D29, E29 and F29, then paste them to B30, C30, D30, E30 and F30, and any other rows below row 30. This can be easily done by selecting cells B29, C29, D29, E29 and F29, and then drag the fill handle down to any row we wanted. If we successfully finish these steps, we should see the following Excel codes in B30, C30, D30, E30 and F30.

B30 (S_2): =B29*(1-G30*(B16-B26)/330000000) or =B29*(1-G30*(B16-B26)/$C$2)

C30 (E_2): =B26-B30

D30 (I_2): =B16-B26

E30 (R_2): =330000000.-B16 or =$C$2-B16

F30 (y_2): =0.0145349*D30 or =$F$2*D30

It can be seen that all row labels increase by 1 after copying cells in row 29 to cell in row 30, but the constant 330000000, the absolute reference $C$2 and $F$2 remain unchanged. Therefore, by simply copying & pasting operation in Excel, we are able to calculate all S_n, E_n, I_n, R_n, and y_n from Eqns. (2a)-(2e), assuming that β_n is a constant (β_n =1 in this example). However, in real situation, β_n varies with time because social interventions are applied for slowing down spread rate of COVID-19. By regulating β_n for fitting the calculated y_n to the reported $\bar{y}_n$, we can determining β_n for each n.

Data availability statement

All data used in this study are publicly available.

REFERENCES

1. United Nations: **WHO chief declares end to COVID-19 as a global health emergency**. 2023:https://news.un.org/en/story/2023/2005/1136367.

2. Worldometer: **Coronavirus Cases**. 2023:https://www.worldometers.info/coronavirus/.
3. Nixon K, Jindal S, Parker F, Marshall M, Reich NG, Ghobadi K, Lee EC, Truelove S, Gardner L: **Real-time COVID-19 forecasting: challenges and opportunities of model performance and translation**. *The Lancet Digital health* 2022, **4**(10):e699-e701.
4. Adam D: **Special report: The simulations driving the world's response to COVID-19**. *Nature* 2020, **580**(7803):316-318.
5. Holmdahl I, Buckee C: **Wrong but Useful - What Covid-19 Epidemiologic Models Can and Cannot Tell Us**. *The New England journal of medicine* 2020, **383**(4):303-305.
6. Wang P, Zheng X, Liu H: **Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics: A review**. *Frontiers in public health* 2022, **10**:1033432.
7. Sofonea MT, Cauchemez S, Boelle PY: **Epidemic models: why and how to use them**. *Anaesthesia, critical care & pain medicine* 2022, **41**(2):101048.
8. James LP, Salomon JA, Buckee CO, Menzies NA: **The Use and Misuse of Mathematical Modeling for Infectious Disease Policymaking: Lessons for the COVID-19 Pandemic**. *Medical decision making : an international journal of the Society for Medical Decision Making* 2021, **41**(4):379-385.
9. Liu X: **Analytical solution of I-i SEIR model-Comparison of I-i SEIR model with conventional SEIR model in simulation of epidemic curves**. *PloS one* 2023, **18**(6):e0287196.
10. Liu X: **Analytical Solution of a New SEIR Model Based on Latent Period-Infectious Period Chronological Order**. *medRxiv, https://doiorg/101101/2021121421267812* 2021:2021.2012.2014.21267812.
11. Liu X: **A simple, SIR-like but individual-based epidemic model: Application in comparison of COVID-19 in New York City and Wuhan**. *Results Phys* 2021, **20**:103712.
12. Liu X, DeVries AC: **Analysis of Current Status and Simulations of Future Spread of COVID-19 in the United States**. *Research Square, https://wwwresearchsquarecom/article/rs-310822/v1* 2021.
13. Staerke NB, Reekie J, Nielsen H, Benfield T, Wiese L, Knudsen LS, Iversen MB, Iversen K, Fogh K, Bodilsen J *et al*: **Levels of SARS-CoV-2 antibodies among fully vaccinated individuals with Delta or Omicron variant breakthrough infections**. *Nature communications* 2022, **13**(1):4466.
14. Chen J, Wang R, Gilby NB, Wei GW: **Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance**. *J Chem Inf Model* 2022, **62**(2):412-422.
15. Tapp T: **Omicron BA.5: One Variant To Rule Them All…For Now**. *Deadline* 2022:https://deadline.com/2022/2007/omicron-ba-2025-covid-variant-dominant-ba-five-one-1235058423/.
16. Tapp T: **Omicron BA.5: One Variant To Rule Them All…For Now**. *Deadline, https://deadlinecom/2022/07/omicron-ba-5-covid-variant-dominant-ba-five-one-1235058423/* 2022.
17. Drake J: **Vaccine Effectiveness In The Omicron Wave**. *Forbes, https://wwwforbescom/sites/johndrake/2022/02/01/vaccine-effectiveness-in-the-omicron-wave/?sh=593d17e54ee6* 2022.
18. CDC: **Monitoring Variant Proportions**. *COVID Data Tracker, https://covidcdcgov/covid-data-tracker/#variant-proportions* 2022.
19. Shaw J: **COVID-19 Common Sense**. *Harvard Magazine https://wwwharvardmagazinecom/2022/11/tripledemic-risks* 2022.
20. CDC: **COVID-19 Data Tracker Weekly Review**. *Centers for Disease Control and Prevention, https://wwwcdcgov/coronavirus/2019-ncov/covid-data/covidview/past-reports/04222022html* 2022.
21. Liu X: **"I made a forecast for the trend of COVID-19 cases between late October 2022 and January 30, 2023."**. In: *Twitter.* https://t.co/is5Hd87o5N 2022.

22. Liu X: **"Continuing the weekly test in Week 6 (Latest reported COVID-19 data updated on 12/2/2022)"**. In: *Twitter.* https://t.co/akVc3wlmjZ; 2022.
23. Haridy R: **Made in USA: XBB.1.5 fastest spreading variant since first Omicron wave**. *New Atlas* 2023:https://newatlas.com/health-wellbeing/xbb15-covid-spreading-variant-made-in-america-explainer/.
24. Liu X: **"I made a new forecast for the trend of daily COVID-19 cases between February 10, 2023 and early Aril, 2023 after considering Omicron reinfection in the model"**. In: *Twitter.* https://t.co/bnRWnkwpXl; 2023.
25. Liu X: **"Eleven weeks after the forecast, the number of reported COVID-19 cases in the US still closely matches the trajectory of COVID-19 transmission predicted by the mathematical model"**. In: *Twitter.* https://t.co/fl1EHa9gl3; 2023.
26. Liu X: **Programming Procedure in Excel for Calculating Model Variables in a New SEIR Epidemic Model Based on the latent-infectious period chronological order**. *Mendeley Data, https://datamendeleycom/datasets/z9jsfq8gbs/3* 2022:Version 3.