Original article

SIMULATIONS AND PREDICTIONS OF COVID-19 PANDEMIC WITH THE USE OF SIR MODEL

I. Nesteruk[*]

Institute of Hydromechanics, NAS of Ukraine, Kyiv, Ukraine
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

*Corresponding author: inesteruk@yahoo.com

Background. The COVID-19 pandemic is of great interest to researchers due to high mortality and a very negative impact to the world economy. A detailed scientific analysis of the phenomenon is yet to come, but the public is already interested in the problems of duration of the epidemic, the expected number of patients, where and when the pandemic started. Correct simulation of the pandemic dynamics needs complicated mathematical models and many efforts for unknown parameters identification. In this article, preliminary estimates for many countries and world will be presented, summarized and discussed.

Objective. We will estimate the epidemic characteristics for USA, Germany, UK, the Republic of Korea and in the world with the use of SIR simulations and compare them with the results obtained before for Italy, Spain, France, the Republic of Moldova, Ukraine and Kyiv. The hidden periods, epidemic durations, final numbers of cases and quarantine measures will be discussed.

Methods. In this study we use the known SIR (susceptible-infected-removed) model for the dynamics of the epidemic, the known exact solution of the linear differential equations and statistical approach developed before.

Results. The optimal values of the SIR model parameters were identified with the use of statistical approach for epidemic dynamics in USA, Germany, UK, the Republic of Korea, and in the world. The actual number of cases and the number of patients spreading the infection versus time were calculated. The hidden periods, durations and final sizes of the epidemic were evaluated. In particular, the pandemic began in China no later than October, 2019. If current trends continue, the end of the pandemic should be expected no earlier than March 2021, the global number of cases will exceed 5 million. A simple method for assessing the risk of premature weakening of quarantines is proposed.

Conclusions. The SIR model and statistical approach to the parameter identification are helpful to make some reliable estimations for the epidemic dynamics, e.g., the real time of the outbreak, final size and duration of the epidemic and the number of persons spreading the infection versus time. This information will be useful to regulate the quarantine activities and to predict the medical and economic consequences of the pandemic.

Keywords: coronavirus pandemic; epidemic outbreak; coronavirus COVID-19; mathematical modeling of infection diseases; SIR model; parameter identification; statistical methods.

Introduction

Here we consider the global COVID-19 pandemic dynamics and epidemic outbreaks in USA, Germany, UK, South Korea other countries and regions with the use of official WHO data sets [1]. The SIR model, connecting the number of susceptible S, infected and spreading the infection I and removed R persons, was applied [2, 3]. The unknown parameters of this model can be estimated with the use of the cumulative number of cases $V = I + R$ and the statistics-based method of parameter identification [4].

This approach was used in [4−11] to estimate the Corona pandemic dynamics in China, the Republic of Korea, Italy, Austria, Spain, Germany, France, the Republic of Moldova, Ukraine and Kyiv. Usually the number of cases registered during the initial period of an epidemic is not reliable, since many infected persons are not detected. That is why the correct estimations of epidemic parameters can be done with the use of data sets obtained for later periods of the epidemic when the number of detected cases is closer to the real one. This fact necessitates a periodic reassessment of the epidemic's characteristics and forecasts for its final size and duration. In this paper we will recalculate the pandemic parameters for Germany and South Korea, provide estimations for UK, USA and the world. The accuracy of the method and the problem of hidden periods of the epidemic outbreaks in different countries will be discussed. Some recommendations about the quarantine mitigation will be proposed.

Materials and methods

Data

The official information regarding the accumulated numbers of confirmed COVID-19 cases V_j in South Korea, Germany, UK, USA and in the world from WHO daily situation reports (numbers 81−109) [1] is presented in Table 1. The corresponding moments of time t_j (measured in days) are also shown in this table. Data sets for the period April 9−29 were used for calculations. Other values were used only for verifications of calculations.

SIR model

The SIR model for an infectious disease [2−5] relates the number of susceptible persons S (persons who are sensitive to the pathogen and **not protected**); the number of infected is I (persons who are sick and **spread the infection**; please don't confuse with the number of still ill persons, so known active cases) and the number of removed R (persons who **no longer spread the infection**; this number is the sum of isolated, recovered, dead, and infected people who left the region); α and ρ are constants:

Table 1: Official cumulative numbers of confirmed cases in the Republic of Korea, Germany, UK, USA, and in the world used for calculations and verifications of predictions [1]

Day in April and May, 2020	Time in days t_j	Number of cases in South Korea	Number of cases in Germany	Number of cases in UK	Number of cases in USA	Global number of cases
9	47	10450	113525	65081	425889	1521252
10	48	10480	117658	70276	461275	1610909
11	49	10512	120479	78995	492881	1696588
12	50	10537	123016	84283	524514	1773084
13	51	10564	125098	88625	553822	1844863
14	52	10591	127584	93877	578268	1914916
15	53	10613	130450	98480	604070	1991562
16	54	10635	133830	103097	632781	2074529
17	55	10653	137439	108696	665330	2160207
18	56	10661	139897	114221	695353	2241359
19	57	10674	141672	120071	723605	2314621
20	58	10683	143457	124747	751273	2397216
21	59	10694	145694	129048	776907	2471136
22	60	10702	148046	133499	800926	2544792
23	61	10708	150383	138082	830053	2626321
24	62	10718	152438	143468	860772	2719897
25	63	10728	154175	148381	899281	2804796
26	64	10738	155193	152844	931698	2878196
27	65	10752	156337	157153	960916	2954222
28	66	10761	157641	161149	983457	3018952
29	67	10765	159119	165225	1003974	3090445
30	68	10774	159119	171257	1035353	3175207
1	69	10780	161703	177458	1067127	3267184
2	70	10793	162496	182264	1093880	3349786
3	71	10801	163175	186603	1125719	3435894
4	72	10804	163860	190588	1154985	3517345
5	73	10806	164897	194994	1171185	3588773
7	75	10822	167300	206719	1215571	3759967

$$\frac{dS}{dt} = -\alpha SI, \tag{1}$$

$$\frac{dI}{dt} = \alpha SI - \rho I, \tag{2}$$

$$\frac{dR}{dt} = \rho I. \tag{3}$$

The parameter α is called the infection rate, since according to (1) it shows how quick the susceptible persons become infected. Large values of this parameter correspond to severe epidemics with many victims. This parameter accumulates many characteristics. First it shows how strong (virulent) is the pathogen and what is the way of its spreading. For airborne droplets transmission (typical for coronavirus COVID-19), the values of α are higher than for AIDS, for example. Parameter α accumulates also the frequency of contacts and the way of contacting. Epidemics use to start at large cities were the average contact rate is much higher, than in small villages. In order to decrease the value of α, 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 value of parameter α and reduce the negative effects of the pandemic.

Usually the parameter ρ is called the immunization rate. This name is reasonable when there is no isolation and all the sick persons recover. Since the removed persons number R is the sum of isolated, recovered and dead people, eq. (3) demonstrates the increase rate for R. So we will call this parameter the removing rate. The inverse value $1/\rho$ is an estimation for time of spreading the infection. So, we are interested in increasing the value of parameter ρ and decreasing $1/\rho$. 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. (1)−(3)), the sum $N = S + I + R$ must be constant for all moments of time and can be treated as the amount on susceptible persons before the moment t_0 of the epidemic outbreak, since $I = R = 0$ at $t < t_0$. It must be noted that the constant N is not the volume of population N_{total}, but only the initial number if people sensitive and not protected to some specific disease.

Unfortunately, the ratio N/N_{total} may be very large for coronavirus COVID-19. We can see this from the situation with the liner Diamond Princess. The total initial number of persons on board was 3711; on February 18, 2020 the cumulative number of confirmed cases was 542. Thus the percentage of susceptible persons can be at least 14.6% and if the people will be not protected and isolated enough, hundred millions in the world can be infected. It means, we have to work hard on protection and isolation to avoid millions of deaths and reduce the number N.

Analytical solution of SIR equations

To determine the initial conditions for the set of eqs. (1)−(3), let us suppose that at the moment of the epidemic outbreak t_0 [4]:

$$I(t_0) = 1, \quad R(t_0) = 0, \quad S(t_0) = N - 1, \quad N = S + I + R.$$

The analytical solution for the set of eqs. (1)−(3) was obtained by introducing the function $V(t) = I(t) + R(t)$, corresponding to the number of victims or cumulative confirmed number of cases [4]:

$$F_1(V, N, \nu) = \alpha(t - t_0), \tag{4}$$

$$F_1 = \int_1^V \frac{dU}{(N - U)[\nu \ln(N - U) + U - \nu \ln(N - 1)]}, \tag{5}$$

$$\nu = \frac{\rho}{\alpha}.$$

Thus, for every set of parameters N, ν, α, t_0 and a fixed value of V the integral (5) can be calculated and the corresponding moment of time can be determined from (4). Then functions $I(t)$ and $R(t)$ can be easily calculated with the formulas [4]:

$$I = \nu \ln S - S + N - \nu \ln(N - 1),$$

$$S = N - V, \quad R = V - I.$$

Function I has a maximum at $S = \nu$ and tends to zero at infinity, see [2, 3]. In comparison, the number of susceptible persons at infinity $S_\infty > 0$, and can be calculated from the non-linear equation [4]:

$$S_\infty = (N - 1)e^{\frac{S_\infty - N}{\nu}}. \tag{6}$$

The final number of victims (final accumulated number of cases) can be calculated from:

$$V_\infty = N - S_\infty . \tag{7}$$

To estimate the duration of an epidemic outbreak, we can use the condition:

$$V(t_{\text{final}}) = 1 . \tag{8}$$

which means that at $t > t_{\text{final}}$ less than one person still spreads the infection.

Parameter identification procedure

In the case of a new epidemic, the values of this independent four parameters are unknown and must be identified with the use of limited data sets. A statistical approach was used in [4−11] to estimate the values of unknown parameters. The registered points for the number of victims V_j corresponding to the moments of time t_j can be used in order to calculate $F_{1j} = F_1(V_j, N, \nu)$ for every fixed values N and ν with the use of (5) and then to check how the registered points fit the straight line (4). For this purpose the linear regression can be used, e.g., [12], and the optimal straight line, minimizing the sum of squared distances between registered and theoretical points, can be defined. Thus we can find the optimal values of α, t_0 and calculate the correlation coefficient r for the linear dependence (4).

Then the F-test may be applied to check how the null hypothesis that says that the proposed linear relationship (4) fits the data set. The experimental value of the Fisher function can be calculated with the use of the formula:

$$F = \frac{r^2(n-m)}{(1-r^2)(m-1)} , \tag{9}$$

where n is the number of observations, $m = 2$ is the number of parameters in the regression equation [12]. The corresponding experimental value F has to be compared with the critical value $F_C(k_1, k_2)$ of the Fisher function at a desired significance or confidence level ($k_1 = m-1$, $k_2 = n-m$) [13]. When the values n 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 and ν, we have to find the maximum of the correlation coefficient for the linear dependence (4). To compare the reliability of different predictions (with different values of n) it is useful to use the ratio $F/F_C(1, n-2)$ at fixed significance level. We will use the level 0.001; corresponding values of $F_C(1, n-2)$

can be taken from [13]. The most reliable prediction yields the highest $F/F_C(1, n-2)$ ratio.

Results

Usually the number of cases during the initial period of an epidemic outbreak is not reliable. To avoid their influence on the results, only V_j values for the period April 9−29, 2020 ($47 \le t_j \le 67$; $n = 21$; $F_C(1, n-2) = 15.2$; see Table 1) were used to calculate the epidemic characteristics. Since during the quarantine, the international people exchange is quite limited, we can apply the SIR model for every country assuming its parameters to be constant (but different for every country) during the fixed period of time. The results of calculations are shown in Tables 2 and 3. To illustrate the influence of data on the results of SIR simulations, the previous estimation for Germany (prediction 1 calculated with the use of V_j values for the period March 28 − April 10, 2020 ($35 \le t_j \le 48$; $n = 14$; $F_C(1, n-2) = 18.6$; see [10]) is also presented in Table 2.

It can be seen that the previous prediction for Germany (see [10]) was more optimistic. Fresh data sets has showed that the final number of cases in this country could reach 177,000 and their appearance can stop only after August 4, 2020 (see Table 2, prediction 2). The presented second prediction for South Korea is also more pessimistic in comparison with the first one [6]. In particular, the epidemic stop is expected after June 29, 2020 (see Table 2). These estimations are valid only when the quarantine measures, isolation rate and the coronavirus activity will be the same as for the period taken for calculations.

Tables 2 and 3 illustrate that real epidemic outbreaks in Germany, USA and the Republic of Korea probably occurred in November-December 2019, in UK − in early 2020. The real beginning of the global COVID-19 pandemic can be attributed to the beginning of October 2019. It happened in China, most likely in Wuhan, the epicenter of its visible course. Unfortunately, official data from China are very contradictory, what makes their analysis impossible using the SIR model. In any case, the estimations of the t_0 values presented in [4, 5] are no longer relevant. The rather long duration of the pandemic is expected. The last cases could stop to appear only in March 2021 after exceeding the value of 5 millions. This long-term prediction is very preliminary, corresponds to

the current situation and does not take into account the repeating outbreaks that are possible and are already happening in many countries.

The results about the epidemic hidden periods (time between t_0 and the day when the first COVID-19 case was confirmed), epidemic durations $t_{final} - t_0$ and the final numbers of cases V_∞ (final sizes) for different countries are presented in Table 4. For the world data, December 8, 2019

was taken as the day of the first laboratory confirmed case in Wuhan, China (according to [14]). The results of previous calculations for Austria, Italy, Spain, France, Moldova, Ukraine and Kyiv from [10, 11] are also shown in Table 4. It can be seen that the longest epidemic durations are expected in the countries with the longest hidden periods (USA, Italy, Germany). Probably, the zero hidden period in France indicates the need of re-calculations after obtaining more recent data on

Table 2: Epidemic characteristics for the Republic of Korea and Germany. Optimal values of parameters, final sizes and days (last two rows)

Country, prediction	South Korea, prediction 2	Germany, prediction 2	Germany, prediction 1 [10]
N	13679.54688	882400	1023648
ν	6892.82635608829	790630.4	946949.0
α	1.91761713114e-05	9.3698628e-07	1.9664814e-06
t_0	−55.5257861059147	−74.7002	−22.83551
ρ	0.132178019	0.7408098	1.862158
$1/\rho$	7.5655544	1.349874	0.537011
r	0.99686044433	0.999140468	0.9990431
F, eq. (9)	3011.66336	11038.27866	6261.40324
$F/F_C(1, n-2)$	198.135747	726.20254	336.63458
S_∞, eq. (6)	2837	705448	874169
V_∞, eq. (7)	10842	176952	149479
t_{final}, eq. (8)	127.8	164.2	104.8

Table 3: Epidemic characteristics for UK, USA and world. Optimal values of parameters, final sizes and days (last two rows)

Country	UK	USA	World
N	479782.4	7082400	6637317.12000000
ν	361758.017613005	6312894.63367680	3537150.79869746
α	9.1371956639e-07	1.08933738707e-07	2.34901375364096e-08
t_0	−47.0679605620665	−85.9627117094321	−144.265486657790
ρ	0.330545378991741	0.687687214511608	0.0830881587484244
$1/\rho$	3.02530322175518	1.45414947216984	12.0354093178044
r	0.999363727388945	0.999378778826037	0.999861835424859
F, eq. (9)	14916.4586251212	15278.2115499352	68744.3303788117
$F/F_C(1, n-2)$	981.345962179029	1005.14549670626	4522.65331439551
S_∞, eq. (6)	264883	5601247	1595881
V_∞, eq. (7)	214899	1481153	5041436
t_{final}, eq. (8)	180.1	213.5	393

Table 4: Epidemic characteristics for different countries and regions. Estimated days of epidemic outbreak, hidden periods, epidemic durations, the final days and number of cases

Country or region, number of prediction, reference	Last day of data set used for calculations	Estimated day of epidemic outbreak t_0	Hidden period, days	Epidemic duration, $t_{final} - t_0$, days	Last day of epidemic t_{final}	Final number of cases V_∞
World	Apr 29	Oct 1	68	537	Mar 21, 2021	5041436
USA	Apr 29	Nov 28	55	300	Sep 23	1481153
Germany, 2	Apr 29	Dec 10	48	238	Aug 4	176952
UK	Apr 29	Jan 6	25	227	Aug 20	214899
South Korea, 2	Apr 29	Dec 28	23	183	June 29	10842
Italy, 6 [10]	Apr 18	Nov 26	65	266	Aug 18	225736
Spain, 2 [10]	Apr 18	Jan 9	22	179	July 6	221011
France, 2 [10]	Apr 18	Jan 24	0	148	June 20	129823
Austria, 4 [10]	Apr 10	Feb 6	19	103	May 19	15108
Moldova, 2 [10]	Apr 18	Feb 26	10	110	June 15	3541
Ukraine, 8 [11]	Apr 24	Feb 16	15	160	Jul 26	15273
Kyiv, 2 [11]	Apr 24	Mar 2	17	122	Jul 2	1947

the number of cases. The predicted saturation value 129823 for this country (see Table 4) is 4.5% lower than real number of cases 135980, registered on May 7 after 20 days of observations (after April 18). The real numbers of cases in Spain, Moldova and Austria are also higher than predicted saturation levels shown in Table 4.

The SIR curves and markers representing the V_j values taken for calculations ("circles"), comparisons ("triangles") and verifications of calculations ("stars") are shown in Figs. 1 and 2 by different colors corresponding to the country or region (World − brown; USA − blue; Italy − green; Spain − yellow; UK − red and South Korea − magenta). Solid lines represent the number of cases (victims) $V(t) = I(t) + R(t)$, dashed lines show the number of infected persons spreading the pathogen $I(t)$. The number of laboratory confirmed cases in Wuhan, China are shown in Fig. 1 by brown "squares". These values was calculated in [5] with the use of information from [14].

Dashed brown curve in Fig. 1 illustrates that more than 600 persons could spread the infection on December 31, 2019 − the day when China notified WHO about the situation in Wuhan (see brown vertical line in Fig. 1). On January 23, 2020 this city was locked down, but the number of infected persons could be estimated by 3000 with hundreds of cases in USA, Italy and Germany (the officially confirmed number of cases was 830 in mainland China, 2 in the Republic of Korea and 1 in USA on this day). The big difference between number of calculated (bold lines in Figs. 1 and 2) and actual cases ("triangles") is explained by the

fact that many infected people do not have symptoms and there were and still are problems with testing. In particular, the hidden periods (time between the estimated epidemic beginning t_0 and the first confirmed case) can be rather long (see Fig. 1 and Table 4).

Recently, there is more and more evidence in the media and literature about the hidden periods of the epidemic. In particular, according to [10] the first COVID-19 could happen in Italy around November 26, 2019 (see Table 4). This result correlates with the information form Giuseppe Remuzzi, director of the Mario Negri Institute for Pharmacological Research that "virus was circulating before we were aware of the outbreak in China" [15]. Probably the spread of the infection was facilitated by the Military World Games held in Wuhan from 18 to 27 October with the participation of 9,300 athletes from more than 140 countries. Many participants became ill with COVID-19 symptoms and passed the infection on to their families [16]. A very fundamental statistical analysis [17] showed that the number of cases of pneumonia in the United States in January and February 2020 exceeded last year's figures and this excess was higher in those states where the actual numbers of reported COVID-19 cases are larger.

Fig. 1 illustrates that before Mach 1, 2020 the estimated numbers of cases in Italy and USA were rather close. Nevertheless, on March 1 the number of confirmed cases in Italy was 1689 in comparison with 62 in USA. This delay with detection of infected persons and later quarantining [18] caused the huge recent number of cases in USA (for

example, in New York the quarantine was announced on March 20, in comparison with March 8 in Lombardy and March 10 in Italy). Fig. 1 shows that in the Republic of Korea the detected number of cases was close to the estimated one (compare solid magenta line and magenta "triangles") already on March 1. Timely testing and isolation of patients allowed this country to avoid a large number of cases.

Discussion

Reliability of predictions

It was already mentioned that during the initial stages of epidemic the registered number of cases is much lower than the real one. This fact reduces the accuracy of any simulation using the registered values. Nevertheless, the prediction have
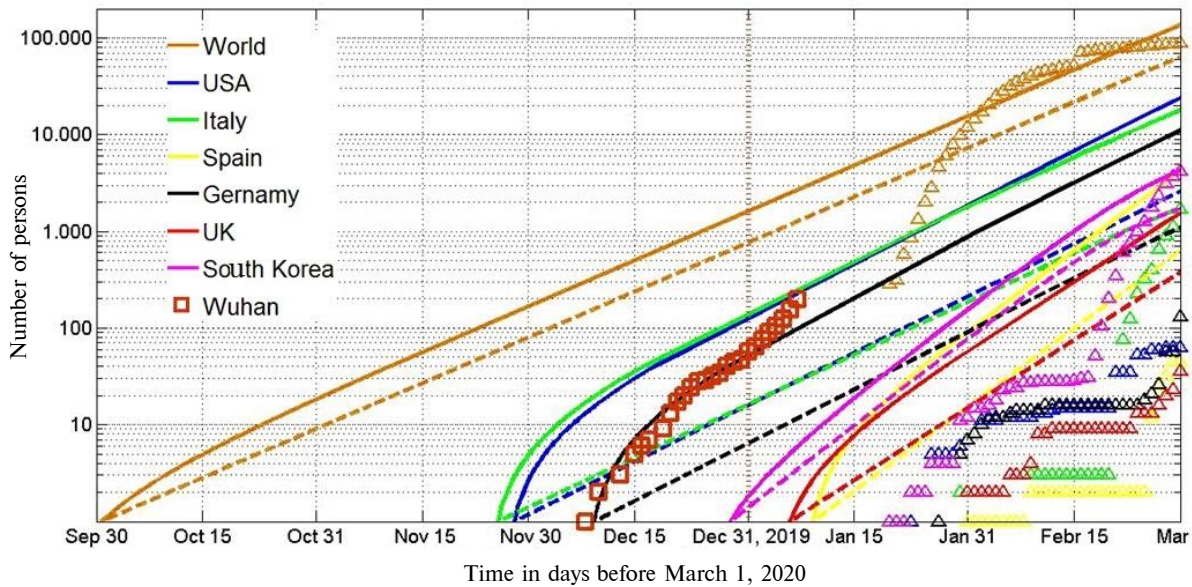
Figure 1: SIR curves (lines) and accumulated number of cases (markers) versus time before March 1, 2020. Numbers of infected and spreading I (dashed) and victims $V = I + R$ (solid)
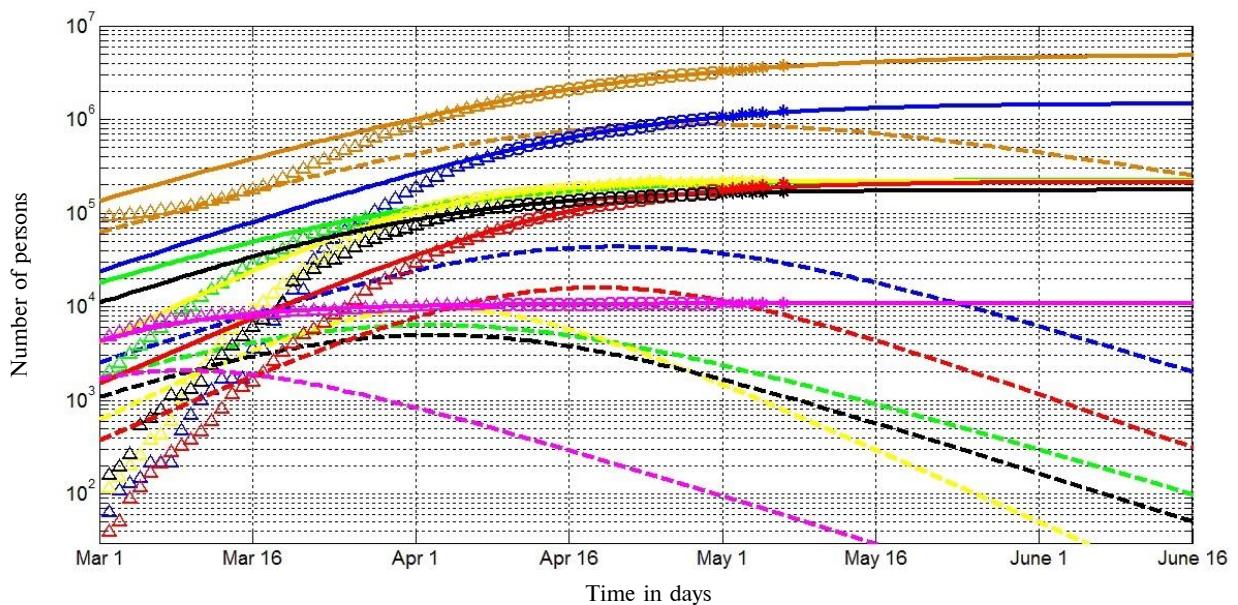
Figure 2: SIR curves (lines) and accumulated number of cases (markers) versus time after March 1, 2020. Numbers of infected and spreading I (dashed) and victims $V = I + R$ (solid)

to be performed in order to estimate the final sizes and the durations of epidemics in different countries even with limited accuracy.

Errors caused by incomplete data can be illustrated by two different predictions for the Republic of Korea. Both of them were performed with the use of SIR model and the same method of parameter identification. The first one used the data for the period February 17 − March 12, 2020 and predicted: the final accumulated number of cases $V_\infty \approx 8117$; $t_0 = -9$ days and the final day of epidemic March 20, 2020 (see [6]). The results of calculations presented in Table 2 illustrate that the difference in final sizes is about 25%, but the epidemic duration predictions differ more than 5 times. Three different estimations for Italy [8−10] yielded the variation in the final sizes from 111548 to 225736; the hidden period estimates increased by 66 days with the use of more recent data sets.

Since the presented forecasts for USA and the world are very long-term, they must be considered as very preliminary and optimistic. The global pandemic dynamics is very unpredictable, since the situation is very different in different countries. In particular, there is no quarantine in Belarus and the word "coronavirus" is prohibited in Turkmenistan. In any case the global dynamics must be updated with the use of new data sets. Unfortunately, the new estimations will probably be more pessimistic. When the number of confirmed cases tends to the real one, the accuracy of SIR simulation may be rather high. An example is the latest forecast for Austria [10]. After 28 days of observation the predicted final size (see Table 4) is only 3.6% lower than the real number 15673 (May 7, 2020).

Comparison of epidemic dynamics in different countries

WHO data [1] and Fig. 1 show that visible periods of epidemic begin in different time in different countries. This fact allows comparing the initial period of the epidemic in one country with the use of SIR curve obtained for another country, where the registered number of cases is already approaching the real one. Such a comparison of the epidemic dynamics in Italy and mainland China was done in [7] (see Fig. 3). To synchronize the time, the time moment t_s corresponding to the number of victims $V=76$ (this cumulative number was confirmed in Italy on March 22, see [1]) was calculated with the optimal values of parameters for the epidemic in mainland China [5]. The result $t_s = -15.82545$ means that all the corresponding time moments for the data set for Italy have to be shifted by the value 52.82545 (since March 22 corresponds to the time moment $t_j = 37$, see Table 1 in [7], zero t_j value corresponds to January 16, 2020). The "stars" in Fig. 3 represent the accumulated number of confirmed cases in Italy. It can be seen that the epidemic in Italy developed much more rapid than it was in China. Thus, the application of SIR model allowed predicting the higher final number of cases in Italy already on March 3, when the preprint [7] was submitted.
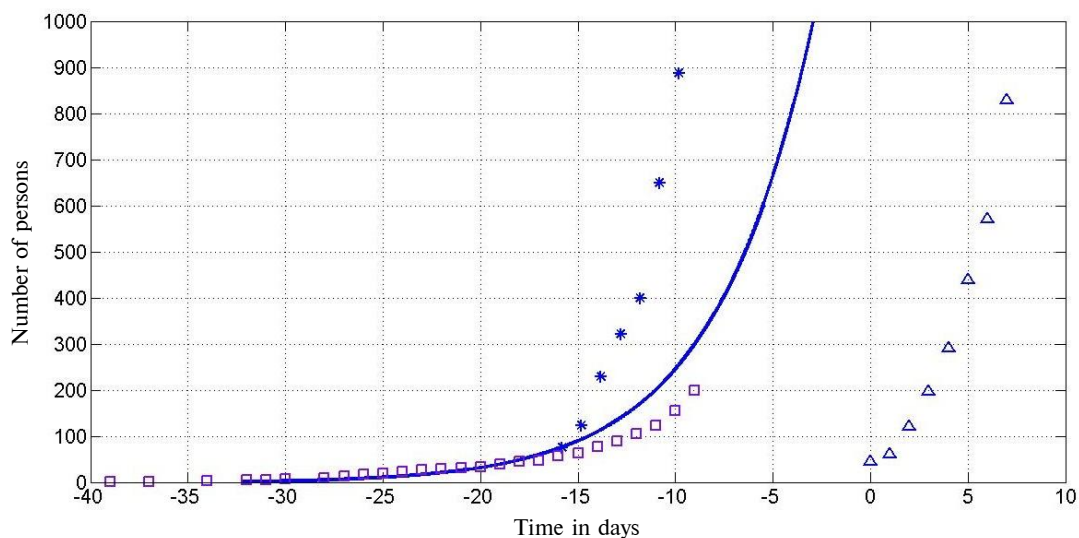

Figure 3: Comparison of the early stage of the epidemic dynamics in Italy with the corresponding SIR curve for mainland China. Number of victims $V = I + R$ (blue line) versus time in days (zero point is January 16, 2020) estimated in [5] for the mainland China. "Triangles" represent the official data set for China from [5], "squares" − the cumulative number of cases in Wuhan calculated with the information presented in [14], "stars" − to the points corresponding to the epidemic in Italy [1], with the corresponding time shift.

Control of testing and relaxation of quarantine

SIR simulation could be used to control the situation with COVID-19 testing. For example, in Ukraine the PLC tests were introduced for pneumonia patients and medical staff only after April 10, 2020. To estimate how the change in testing algorithm affected the epidemic dynamics in Ukraine and Kyiv, two series of calculations were performed in [10, 11] with the use of data sets for the periods March 28 − April 10 and April 11−24.

Figs. 4 and 5 illustrate the results of these estimations. The corresponding SIR curves (solid for the second data set and dashed for the first one) are very different. Delayed proper testing could cost Ukraine at least 9,000 additional patients and an epidemic duration increase of 47 days (compare blue lines in Fig. 4). That is why the maximum PCR testing (especially for medical staff and patients with pneumonia) can be recommended in all countries as an effective means of reducing the scale and duration of the pandemic.
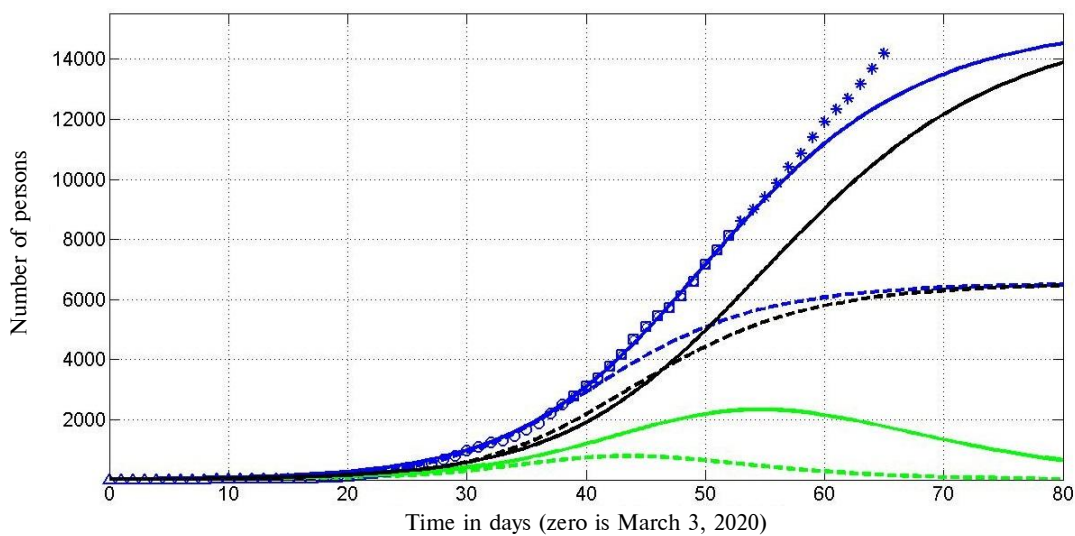

Figure 4: Ukraine: SIR curves (lines) and accumulated number of cases (markers) versus time. Numbers of infected and spreading *I* (green), removed *R* (black) and victims *V = I + R* (blue). Dashed lines correspond to the fist data set (March 28 − April 10) shown by "circles", solid − to the second data set (April 11−24), shown by "squares". "Triangles" show the accumulated numbers of cases before March 28, "stars" − after April 24
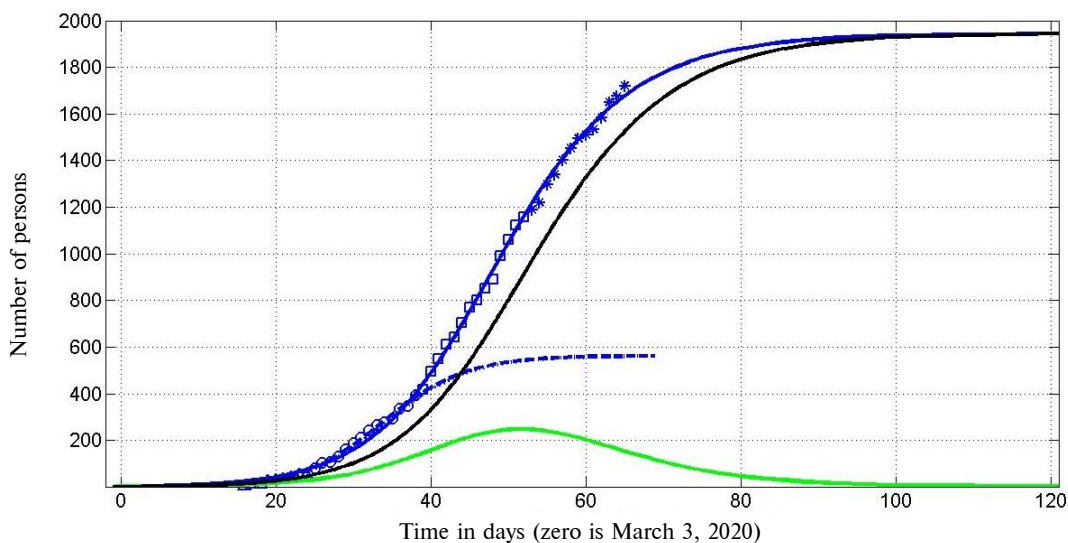

Figure 5: Kyiv: SIR curves (lines) and accumulated number of cases (markers) versus time. Numbers of infected and spreading *I* (green), removed *R* (black) and victims *V = I + R* (blue). Dashed line corresponds to the fist data set (March 28 − April 10) shown by "circles", solid − to the second data set (April 11−24), shown by "squares". "Triangles" show the accumulated numbers of cases before March 28, "stars" − after April 24

The calculated SIR curved were used to control the epidemic dynamics after April 24, 2020 (see "stars" in Figs. 4 and 5). It can be seen, that the situation in Kyiv develops according to the second prediction (see Fig. 5), but the number of cases in Ukraine increases much rapidly in comparison with the second prediction shown in the Fig. 4 by solid blue line. The obvious differences in the epidemic dynamics can be explained by the fact that the situation with testing in the Ukrainian province is much worse than in the capital. It is possible that the dynamics was also affected by the large number of quarantine violations in the province during the orthodox Easter celebrations.

The problem of easing quarantine has become urgent in many countries. The SIR $I(t)$ curves can be used to estimate possible risks. For example, green line in Fig. 5 shows that in Kyiv the number of infected persons (they may feel completely healthy or have mild symptoms) is estimated by 100 on May 11, 2020 (the day of quarantine easing, $t_j = 68$). It means that the probability to meet such person is rather low $p \approx 3,4 \times 10^{-5}$. But if you have M contacts, the probability of meeting at least one infected person increases according to the formula

$$P = 1 - (1 - p)^M \approx pM. \qquad (10)$$

Therefore, a person on duty in the subway or a trolleybus driver is quite likely to meet an infected person during, for example, 10 days of work (until the value of p decreases). Meeting an infected person does not mean getting infected, so masked mode and distance in transport and other public places must remain mandatory. Workers in transport, trade, pharmacies, police (all of whom are forced to have many contacts) must be provided with enhanced protection. People at risk should continue to refrain from traveling, visiting indoors and minimizing visits to medical facilities. Where possible, distant work and study should be maintained.

For ordinary Kyiv citizens working in small groups, the risk of meeting an infected person depends on the transport situation. If cars and metro stations, land transport will be regularly decontaminated, the values of M and P for passengers will be small. Everyone (not only in Kyiv) can assess their own level of risk (allowable probability) using formula (10).

Could the pandemic be avoided?

The answer to this question may be rather positive. This is evidenced by the experience of Hong Kong, which introduced the relevant measures on December 31, 2019 as soon as alarming information from Wuhan appeared on a social media platform. The Hong Kong authorities immediately closed "transport ties with Wuhan"; increased "vigilance and temperature screenings at every border checkpoint, including the city's international airport and high-speed railway station in West Kowloon", the hospitals were told "to report further cases of "pneumonia of unknown origin" [19]. As a result the accumulated number of cases in Hong Kong was 1045 on May 9, 2020. For example, the number of infected people in Kyiv was 1.7 times higher.

Quarantine has been announced all over Ukraine on March 12, 2020. It happened 9 days after the first COVID-19 case was confirmed in Chernivtsi region and 7 days before the first case in Kyiv. On May 11 the corresponding numbers of cases were 2396 and 1930 respectively. Taking into account that the population of Kyiv exceeds the population of Chernivtsi region more than three times, it can be concluded that the timely introduction of quarantine can effectively slow down spread of the infection.

Conclusions

The SIR (susceptible-infected-removed) model and statistical approach to the parameter identification provide a possibility for making reliable estimations for the epidemic dynamics, e.g., the real time of the outbreak, final size and duration of the epidemic and the number of persons spreading the infection versus time. This information may be useful to regulate the quarantine activities and to predict the medical and economic consequences of the pandemic. The pandemic outbreak probably occurred in China not later than the end of September 2019, it could continue beyond mid-March 2021, and the number of infected people in the world could exceed 5 million.

Acknowledgements

References

[1]   Coronavirus Disease (COVID-19) Situation Reports [Internet]. Who.int. 2020 [cited 2020 May 9]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

[2]   Kermack WO, McKendrick AG. A Contribution to the mathematical theory of epidemics. J Royal Stat Soc Ser A. 1927;115:700-21.

[3]   Murray JD. Mathematical biology I/II. New York: Springer; 2002.

[4]   Nesteruk I. Statistics-based predictions of coronavirus epidemic spreading in mainland China. Innov Biosyst Bioeng. 2020;4(1):13-8. DOI: 10.20535/ibb.2020.4.1.195074

[5]   Nesteruk I. Characteristics of coronavirus epidemic in mainland China estimated with the use of official data available after February 12, 2020. ResearchGate [Preprint] 2020. DOI: 10.13140/RG.2.2.19667.32804

[6]   Nesteruk I. Estimations of the coronavirus epidemic dynamics in South Korea with the use of SIR model. ResearchGate [Preprint] 2020. DOI: 10.13140/RG.2.2.15489.40807

[7]   Nesteruk I. Comparison of the coronavirus epidemic dynamics in Italy and mainland China MEDRXIV [Preprint] 2020. Available from: https://medrxiv.org/cgi/content/short/2020.03.18.20038133v1

[8]   Nesteruk I. Stabilization of the coronavirus pandemic in Italy and global prospects. MEDRXIV [Preprint] 2020. DOI: 10.1101/2020.03.28.20045898

[9]   Nesteruk I. Long-term predictions for COVID-19 pandemic dynamics in Ukraine, Austria and Italy. MEDRXIV [Preprint] 2020. DOI: 10.1101/2020.04.08.20058123

[10] Nesteruk I. SIR-simulation of Corona pandemic dynamics in Europe. MEDRXIV [Preprint] 2020. DOI: 10.1101/2020.04.22.20075135

[11] Nesteruk I. COVID-19 epidemic dynamics in Ukraine and Kyiv after testing has improved. ResearchGate [Preprint] 2020. DOI: 10.13140/RG.2.2.36705.86885

[12] Draper NR, Smith H. Applied regression analysis. 3rd ed. John Wiley; 1998.

[13] Statistical tables [Internet]. Onlinepubs.trb.org. 2020 [cited 2020 May 9]. Available from: https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf

[14] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England J Med. 2020;382:1199-207. DOI: 10.1056/NEJMoa2001316

[15] Italian doctors saw 'strange pneumonia' in Lombardy in November [Internet]. South China Morning Post. 2020 [cited 2020 May 9]. Available from: https://www.scmp.com/news/china/society/article/3076334/coronavirus-strange-pneumonia-seen-lombardy-november-leading

[16] GmbH F. Militärweltspiele in Wuhan: „Wir sind alle erkrankt" [Internet]. FAZ.NET. 2020 [cited 2020 May 9]. Available from: https://m.faz.net/aktuell/sport/mehr-sport/militaerweltspiele-2019-in-wuhan-damals-schon-corona-faelle-16758894.html

[17] Weinberger DM, Cohen T, Crawford F, Mostashari F, Olson D, Pitzer VE, et al. Estimating the early death toll of COVID-19 in the United States. MEDRXIV [Preprint] 2020. DOI: 10.1101/2020.04.15.20066431

[18] Trump 'considering quarantining New York' [Internet]. BBC News. 2020 [cited 2020 May 9]. Available from: https://www.bbc.com/news/world-us-canada-52079121

[19] Hong Kong takes emergency steps as mystery 'pneumonia' infects 27 in Wuhan [Internet]. South China Morning Post. 2020 [cited 2020 May 9]. Available from: https://www.scmp.com/news/china/politics/article/3044050/mystery-illness-hits-chinas-wuhan-city-nearly-30-hospitalised

І.Г. Нестерук

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ПАНДЕМІЇ COVID-19 ІЗ ВИКОРИСТАННЯМ SIR-МОДЕЛІ

Проблематика. Пандемія COVID-19 становить великий інтерес для дослідників через високу смертність і дуже негативний вплив на світову економіку. Детальний науковий аналіз цього явища ще попереду, але громадськість вже цікавиться питаннями тривалості епідемії, очікуваною кількістю хворих, де і коли почалася пандемія. Для правильного моделювання динаміки пандемії потрібні складні математичні моделі та багато зусиль для визначення невідомих параметрів. У цій статті будуть представлені, узагальнені та обговорені попередні оцінки для багатьох країн і світу в цілому.

Мета. Ми оцінимо характеристики епідемії для США, Німеччини, Великобританії, Республіки Корея й у світі за допомогою SIR-моделі та порівняємо їх з результатами, отриманими раніше для Італії, Іспанії, Франції, Республіки Молдова, України і Києва. Буде обговорено приховані періоди, тривалість епідемій, остаточну кількість випадків і карантинні заходи.

Методика реалізації. У цьому дослідженні ми використовуємо відому SIR (сприйнятливі–заражені–видалені) модель динаміки епідемії, відомий точний розв'язок лінійних диференціальних рівнянь і статистичний підхід, розроблений раніше.

Результати. Із застосуванням статистичного підходу визначено оптимальні значення параметрів SIR-моделі для динаміки епідемії у США, Німеччині, Великобританії, Республіці Корея та у світі. Розраховано часові залежності для фактичної кількості випадків та кількості хворих, що поширюють інфекцію. Оцінено приховані періоди, тривалість і остаточні розміри епідемій. Зокрема, пандемія почалася в Китаї не пізніше жовтня 2019 року. Якщо нинішні тенденції триватимуть, то кінця пандемії слід

очікувати не раніше березня 2021 року, загальна кількість випадків перевищить 5 мільйонів. Запропоновано простий метод оцінки ризиків передчасного ослаблення карантинних заходів.

Висновки. SIR-модель і статистичний підхід для ідентифікації параметрів допомагають зробити достовірні оцінки динаміки епідемії, наприклад реальний час спалаху, остаточний розмір і тривалість епідемії та кількість людей, що поширюють інфекцію, залежно від часу. Ця інформація буде корисною для регулювання карантинної діяльності та прогнозування медичних і економічних наслідків пандемії.

Ключові слова: пандемія коронавірусу; спалах епідемії; коронавірус COVID-19; коронавірус 2019-nCoV; математичне моделювання інфекційних захворювань; SIR-модель; ідентифікація параметрів; статистичні методи.

....

И.Г. Нестерук

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ПАНДЕМИИ COVID-19 С ИСПОЛЬЗОВАНИЕМ SIR-МОДЕЛИ

Проблематика. Пандемия COVID-19 представляет большой интерес для исследователей из-за высокой смертности и очень негативного влияния на мировую экономику. Подробный научный анализ этого явления еще впереди, но общественность уже интересуется вопросами продолжительности эпидемии, ожидаемого числа пациентов, где и когда началась пандемия. Правильное моделирование динамики пандемии требует сложных математических моделей и много усилий для идентификации неизвестных параметров. В этой статье будут представлены, обобщены и обсуждены предварительные оценки для многих стран и мира в целом.

Цель. Мы оценим характеристики эпидемии для США, Германии, Великобритании, Республики Корея и в мире с использованием SIR-модели и сравним их с результатами, полученными ранее для Италии, Испании, Франции, Республики Молдова, Украины и Киева. Будут обсуждены скрытые периоды, продолжительность эпидемии и карантинные меры.

Методика реализации. В этом исследовании мы используем известную SIR (восприимчивые–инфицированные–удаленные) модель для динамики эпидемии, известное точное решение линейных дифференциальных уравнений и статистический подход, разработанный ранее.

Результаты. С использованием статистического подхода определены оптимальные значения параметров SIR-модели для динамики эпидемии в США, Германии, Великобритании, Республике Корея и в мире. Рассчитаны фактическое количество случаев и количество пациентов, распространяющих инфекцию, в зависимости от времени. Были оценены скрытые периоды, длительности и окончательные размеры эпидемий. В частности, пандемия началась в Китае не позднее октября 2019 года. Если нынешние тенденции сохранятся, конец пандемии следует ожидать не ранее марта 2021 года, общее число случаев в мире превысит 5 миллионов. Предложен простой метод оценки риска преждевременного ослабления карантинных мер.

Выводы. SIR-модель и статистический подход для идентификации параметров помогают сделать достоверные оценки динамики эпидемии, например реального времени вспышки, окончательного размера и продолжительности эпидемии, а также числа людей, распространяющих инфекцию, в зависимости от времени. Эта информация может быть полезна для регулирования карантинной деятельности и прогнозирования медицинских и экономических последствий пандемии.

Ключевые слова: пандемия коронавируса; вспышка эпидемии; коронавирус COVID-19; коронавирус 2019-nCoV; математическое моделирование инфекционных заболеваний; SIR-модель; идентификация параметров; статистические методы.