IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2023.0322000*

# Method and Application of Estimating Epidemiological Parameters Based on Data-driven Approach

**YUQING SUN[1], ZHONGHUA ZHANG[1], and GAOCHANG ZHAO[1]**
[1]School of Science, Xi'an University of Science and Technology, Xi'an, 710054, Shaanxi, China
Corresponding author: ZHONGHUA ZHANG (e-mail: wwwzhangzhonghua@xust.edu.cn).

**ABSTRACT** Determining initial variables and key parameters, such as case fatality ratio (CFR), dynamic case fatality ratio (DCFR), reproduction number ($R_0$), and so on, helps shed more light on the transmission and control of emerging and re-emerging infectious diseases. Here, we established a SAIUHR model, which describes the dynamic changes of susceptible, asymptomatic infectious, under-reported symptomatic infectious, hospitalized and recovered individuals. And we proposed a novel approach based on our model to calculate the report rate, starting time, basic reproduction number, the initial conditions for the compartments, CFR and DCFR. Finally, we apply our method to epidemiological datasets from China, Italy, Germany, and France. The results show that the goodness of fit for the cumulative confirmed cases is greater than $97.45\%$ in each of the countries, DCFR is more effective than CFR in predicting the future tend of infectious disease, and improving the report rate, raising the control strength and shortening the wait time are the effective strategies against infectious diseases. This study highlights the implications of taking proper restrictions and strong policies to deal with emerging and re-emerging infectious diseases from their spread in the early stage.

**INDEX TERMS** Basic reproduction number, data-driven, epidemic model, parameter estimation.

## I. INTRODUCTION

Emerging and re-emerging infectious diseases not only have led to severe public health crisis in most of countries in the world, but also brought negative influence on economic growth [1].For example, more than 600 million individuals worldwide have been diagnosed and over 6.5 million have died due to the ongoing global COVID-19 by late-October 2022 [2]. The most urgent question to be solved for the global is to take scientific and effective strategies to prevent and control various contagious diseases. Mathematical model helps explore the propagation mechanism and provide theoretical support for the public authorities to make prevention and control measures. Lots of differential equation models have been formulated to character the dynamics of different diseases such as Ebola, COVID-19 and Plague, and to evaluate the availability of the prevention measures, see [3] – [12] for examples. And many scholars had modified the early compartment epidemic models, such as SIR and SEIR et al. to simulate the evolution process of diseases and to analyze their epidemiological characteristics, see [13] – [21]. For examples, Cooper et al. developed a SIR model that provides a theoretical framework to investigate the spread of coronavirus virus within communities and found that the disease can be under control in all communities considered by comparing the recorded data with the data from their modelling approaches if proper restrictions and strong policies are implemented to control the infection rates early [13]; Osemwinyen et al. proposed two modified SIR models to simulate the transmission dynamics of Ebola, and the results revealed that although there are no particular drugs to treat it currently, effective segregating measures can help manage and control its spread [14]; Ngeleja et al. modified the conventional SEIR model to study the dynamics of Plague, the results showed that the infected flea population plays a decisive role in the spread of the epidemic to humans and rodents [19].

As well known, suitable models and reliable data are necessary to ascertain the spread mechanism of infectious diseases. Though a large number of works, such as [13] – [21] and the references therein, have made it possible to research on the development of diseases with the previous models quantitatively or qualitatively, there still remains some uncertainties in the theoretical framework, especially for how to estimate the number of the undocumented infected individuals in the population, or what kind of prevention and control measure

should be extensively adopted by the government to deal with the disease.

To explore the potential influence of under-reported symptomatic on disease transmission, the corresponding compartments were soon included in mathematical models in [22] – [26] and the references therein. Nevertheless, it is concomitantly an additional source of uncertainty in the initial conditions. Bayesian statistical inference methods are developed to estimate the percentage of undocumented infections based on epidemic models in [22] – [25]. However, the predictions deeply rely on prior and posterior probabilities. Additionally, due to a lack of complete data, estimating the parameters and initial conditions of the models to assess the effectiveness of each prevention and control measure remains a major challenge. A study presented a mechanistic approach using a compartmental model including asymptomatic and pre-symptomatic infectious that allows to estimate the level of undocumented infections and the effective reproduction number from the reported cases, deaths, and epidemiological parameters [26]. Mechanistic approaches provide a new way of drawing conclusions from the exiting data, without referring to prior and posterior probabilities, as well as they are considerably more robust as predictive tools than are purely empirical method.

CFR, defined as the ratio between the numbers of diagnosed deaths and cases, provides useful information on clinical types and presentations of the disease, see [26] – [31]. However, its definition ignores the role of the cases cured by medical treatment in forecasting transmission tend of disease. Here, we present another variable named DCFR which is the ratio between the cumulative deaths and sum of the cumulative deaths and the cumulative cured. It differs substantially from CFR and has advantage in predicting the future tend of epidemic.

Referring to [32], we assume the transmission rate is a piecewise function of time, and formulate a SAIUHR model including the reported, the under-reported symptomatic infectious and hospital-treated compartments. A method similar to [26] is used to estimate the parameters and initial conditions of the model on the basis of the epidemic surveillance data for the confirmed cases. We report the starting time, report rate, under-report rate, and basic reproduction number across different geographical areas. Additionally, the impact of the control measures with different intensities is analyzed by using the basic reproduction number and the time-varying transmission rate. The method can be rapidly altered or universally applied to work with different types of data, pandemic compartments, and population stratification.

The remainder of this paper is organized as follows: the formulation process by which an epidemic compartmental model is presented in subsection II-A, the approach to estimate initial conditions and parameters is described in subsection II-B, while the method to calculate DCFR is presented in subsection II-C, and the quantitative analysis of prevention and control measures is detailed in subsection II-D. Section III is the application to epidemiological time-series datasets.

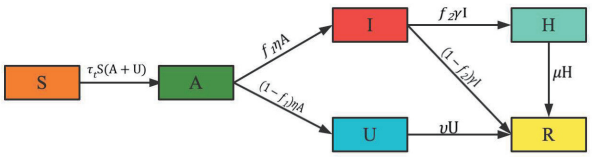



**FIGURE 1. Schematic diagram of the SAIUHR model.**

And main conclusions of the study are made in section IV.

## II. MATERIALS AND METHODS

### A. EPIDEMIC MODEL

We divide the population into six compartments: susceptible individuals (S), asymptomatic infectious individuals (A), reported symptomatic infectious individuals (I), under-reported symptomatic infectious individuals (U), hospitalized individuals (H) and recovered individuals (R). $S(t), A(t), I(t), U(t), H(t), R(t)$ denote at time $t$ the number of the six population categories S, A, I, U, H, and R, respectively, all of them are continuous variables. The following reasonable assumptions are made to simplify the problem.

**Hypothesis 1.** The population is generally susceptible to infection at the start of the epidemic, and the change in $S(t)$ is negligible. A and U are infectious, thus S may be infected by them. Assuming the transmission rate is $\tau_t$, as well as the increment content of A at time $t$ equals $\tau_t S(t)[A(t) + U(t)]$.

**Hypothesis 2.** Assuming that symptomatic could only be converted through asymptomatic infectious individuals, the average incubation period of asymptomatic infections is determined by the nature of the virus itself and not affected by elements like geographic location. Let $f_1$ and $\eta$ be report rate of symptomatic infectious individuals and the removal rate of asymptomatic infectious individuals, so that the increment content of I and U at time $t$ are $f_1\eta A(t)$ and $(1 - f_1)\eta A(t)$, respectively.

**Hypothesis 3.** Infected individuals who have reported as well as hospitalized patients no longer have the ability to infect others. Due to local medical resource constraints, a proportion of mildly ill patients cannot enter hospitals for treatment. Let $f_2$ and $\gamma$ be hospitalization rate and the removal rate of reported symptomatic infectious individuals, so that the increment content of H is $f_2\gamma I(t)$ at time $t$.

**Hypothesis 4.** Assuming that $\upsilon$ is the direct removal rate of symptomatic infections and $\mu$ is the removal rate of inpatients, so that the increment content of R at time $t$ is $(1 - f_2)\gamma I(t) + \upsilon U(t) + \mu H(t)$.

Under the above hypotheses, the compartment diagram is shown in Fig. 1.

Our model is governed by following nonautonomous ordinary differential equations

$$\begin{cases} S^{'}(t) &= -\tau_t S(t)[A(t)+U(t)], \\ A^{'}(t) &= \tau_t S(t)[A(t)+U(t)] - \eta A(t), \\ I^{'}(t) &= f_1\eta A(t) - \gamma I(t), \\ U^{'}(t) &= (1-f_1)\eta A(t) - \upsilon U(t), \\ H^{'}(t) &= f_2\gamma I(t) - \mu H(t), \\ R^{'}(t) &= (1-f_2)\gamma I(t) + \upsilon U(t) + \mu H(t). \end{cases} \tag{1}$$

Because the state variable $R$ is decupled from the rest of (1), we will not focus on its dynamic in the following. Then, (1) is reduced to

$$\begin{cases} S^{'}(t) &= -\tau_t S(t)[A(t)+U(t)], \\ A^{'}(t) &= \tau_t S(t)[A(t)+U(t)] - \eta A(t), \\ I^{'}(t) &= f_1\eta A(t) - \gamma I(t), \\ U^{'}(t) &= (1-f_1)\eta A(t) - \upsilon U(t), \\ H^{'}(t) &= f_2\gamma I(t) - \mu H(t). \end{cases} \tag{2}$$

For convenience, set the initial conditions

$$\begin{aligned} S(t_0) = S_0 > 0, A(t_0) = A_0 > 0, \\ I(t_0) = I_0 = 0, U(t_0) = U_0 \geq 0, H(t_0) = H_0 \geq 0. \end{aligned} \tag{3}$$

### *B. METHOD TO ESTIMATE INITIAL CONDITIONS AND PARAMETERS*

The parameters and initial conditions are typically unknown. Our goal is to determine the remaining parameters as well as the initial conditions for (2) from time-series data of the cumulative reported cases. By using the method from the literature [33], we suppose that the number of the cumulative reported cases is approximately exponential growth with time evolving. The cumulative reported confirmed data at time $t$, denoted by $CI(t)$, is formatted as follows

$$CI(t) = aexp(bt) - c, t \geq t_0. \tag{4}$$

We estimate parameters $a$, $b$, and $c$ by using the least squares estimation method. It can be seen that the cumulative reported symptomatic confirmed data is zero at $t_0$. Then, by using (4), we obtain

$$t_0 = \frac{1}{b} ln \frac{c}{a}. \tag{5}$$

Using a method of Constant Variation to solve the last three equations of system (2), we obtain the formula for cumulative individuals of the reported symptomatic

$$CI(t) = f_1\eta \int_{t_0}^{t} A(\theta)d\theta, t \geq t_0, \tag{6}$$

the under-reported symptomatic

$$CU(t) = (1-f_1)\eta \int_{t_0}^{t} A(\theta)d\theta, t \geq t_0, \tag{7}$$

and the hospitalized

$$CH(t) = f_2\gamma \int_{t_0}^{t} I(\theta)d\theta, t \geq t_0. \tag{8}$$

Differentiating (4) and (6) with respect to $t$ yields

$$\begin{cases} CI^{'}(t) &= abexp(bt), \\ CI^{'}(t) &= f_1\eta A(t). \end{cases} \tag{9}$$

Solving (9) leads to

$$A(t) = A_0 exp[b(t-t_0)]. \tag{10}$$

The derivative of $A(t)$ is

$$A^{'}(t) = bA_0 exp[b(t-t_0)]. \tag{11}$$

From the second equation of system (2), we obtain

$$U(t) = \frac{1}{\tau_t S_0}[A^{'}(t) + \eta A(t)] - A(t). \tag{12}$$

By using of (10), (11), (12) we obtain

$$U(t) = (\frac{bA_0}{\tau_t S_0} + \frac{\eta A_0}{\tau_t S_0} - A_0)exp[b(t-t_0)]. \tag{13}$$

Let $t = t_0$. Then, we obtain

$$U_0 = \frac{bA_0}{\tau_t S_0} + \frac{\eta A_0}{\tau_t S_0} - A_0. \tag{14}$$

Substituting (14) into (13), we get

$$U(t) = U_0 exp[b(t-t_0)], \tag{15}$$

and

$$U^{'}(t) = bU_0 exp[b(t-t_0)]. \tag{16}$$

It can be seen that the infection state system for (2) is

$$\begin{cases} A^{'}(t) &= \tau_t S(t)[A(t)+U(t)] - \eta A(t), \\ U^{'}(t) &= (1-f_1)\eta A(t) - \upsilon U(t). \end{cases} \tag{17}$$

By using of (11), (16) and (17), and set $t = t_0$, we obtain

$$\begin{cases} bA_0 &= \tau_t S_0[A_0 + U_0] - \eta A_0, \\ bU_0 &= (1-f_1)\eta A_0 - \upsilon U_0. \end{cases} \tag{18}$$

Solving (14) and (18), we obtain

$$\begin{cases} \frac{U_0}{A_0} &= \frac{b+\eta}{\tau_t S_0} - 1, \\ \frac{U_0}{A_0} &= \frac{(1-f_1)\eta}{\upsilon+b}. \end{cases} \tag{19}$$

From (19), it can be derived that the report rate of symptomatic infectious individuals

$$f_1 = 1 - \frac{(\upsilon+b)(b+\eta-\tau_t S_0)}{\tau_t S_0 \eta}. \tag{20}$$

And from (5), (9) and (20), we obtain the initial values of asymptomatic and under-reported symptomatic infectious individuals as follows

$$A_0 = \frac{bc\tau_t S_0}{\tau_t S_0\eta - (\upsilon+b)(b+\eta-\tau_t S_0)}, \tag{21}$$

$$U_0 = \frac{bc(b+\eta-\tau_t S_0)}{\tau_t S_0\eta - (\upsilon+b)(b+\eta-\tau_t S_0)}. \tag{22}$$

Using (4), (8) and (20), it follows that

$$H_0 = \frac{f_2\upsilon ab}{1-f_2}. \tag{23}$$

Gathering up above discussions, the expressions for unknown parameters are given in (5), (20) and the initial conditions are governed by (21), (22), (23), respectively.

### C. METHOD TO CALCULATE DCFR

Here, we introduce a function named the dynamic case fatality ratio (DCFR) to help us predict the evolving trend of disease, and it is defined as

$$\frac{x(t)}{x(t)+y(t)}, \tag{24}$$

where $x(t), y(t)$ are the numbers of cumulative death cases and the cumulative cured cases, respectively.

Obviously, the dynamic case fatality ratio differs substantially from the case fatality ratio [28], and it provides a simple way to give us information for the spreading tend of the disease.

### D. QUANTITATIVE ANALYSIS OF PREVENTION AND CONTROL STRATEGIES

Basic reproduction number is an important parameter in virus research from an epidemiological standpoint, as it allows us to understand how many individuals infected with a disease will transmit it to the rest of the population on average, without external intervention and in the absence of immunity in the population. Usually, the basic reproduction number is denoted as $R_0$, and it is calculated by NGN [34].

Obviously, the disease-free equilibrium of (2) is

$$E_0 = (S_0, 0, 0, 0, 0),$$

and the infection state is

$$\{A, U\}.$$

The transmission matrix and transition matrix are respectively as

$$F = \begin{bmatrix} \tau_t S_0 & \tau_t S_0 \\ (1-f_1)\eta & 0 \end{bmatrix},$$

and

$$V = \begin{bmatrix} \eta & 0 \\ 0 & \upsilon \end{bmatrix}.$$

Therefore, the next generation matrix is governed by

$$K = FV^{-} = \begin{bmatrix} \frac{\tau_t S_0}{\eta} & \frac{\tau_t S_0}{\upsilon} \\ 1-f_1 & 0 \end{bmatrix}.$$

It can be calculated that the basic reproduction number is as follows

$$R_0 = \rho(FV^{-}) = \frac{\tau_t S_0}{2\eta}\left(1+\sqrt{1+\frac{4(1-f_1)\eta^2}{\upsilon\tau_t S_0}}\right). \tag{25}$$

Suppose that the number of cumulative cases exponentially increases with a constant transmission rate at the early stage of the pandemic. With the development of the disease, the transmission rate $\tau_t$ will be weakened once a prevention and control measure takes effect. Denoting the days from the disease outcoming to a kind of measure taking effect as $T$ (here we called it as wait time), we employ an exponential function to express this decline, and the piecewise transmission rate is governed by

$$\tau_t = \begin{cases} \tau_0, 0 \le t \ge T, \\ \tau_0 exp[-\delta(t-T)], T < t, \end{cases} \tag{26}$$

**TABLE 1. The cumulative data of confirmed cases in China, Italy,Germany and France.**

| Date | Cases in China | Date | Cases in Italy | Date | Cases in Germany | Date | Cases in France |
|---|---|---|---|---|---|---|---|
| 01/19 | 198 | 02/23 | 155 | 02/28 | 48 | 02/27 | 38 |
| 01/20 | 291 | 02/24 | 229 | 02/29 | 79 | 02/28 | 57 |
| 01/21 | 440 | 02/25 | 322 | 03/01 | 130 | 02/29 | 100 |
| 01/22 | 571 | 02/26 | 453 | 03/02 | 159 | 03/01 | 130 |
| 01/23 | 830 | 02/27 | 655 | 03/03 | 196 | 03/02 | 191 |
| 01/24 | 1287 | 02/28 | 888 | 03/04 | 262 | 03/03 | 212 |
| 01/25 | 1975 | 02/29 | 1128 | 03/05 | 482 | 03/04 | 288 |
| 01/26 | 2744 | 03/01 | 1694 | 03/06 | 670 | 03/05 | 426 |
| 01/27 | 4515 | 03/02 | 2036 | 03/07 | 799 | 03/06 | 616 |
| 01/28 | 5974 | 03/03 | 2502 | 03/08 | 1040 | 03/07 | 948 |
| 01/29 | 7711 | 03/04 | 3089 | 03/09 | 1176 | 03/08 | 1125 |
| 01/30 | 9692 | | | | | | |
| 01/31 | 11791 | | | | | | |

where $\delta$ denotes the strength of a prevention and control measure.

## III. APPLICATION

### A. DATA PREPARATION

Our analysis focuses on four typical countries (China, Italy, Germany, and France) to estimate the values of key parameters and initial variables of our model. Epidemiological time-series datasets include the confirmed cases, the deaths, and the cured. The data in China is collected and compiled by the National Health Commission of the People's Republic of China [35], and in other countries is obtained from the Center for Systems Science and Engineering at Johns Hopkins University [36].

### B. INITIAL CONDITIONS AND PARAMETERS

Time windows for the four countries are respectively chosen as: (1) from January 19 to January 31, 2020, for China, (2) from February 23 to March 4, 2020, for Italy, (3) from February 28 to March 9, 2020, for Germany, (4) from February 27 to March 8, 2020, for France. The actual epidemic data in each country during the time window is shown in Table 1.

Firstly, using the cumulative confirmed cases as showed in Table 1, we fix the value of parameter $c$ as 1 for the selected countries, and then estimate the values of parameters $a$ and $b$ by the least square method. The values of $\eta$, $\upsilon$ and $\tau_t$, referring to literatures [32], [37], [38] are $\eta = \frac{1}{5}$, $\upsilon = \frac{1}{7}$, and $\tau_t$ equals to: (1) $4.44 \times 10^{-8}$ in China; (2) $6.15 \times 10^{-9}$ in Italy; (3) $5.18 \times 10^{-9}$ in Germany; (4) $6.18 \times 10^{-9}$ in France. Next, we take the initial values of susceptible individuals $(S_0)$ as the total population who lived in the outbreak region at the early stage of the disease, *i.e.* (1) $1.1212 \times 10^7$ in China; (2) $5.9066 \times 10^7$ in Italy; (3) $8.32 \times 10^7$ in Germany; (4) $6.74 \times 10^7$ in France. We assume that the hospitalization rate of reported symptomatic infectious individuals in the early stages is $85\%$ $(f_2 = 85\%)$. At last, the values of $t_0, f_1, A_0, U_0, H_0$ are calculated from (5), (20), (21), (22), (23) and they are exhibited in Table 2.

Obviously, if further epidemiological data becomes available, these parameters may take new values. Based on the

**TABLE 2. Parameters values of (2).**

| Country | $a$ | $b$ | $c$ | $t_0$ | $f_1$ |
|---|---|---|---|---|---|
| China | 0.32 | 0.36 | 1 | 3.17 | 68.59% |
| Italy | 0.7 | 0.26 | 1 | 1.37 | 46.35% |
| Germany | 0.01 | 0.32 | 1 | 14.39 | 52.2% |
| France | 0.01 | 0.33 | 1 | 13.96 | 62.22% |

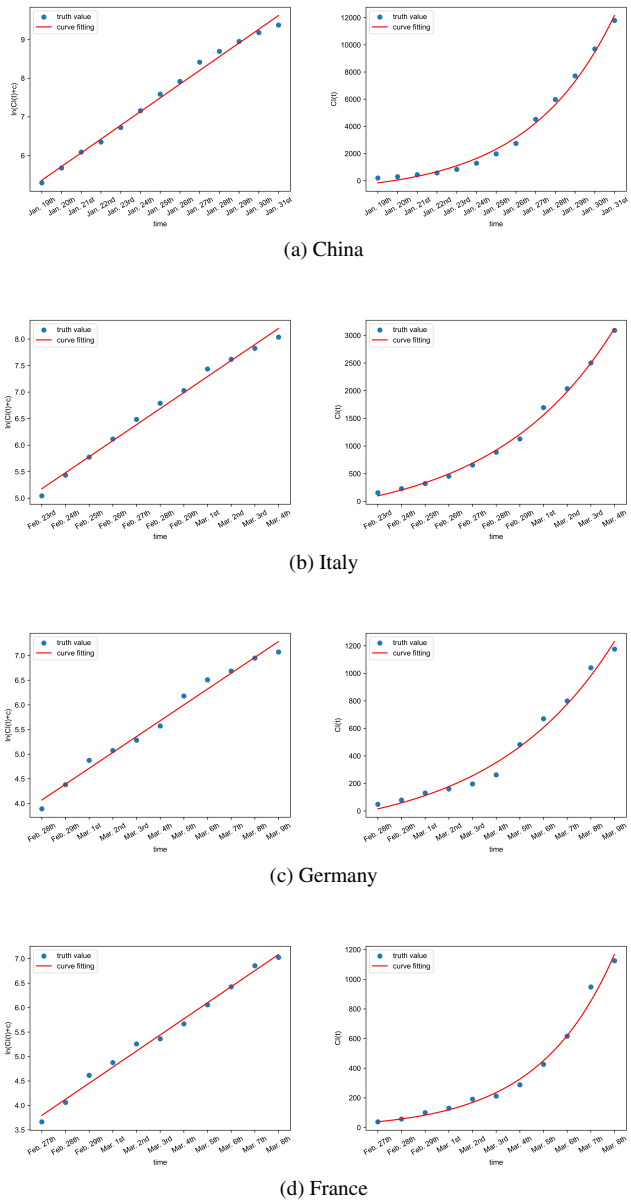
**TABLE 3. The effect evaluation of cumulative confirmed cases prediction in China.**

| Date | Truth Value | Predictive Value | Absolute Error | Relative Error | Goodness of Fit |
|---|---|---|---|---|---|
| 01/27 | 4515 | 4261 | 254 | 5.63% | |
| 01/28 | 5974 | 5619 | 355 | 5.63% | |
| 01/29 | 7711 | 7327 | 384 | 4.98% | 98.63% |
| 01/30 | 9692 | 9477 | 215 | 2.22% | |
| 01/31 | 11791 | 12182 | 391 | 3.32% | |

values in Table 2, it can be inferred that the starting time of the epidemic for the four countries are respectively January 4, 2020, February 2, 2020, February 15, 2020, and February 14, 2020. How the cumulative confirmed cases curve $CI(t)$ in (4) fits the cumulative confirmed data in Table 2 is shown in Fig. 2. It can be observed that the data for four countries has a good fit for $CI(t)$.

In what follows, the predictive value, absolute error, relative error and goodness of fitting for each country are presented in Table 3, Table 4, Table 5, Table 6 respectively. It can be seened that the predictive values do not differ significantly from the actual values, and all of the goodness's of fit are greater than 97.44%. The results indicate that the data for each country fit the curve of (4) well.

Next, we will predict the cumulative cases of the asymptomatic infectious, the under-reported symptomatic infectious and the hospitalized individuals, during the time window, by the ODE model (2) with initial values (obtained from (21)), (22) and (23)): (1) $A_0 = 2.62, U_0 = 0.33, H_0 = 0.0933$ for China, (2) $A_0 = 2.80, U_0 = 0.75, H_0 = 0.1474$ for




(a) China

(b) Italy

(c) Germany

(d) France

**FIGURE 2. In the left side, the dots (blue) correspond to $t \to ln(CI(t) + c)$, and the straight line (red) corresponds to $t \to ln(a) + bt$. In the right side, the dots (blue) correspond to $t \to CI(t)$, and the fitting curve (red) corresponds to $t \to aexp(bt) - c$, where $CI(t)$ is taken from the cumulative confirmed cases in Table 1.**

**TABLE 4. The effect evaluation of cumulative confirmed cases prediction in Italy.**

| Date | Truth Value | Predictive Value | Absolute Error | Relative Error | Goodness of Fit |
|---|---|---|---|---|---|
| 02/29 | 1128 | 1213 | 85 | 7.54% | |
| 03/01 | 1694 | 1562 | 132 | 7.79% | |
| 03/02 | 2036 | 1986 | 50 | 2.46% | 98.96% |
| 03/03 | 2502 | 2502 | 0 | 0% | |
| 03/04 | 3089 | 3129 | 40 | 1.29% | |

**TABLE 5. The effect evaluation of cumulative confirmed cases prediction in Germany.**

| Date | Truth Value | Predictive Value | Absolute Error | Relative Error | Goodness of Fit |
|---|---|---|---|---|---|
| 03/05 | 482 | 467 | 15 | 3.11% | |
| 03/06 | 670 | 607 | 63 | 10.37% | |
| 03/07 | 799 | 777 | 22 | 2.75% | 97.45% |
| 03/08 | 1040 | 983 | 57 | 5.48% | |
| 03/09 | 1176 | 1233 | 57 | 4.85% | |

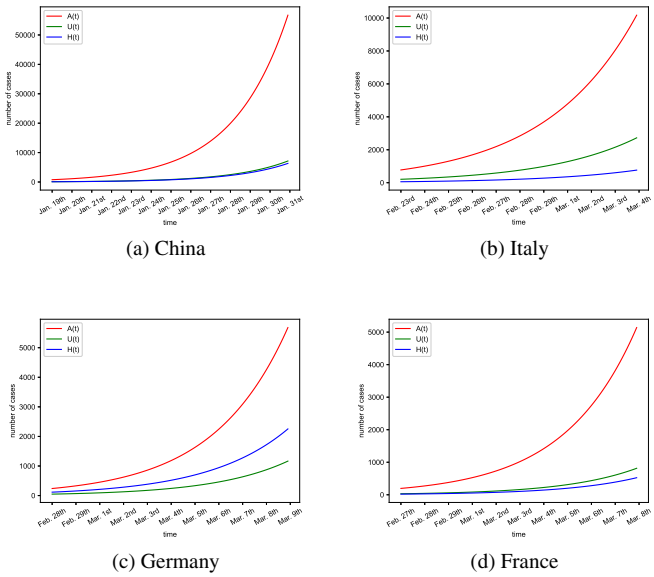
Italy, (3) $A_0 = 3.07, U_0 = 0.63, H_0 = 0.0026$ for Germany, (4) $A_0 = 2.65, U_0 = 0.42, H_0 = 0.0027$ for France. The numerical results are showed in Fig. 3.

In Fig. 3, the cases $A(t), U(t)$, and $H(t)$ in all counties increase with time evolving, and the raising speeds become more quickly at the late stage of time windows.

Using the value of $f_1$ in Table 2, it can be calculated that the under-report rates in the selected countries are 31.41%, 53.65%, 47.80%, 37.78%, respectively. For the sake of comparation, the report rate and the under-report rate are shown as a stacked histogram in Fig. 4. Clearly, the under-report rate

**TABLE 6. The effect evaluation of cumulative confirmed cases prediction in France.**

| Date | Truth Value | Predictive Value | Absolute Error | Relative Error | Goodness of Fit |
|---|---|---|---|---|---|
| 03/04 | 288 | 327 | 39 | 13.54% | |
| 03/05 | 426 | 451 | 25 | 5.87% | |
| 03/06 | 616 | 621 | 5 | 0.81% | 98.92% |
| 03/07 | 948 | 854 | 94 | 9.92% | |
| 03/08 | 1125 | 1170 | 45 | 4.00% | |




(a) China (b) Italy

(c) Germany (d) France

**FIGURE 3. The graphs of $t \to A(t), t \to U(t)$, and $t \to H(t)$ in each country.**

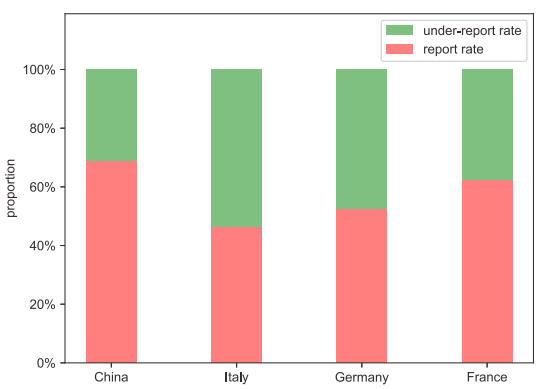
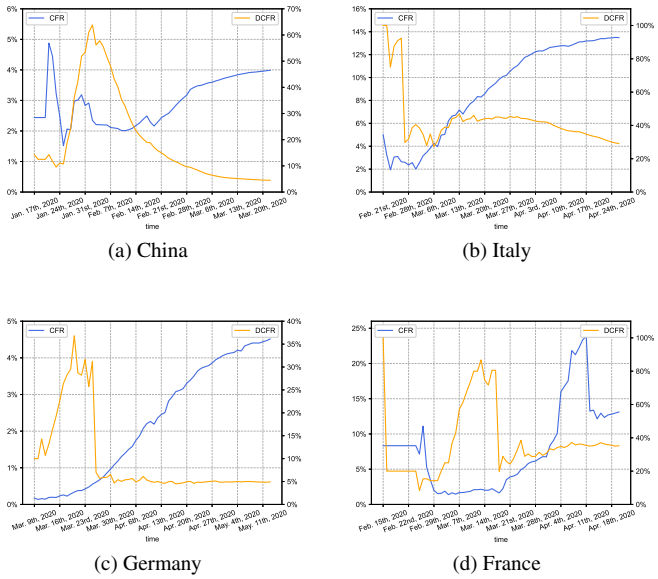



**FIGURE 4. The stacked histogram of report rate and under-report rate in four countries.**

in Italy is higher than those in the other countries at the early phase.




(a) China (b) Italy

(c) Germany (d) France

**FIGURE 5. The CFR and DCFR in 66 days after the first case confirmed in each country.**

### C. COMPARISON OF CFR AND DCFR

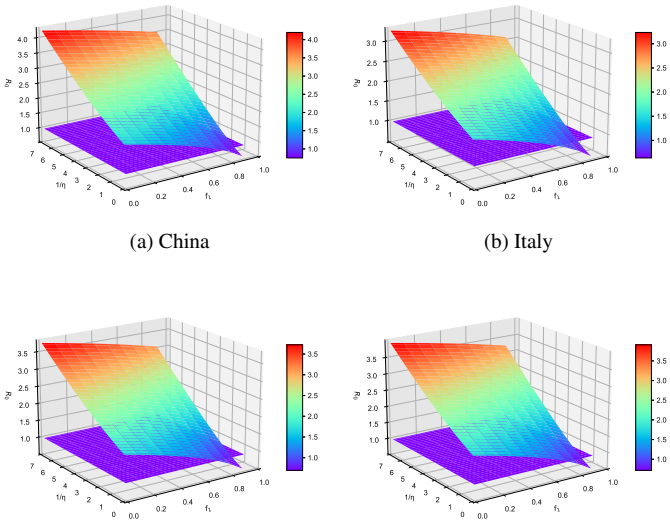
By using of (24) and the data of the confirmed cases, the death cases and the cured cases presented in [35], [36], we fit both the CFR and DCFR in each country, and the curves are shown Fig. 5, respectively. We choose the data (1) from January 17 to March 22, 2020 in China, (2) from February 21 to April 26, 2020 in Italy, (3) from March 9 to May 13, 2020 in Germany, and (4) from February 15 to April 20, 2020 in France.

Seen from Fig. 5 that CFR increases with the increasing of the number of deaths at the early stage of the pandemic, and then grows slowly at the end of time and ends at 4%, 13.5%, 4.5% and 13%, respectively. However, even though CFR increases continuously, DCFR becomes stable at 4.5%, 29%, 5%, and 35.5%, respectively, after the first oscitation. The decrease and stabilization of DCFR curve indicates a decrease in the number of deaths and an increase in the number of cured individuals. Calculating DCFR can evaluate the medical situation in different regions, provide scientific guidance, and reasonably arrange subsequent medical matters.

### D. PREVENTION AND CONTROL STRATEGIES

By fixing $\frac{1}{\upsilon} = 7$ , we estimate how the parameters $f_1$ and $\frac{1}{\eta}$ affect the value of the basic reproduction of $R_0$ using (25) in the domain $0 \leq \frac{1}{\eta} \geq 7, 0 \leq f_1 \geq 1$, to demonstrate the significance of report rate in the progression of the epidemic. The numerical results are shown in Fig. 6.

In Fig. 6, the maximum value of $R_0$ is larger than 3 for each of the selected countries, and then decreases to the number less than threshold value 1 as the decreasing of $\eta$ and increasing of $f_1$. That is to see that disease can be controlled and then extinguish if the government takes effective measures to



(a) China (b) Italy

(c) Germany (d) France

**FIGURE 6.** **The range of $R_0$ with respect to $f_1$ and $\frac{1}{\eta}$ in each country.**

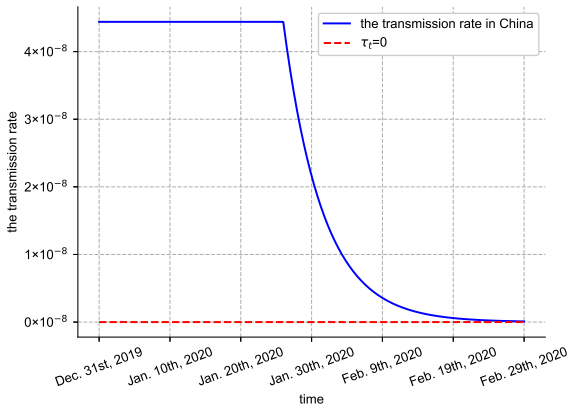


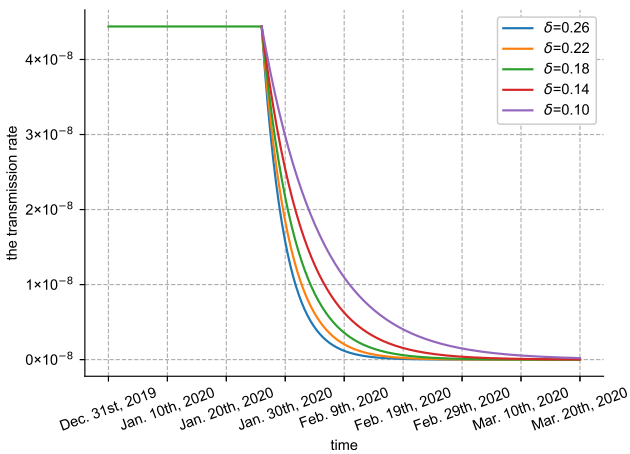

**FIGURE 7.** **The change of transmission rate in China with $T = 26$ and $\alpha = 0.18$.**

improve the report rate of symptomatic infectious individuals and reduces the removal rate of asymptomatic infectious individuals in a country. In most cases, whether an asymptomatic patient turns to a symptomatic one is unpredictable. Therefore, we confirm that report rate is an important parameter can be adjusted to affect the development of disease.

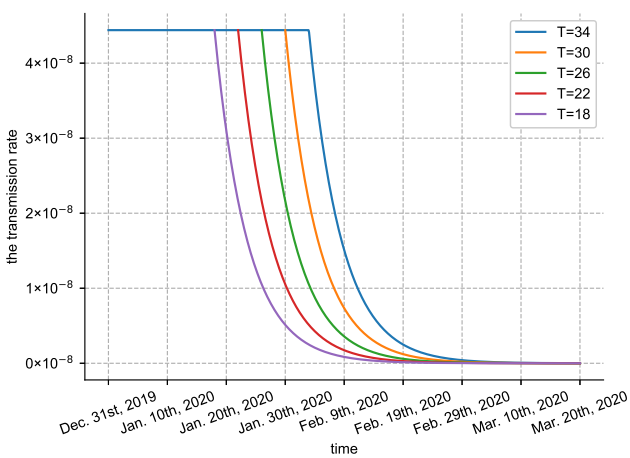
Considering China as an example, in addition, we will discuss the effect of control strength and the wait time on the control and prevention by using (26). Chinese National Health and Wellness Commission released the first version of Prevention and control of COVID-19 on January 15, 2020, which was implemented on January 16, 2020 [39]. It needs about 7 to 14 days for a prevention and control measure to take effect after it is implemented (see [40]). For simplicity, we assume that the time period for a control measure taking effect is 10 days. So, we deduce that the wait time is January 26, 2020 ($i.e. T = 26$) in China. In follows, we illustrate time-varying transmission rate $\tau_t$ in Fig. 7 with $\tau_0 = 4.44 \times 10^{-8}$, $\delta = 0.18$ (estimated in [32]) and $T = 26$.

It can be seen from Fig. 7 that the transmission rate declines sharply after the prevention and control measure taking effect and then tends to zero asymptotically. In the follows, Fig. 8a and Fig. 8b show the sensibility of $\tau_t$ with respect to the parameters $\delta$ and $T$, respectively.

From Fig. 8, the bigger the value of control strength and the earlier the prevention and control measure taking effect, the more quickly the transmission rate $\tau_t$ tends to 0. These mean that control strength and wait time may have a significant impact on the spread of the epidemic.




(a) Plot with $\delta = 0.10, 0.14, 0.18, 0.22, 0.26$ and $T = 26$




(b) Plot with $T = 18, 22, 26, 30, 34$, and $\delta = 0.18$

**FIGURE 8.** **Influence of $\delta$ and $T$ on $\tau_t$.**

## IV. CONCLUSION

In the present study, we formulate a SAIUHR compartment model to estimate the report rate of the infected cases which

cannot be directly accessible from the surveillance systems, and to calculate the starting time, the basic reproduction number and the initial conditions for the compartments by using the cumulative confirmed cases across different geographical areas. A simple data-driven approach is adopted to estimate the variables of CFR and DCFR in terms of time series with confirmed cases, death cases and cured cases. A piecewise function $\tau_t$ is used to express the transmission rate, which is perturbated by the wait time $T$.

The designed method was applied to epidemiological time-series datasets in the China, Italy, Germany, and France. Fig. 2 shows that our model fits the early cumulative data of confirmed cases in four countries well. With the actual data from the National Health Commission of the People's Republic of China and the Johns Hopkins Coronavirus Resource Center, we project the number of cases, asymptomatic, underreported symptomatic, and hospitalized, in each country during the time window. The results in Fig. 4 exhibit that the report rate in Italy is lower than those in other countries at the early phase of disease. Even though CFR increases continuously, DCFR tends to a stable value after significant fluctuations (as evident in Fig. 5). This means DCFR is more effective in predicting the controllability of the pandemic than the rate CFR, which also gives positive information the public and encourages them a lot to deal with the disease. By analyzing the impact of report rate $f_1$ on the basis reproduction number $R_0$ and the impact of controlling strength $\delta$ and the wait time $T$ on the time-dependent transmission rate, as illustrated in Fig. 6, Fig. 7, Fig. 8, we find that report rate, controlling strength and wait time affect the future tend of the disease tremendously. It is evident that the government authorities and the public health department should take the stricter containment measures and improve ability to identify as many as possible existing cases at the early stage to deal with the emerging and re-emerging infectious diseases.

This paper does have certain limitations. First, a few research findings have discovered that population movement, climate, and various mutated viruses may have some influence on the emergence of pandemics [18], [41], [42]. However, our model does not take these factors into account, and we will consider these factors in subsequent studies to make the model more consistent with reality. Second, We will study the model's stability analysis, method analysis and other factors, such as delayed and take geographic proximity into consideration as a risk factor in subsequent research. Third, because most of the parameters will change over time as the epidemic evolves, we fixed a few of them based on existing studies to reduce the uncertainty, and the validity of each parameter value will need to be improved further in future research.

## REFERENCES

[1] D. L. Heymann, G. R. Rodier, "Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases," *The Lancet infectious diseases*, vol. 1, no. 5, pp. 345–353, 2001.

[2] Johns Hopkins Coronavirus Resource Center. Accessed on: Oct. 24 2022. [Online]. Available: https://coronavirus.jhu.edu/map.html

[3] K. H. Green, "Fitting a Three-Phase Discrete SIR Model to New Coronavirus Cases in New York State," *International Journal of Data Analytics (IJDA)*, vol. 2, no. 2, pp. 59–74, 2021.

[4] S. Paul, A. Mahata, S. Mukherjee, B. Roy, M. Salimi, and A. Ahmadian, "Study of fractional order SEIR epidemic model and effect of vaccination on the spread of COVID-19," *International Journal of Applied and Computational Mathematics*, vol. 8, no. 5, pp. 237, Aug. 2022.

[5] M. Cai, G. Em Karniadakis, and C. Li, "Fractional SEIR model and data-driven predictions of COVID-19 dynamics of Omicron variant," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 7, pp. 071101, Jul. 2022.

[6] O. Zakary, A. Larrache, M. Rachik, and I. Elmouki, "Effect of awareness programs and travel-blocking operations in the control of HIV/AIDS outbreaks: a multi-domains SIR model," *Adv. Differ. Equ*, vol. 2016, no. 1, pp. 1–17, Jun. 2016.

[7] M. T. Hossain, M. M. Miah, and M. B. Hossain, "Numerical study of kermack-mckendrik SIR model to predict the outbreak of ebola virus diseases using euler and fourth order runge-kutta methods," *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 37, no. 1, pp. 1–21, Oct. 2017.

[8] P. E. Lekone, and B. F. Finkenstädt, "Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study," *Biometrics*, vol. 62, no. 4, pp. 1170–1177, Jun. 2006.

[9] O. Zakary, M. Rachik, and I. Elmouki, "On the impact of awareness programs in HIV/AIDS prevention: an SIR model with optimal control," *Int. J. Comput. Appl*, vol. 133, no. 9, pp. 1–6, Jan. 2016.

[10] V. Guseva, N. Doktorova, O. Krivorotko, O. Otpushchennikova, L. Parolina, I. Vasilyeva, M. Sosnovskaya, and A. Nemerov, "BUILDING A SEIR-MODEL FOR PREDICTING THE HIV/TUBERCULOSIS COINFECTION EPIDEMIC FOR RUSSIAN TERRITORIES WITH LOW TB BURDEN," *INT J INFECT DIS*, vol. 134, pp. S4–S5, Aug. 2023.

[11] L. A. White, and L. Mordechai, "Modeling the Justinianic Plague: Comparing hypothesized transmission routes," *PloS one*, vol. 15, no. 4, pp. e0231256, Apr. 2020.

[12] C. L. Mitchell, A. R. Schwarzer, A. Miarinjara, C. O. Jarrett, A. D. Luis, and B. J. Hinnebusch, "A Role for Early-Phase Transmission in the Enzootic Maintenance of Plague," *PLoS Pathogens*, vol. 18, no. 2, pp. e1010996, Dec. 2022.

[13] I. Cooper, A. Mondal, and C.G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," *Chaos, Solitons Fractals*, vol. 139, pp. 110057, Oct. 2020.

[14] A. C. Osemwinyen, and A. Diakhaby, "Mathematical modelling of the transmission dynamics of ebola virus," *APPL COMPUT MATH-BAK*, vol. 4, no. 4, pp. 313–320, Jul. 2015.

[15] D. E. Clark, G. Welch, and J. S. Peck, "A modified SIR model equivalent to a generalized logistic model, with standard logistic or log-logistic approximations," *IISE Trans Healthc Syst Eng*, vol. 12, no. 2, pp. 130–136, Aug. 2021.

[16] J. Wang, Y. Liu, X. Liu, and K. Shen, "A Modified SIR model for the COVID-19 epidemic in China," *J. Phys.: Conf. Ser.*, vol. 12, no. 2148, pp. 1, 2022.

[17] S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, and R. Mbogo, "SEIR model for COVID-19 dynamics incorporating the environment and social distancing," *BMC Res. Notes*, vol. 13, no. 1, pp. 1–5, Jul. 2020.

[18] Y. Wei, Z. Lu, Z. Du, Z. Zhang, Y. Zhao, S. Shen, B. Wang, Y. Hao, and F. Chen, "Fitting and forecasting the trend of COVID-19 by SEIR (+ CAQ) dynamic model," *Zhonghua Liuxingbingxue Zazhi*, vol. 41, no. 4, pp. 470–475, Apr. 2020.

[19] R. C. Ngeleja, L. S. Luboobi, and Y. Nkansah-Gyekye, "Modelling the dynamics of bubonic plague with yersinia pestis in the environment," *Commun. Math. Biol. Neurosci.*, vol. 2016, pp. Article ID 10, Jul. 2016.

[20] B. Gu, "Forecast and analysis of COVID-19 epidemic based on improved SEIR model," *J. Phys.: Conf. Ser.*, vol. 1802, no. 4, pp. 042050, 2021.

[21] H. M. Youssef, N. A. Alghamdi, M. A. Ezzat, A. A. El-Bary, and A. M. Shawky, "A modified SEIR model applied to the data of COVID-19 spread in Saudi Arabia," *AIP Adv.*, vol. 10, no. 12, pp. 125210, Dec. 2020.

[22] S. Djilali, L. Benahmadi, A. Tridane, and K. Niri, "Modeling the impact of unreported cases of the COVID-19 in the North African countries," *Biology*, vol. 9, no. 11, pp. 373, Nov. 2020.

[23] C. M. Batistela, D. P. Correa, Á. M. Bueno, and J. R. C. Piqueira, "SIRSi compartmental model for COVID-19 pandemic with immunity loss," *Chaos, Solitons Fractals*, vol. 142, pp. 110388, Jan. 2021.

[24] Z. Peng, S. Ao, L. Liu, S. Bao, T. Hu, H. Wu, and R. Wang, "Estimating unreported COVID-19 cases with a time-varying SIR regression model," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, pp. 1090, Jan. 2021.
[25] M. S. Aronna, R. Guglielmi, and L.M. Moschen, "Estimate of the rate of unreported COVID-19 cases during the first outbreak in Rio de Janeiro," *Infectious Disease Modelling*, vol. 18, no. 3, pp. 317–332, Jan. 2021.
[26] G. S. Costa, W. Cota, and S. C. Ferreira, "Data-driven approach in a compartmental epidemic model to assess undocumented infections," *Chaos, Solitons Fractals*, vol. 163, pp. 112520, Oct. 2022.
[27] S. S. Musa, A. Tariq, L. Yuan, W. Haozhen, and D. He, "Infection fatality rate and infection attack rate of COVID-19 in South American countries," *Infect. Dis. Poverty*, vol. 11, no. 1, pp. 1–11, Apr. 2022.
[28] G. Luo, X. Zhang, H. Zheng, and D. He, "Infection fatality ratio and case fatality ratio of COVID-19," *Int. J. Infect. Dis.*, vol. 113, pp. 43–46, Dec. 2021.
[29] L. J. Wolfson, R. F. Grais, F. J. Luquero, M. E. Birmingham, and P. M. Strebel, "Estimates of measles case fatality ratios: a comprehensive review of community-based studies. International journal of epidemiology," *INT J EPIDEMIOL*, vol. 38, no. 1, pp. 192–205, Feb. 2009.
[30] T. W. Russell, J. Hellewell, S. Abbott, N. Golding , H, Gibbs, C. Jarvis, K. van Zandvoort, S. Flasche, R. Eggo, and W. Edmunds, "Using a delay-adjusted case fatality ratio to estimate under-reporting," *Centre for Mathematical Modeling of Infectious Diseases Repository*, vol. 22, Mar. 2020.
[31] N. Wilson, and M. G. Baker, "The emerging influenza pandemic: estimating the case fatality ratio," *Eurosurveillance*, vol. 14, no. 26, pp. 19255, Jul. 2009.
[32] Z. Liu, P. Magal, and G. Webb, "Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom," *J. Theor. Biol.*, vol. 509, pp. 110501, Jan. 2021.
[33] Q. Griette, Z. Liu, and P. Magal, "Estimating the end of the first wave of epidemic for COVID-19 outbreak in mainland China," , Jul. 2020.
[34] O. Diekmann, J. Heesterbeek, and M. G. Roberts, "The construction of next-generation matrices for compartmental epidemic models," *J. R. Soc., Interface*, vol. 7, no. 47, pp. 873–885, Nov. 2009.
[35] National Health Commission of the People's Republic of China. Accessed on: Jun. 30 2022. [Online]. Available: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml
[36] Johns Hopkins Coronavirus Resource Center. Accessed on: Jun. 30 2022. [Online]. Available: https://coronavirus.jhu.edu/
[37] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong et al., "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, pp. 1199–1207, Jan. 2020.
[38] H. Xie, J. Tao, G. J. McHugo, and R. E. Drake, "Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation," *J. Subst. Abuse Treat.*, vol. 45, no. 1, pp. 99–108, Jul. 2013.
[39] National Health Commission of the People's Republic of China. Accessed on: Jun. 30 2022. [Online]. Available: http://www.nhc.gov.cn/wjw/xwdt/202004/ef0708c0e01f452492dbf6c418ef41a0.shtml
[40] J. Sun, Y. Zheng, W. Liang, Z. Yang, Z. Zeng, T. Li, J. Luo, M. T. A. Ng, J. He, and N. Zhong, "Quantifying the effect of public activity intervention policies on COVID-19 pandemic containment using epidemiologic data from 145 countries," *Value Health*, vol. 25, no. 5, pp. 699–708, May. 2022.
[41] K. Liu, Y. Yuan, Z. J. LIU, and M. LIU, "The impact of meteorological and environmental conditions on the spread of COVID-19," *Journal of Public Health and Preventive Medicine*, no. 6, pp. 9–13, 2020.
[42] E. B. Pathak, J. L. Salemi, N. Sobers, J. Menard, and I. R. Hambleton, "COVID-19 in children in the United States: intensive care admissions, estimated total infected, and projected numbers of severe pediatric cases in 2020," *Journal of Public Health Management and Practice*, pp. 325–333, Apr. 2020.

**YUQING SUN** received the B.Sc. degree in science from Xi'an University of Science and Technology, Xi'an, China, in 2021.

She is currently pursuing the M.Sc. degree in mathematics with Xi'an University of Science and Technology under the supervision of Dr. Zhonghua Zhang. Her research interests include parameter estimation, and modeling of epidemics.

**ZHONGHUA ZHANG** received the M.Sc., and Ph.D. degrees in school of mathematics and statistics, Xi'an Jiaotong University, Xi'an, China, in 2002, and 2007, respectively.

He is currently dean of the School of Science Xi'an University of Science and Technology. His research interests include differential dynamical system and biomathematics.

**GAOCHANG ZHAO** received the B.Sc. degree in science from Sun Yat sen University, Xi'an, China, in 1987.

He is currently a Professor with the Xi'an University of Science and Technology. His research interests include bayesian classification and application of catastrophe theory.