Improvement of the software for modeling the dynamics of epidemics and developing a user-friendly interface

Igor Nesteruk [a, b]

[a] Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine
[b] Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

ARTICLE INFO

ABSTRACT

The challenges humanity is facing due to the Covid-19 pandemic require timely and accurate forecasting of the dynamics of various epidemics to minimize the negative consequences for public health and the economy. One can use a variety of well-known and new mathematical models, taking into account a huge number of factors. However, complex models contain a large number of unknown parameters, the values of which must be determined using a limited number of observations, e.g., the daily datasets for the accumulated number of cases. Successful experience in modeling the COVID-19 pandemic has shown that it is possible to apply the simplest SIR model, which contains 4 unknown parameters. Application of the original algorithm of the model parameter identification for the first waves of the COVID-19 pandemic in China, South Korea, Austria, Italy, Germany, France, Spain has shown its high accuracy in predicting their duration and number of diseases. To simulate different epidemic waves and take into account the incompleteness of statistical data, the generalized SIR model and algorithms for determining the values of its parameters were proposed. The interference of the previous waves, changes in testing levels, quarantine or social behavior require constant monitoring of the epidemic dynamics and performing SIR simulations as often as possible with the use of a user-friendly interface. Such tool will allow predicting the dynamics of any epidemic using the data on the number of diseases over a limited period (e.g., 14 days). It will be possible to predict the daily number of new cases for the country as a whole or for its separate region, to estimate the number of carriers of the infection and the probability of facing such a carrier, as well as to estimate the number of deaths. Results of three SIR simulations of the COVID-19 epidemic wave in Japan in the summer of 2022 are presented and discussed. The predicted accumulated and daily numbers of cases agree with the results of observations, especially for the simulation based on the datasets corresponding to the period from July 3 to July 16, 2022. A user-friendly interface also has to ensure an opportunity to compare the epidemic dynamics in different countries/regions and in different years in order to estimate the impact of vaccination levels, quarantine restrictions, social behavior, etc. on the numbers of new infections, death, and mortality rates. As example, the comparison of the COVID-19 pandemic dynamics in Japan in the summer of 2020, 2021 and 2022 is presented. The high level of vaccinations achieved in the summer of 2022 did not save Japan from a powerful pandemic wave. The daily numbers of cases were about ten times higher than in the corresponding period of 2021. Nevertheless, the death per case ratio in 2022 was much lower than in 2020.

E-mail address: inesteruk@yahoo.com.
Peer review under responsibility of KeAi Communications Co., Ltd.

1. Introduction

The challenges humanity is facing due to the COVID-19 pandemic require timely forecasting of the dynamics of various epidemics in order to minimize the negative consequences for public health and the economy. For this purpose, a variety of well-known mathematical models can be used, taking into account many factors (Brauer & Castillo-Chávez, 2001; Daley & Gani, 2005; de Andres et al., 2021; Enrique Amaro et al., 2021; Gang, 2020; Hethcote, 2000; Huppert & Katriel, 2013; Kermack & McKendrick, 1927; Liu et al., 2021; Mohammadi et al., 2021; Nakamura, Cardoso, & Martinez, 2020; Costris-Vas et al., 2020; Vicuña et al., 2021; Weiss, 2013). However, complex models contain large numbers of unknown parameters, the values of which must be determined using a limited number of observations, e.g., the daily datasets for the accumulated number of cases. The limited amount of this information is connected not only with the need to make estimates and forecasts as quickly as possible, but also with constant changes in quarantine conditions, testing and isolation of patients, social behavior, mutations in the pathogen etc.

Successful experience in modeling the COVID-19 pandemic has shown that it is possible to apply the simplest SIR model, which links the number of susceptible (and unprotected) people S, infectious (infected and spreading the infection) I and removed R (immunized, isolated and dead), (Kermack & McKendrick, 1927). For the first waves of epidemics, the SIR model contains 4 unknown parameters, the values of which can be determined using the original algorithm (Nesteruk, 2021a). Its application for the first waves of the COVID-19 pandemic in China, South Korea, Austria, Italy, Germany, France, Spain has shown its high accuracy in predicting the first wave durations and numbers of cases.

To simulate the next epidemic waves and to take into account the incompleteness of statistical data, a generalized SIR model and algorithms for determining the values of additional parameters were proposed (Nesteruk, 2020, 2021a, 2021b). With the use of this approach, fourteen epidemic waves in Ukraine (Nesteruk, 2021a; Nesteruk, 2020; Nesteruk, 2021b; Nesteruk & Benlagha, 2021; Nesteruk, 2021c; Nesteruk, 2021d; Nesteruk, 2022e) and seven pandemic waves in the whole world (Nesteruk, 2021a; Nesteruk, 2022e) were successfully simulated. Nevertheless, the accuracy of the parameter identification procedure can be not high enough for log-term predictions, especially when the numbers of cases registered at the beginning of a new wave are used. The corresponding algorithm has to be modified to improve the accuracy of long-term predictions immediately after outbreak of a new epidemic wave. The interference of the previous waves, changes in testing levels, quarantine or social behavior require constant monitoring of the epidemic dynamics and performing SIR simulations as often as possible with the use of a user-friendly interface. Such tool will allow predicting the dynamics of any epidemics of infectious diseases using the data on the number of diseases over a limited period (e.g., 14 days). It will be possible to estimate the daily number of new cases for the country as a whole or its separate region, to calculate the numbers of new daily cases, the numbers of carriers of the infection and the probability of facing an infectious person, as well as to estimate the numbers of deaths.

Since many epidemic characteristics (e.g., daily numbers of new cases, deaths or tests) are random and characterized by some weekly periodicity, the smoothed characteristics should be used (Nesteruk, 2020; Nesteruk, 2021b; COVID-19 Data Repository by the). Their numerical derivatives can be applied in the user-friendly interface for detection of the data irregularities and moments of starting new epidemic waves. The smoothed and relative characteristics (e.g. daily numbers of cases per capita) can be also used to compare the epidemic dynamics in different years and countries. As an example, the COVID-19 pandemic dynamics in Japan in the summer of 2020, 2021 and 2022 will be investigated in order to estimate the impact of vaccination levels on the numbers of new infections, death, and mortality rates. Some other studies of COVID-19 dynamics in Japan can be found in (Chen et al., 2021; Fukumoto et al., 2021; Nesteruk, 2022a; Nesteruk, 2022b; Nesteruk, 2022c; Niwa et al., 2020; Sala et al., 2022; Shimizu & Negita, 2020; Torjesen, 2022; Uehara, 2020; Yamauchi et al., 2022; Yoneoka et al., 2021; Ko et al., 2022; Kato et al., 2022; Zeyu & Takikawa, 2021). An example of a user-friendly interface is presented in (Dong et al., 2020).

The accumulated numbers of laboratory-confirmed cases in Japan in July 2022 were applied to simulate and predict the pandemic dynamics with the use of the generalized SIR model and the modified algorithm of its parameter identification. Results of two SIR simulations first published on July 29, 2022 were confirmed by further observations of the epidemic dynamics in Japan (Nesteruk, 2022b, 2022c). The results of the third SIR simulation based on the earlier dataset corresponding to the period from July 1 to July 14, 2022 will be presented and discussed.

The quality of data is very important for the accuracy of predictions and comparisons. SIR simulations need the daily information about the cumulative numbers of cases V_j. For comparisons, accumulated and daily, absolute and relative (per capita) characteristics must be used. Unfortunately, even official datasets (e.g., WHO daily situation reports (World Health Organization, 2019)) may contain some mistakes. In particular, some examples of WHO datasets with $V_{j+1} < V_j$ are mentioned in (Nesteruk, 2021a). The official national sources of information can also be incomplete or contain some irregularities. For example, on February 12, 2022, National Health Commission (NHC) of the People's Republic of China (https://en.wikipedia.org/wiki/Timeline_of_the_2019%E2%80%9320_Wuhan_coronavirus_outbreak. Retrieved March 4, 2019) has added 12,289 new cases (not previously included in official counts). This fact made SIR simulations based on the previous V_j values irrelevant, (Nesteruk, 2021a).

Unfortunately, WHO stopped reporting the daily figures of the accumulated numbers of SARS-CoV-2 cases in the summer of 2020, but this information and many other useful absolute and relative characteristics are still available in COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), (COVID-19 Data Repository by the). It must be noted, that there are some differences between JHU and national datasets. For example, the results of SIR simulations based on JHU dataset and Ukrainian national statistics are slightly different (Nesteruk, 2021b). In some urgent situations (e.g., after the COVID-19 epidemic outbreak in Italy in February 2020), the information from media can be used, (Nesteruk, 2021a). It must be taken into account the time when the V_j values are registered. Usually they correspond to the end of the previous day and does not represent the real numbers of the infected people due to delays between the day of testing and day when the results of tests are ready.

The real numbers of cases can be much higher than the number of laboratory-confirmed and registered ones. Especially large differences between registered and actual numbers of patients occur immediately after an epidemic outbreak caused by a new pathogen, since testing and diagnosis procedures are imperfect. In the case of the COVID-19 pandemic, many patients probably were not detected in 2019 and early 2020 (Weinberger et al., 2020). Since many patients have no symptoms or make no tests, the large differences between actual and registered COVID-19 cases existed during next stages of the pandemic as well (Nesteruk, 2021c; Nesteruk, 2021d; https://edition.cnn.com/2020/11/02/europe/slovakia-mass-coronavirus-test-intl/index.html, 2020; https://podillyanews.com/2020/12/17/u-shkolah-hmelnytskogo-provely-eksperyment-z-testuvannyam-na-covid-19/, 2020). These differences can complicate the SIR simulations of the first epidemic wave. In particular, the first reliable results for Italy were obtained only when the first wave became close to its stabilization (Nesteruk, 2021a). To have timely predictions, the attempts of SIR simulations have to be repeated every day after obtaining new V_j values. This is why the reliable and comfortable interface is very necessary.

The aim of the study is improving the algorithm for determining the parameters of the generalized SIR model, its approbation on the example of the new epidemic wave in Japan, and the discussion of the prospects for creating a user-friendly interface.

2. Materials and methods

2.1. Data

In a user-friendly interface, some data sets can be used for calculations, some of them only for verifications of calculations and comparisons. As an example, we will use the data sets regarding the accumulated numbers of laboratory-confirmed COVID-19 cases V_j, deaths D_j, the percentages of fully vaccinated people and boosters in Japan in the summer of 2020, 2021 and 2022 reported by JHU, (COVID-19 Data Repository by the) (the version of the files corresponding to September 4, 2022, see Tables A.1-A.4). To check the influence of the testing levels, we will use the daily numbers of tests T_j in Japan from July 1 and August 31, 2022 reported by Japanese Ministry of Health, Labor and Welfare (https://www.mhlw.go.jp/stf/covid-19/open-data_english.html) (the version of the files corresponding to September 4, 2022, see Table A.3).

2.2. Data smoothing procedure

Since daily numbers of new cases, deaths and tests are random and characterized by some weekly periodicity, the use of the smoothed characteristics is recommended, (Nesteruk, 2020, 2021b):

$$\overline{C}_i = \frac{1}{7}\sum_{j=i-3}^{j=i+3} C_j, \quad C_j = V_j;\ D_j;\ T_j \tag{1}$$

Some different smoothing procedure is used in (COVID-19 Data Repository by the) based on the values corresponding to the fixed and previous 6 days:

$$\tilde{C}_i = \frac{1}{7}\sum_{j=i-6}^{j=i} C_j.$$

To estimate the smoothed numbers of new daily cases DV_i, deaths DD_i, tests, and the average daily death per case ratio $m_i = DD_i/DV_i$, the numerical derivatives of the smoothed values (1) will be used as follows (Nesteruk, 2020, 2021b):

$$DC_i \equiv \frac{d\overline{C}}{dt}\bigg|_{t=t_i} \approx \frac{1}{2}(\overline{C}_{i+1} - \overline{C}_{i-1}). \tag{2}$$

To detect changes in the pandemic dynamics the DV_i values must be calculated every day. The increase in DV_i figures indicates the beginning of a new epidemic wave. It would be also useful to provide daily calculations of the second derivatives (Nesteruk, 2020, 2021b):

$$\frac{d^2\overline{V}}{dt^2}\bigg|_{t=t_i} \approx \overline{V}_{i+1} - 2\overline{V}_i + \overline{V}_{i-1}$$

2.3. Generalized SIR model

The generalized SIR-model relates the numbers of susceptible $S(t)$, infectious $I(t)$ and removed persons $R(t)$ over time t for a particular epidemic wave i, (Nesteruk, 2021c):

$$\frac{dS}{dt} = -\alpha_i SI, \tag{3}$$

$$\frac{dI}{dt} = \alpha_i SI - \rho_i I, \tag{4}$$

$$\frac{dR}{dt} = \rho_i I. \tag{5}$$

Compartment $S(t)$ includes the persons who are sensitive to the pathogen and **not protected**. Compartment $I(t)$ includes the persons who **spread the infection**; it is not the number of so known active cases, since people can be ill, but isolated and unable to spread the infection. $R(t)$ includes persons who **no longer spread the infection**; this number is the sum of isolated, recovered, dead, and infected people who left the region under consideration. Parameters α_i and ρ_i are supposed to be constant for every epidemic wave, i.e. for the time periods: $t_i^* \leq t \leq t_{i+1}^*, i = 1, 2, 3, \ldots$. The dimension of these parameters is [time]$^{-1}$. In particular, if time is measured in days, parameters α_i and ρ_i are measured in [day]$^{-1}$.

Values of α_i show how quick the susceptible persons become infected (see (3)). Large values of this parameter correspond to severe epidemics with many victims. These parameters accumulate many characteristics. First they shows how strong (virulent) is the pathogen and what is the way of its spreading. Parameters α_i accumulate also the frequency of contacts and the way of contacting. In order to decrease the values of α_i, we have to minimize the number of our contacts and change our contacting habits. For example, we have to avoid the public places and use masks there, minimize or cancel traveling. We have to change our contact habits: to avoid handshakes and kisses. First, all these simple things are very useful to protect yourself. In addition, if most people follow these recommendations, we have chance to diminish the values of parameters α_i and reduce the negative effects of an epidemic.

The parameters ρ_i characterize the patient removal rates (according to eq. (5)). The inverse values $1/\rho_i$ are the estimations for the average time of spreading infection τ_i or the generation time (Hart et al., 2022) during i-th epidemic wave:

$$\tau_i \approx \frac{1}{\rho_i}. \tag{6}$$

So, we are interested in increasing the values of parameters ρ_i and decreasing $1/\rho_i$. People and public authorities should work on this and organize immediate isolation of suspicious cases.

Since the derivative $d(S+I+R)/dt$ is equal to zero (it follows from summarizing eqs. (3)–(5)), the sum

$$N_i = S + I + R \tag{7}$$

must be constant for every wave. Value N_i is an unknown parameter of the SIR model corresponding to the i-th wave and is not the volume of the population.

To determine the initial conditions for the set of equations (3)–(5), let us suppose that at the beginning of every epidemic wave t_i^*:

$$I(t_i^*) = I_i, \quad R(t_i^*) = R_i, \quad S(t_i^*) = N_i - I_i - R_i. \tag{8}$$

The exact solution of the set of non-linear differential equations (3)–(5) uses the function

$$V(t) = I(t) + R(t), \tag{9}$$

corresponding to the number of victims or the cumulative numbers of cases over time t (Nesteruk, 2022c).

$$F_i^*(V, N_i, I_i, R_i, \nu_i) = \alpha_i(t - t_i^*), \tag{10}$$

$$F_i^* = \int\limits_{R_i+I_i}^{V} \frac{dU}{(N_i - U)[\nu_i \ln(N_i - U) + U - R_i - \nu_i \ln(N_i - R_i - I_i)]}; \quad \nu_i = \frac{\rho_i}{\alpha_i}. \tag{11}$$

Thus, for every set of parameters $N_i, I_i, R_i, \nu_i, \alpha_i, t_i^*$ and a fixed value of V, integral (11) can be calculated and a corresponding moment of time can be determined from (10).

In the case of the first epidemic wave, we can use the values $I_1 = 1; R_1 = 0$. Then the number of parameters reduces to four and t_1^* can be treated as the moment of appearance of "zero" patient. Immediately after an epidemic outbreak, $N_1 \gg V \geq 1$, and integral (11) can be simplified as follows $F_1^* \approx \ln V/N_1$. Then eq. (10) yields

$$V = e^{\gamma_1(t-t_1^*)}, \quad \gamma_1 = \alpha_1 N_1. \tag{12}$$

I.e., the exponential growth of the accumulated numbers of cases and the daily numbers of cases dV/dt is typical for the initial stages of epidemics. In a user-friendly interface, the logarithmic scale can be used to present the daily numbers of new cases, which follow the straight lines after the outbreaks. The deviations from the straight lines indicate the epidemic wave stabilization (Nesteruk, 2021a).

2.4. Parameter identifications

Let us illustrate the procedure of the parameter identification using the example of the exponential growth model (12), which can be rewritten as follows:

$$z \equiv \log V = \gamma_1(t - t_1^*) \tag{13}$$

and yields the linear correlation between the random values $z_j \equiv \log V_j$ and corresponding moments of time t_j. It means that standard formulas (Draper & Smith, 1998) can be used to calculate the correlation coefficient r and estimate the optimal values of parameters γ_1 and t_1^*, which ensure the best fitting between the results of observations V_j and the straight line (13). Then the corresponding $V(t)$ curve and the time of cases duplication:

$$\tau_d = \frac{\ln 2}{\gamma_1}$$

can be calculated (see examples in (Nesteruk, 2021a)) and displayed in a user-friendly interface. When a non-linear dependence can be reduced to the linear one (as in the case of eqs. (10), (11)), the optimal values of parameters correspond to the maximum of value of $|r|$. Similar parameter identification methods was successfully used in (Nesteruk & Rodionov, 2022) for different non-linear correlations.

For many epidemics (including the COVID-9 pandemic) we cannot register dependencies $S(t), I(t)$ and $R(t)$ but observations of the accumulated number of cases V_j corresponding to the moments t_j provide information for direct assessments of the dependence $V(t)$. Nevertheless, as mentioned earlier the real numbers of cases can be much higher than the numbers of laboratory-confirmed patients. If we assume, that data set V_j is incomplete and there is a constant coefficient $\beta_i \geq 1$, relating the registered and real number of cases during the i-th epidemic wave $V(t_j) \cong \beta_i V_j$, then the number of unknown parameters increases by one.

The registered numbers of victims V_j corresponding to the moments of time t_j can be used in eq. (11) in order to calculate $F_{i,j} = F_i^*(V_j, N_i, \nu_i, I_i, R_i, \beta_i)$ for every fixed values of $N_i, \nu_i, I_i, R_i, \beta_i$ and to check how the registered points fit the straight line (10) with the use of the linear regression (Draper & Smith, 1998). The reliability of the model can be checked by calculating the correlation coefficients r_i and/or with the use of the F-test for the null hypothesis that says that the proposed linear relationship (10) fits the data set. The experimental values of the Fisher function can be calculated for every epidemic wave with the use of the formula, (Draper & Smith, 1998):

$$F_i = \frac{r_i^2(n_i - m)}{(1 - r_i^2)(m - 1)}, \tag{14}$$

where n_i is the number of observations for the i-th epidemic wave, $m = 2$ is the number of parameters in the regression equation. The corresponding experimental value F_i has to be compared with the critical value $F_C(k_1, k_2)$ of the Fisher function at a desired significance level ($k_1 = m - 1; k_2 = n_i - m$), (https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf). When the values n_i and m are fixed, the maximum of the Fisher function coincides with the maximum of the correlation coefficient. Therefore, to find the optimal values of parameters $N_i, \nu_i, I_i, R_i, \beta_i$, we must find the maximum of the correlation coefficient (or the highest $F_i/F_C(1, n_i - 2)$ ratio) for the linear dependence (10).

To calculate these values, any method of finding the maximum in the domain of independent parameters can be applied. In this study, the simplest one will be used: create a grid; calculate values of correlation coefficient r_i at nodes, find one providing the maximum of r_i, then take the vicinity of this node, reduce the grid step and calculate the r_i values in the new nodes; the procedure is repeating until the r_i value stops to change.

In the case of sequential calculation of epidemic waves $i = 1,2,3 \ldots$ and when $\beta_i = 1$, it is possible to avoid determining the four optimal unknown parameters N_i, v_i, I_i, R_i, thereby reducing the amount of calculations and difficulties in isolation a maximum of the correlation coefficient. For parameters I_i, R_i it is possible to use the numbers of I and R calculated for the previous epidemic wave at the moment of time when the following wave began. Then we need to calculate values $F_i^*(V_i, N_i, v_i)$, linear regression coefficients in (10), correlation coefficient r_i or $F/F_C(1, n-2)$ and to isolate the values of parameters N_i and v_i corresponding to the maximum of r_i or $F/F_C(1, n-2)$. This approach has been successfully used in (Nesteruk, 2020, 2021a, 2021b) to simulate six waves of the COVID-19 epidemic in Ukraine and four pandemic waves in the world.

Segmentation of epidemic waves and their sequential SIR simulations need many efforts. To avoid this, the new method of obtaining the optimal values of SIR parameters was proposed in (Nesteruk, 2021a) for the case $\beta_i = 1$, which supposed that the day t_i^* of starting of the i-th epidemic wave and corresponding values $V_i = I_i + R_i$ (follows from (9)) are known. To estimate the value V_i and the average number of new cases dV/dt at the moment of time t_i^*, formulas (1) and (2) and the relationship

$$\frac{dV}{dt} = \alpha_i I(N_i - V) \tag{15}$$

(it follows from (4), (5), (7), (9)) have been used. This approximate method was successfully applied to simulate the COVID-19 pandemic waves in Ukraine, Qatar and the whole world (Nesteruk, 2021a; Nesteruk & Benlagha, 2021; Nesteruk, 2021c; Nesteruk, 2021d; Nesteruk, 2022e).

To improve the accuracy, we suppose that the time t_i^*, corresponding to the number of victims $V_i = R_i + I_i$ on the optimal SIR curve, is not given and has to be calculated with the use of (10). The same equation and an iteration procedure allows calculating the value of parameter α_i. In this study the following procedure will be applied: a starting value $\alpha_i^{(1)}$ will be used to calculate a new value of the coefficient $\alpha_i^{(2)}$ in the linear regression presented by eq. (10) (with the use of the standard formula (Draper & Smith, 1998)); procedure is be repeated until the relative error:

$$\varepsilon_\alpha^{(k)} = \frac{\left| \alpha_i^{(k)} - \alpha_i^{(k-1)} \right|}{\alpha_i^{(k)}}, \quad k = 2, 3, 4, ..$$

becomes small enough. If $\beta_i = 1$, the independent parameters are N_i, v_i, I_i and their optimal values must ensure the maximal value of the correlation coefficient r_i.

Calculations were performed with the help of the original program code using MATLAB. The same computing software can be applied for developing the user-friendly interface. The exact solution (10)-(11) allows avoiding the numerical integration of the set of differential equations (3)–(5) and reducing the time required for calculations. In comparison, no analytical solutions were proposed for more complicated SEIR and SEIHDRS$_V$ models (Feng et al., 2022; Lin et al., 2022). All calculations given in this study were carried out on a regular laptop. For massive numerical experiments, fast and reliable calculations of the optimal values of parameters (especially for the incomplete data case $\beta_i > 1$) it is necessary to use more powerful computers.

2.5. Characteristics of an epidemic wave

After the optimal values of the model parameters are determined, the corresponding SIR dependences can be calculated with the use of (10), (11), and the relationships, (Nesteruk, 2021a):

$$I = v_i \ln S - S + N_i - R_i - v_i \ln(N_i - I_i - R_i), \quad S = N_i - V, \quad R = V - I.$$

In particular, the accumulated and daily numbers of cases can be predicted and displayed with the use of eqs. (9), (15) and a user-friendly interface. The corresponding number of susceptible persons at infinity $S_{i\infty} > 0$ can be calculated from a nonlinear equation, (Nesteruk, 2021a):

$$S_{i\infty} = (N_i - I_i - R_i)e^{\frac{S_{i\infty} - N_i - R_i}{v}}. \tag{16}$$

The final numbers of victims (final accumulated number of cases corresponding to the i-th epidemic wave) can be calculated from:

$$V_{i\infty} = N_i - S_{i\infty}. \tag{17}$$

To estimate the final day of the i-th epidemic wave, we can use the condition:

$$I\left(t_{if}\right) = 1. \tag{18}$$

which means that at $t > t_{if}$ less than one person still spreads the infection. Similar condition can be used to calculate the time moment t_{i0}, when a "zero" patient has appeared:

$$I(t_{i0}) = 1. \tag{19}$$

Usually new epidemic waves are connected with the appearance of new pathogen strains (see e.g. (Lin et al., 2021; World Health Organization, 2019)). Then the value t_{i0} could indicate the time when the new pathogen strain that caused the *i-th* epidemic wave started to circulate in a population. In the case of the COVID-19 pandemic, the time moments t_{10}, when a "zero" patient of the first wave has appeared, have been calculated in (Nesteruk, 2021a) with the use of global datasets. The most reliable estimation yielded that SARS-CoV-2 started to circulate in early August 2019.

The effective reproduction number shows the average number of people infected by one person (https://en.wikipedia.org/wiki/Basic_reproduction_number; https://www.r-bloggers.com/2020/04/effective-reproduction-number-estimation/, 2020; an der Heiden & Hamouda, 2020; Cori et al., 2013; Arroyo-Marioli et al., 2021). For the *i-th* epidemic wave this function can be estimated as follows:

$$R_{ti}(t) = \frac{V(t + 0.5\tau_i) - V(t - 0.5\tau_i)}{I(t)} = \frac{1}{I(t)} \int_{t-0.5\tau_i}^{t+0.5\tau_i} \frac{dV(t^*)}{dt^*} dt^* \approx \frac{\tau_i}{I(t)} \frac{dV(t)}{dt} \tag{20}$$

Where τ_i is the average time of spreading the infection (estimated by (6)). Finally, with the use of (6), (7), (15), and second equation (11), formula (20) can be rewritten as follows:

$$R_{ti}(t) \approx \frac{\tau_i}{I(t)} \frac{dV(t)}{dt} = \frac{\alpha_i}{\rho_i}[N_i - V(t)] = \frac{N_i - V(t)}{\nu_i} = \frac{S(t)}{\nu_i} \tag{21}$$

3. Calculations

Table 1 represent the results of three SIR simulations for Japan with the use of three different periods taken for calculations: from July 3 to July 16; from July 14 to July 27; and from July 1 to July 14, 2022. The number of observations $n_i = 14$ and $\beta_i = 1$ for all simulations. The V_j values for these periods have been used for calculations in integral (11). The calculated t_i^* values are close to the to the values 896; 907; and 894 corresponding to the days July 3; July 14; and July 1 taken as the starting ones for three simulations (zero point is January 19, 2020). The experimental values of the Fisher function F are much higher then the critical value $F_c(12,1) = 16.8$ for the significance level 0.001, (https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf). These results illustrate rather high accuracy of simulations.

4. Results

We will illustrate the advantages of comparing the epidemic dynamics and vaccination levels in a fixed season over several years. This simple method allows revealing seasonal trends and the influence of the vaccination levels and different

Table 1
Optimal values of SIR parameters and other characteristics of the COVID-19 pandemic wave in Japan that started in 2022 ($\beta_i = 1$).

Characteristics	First simulation, (Nesteruk, 2022b)	Second simulation, (Nesteruk, 2022b)	Third simulation
Period taken for calculations	July 3 − July 16, 2022	July 14 − July 27, 2022	July 1 − July 14, 2022
I_i	96,886.45104	199,983.62	65,686.9075912
R_i	9,291,335.54896	9,799,197.38	9,274,392.0924088
N_i	26,670,478.56	16,071,723	13,042,931.6964
ν_i	10,743,680.940799	3,491,870.55584211	1,927,229.2338353
α_i, [day]$^{-1}$	1.75618809161137e-08	5.84447875112661e-08	8.31537875426961e-08
t_i^*, days	896.188730336028	906.779042796485	894.004059944712
Day of the appearance of "zero" patient, t_{i0} , eq. (19)	March 2, 2022	March 24, 2022	April 19, 2022
ρ_i, [day]$^{-1}$	0.188679245283032	0.204081632653039	0.160256410256414
$1/\rho_i$, days	5.29999999999962	4.90000000000054	6.23999999999987
r_i	0.999059537295348	0.998335488255617	0.998939919410355
F_i, eq. (14)	6370.83972958245	3595.66315699370	5650.94863518777
$S_{i\infty}$, eq. (16)	5,983,383	1,587,272	789,094
$V_{i\infty}$, eq. (17)	20,687,095	14,484,451	12,253,838
Final year of the epidemic wave, eq. (18)	2037	2029	2028

coronavirus variants (Lin et al., 2021) on the number of disease cases, deaths and the mortality rate. The results presented in (Nesteruk, 2022a) showed that the implementation of this comparison into an interface looks very useful. The comparative analysis of the summer COVID-19 pandemic dynamics in Japan gives another example.

Fig. 1 shows the results of monitoring the COVID-19 pandemic dynamics in Japan with the use of smoothed values calculated according to equations (1) and (2). Markers represent the corresponding daily characteristics in the summer of 2020 (red), 2021 (blue) and 2022 (black). "Circles" show the smoothed daily numbers of new cases DC_i calculated with the use of accumulated numbers of cases V_j listed in Tables A.1-A.3. "Crosses" show the smoothed daily numbers of deaths DD_i calculated with the use of accumulated numbers of deaths D_j listed in Table A.4. "Triangles" represent the averaged daily mortality rates $m_i = DD_i/DC_i$. The smoothed daily numbers of cases started to increase approximately on the same day (on June 18, 19 and 20 for 2022, 2021 and 2020, respectively). The periods of increasing were 46 and 64 days 2020 and 2021, respectively (see red and blue "circles" in Fig. 1). In 2022 three local maxima of smoothed daily numbers of cases were registered (after 47, 50 and 63 days, see green "crosses" in Fig. 2). The reasons of this phenomenon we will discuss later.

Solid and dashed lines in Fig. 1 respectively show the percentages of fully vaccinated people and boosters listed in Tables A.2 (2021, the blue solid line) and A.3 (2022, the black solid and dashed lines). It is clear that the high level of vaccinations achieved in the summer of 2022 did not save Japan from a powerful pandemic wave, with the daily number of cases about ten times higher than on the corresponding period in 2021 (compare "circles" in Fig. 1). The same result was reported for Israel already in 2021, (Nesteruk, 2021d). As of February 1, 2022 the daily numbers of new COVID-19 cases per capita in European countries even increased with the increase of the vaccination level (probably due to canceling the quarantine restrictions for vaccinated people) (Nesteruk & Rodionov, 2022).

The averaged daily numbers of deaths in July and August 2022 were also much higher than in the summer of 2020 and 2021 (compare "crosses" in Fig. 1). Nevertheless, the values of the death per case ratio in 2022 are much lower (compare "triangles" in Fig. 1) demonstrating the positive effect of vaccination and/or the lower infection mortality due to different variants (Lin et al., 2021). It should be noted that in August 2021 and 2022 these values became closer (probably it is connected with the close vaccination levels shown by the blue and black solid lines). It is difficult to investigate the effect of boosters, but the high BC_j values achieved in July and August 2022 (see the dashed black line) have not significantly diminished the mortality rates in August 2022. Some effects of the booster campaign in Israel are reported in (Feng et al., 2022; Gavish, Yaari, Huppert, & Katriel, 2022). Probably we can say that existing vaccines cannot stop new waves of the COVID-19 pandemic, but significantly reduce the probability of death of infected persons. Nevertheless, the COVID-19 related death per case ratio (around 0.001) is still approximately 4 times higher than in the case of influenza (12 millions of cases and 3400 excess deaths were registered in Japan in season 2018–2019, (https://www.niid.go.jp/niid/en/iasr/865-iasr/9288-477te.html)).
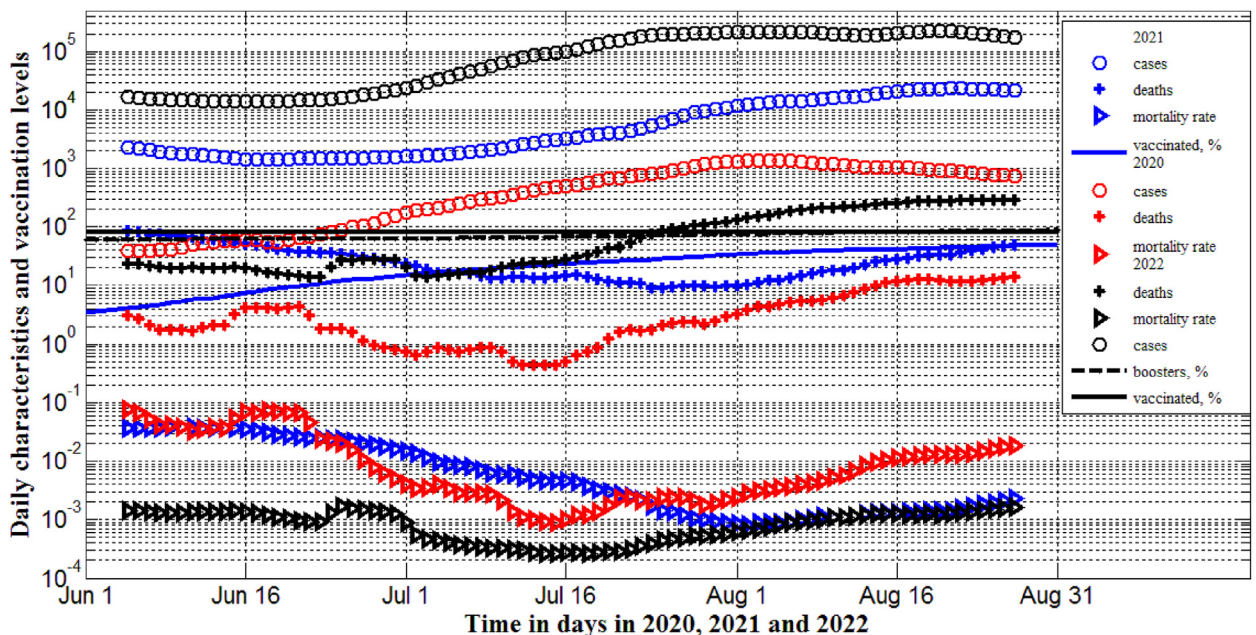
Fig. 1. The COVID-19 pandemic waves in Japan in the summer of 2020 (red), 2021 (blue), and 2022 (black).
"Circles" and "triangles" respectively represent daily numbers of new cases DV_i and daily numbers of deaths DD_i (values from Tables A.1-A.4 smoothed according to formulae (1) and (2)). "Triangles" show the averaged daily deaths per case ratio (mortality rate) $m_i = DD_i/DV_i$. Solid and dashed lines respectively represent the percentages of fully vaccinated people and boosters.

Fig. 2. The results of the first (red curves) and second (blue curves) SIR simulations of the summer COVID-19 pandemic wave in Japan in 2022 and comparisons with the registered cases (markers) and testing levels (magenta line and "dots").
Red lines correspond to the first simulation with the use of observations in period July 3 − July 16, 2022, blue ones − in period July 14- July 27, 2022, (Nesteruk, 2022b). Solid lines show the number of victims $V(t) = I(t) + R(t)$, dotted lines − the theoretical estimations of the daily nu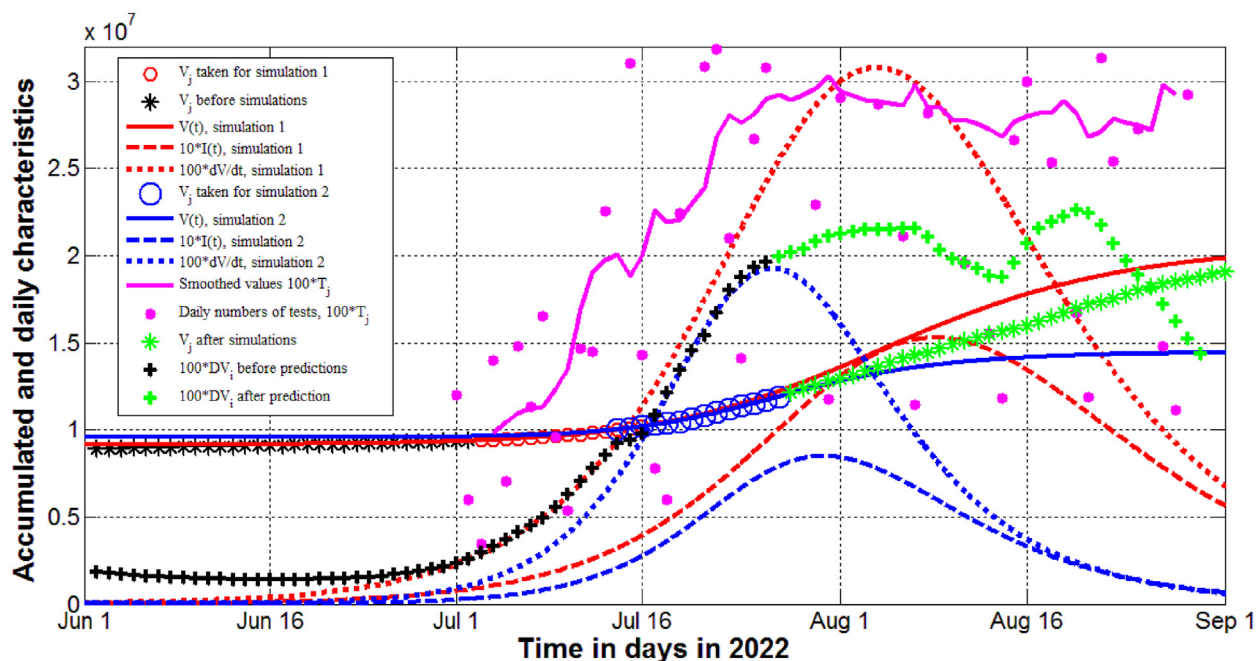mbers of new cases multiplied by 100, i.e. $100*dV/dt$ (eq. (15)), dashed lines − the numbers of infectious persons multiplied by 10, i.e. $10*I(t)$. "Crosses" represent the averaged daily numbers of cases DV_i calculated with the use of accumulated numbers of cases listed in Table A3, formulas (1), (2) and multiplied by 100 (the green color corresponds to figures obtained after predictions presented in (Nesteruk, 2022b)). "Circles" and "stars" show the accumulated numbers of cases taken from Table A3; red and blue "circles" represent the V_j values taken for the first and second SIR simulations, respectively; black "stars" − V_j values before July 3, 2022; green "stars" show the V_j values registered after the predictions of preprint (Nesteruk, 2022b) were completed. Magenta "dots" and line respectively represent the daily numbers of tests T_j (figures from Table A3 multiplied by 100) and their smoothed values multiplied by 100 (according to eq. (1)).

The maximum of smoothed daily numbers of death DD_i caused by a new SARS-CoV-2 wave occurred approximately three weeks later after the maximum of smoothed daily numbers of new cases DV_i has been achieved. These maxima of DD_i values are not visible in Fig. 1 since they occurred in September. To check the time difference between the maxima of deaths and cases let us use the smoothed values calculated in (COVID-19 Data Repository by the). For the summer waves in Japan in 2020 and 2021 this difference was 23 and 20 days, respectively; for the Omicron wave in Japan in February 2022 it was 19 days, i.e., the averaged delay is approximately 21 days.

Table 1 and Figs. 2 and 3 represent the results of three SIR simulations with the use of three different periods taken for calculations: from July 3 to July 16 (red lines); from July 14 to July 27 (blue lines); and from July 1 to July 14, 2022 (black lines). The V_j values for these periods (respectively shown in Figs. 2 and 3 by red, blue and black "circles") have been used for calculations in integral (11). The V_j values corresponding to the period before July 1, 2022 (shown by black "stars") are close to the corresponding solid lines.

Nevertheless, the optimal values of the SIR model parameters and other characteristics of the pandemic wave (listed in Table 1) are rather different. In particular, there are significant differences in the saturation levels ($V_{i\infty}$ is 41% lower for the third simulation in comparison with the first one, see Table 1, red and black solid lines in Fig. 3). The estimated pandemic wave durations (calculated with the assumption that the number of infectious persons becomes less than one, eq. (18)) differ by 9 years (see the last row of Table 1).

Dotted lines in Figs. 2 and 3 illustrate rather large differences in theoretical estimations of the daily numbers of new cases (15). Before July 15 the registered smoothed daily numbers of new cases DV_i calculated with the use of eqs. (1) and (2)) were closer to the first SIR simulation (compare black "crosses" with the red dotted line in Figs. 2 and 3). From July 15 to July 27, the DV_i values followed the second SIR simulation (compare black and green "crosses" with the blue dotted line). After July 27, 2022, the registered and theoretical numbers of new daily cases started to deviate (compare green "crosses" and the blue dotted line in Fig. 2). In particular, the DV_i values continued to increase after the theoretical maximum on the blue dotted curve was achieved. For the third simulation, the theoretical and registered daily numbers of new cases were close only during the period taken for calculations (compare the black dotted line and black "crosses" in Fig. 3). The reasons of these differences will be discussed in the next Section.
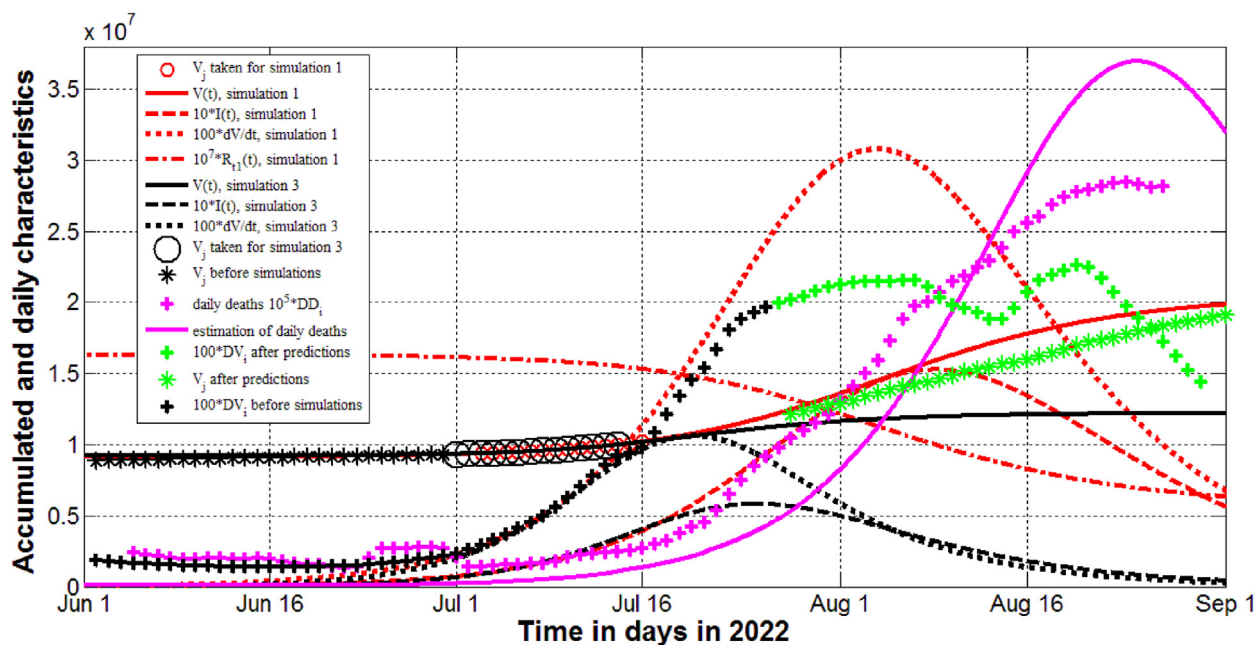
Fig. 3. The results of the first (red curves) and third (black curves) SIR simulations of the summer COVID-19 pandemic wave in Japan in 2022 and the daily numbers of deaths (magenta).

Red lines correspond to the first simulation with the use of observations in period from July 3 to July 16, 2022, (Nesteruk, 2022b), black ones - to the third simulation based on dataset July 1- July 14, 2022. Solid lines show the number of victims $V(t) = I(t) + R(t)$; dotted lines − the theoretical estimations of the daily numbers of new cases multiplied by 100, i.e. *100*dV/dt* (eq. (15)); dashed lines − the numbers of infectious persons multiplied by 10, i.e. *10*I(t)*; the dashed-dotted line represent the theoretical estimations of the effective reproduction number multiplied by 10 million (eq. (21)). "Crosses" represent the averaged daily numbers of cases DV_i calculated with the use of accumulated numbers of cases listed in Table A3, formulas (1), (2) and multiplied by 100 (the green color corresponds to figures obtained after predictions presented in (Nesteruk, 2022b)). "Circles" and "stars" show the accumulated numbers of cases taken from Table A3; red and black "circles" represent the V_j values taken for the first and third SIR simulations, respectively; black "stars" − V_j values before July 1, 2022; green "stars" show the V_j values registered after the predictions of preprint (Nesteruk, 2022b) were completed. Magenta "crosses" and line respectively represent the smoothed daily numbers of deaths DD_i (Table A4, eqs. (1) and (2)) and its theoretical estimation (both multiplied by 100,000).

The mortality rate in the summer of 2022 was more or less stable (see black "triangles" in Fig. 1), thus we can use its value $m^* = 0.0012$ corresponding the beginning of a new COVID-19 (June 18, 2022) to predict the average daily numbers of deaths. This information is important to evaluate the necessary numbers of places in the intensive care units and have to be displayed in an interface. We have used the *dV(t)/dt* curve of the most reliable first SIR simulation, applied delay 21 days and multiplied by m^*(see the magenta line in Fig. 3). Magenta "crosses" represent the registered smoothed daily numbers of deaths DD_i (calculated according to eqs. (1) and (2)).

Two first simulations were completed on July 27, 2022 and corresponding preprint (Nesteruk, 2022b) was uploaded two days later. Red and blue solid lines in Fig. 2 show the predicted accumulated numbers of cases. Green "stars" represent the accumulated numbers of cases registered after July 27, 2022 and are located between theoretical blue and red solid lines. The same result is valid for the third prediction shown in Fig. 3 (green "stars" are located between solid curves, but closer to the red line of the first simulation).

5. Discussion

The first local maximum of DV_i values (registered on August 4, 2022, see green "crosses" in Fig. 2) agrees very well with the day corresponding to the maximum point on the red dotted theoretical curve for the first prediction. Thus, the accuracy of the first simulation turned out to be very good only in relation to the time corresponding to the maximum number of daily cases, but the maximal theoretical value is approximately 1.44 times higher than the results of observations (compare the dotted red line and green "crosses").

Let us analyze a possible reason of this discrepancy with the use of the daily number of COVID-19 tests T_j in Japan, shown in Table A3 and in Fig. 2 (magenta "dots"). You can see that after July 16 many T_j values are less than smoothed numbers of daily cases shown by black and green "crosses". The smoothed values of daily numbers of tests calculated with the use of formula (1) and shown by magenta curve became less than the average expected daily number of cases (the dotted red line) in early August 2022. It means that the testing level in Japan became too low to reveal all COVID-19 cases. It should be noted that the first local minimum on the magenta curve in Fig. 2 coincides with the visible changes in the pandemic dynamics on July 15,

2022 (see black "crosses"). It looks that second SIR simulation was based on the incomplete V_j numbers registered from July 14 to July 27. That is why the corresponding prediction was more optimistic than the first one based on V_j numbers registered from July 3 to July 16. The reason of second maximum of DV_i values (corresponding to August 20, see "green" crosses in Fig. 2) is likely to be improved testing, as the average numbers of tests (the magenta line) have become higher than predicted daily numbers of new cases (the red dotted line).

It was expected that the third simulation based on the earlier dataset (from July 1 to July 14) would be most precise, but presented in Fig. 3 calculation results do not confirm this expectation. The registered accumulated and daily numbers of cases (green "stars" and "crosses") are much higher than the theoretical estimations shown by black solid and dotted curves. The reason for this unexpected result may be due to the influence of the large number of cases associated with previous pandemic waves, as evidenced by the large difference between the dotted black line and the black "crosses" in June 2022.

The summer epidemic wave can be caused by new variant/variants of SARS-CoV-2, which appeared between March 2 and April 19, 2022 (according to SIR simulations the number of infectious persons became grater than one on these days, see eq. (19), Table 1 and Figs. 2 and 3). In early July 2022 the numbers of cases connected with this strain was not large enough in comparison with the numbers of Delta and Omicron cases, which were still present in the population. Since the generalized SIR model supposes constant values of parameters, interference of previous epidemic waves decreases its accuracy.

This fact reduces the value of forecasts based on the number of cases registered immediately after the start of a new epidemic wave. However, in the case of constant monitoring of the average daily number of new cases and daily calculations using the generalized SIR model, timely and accurate forecasts can be obtained. That is why it is very important to improve methods for finding optimal parameter values and to use them in user-friendly interface. Such tool will allow non-mathematicians to make calculations and visualize results. For example, to estimate the real pandemic dynamics in Japan in the summer 2022, the first simulation (red curves in Figs. 2 and 3) can be used. In particular, the difference between the solid red line and green "stars" corresponds to the numbers of unregistered cases. The predicted huge amount of new cases shown by the dotted red curve allows estimating the need for tests, which are necessary to detect them.

Timely forecasts with the use of a user-friendly interface also make it possible to evaluate the number of infectious people $I(t)$ and the probability of meeting them in a country or region:

$$p(t) = \frac{I(t)}{N_{pop}} \tag{22}$$

where N_{pop} is the volume of population. This information could help to regulate the quarantine measures inside the region and the traveling restrictions between countries or regions. For example, the number of infectious persons in Japan has achieved its maximum on August 8, 2022 (see the red dashed line in Fig. 2). With the use of eq. (22), the maximal probability of meeting an infected person can be estimated as 0.012 taking into account the population N_{pop} of Japan. This probability remained very high in August 2022.

Presented approach allows estimating the effective reproduction number with the use of eq. (21) The red dashed-dotted curve in Fig. 3 (corresponding to the most reliable first simulation) illustrates that the R_{ti} values were long time close to 1.6 and tend to 0.557 at infinity. The internet resource (https://toyokeizai.net/sp/visual/tko/covid19/en.html) showed values 1.0–1.25 in July and August 2022 for Japan; JHU values (calculated with the use of method proposed in (an der Heiden & Hamouda, 2020)) ranged between 1.21 and 1.83 in July and between 0.86 and 1.21 in August (COVID-19 Data Repository by the). If we suppose that the summer epidemic wave was cased by a new variant (or variants) of SARS-CoV-2, the first theoretical estimation shows, that this strain will continue to circulate 15 more years and even after this the susceptible persons will continue to be present in the population. We can also estimate the time of appearance of this strain t_{i0} with the use of eq. (19). The most reliable first simulation yields March 2, 2022 as the day, when the new variant has started to spread in Japan. Large differences between the theoretical estimation of the daily numbers of cases (the red dotted line in Fig. 2) and real numbers (black "crosses") during almost 4 month are connected with spreading previous strains, which are still present in the population.

The slow decline of the COVID-19 pandemic waves is characteristic not only of Japan. Thus, the end of the global Omicron wave is expected only in 2085, (Nesteruk). The presence of large numbers of infected persons (even between waves) increases the probability of appearance of new dangerous variants. That is why the detection an isolation of infected persons remains very important. To solve this problem, the numbers of tests have to be significantly increased. The huge value of test per case ratio allowed China to control the pandemic completely, (Nesteruk, 2022a). If the test per case ratio in Japan could exceeds 500, the average daily number of new cases per million might be less than 0.36 and daily deaths around 0.01 per million (Nesteruk, 2022a).

6. Conclusions

The parameter identification procedure for generalized SIR-model was improved to simulate the dynamics of any epidemic wave based on the accumulated numbers of cases registered immediately after a new wave outbreak. The algorithm was tested for the COVID-19 pandemic wave in Japan occurred in the summer of 2022. Predictions first published on July 29, 2022 were confirmed by further observations of the epidemic dynamics in Japan. The presented approach can be used for simulations in other countries and/or regions (see, e.g. (Nesteruk, 2022f; Nesteruk, 2022d),) and for other epidemics. It could

be also the basis of a user-friendly interface for regular and timely predictions of the accumulated and daily numbers of cases and deaths, numbers of infectious persons, the probability of meeting such a person in a country or a region, effective reproduction numbers, epidemic waves durations, the average time of spreading the infection, etc.

The software should calculate daily smoothed numbers of new cases, deaths and mortality rate for timely detection of the beginning of new waves and forecasting of the number of deaths. The interface should yield an opportunity to compare the characteristics in different regions and in different years and has to be as simple as possible for wide and regular use in medical statistics bodies for timely and accurate forecasting of the epidemic dynamics.

The proposed approaches and their future implementation in user-friendly interfaces will allow improving the forecasting of various epidemics and reducing their negative impact on the public health and the economy.

Funding and/or conflicts of interests/competing interests

Data availability statement

The data sets regarding the accumulated numbers of laboratory-confirmed COVID-19 cases, deaths, the percentages of fully vaccinated people and boosters in Japan in the summer of 2020, 2021 and 2022 reported by JHU, (COVID-19 Data Repository by the) (the version of the files corresponding to September 4, 2022) and the daily numbers of tests in Japan from July 1 and August 31, 2022 reported by Japanese Ministry of Health, Labor and Welfare (https://www.mhlw.go.jp/stf/covid-19/open-data_english.html) (the version of the files corresponding to September 4, 2022) were used in this study.

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.

Acknowledgement

Appendix A

Table A.1
Cumulative numbers of laboratory-confirmed Covid-19 cases in Japan for the period of June 1 to August 31, 2020 according to JHU report on September 4, 2022, (COVID-19 Data Repository by the).

Day in corresponding month of 2020	Numbers of cases in 2020, V_j		
	June	July	August
1	16,778	18,732	37,790
2	16,829	18,926	39,121
3	16,858	19,176	40,086
4	16,902	19,450	41,332
5	16,949	19,658	42,687
6	16,995	19,834	44,172
7	17,033	20,042	45,777
8	17,054	20,249	47,348
9	17,099	20,604	48,789
10	17,135	21,034	49,627
11	17,179	21,420	50,324
12	17,240	21,827	51,303
13	17,285	22,086	52,475
14	17,360	22,419	53,831
15	17,432	22,869	55,062
16	17,474	23,493	56,079
17	17,520	24,090	56,726
18	17,589	24,751	57,642
19	17,650	25,262	58,712

Table A.1 (*continued*)

Day in corresponding month of 2020	Numbers of cases in 2020, V_j		
	June	July	August
20	17,714	25,680	59,890
21	17,769	26,312	60,927
22	17,812	27,107	61,911
23	17,869	28,088	62,655
24	17,965	28,867	63,148
25	18,047	29,668	63,863
26	18,152	30,507	64,766
27	18,244	31,107	65,629
28	18,357	32,092	66,505
29	18,467	33,362	67,349
30	18,605	34,670	67,949
31	–	36,256	68,385

Table A.2
Cumulative numbers of laboratory-confirmed Covid-19 cases and percentage of fully vaccinated people in Japan for the period of June 1, 2021 to August 31, 2021 according to JHU report on September 4, 2022, (COVID-19 Data Repository by the).

Day in corresponding month of 2021	Numbers of cases in 2021, V_j			Percentage of fully vaccinated people in 2021, VC_j		
	June	July	August	June	July	August
1	749,126	801,337	936,852	3.44	15.3	33.74
2	752,160	803,112	945,330	3.63	15.9	34.4
3	754,991	804,990	957,431	3.82	16.41	35.1
4	757,583	806,475	971,743	4	16.86	35.79
5	760,233	807,504	987,060	4.1	17.42	36.51
6	762,253	809,172	1,002,867	4.25	17.99	37.26
7	763,533	811,361	1,018,688	4.51	18.54	37.98
8	765,414	813,607	1,033,214	4.77	19.09	38.5
9	767,652	815,883	1,045,392	5.09	19.64	38.67
10	769,694	818,341	1,056,125	5.41	20.17	39.34
11	771,628	820,371	1,072,153	5.7	20.71	39.94
12	773,569	821,876	1,091,187	5.92	21.29	40.29
13	774,951	824,260	1,111,683	6.18	21.86	40.55
14	775,884	827,452	1,131,986	6.62	22.44	40.89
15	777,301	830,870	1,149,874	7.07	23.01	41.19
16	779,008	834,303	1,164,881	7.56	23.56	41.66
17	780,558	838,189	1,184,986	8.05	24.12	42.23
18	782,178	841,290	1,209,082	8.52	24.7	42.78
19	783,696	843,618	1,234,423	8.88	25.28	43.29
20	785,002	847,377	1,260,415	9.23	25.89	43.85
21	785,870	852,326	1,286,077	9.74	26.55	44.43
22	787,305	857,725	1,308,422	10.28	26.83	44.93
23	789,085	861,959	1,325,346	10.84	27.05	45.45
24	790,759	865,553	1,347,015	11.41	27.73	45.98
25	792,469	870,592	1,371,426	11.98	28.37	46.49
26	794,100	875,301	1,396,464	12.46	29.15	46.97
27	795,382	882,971	1,420,783	12.9	29.93	47.51
28	796,383	892,599	1,443,642	13.48	30.74	48.11
29	797,763	903,351	1,462,979	14.1	31.53	48.62
30	799,583	914,182	1,476,653	14.71	32.37	48.85
31	–	926,604	1,494,372	–	33.17	49.28

Table A.3
Cumulative numbers of laboratory-confirmed Covid-19 cases, vaccination and testind levels in Japan in June, July, and August 2022 according to the reports of JHU (COVID-19 Data Repository by the) and Japanese Ministry of Health, Labor and Welfare (https://www.mhlw.go.jp/stf/covid-19/open-data_english.html) (the versions of the files corresponding to September 4, 2022)

Day in corres-pon-ding month of 2022	Accumulated numbers of cases V_j, (COVID-19 Data Repository by the)			Percentage of fully vaccinated people VC_j, (COVID-19 Data Repository by the)			Total busters per hundred BC_j, (COVID-19 Data Repository by the)			Daily numbers of tests T_j, (https://www.mhlw.go.jp/stf/covid-19/open-data_english.html)	
	June	July	August	June	July	Aug.	June	July	Aug.	July	Aug.
1	8,861,405	9,340,079	12,918,780	81.94	82.17	82.3	60.63	63.58	75.45	120,244	290,320
2	8,881,491	9,364,954	13,129,678	81.95	82.19	82.3	60.69	63.79	75.95	59,903	330,933
3	8,900,748	9,388,222	13,379,386	81.96	82.19	82.3	60.86	63.86	76.43	34,144	429,491
4	8,918,984	9,405,007	13,618,037	81.99	82.19	82.3	61.06	64.06	76.9	139,927	286,766
5	8,934,078	9,441,171	13,851,692	81.99	82.19	82.31	61.11	64.28	77.56	70,601	357,678
6	8,943,156	9,486,965	14,079,172	82	82.19	82.32	61.17	64.49	78.24	148,047	210,929
7	8,960,177	9,534,908	14,285,593	82	82.2	82.32	61.23	64.71	78.57	113,303	114,699
8	8,978,571	9,584,835	14,423,315	82.01	82.2	82.32	61.29	65.09	79.04	165,011	281,562
9	8,995,359	9,639,819	14,635,748	82.01	82.21	82.32	61.35	65.44	79.55	95,578	323,281
10	9,010,943	9,693,879	14,886,036	82.03	82.22	82.33	61.5	65.56	80.06	53,736	509,327
11	9,026,281	9,730,993	15,126,145	82.05	82.22	82.33	61.69	65.88	80.22	146,743	195,561
12	9,039,662	9,806,971	15,294,863	82.05	82.22	82.33	61.73	66.24	80.63	144,967	362,219
13	9,047,604	9,901,424	15,478,389	82.06	82.22	82.33	61.79	66.57	80.87	225,407	155,645
14	9,062,918	9,999,181	15,656,681	82.06	82.23	82.34	61.84	66.92	80.98	357,914	118,230
15	9,079,462	10,102,454	15,795,156	82.07	82.23	82.34	61.9	67.43	81.13	310,435	266,466
16	9,094,948	10,213,101	15,961,249	82.07	82.24	82.34	61.95	67.95	81.42	142,974	299,664
17	9,109,632	10,318,639	16,192,610	82.08	82.24	82.35	62.1	68.2	81.8	77,567	481,646
18	9,124,454	10,394,789	16,448,037	82.1	82.25	82.35	62.26	68.28	82.21	59,865	253,190
19	9,137,595	10,461,481	16,708,985	82.1	82.25	82.36	62.31	68.74	82.8	224,068	386,840
20	9,145,373	10,613,947	16,962,177	82.11	82.25	82.37	62.37	69.2	83.4	409,922	167,019
21	9,160,740	10,800,103	17,188,265	82.11	82.25	82.38	62.43	69.65	83.7	308,436	118,488
22	9,178,003	10,995,193	17,329,185	82.12	82.26	82.38	62.5	70.28	84.09	318,534	313,442
23	9,194,660	11,196,063	17,537,636	82.12	82.27	82.39	62.57	70.92	84.51	209,938	254,220
24	9,210,607	11,372,555	17,781,018	82.13	82.27	82.39	62.74	71.27	84.9	140,953	385,874
25	9,227,180	11,499,018	18,001,865	82.15	82.27	82.4	62.92	71.76	85.26	266,818	272,714
26	9,241,386	11,695,380	18,194,170	82.15	82.27	82.41	62.97	72.29	85.79	307,768	440,793
27	9,250,940	11,904,980	18,374,230	82.16	82.28	82.42	63.07	72.8	86.31	379,977	148,087
28	9,270,305	12,137,977	18,531,986	82.16	82.28	82.43	63.17	73.3	86.55	346,233	111,230
29	9,293,629	12,359,275	18,627,815	82.16	82.28	82.43	63.27	73.97	86.85	377,065	292,644
30	9,316,954	12,581,505	18,780,302	82.17	82.29	82.43	63.37	74.66	87.15	229,110	432,524
31	-	12,779,171	18,949,793	-	82.29	82.44	-	75.01	87.4	117,747	348,279

Table A.4
Cumulative numbers of Covid-19 related deaths D_j in Japan in June, July, and August 2020, 2021, and 2022 according to JHU report on September 4, 2022, (COVID-19 Data Repository by the).

Day in corres-pon-ding month	2020			2021			2022		
	June	July	August	June	July	August	June	July	August
1	900	976	1013	13,160	14,807	15,196	30,655	31,298	32,703
2	903	977	1013	13,273	14,832	15,207	30,678	31,309	32,846
3	908	977	1018	13,384	14,841	15,217	30,710	31,314	33,015
4	913	977	1023	13,470	14,847	15,231	30,733	31,328	33,176
5	917	977	1030	13,534	14,866	15,239	30,749	31,348	33,390
6	919	978	1036	13,584	14,888	15,259	30,773	31,360	33,542
7	919	980	1043	13,659	14,902	15,273	30,799	31,375	33,694
8	922	982	1044	13,758	14,919	15,282	30,820	31,404	33,844
9	923	982	1049	13,854	14,939	15,294	30,843	31,416	34,122
10	925	982	1054	13,925	14,950	15,313	30,864	31,428	34,373
11	925	983	1061	13,988	14,956	15,333	30,879	31,443	34,579
12	929	983	1066	14,043	14,959	15,357	30,888	31,466	34,789
13	931	984	1077	14,075	14,977	15,382	30,909	31,497	35,045

(continued on next page)

Table A.4 (*continued*)

Day in corres-pon-ding month	2020			2021			2022		
	June	July	August	June	July	August	June	July	August
14	931	984	1086	14,135	14,997	15,400	30,938	31,530	35,198
15	933	985	1093	14,202	15,019	15,410	30,961	31,561	35,402
16	938	985	1104	14,282	15,029	15,437	30,981	31,581	35,708
17	939	985	1119	14,329	15,045	15,484	30,999	31,598	35,998
18	940	986	1135	14,377	15,049	15,514	31,019	31,617	36,285
19	959	986	1149	14,405	15,061	15,542	31,029	31,645	36,579
20	959	988	1160	14,425	15,081	15,578	31,042	31,698	36,833
21	960	989	1175	14,460	15,101	15,612	31,059	31,746	37,059
22	961	990	1181	14,504	15,107	15,636	31,074	31,798	37,304
23	965	992	1190	14,562	15,115	15,668	31,089	31,873	37,647
24	969	995	1203	14,604	15,123	15,710	31,104	31,898	37,948
25	971	998	1217	14,635	15,127	15,755	31,113	31,946	38,244
26	971	998	1230	14,664	15,139	15,807	31,125	32,061	38,565
27	972	999	1241	14,674	15,151	15,864	31,138	32,190	38,826
28	972	1002	1261	14,712	15,159	15,906	31,246	32,304	39,047
29	974	1006	1272	14,742	15,173	15,955	31,263	32,426	39,281
30	974	1007	1286	14,783	15,182	16,001	31,277	32,527	39,600
31	–	1013	1300	–	15,191	16,065	–	32,609	39,938

References

de Andres, P. L., de Andres-Bragado, L., & Hoessly, L. (2021). Monitoring and forecasting COVID-19: Heuristic regression, susceptible-infected-removed model and, spatial stochastic, 21 May *Front Appl Math Stat*. https://doi.org/10.3389/fams.2021.650716.

Arroyo-Marioli, F., Bullano, F., Kucinskas, S., & Rondón-Moreno, C. (2021). Tracking R of COVID-19: A new real-time estimation using the kalman filter. *PLoS One, 16*(1), Article e0244474. https://doi.org/10.1371/journal.pone.0244474

Brauer, F., & Castillo-Chávez, C. (2001). *Mathematical models in population biology and epidemiology.* New York: Springer.

Chen, Z., Shu, Z., Huang, X., Peng, K., & Pan, J. (2021). Modelling analysis of COVID-19 transmission and the state of emergency in Japan. Jul *Int J Environ Res Public Health, 18*(13), 6858. https://doi.org/10.3390/ijerph18136858. Published online 2021 Jun 26.

Cori, A., Ferguson, N. M., Fraser, C., & Simon, C. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

Costris-Vas, C., Schwartz, E. J., & Smith, R. J. (2020). Predicting COVID-19 using past pandemics as a guide: How reliable were mathematical models then, and how reliable will they be now?. November *Mathematical Biosciences and Engineering, 17*(6), 7502–7518. https://doi.org/10.3934/mbe.

COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://github.com/owid/covid-19-data/tree/master/public/data.

Daley, D. J., & Gani, J. (2005). *Epidemic modeling: An introduction.* New York: Cambridge University Press.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). John Wiley.

Enrique Amaro, J., Dudouet, J., & Nicolás Orce, J. (2021). Global analysis of the COVID-19 pandemic using simple epidemiological models. *Applied Mathematical Modelling, 90*, 995–1008. https://doi.org/10.1016/j.apm.2020.10.019

Feng, A., Obolski, U., Stone, L., & He, D. (2022). Modelling COVID-19 vaccine breakthrough infections in highly vaccinated Israel—the effects of waning immunity and third vaccination dose. *PLOS Glob Public Health, 2*(11), Article e0001211. https://doi.org/10.1371/journal.pgph.0001211

Fukumoto, K., McClean, C. T., & Nakagawa, K. (2021). No causal effect of school closures in Japan on the spread of COVID-19 in spring 2020. *Nature Medicine, 27*, 2111–2119. https://doi.org/10.1038/s41591-021-01571-8

Gang, X. (2020). A novel Monte Carlo simulation procedure for modelling covid-19 spread over time. *Scientific Reports, 10*, Article 13120. https://doi.org/10.1038/s41598-020-70091-1

Hart, W. S., Miller, E., Andrews, N. J., Waight, P., Maini, P. K., Funk, S., & Thompson, R. N. (2022). Generation time of the alpha and delta SARS-CoV-2 variants: An epidemiological analysis. *The Lancet Infectious Diseases, 22*(5), 603–610. https://doi.org/10.1016/S1473-3099(22)00001, 9. May01.

an der Heiden, M., & Hamouda, O. (2020). Schätzung der aktuel-len entwicklung der sars-cov-2-epidemie in deutsch-land − nowcasting. *Epidemiologisches Bulletin, 17*, 10–15. https://doi.org/10.25646/669

Gavish, N., Yaari, R., Huppert, A., & Katriel, G. (2022). Population-level implications of the Israeli booster campaign to curtail COVID-19 resurgence. *Science Translational Medicine, 14*, eabn9836. https://doi.org/10.1126/scitranslmed.abn9836

Hethcote, H. W. (2000). *The mathematics of infectious diseases. SIAM Review, 42*(4), 599–653.

Huppert, A., & Katriel, G. (2013). Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection, 19*(11), 999–1005. https://doi.org/10.1111/1469-0691.12308

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the mathematical theory of epidemics. *J Royal Stat Soc Ser A, 115*, 700–721.

Lin, L., Liu, Y., Tang, X., & He, D. (2021). The disease severity and clinical outcomes of the SARS-CoV-2 variants of concern. *Frontiers in Public Health, 9*. https://www.frontiersin.org/articles/10.3389/fpubh.2021.775224.

Lin, L., Zhao, Y., Chen, B., & He, D. (2022). Multiple COVID-19 waves and vaccination effectiveness in the United States. *International Journal of Environmental Research and Public Health, 19*(4), 2282. https://doi.org/10.3390/ijerph19042282

Liu, J., Wang, L., Zhang, Q., & Yau, Shing-Tung (2021). The dynamical model for COVID-19 with asymptotic analysis and numerical implementations. *Applied Mathematical Modelling, 89*, 1965–1982. https://doi.org/10.1016/j.apm.2020.07.057. Part 2.

Mohammadi, Alireza, Meniailov, Ievgen, Bazilevych, Kseniia, Yakovlev, Sergey, & Chumachenko, Dmytro (2021). Comparative study of linear regression and SIR models of COVID-19 propagation in Ukraine before vaccination. *Radioelectronic and Computer Systems, 0*(3), 5–18. https://doi.org/10.32620/reks.2021.3.01

Nakamura, Gilberto M., Cardoso, George C., & Martinez, Alexandre S. (2020). Improved susceptible−infectious−susceptible epidemic equations based on uncertainties and autocorrelation functions. *Royal Society Open Science, 7*(2), 191504.

Nesteruk, I. (2020). Simulations and predictions of COVID-19 pandemic with the use of SIR model. *Innov Biosyst Bioeng, 4*(2), 110–121. https://doi.org/10.20535/ibb.2020.4.2.204274. http://ibb.kpi.ua/article/view/2042748

Nesteruk, I. (2021a). *COVID-19 pandemic dynamics*. Springer Nature. https://doi.org/10.1007/978-981-33-6416-5. https://link.springer.com/book/10.1007/978-981-33-6416-5

Nesteruk, I. (2021b). Detections and SIR simulations of the COVID-19 pandemic waves in Ukraine. *Computer Math Biophysics, 9*, 46–65. https://doi.org/10.1515/cmb-2020-0117

Nesteruk, I. (2021c). Visible and real sizes of new COVID-19 pandemic waves in Ukraine. *Innov Biosyst Bioeng, 5*(2), 85–96. https://doi.org/10.20535/ibb.2021.5.2.230487. http://ibb.kpi.ua/article/view/230487

Nesteruk, I. (2021d). Influence of possible natural and artificial collective immunity on new COVID-19 pandemic waves in Ukraine and Israel. *Explor Research Hypothesis Medical*. https://doi.org/10.14218/ERHM.2021.00044. Nov 11.

Nesteruk, I. (2022a). Vaccination and testing as a means of ending the COVID-19 pandemic: Comparative and statistical analysis. *MedRxiv. Posted June, 21*. https://doi.org/10.1101/2022.06.16.22276531

Nesteruk, I. (2022b). Summer COVID-19 waves in Japan: What we can expect in 2022? Preprint. Research gate. *Posted July, 29*. https://doi.org/10.13140/RG.2.2.36687.56483

Nesteruk, I. (2022c). Verification of SIR simulations of the new summer COVID-19 wave in Japan. *Preprint Research Gate Posted August, 11*. https://doi.org/10.13140/RG.2.2.28528.33285

Nesteruk, I. (2022d). The COVID-19 epidemic wave in mainland China at the end of 2022: Monitoring and predicting with the use of the generalized SIR model. *Preprint Resarch Gate Posted December, 19*. https://doi.org/10.13140/RG.2.2.19479.24488

Nesteruk, I. (2022e). Epidemic waves caused by SARS-CoV-2 omicron (B.1.1.529) and pessimistic forecasts of the COVID-19 pandemic duration. March *MedComm, 3*(1). https://doi.org/10.1002/mco2.122.

Nesteruk, I. (2022f). Simulations of new COVID-19 pandemic waves in Ukraine and in the world by generalized SIR model. *System Research & Information Technologies, 2*, 94–103. https://doi.org/10.20535/SRIT.2308-8893.2022.2.07

Nesteruk, I., & Benlagha, N. (2021). Predictions of COVID-19 pandemic dynamics in Ukraine and Qatar based on generalized SIR model. *Innov Biosyst Bioeng, 5*(1), 37–46. https://doi.org/10.20535/ibb.2021.5.2.230487. http://ibb.kpi.ua/article/view/230487

Nesteruk, I., & Rodionov, O. (2022). Omicron waves of the COVID-19 pandemic and efficacy of vaccinations and testing. May 24 *Journal Biomed Research Environment Science, 3*(5), 588–594. https://doi.org/10.37871/jbres1484.

Niwa, Makoto, Hara, Yasushi, Sengoku, Shintaro, & Kodama, Kota (2020). Effectiveness of social measures against COVID-19, outbreaks in selected Japanese regions analyzed by system dynamic modeling. *International Journal of Environmental Research and Public Health, 17*, 6238. https://doi.org/10.3390/ijerph17176238

Sala, A., & Scoccimaro, R. (2022). Mapping COVID-19 in Japan and greater tokyo area, socio-spatial and political analysis of the epidemic. In M. Yano, F. Matsuda, A. Sakuntabhai, & S. Hirota (Eds.), *Socio-life science and the COVID-19 outbreak. Economics, law, and institutions in asia pacific*. Singapore: Springer. https://doi.org/10.1007/978-981-16-5727-6_5.

Shimizu, Kazuki, & Negita, Masashi (2020). Lessons learned from Japan's response to the first wave of COVID-19: A content analysis. *Healthcare, 8*, 426. https://doi.org/10.3390/healthcare8040426

Torjesen, I. (2022). Covid-19: Peak of viral shedding is later with omicron variant, Japanese data suggest. *BMJ, 376*, o89. https://doi.org/10.1136/bmj.o89

Uehara, T. (2020). Understanding the spread of COVID-19 in Japan: Preliminary results from a system dynamics model. *Preprints*. , Article 2020050499. https://doi.org/10.20944/preprints202005.0499.v1

Vicuña, María Ignacia, Vásquez, Cristián, & Quiroga, Bernardo F. (2021). Forecasting the 2020 COVID-19 epidemic: A multivariate quasi-Poisson regression to model the evolution of new cases in Chile, 23 April *Frontiers in Public Health*. https://doi.org/10.3389/fpubh.2021.610479.

Weiss, H. (2013). The SIR model and the foundations of public health. *MatMat, 3*, 1–17.

World Health Organization. "Coronavirus disease (COVID-2019) situation reports". https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

Yamauchi, T., Takeuchi, S., Uchida, M., Saito, M., & Kokaze, A. (2022). The association between the dynamics of COVID-19, related measures, and daytime population in Tokyo, 23 Feb *Scientific Reports, 12*(1), 3063. https://doi.org/10.1038/s41598-022-06716-4.

Zeyu, Lyu, & Takikawa, Hiroki (2021). *Disparity and dynamics of social distancing behaviors in Japan: An investigation of mobile phone mobility data*. July. https://doi.org/10.2196/preprints.31557.

Kato, M., Sakihama, T., Kinjo, Y., Itokazu, D., & Tokuda, Y. (2022). Effect of climate on COVID-19 incidence: A cross-sectional study in Japan. *Jan. Korean J Fam Med, 43*(1), 37–41. Article in English | MEDLINE | ID: covidwho-1667450.

Ko, Y., Kinoshita, R., Yamauchi, M., Otani, K., Kamigaki, T., Kasuya, K., Yoneoka, D., Arima, Y., Kobayashi, Y., Arashiro, T., Otsuka, M., Shimbashi, R., & Suzuki, M. (2022). Impact of the coming-of-age day and ceremony on the risk of SARS-CoV-2 transmission in Japan: A natural-experimental study based on national surveillance data 27 July. https://doi.org/10.1111/irv.13027.

Yoneoka, D., Shi, S., Nomura, S., Tanoue, Y., Kawashima, T., Eguchi, A., Matsuura, K., Makiyama, K., Uryu, S., Ejima, K., Sakamoto, H., Taniguchi, T., Kunishima, H., Gilmour, S., Nishiura, H., & Miyata, H. (2021). Assessing the regional impact of Japan's COVID-19 state of emergency declaration: A population-level observational study using social networking services. *BMJ Open, 11*(2), Article e042002. https://doi.org/10.1136/bmjopen-2020-042002. Published online 2021 Feb 15.

Weinberger, D. M., Cohen, T., Crawford, F. W., Mostashari, F., Olson, D., Pitzer, V. E., Reich, N. G., Russi, M., Simonsen, L., Watkins, A., & Viboud, C. (2020). *Estimating the early death toll of COVID-19 in the United States* [Preprint.] MEDRXIV Apr https://doi.org/10.1101/2020.04.15.2006643.

https://edition.cnn.com/2020/11/02/europe/slovakia-mass-coronavirus-test-intl/index.html.

https://en.wikipedia.org/wiki/Basic_reproduction_number.

https://en.wikipedia.org/wiki/Timeline_of_the_2019%E2%80%9320_Wuhan_coronavirus_outbreak. (Accessed 4 March 2020).

https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf.

https://podillyanews.com/2020/12/17/u-shkolah-hmelnytskogo-provely-eksperyment-z-testuvannyam-na-covid-19/.

https://toyokeizai.net/sp/visual/tko/covid19/en.html.

https://www.mhlw.go.jp/stf/covid-19/open-data_english.html.

https://www.niid.go.jp/niid/en/iasr/865-iasr/9288-477te.html.

https://www.r-bloggers.com/2020/04/effective-reproduction-number-estimation/.