Information

Dynamics of coronavirus spread in terms of chemical reaction kinetics

V. M. Goldberg

N. M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, 4 ul. Kosygina, 119334 Moscow, Russian Federation. E-mail: goldberg@sky.chph.ras.ru

The data on coronavirus spread in Russia (as a whole), Sweden, Moscow, and Yakutia were analyzed within the framework of kinetics of autocatalytic branching chain reactions. It was shown that in certain time intervals, the dynamics of this process can be described by a logistic function with a graph similar to the kinetic curve of product accumulation in a first-order autocatalytic reaction both in the consumption of the reactant and in the autocatalysis by the product. The logistic curve expression has a certain predictive ability and can be used to evaluate the effectiveness of anti-epidemic measures; it also poses the problem of elucidating the causes of unexpected abrupt changes in the logistic function parameters in the description of actual statistical data.

Key words: kinetics, chain reactions, logistic function.

The coronavirus disease pandemic is an enormous challenge for the present-day mankind. Quite a number of mathematical models have been proposed for describing the pandemic. One can get acquainted with the models by typing "mathematical modeling of COVID-19" in any search engine. As an example, see Ref. 1.

It is very important now to comprehend this phenomenon. Evidently, analysis of the phenomenon from different standpoints could be useful. This study is an attempt to consider the pandemic from the standpoint of chemical kinetics. We believe that this analysis would provide more in-depth understanding of the causes for the high rate and the enormous area of its spread in each country, and, for large countries such as Russia, also for particular regions. Furthermore, there is a certain probability to predict both the rate of spread and the scale of pandemic, which has an obvious practical value.

In terms of description of chemical processes, the increase in the number of infected people corresponds to a typical autocatalytic reaction, because every infected person (reaction product) is the source of infection for one or several persons. In this sense, every epidemic is a branching chain reaction. Since, in reality, this takes place over rather long time, among physicochemical processes, it should not be associated with a nuclear explosion, but rather with the consumption of a hydrocarbon during liquid-phase oxidation autocatalyzed by hydroperoxides, in which each hydroperoxide molecule results from the reaction of one peroxide radical, but is converted to two radicals, which are again converted to peroxide radicals.

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya*, No. 10, pp. 2022–2028, October, 2020. 1066-5285/20/6910-2022 © 2020 Springer Science+Business Media LLC It is known that such a process is called degenerate branching chain reaction. Evidently, the branching factor, which amounts to three for a thermonuclear fusion and two for the oxidation of hydrocarbons, is much greater for the pandemic, as one person can infect hundreds of other persons.

However, the obvious difference between the spread of a disease and a chemical reaction is that the conditions for the end of a chemical reaction are known: complete consumption of the reactant with its conversion to products. The amounts of the reactants and products can be specified in advance. For an epidemic, it is considered that the corresponding value, that is, the total number of people who will be infected during this disease outbreak in some region, cannot be determined in advance, because it depends on a large number of factors, some of which cannot be taken into account at all. The attempts to estimate this value are rather rough and, most often, it is unclear what scientific grounds served for estimation.

The basic concept behind this study is the assumption that a limited number of people are infected during every epidemic outbreak. Then, in combination with the obvious exponent in the initial stage, the infection kinetics should be a logistic or S-shaped function. This function has long been used to describe a variety of processes ranging from growth of microorganisms to growth of Earth's population.^{2,3}

It is known that an expression describing the rate of an autocatalytic reaction should have two multipliers, one responsible for the exponential growth and one responsible for growth retardation. The latter term is related to the decrease in the concentration of the starting reactant due to consumption. For a simplest reaction of this type, in which both the autocatalysis and the reaction rate are first order with respect to the reactant, the reactant consumption rate can be expressed by the equation

$$-d[A]_t/dt = k[A]_t([A]_0 - [A]_t),$$
(1)

where $[A]_0$ is the initial reactant concentration, $[A]_t$ is the current reactant concentration at time point *t*, *t* is time, and *k* is the rate constant of the reaction.

For the subsequent interpretation of experimental results, it should be borne in mind that in this case, the rate constant of the reaction has second order with respect to the reactant concentration.

This expression can be converted, without any problems, to an equation for the spread of the coronavirus infection:

$$d(N_t)/dt = kN_t(N_{\max} - N_t), \qquad (2)$$

where N_{max} is the maximum number of people that will be infected during the epidemic outbreak in a specified country or region, N_t is the number of infected people at time point t.

The integration of Eq. (2) for more convenient processing of experimental data leads to the following equation for linearization of statistical data:

$$\ln(N_t / (N_{\max} - N_t)) = k N_{\max} t + C,$$
(3)

where *C* is the integration constant.

It follows from Eq. (3) that

$$N_t = N_{\max}(\exp(kN_{\max}t + C)/(1 + (kN_{\max}t + C)))$$
(4)

As long as the $(kN_{max}t + C)$ value is much less than unity, it can be neglected, and function (4) is converted to a common exponent, which is observed in the beginning of any self-accelerating process, including autocatalytic chemical reactions.

Thus, in terms of Eq. (3), the growth of the number of infected people should be represented by a straight line with a slope of kN_{max} . A characteristic feature of this function is that it intersects the x axis at $N_t = 1/2N_{\text{max}}$. In terms of kinetics of a complex chemical reaction, kN_{max} is a parameter describing the general time variation, but providing no possibility to determine each of the values separately.

It follows from the above that the exact form of the logistic function suitable for calculating the number of people infected up to a particular time is determined by only one value, that is, $N_{\rm max}$, which is unknown in advance. It also cannot be determined from the time course of the infection spread.

For estimating this value, one should exactly know at which degree of conversion and how much the exponent will deviate from the logistic function.

Figure 1 shows the dimensionless functions ln(x/(1-x)) = k(t+3) (curve *I*) and lnx = k(t+3) (curve 2). The addition of some constant value (in this case 3) to the dimensionless time was due to the fact that half of the values of the logistic function (whose maximum value is 1) are negative, so their logarithms cannot be taken.

Figure 1 also shows the equation of the correlation straight line for the initial part of the logistic curve plotted on semilogarithmic coordinates and the correlation coefficient R^2 , which indicates how close the set of points is



Fig. 1. Dimensionless functions $\ln(x/(1-x)) = k(t+3)$ (curve *I*) and $\ln x = k(t+3)$ (curve *2*) (*x* varies from 0.05 to 0.95).

Table 1. Dependence of the correlation coefficients for the semilogarithmic transformation ofthe logistic function on the degree of completionof the process

Degree of completion (%)	Correlation coefficient
20	0.9994
25	0.9989
30	0.9981
35	0.9971
40	0.9959
45	0.9943

to a straight line. It can be seen from Fig. 1 that the values deviate noticeably from the correlation line when the degree of completion of the process is 40%. This corresponds to the correlation coefficient of 0.9943.

Table 1 presents the correlation coefficients for description of the initial part of the logistic curve by an exponent for different degrees of conversion of the process, which is actually described by a logistic function.

It can be seen that with increasing degree of completion of the process, the correlation coefficient of the semilogarithmic straight line derived from the initial part of the logistic function markedly decreases. The dependence presented in Table 1 can be used to estimate $N_{\rm max}$ starting from the point when ~30% of this value is attained. Certainly, this method is applicable when the data are relatively accurate, so that the correlation coefficients for semilogarithmic plots are rather high. In the case of the current coronavirus pandemic, this requirement is well met.

Figure 2 shows the semilogarithmic transformation of the dynamics of the number of people infected in Russia from April 3 to April 19, 2020. It can be seen that the experimental points are very well described by a linear function. The correlation coefficient is very high and



Fig. 2. Semilogarithmic transformation of the dynamics of coronavirus infection spread in Russia for the period from April 3 to April 19, 2020 (here and in Fig. 3 the *x* axis indicates days elapsed since the onset of coronavirus spread in Russia, March 3, 2020).



Fig. 3. Transformation of the dynamics of coronavirus infection spread in Russia from April 3 to April 29, 2020, in the $\ln[(N_t/(150000 - N_t)]$ —time (t) coordinates.

amounts to $R^2 = 0.9995$. By analysis of further reaction dynamics, the maximum number of people infected in Russia during this epidemic outbreak is estimated as $N_{\text{max}} = 150000$.

Figure 3 presents the corresponding transformation, from which it follows that the statistic data are described, with a high degree of correlation, by the logistic function with $N_{\text{max}} = 150000$. It is worth noting that these data represent the second way of using this function. The very beginning of the epidemic, from March 7 to April 4, was also perfectly described by a logistic function with $R^2 = 0.999$, but $N_{\text{max}} = 12500$, *i.e.*, an order of magnitude lower value. After April 4, the rate of growth of the number of infected cases sharply increased and was again described by a common exponent, as if it had been the very beginning of the epidemic. Obviously, the noted phenomenon should be an important point for thinking over and drawing conclusions for both people and organizations who develop anti-epidemic measures.

It seemed quite likely that no such breaks in the infection dynamics would occur in the future. Using Eq. (4), in which the exponent was taken from the correlation straight line, we calculated the dynamics of infection in Russia both for the past period (before April 29) and for the subsequent stage, *i.e.*, the hypothesized number of infected people in the future.

Figure 4 shows the curve reflecting the dynamics of growth of the number of infected cases in Russia (calculated data and the real statistical values). The calculation was done on April 29, 2020. The subsequent values were derived from the logistic function with $N_{\text{max}} = 150000$. According to this calculation, by May 21, 2020, approximately 148000 infected people were expected in Russia, which would be equal to 98.7% of the maximum number. The daily increase in the infected cases expected by this time was 360, which is almost 30 times lower than the number of cases observed in reality (10000).



Fig. 4. Calculated (1) and real (2) number of coronavirus cases in Russia.

However, the reality disproved these forecasts. After April 30, the story of sharp acceleration and initial exponent happened again. A new N_{max} value equal to 400000 people was determined by the above-described procedure. Figure 5 depicts the coronavirus spread dynamics with the new found N_{max} value. It can be seen that this stage of growth of the number of coronavirus-infected cases in Russia from April 30 to May 20, 2020, obeys the logistic function with the correlation coefficient $R^2 = 0.9999$. Nevertheless, the latest data on the number of infected people in Russia suggest that, despite this high R^2 , it will again be necessary to adjust N_{max} to higher values. Indeed, a new acceleration stage started on May 21.

The latest data on the coronavirus spread in Russia, known by the time the paper was submitted, indicated that



Fig. 5. Transformation of the dynamics of coronavirus infection spread in Russia from April 30 to May 20, 2020, in the $\ln[N_t/(40000 - N_t)]$ —time (t) coordinates.



Fig. 6. Calculated (1) and real (2) dynamics of coronavirus spread in Russia.

the number of the infected people corresponds to the logistic function with N_{max} of 550000. This can be used to calculate the time curve for the subsequent spread of the coronavirus (Fig. 6).

According to this calculation, the rate of infection will decrease by July 2, 2020, to 1446 cases per day, which is almost 30 times lower than the maximum value. A different matter is whether or not this would be the end of the epidemic.

For the accurate calculation of the time dependence of the number of people that may be infected during this epidemic outbreak, an important issue is the maximum rate of infection, which is called "the peak of epidemic" in mass media and in public statements of administrative officials.

Figure 7 shows the real daily increase in the number of infected cases in Russia and the data calculated from the function that was used to derive the data shown in



Fig. 7. Comparison of real (1) and calculated (2) infection rates in Russia for the highest infection rate period.

Figs 3 and 5 and that perfectly describes the real situation in the indicated time range.

It can be seen that the calculated maximum value is approximately 9343 people per day. Using the real data (see Fig. 7, curve 1) on the daily increase in the number of infected people, it is difficult to determine the maximum rate and the exact time when the maximum rate was attained. Only from the calculated data can one understand that the maximum rate was reached on May 17.

It is noteworthy that in reality, the maximum infection rate was reached on May 11 and amounted to 11656 people. Meanwhile, the maximum of the logistic function, which perfectly describes the time course of epidemics, fell on May 17. It seems that, when choosing a method for determining these two values, the maximum rate and the time when it is reached, which are of obvious practical importance, one should prefer the theoretical calculation, which is seen to be free from statistical scattering inherent in real data.

The shape of approaching both the theoretical and real maximum infection rates can by no means be called a peak; the more so, there is no sense to speak about "smoothing" or "preventing the explosive nature", as was stated in mass media. This maximum is always "smooth", irrespective of the presence or absence of any measures. The phrase "we managed to avoid the explosive peak of infection" has no real content either. Thus, an elementary logistic function, analogous to the function for the kinetics of autocatalytic first-order reaction, describes satisfactorily the dynamics of spread of coronavirus in Russia in four separate parts with the maximum numbers of infected cases being 12.5, 150, 400, and 550 thousand people, respectively. The causes for this sharp increase in the infection rates are obviously beyond the scope of kinetic analysis. However, it can be assumed that the sharp increase in the growth rate at the end of April is attributable to the large gatherings of people at the entrances of metro stations in Moscow on April 15 due to the check of passes, with the situation in Moscow strongly influencing the all-Russian statistics.

The fact that the logistic function of coronavirus infection in Russia has several sharp breaks attests in favor of the analogy with branching chain reactions. These reactions are subject to so-called "critical phenomena" in which the reaction rate sharply changes as a result of rather small changes in the reaction conditions. In the case of nuclear reactions, a possible factor is the weight (actually, the size) of the piece of uranium; in the case of gas-phase phosphorus oxidation, this is the size of the reaction vessel; and in the case of liquid-phase oxidation of hydrocarbons, this is the concentration of free radical inhibitor. The following analogy applies to an epidemic: after completion of every next stage of correspondence of the infection dynamics to a particular N_{max} in the logistic function, an explosion, that is an abrupt increase in N_{max} , takes place as if two halves of an uranium sphere were put together.

1.6 y = 0.0741x - 4.6594 $R^2 = 0.9995$ 1.4 1.2 • $\ln[N_t/(210000 - N_t)]$ 1.00.8 Linear $\ln[N_t/(210000 - N_t)]$ 0.6 74 76 78 80 72 82 84 t/days

 $\ln[N_t/(210000 - N_t)]$

1.8

Fig. 8. Transformation of the dynamics of coronavirus infection spread in Moscow for the period from May 15 to May 28, 2020, in the $\ln[N_t/(210000 - N_t)]$ —time (*t*) coordinates.

As has already been noted, the accepted model has a predictive power. The detected deviations of the exponent from the logistic function, which allow the prediction of the further course of the infection, start at a 20-30% degree of completion of the process, *i.e.*, at a relatively early stage.

The situation in Moscow exactly reproduced the growth dynamics in Russia with a factor of 1/2. However, for describing the recent data, it is necessary to reduce N_{max} to 210 thousand people. According to the data shown in Fig. 8, the correlation coefficient for this N_{max} is rather high ($R^2 = 0.9999$).

It was also of interest to analyze the situation for one of remote regions of Russia with the lowest population density, namely, the Republic of Sakha (Yakutia).

By March 19, seven cases of coronavirus infection were confirmed in Yakutia. The first point of the plot in Fig. 9 corresponds to April 11 when the number of cases increased to 19. The gap between the points in the plot corresponds



Fig. 9. Transformation of the dynamics of coronavirus infection spread in Yakutia for the period from March 18 to May 20, 2020, in the $\ln[N_t/(1550 - N_t)]$ —time (t) coordinates.

to the period of time from April 18 to 27 when the number of cases increased from 47 to 88 in only one day. This may be attributable to arrival of rotating shift workers. Later, the situation started to approach the primary logistic function. It can be seen that up to May 21, 2020, there are no breaks in the plot, similar to those inherent in the dynamics of infection spread in Russia as a whole and in Moscow, in particular. The correlation coefficient for the logistic function is, in this case, relatively high: $R^2 = 0.9993$.

Attention is drawn by the fact that in these absolutely different regions, the parameter describing the time course of the spread of coronavirus infection is approximately the same (the number of cases per day): 0.14 in Moscow, 0.11 in Russia (as a whole), and 0.13 in Yakutia. It is also possible to estimate the real rate constants for the spread of coronavirus in these regions. For this purpose, we will divide the slope of the linearized function by the maximum number of people that can be infected in this period of time in a particular region. Then the rate constants are $4.1 \cdot 10^{-7}$, $8 \cdot 10^{-7}$, and $8.7 \cdot 10^{-5}$ ((day)⁻¹ (person)⁻¹) for Russia (as a whole), Moscow, and Yakutia, respectively. The specific rate of coronavirus spread in Moscow is somewhat higher than in Russia as a whole, since the population density and, hence, the number of contacts are, on average, greater in Moscow than over the whole country. However, in Yakutia, in which the population density is much lower than in Moscow, the specific rate of coronavirus spread is almost 100-fold higher. Probably, the population density is only one of the factors and its influence on the rate of infection spread is not crucial in comparison with other factors.

This work is mainly devoted to the situation in Russia. However, for the kinetic analysis of coronavirus spread, it is of interest to consider the situation in Sweden, because the quarantine restrictions of contacts and people mobility in Sweden were very weak.

Figure 10 shows the transformation of the dynamics of coronavirus spread in Sweden. It can be seen that description of coronavirus spread in this country is in poorer agreement with the elementary logistic function than the description of separate time periods for Russia (as a whole) or Moscow. However, note that for the above-mentioned regions, we did not consider long time periods such as the period considered for Sweden (more than two months), which covers virtually the whole period of development of coronavirus infection in this country.

The noticeable deviation of recent data for the number of cases attests to retardation rather than acceleration of the infection spread, that is, the decrease in the infection rate in comparison with the previous parameter $N_{\rm max}$, although in this case, too, the end of the epidemic is postponed and $N_{\rm max}$ increases.

The period from March 5 to March 13 in which the number of cases increased from 35 to 814 was excluded from consideration of the coronavirus spread in Sweden.



Fig. 10. Transformation of the dynamics of coronavirus infection spread in Sweden for the period from March 10 to May 20, 2020, in the $\ln[N_t/(36000 - N_t)]$ —time (*t*) coordinates.

The results of semilogarithmic transformation of these data for the initial stage of coronavirus epidemic in Sweden indicate (Fig. 11) that in the very beginning, growth acceleration was approximately the same as in Moscow and in Russia as a whole, with the number of cases being doubled between two and three days.

A comparison of the kinetic parameters of coronavirus spread shows that the spread parameter is more than two times lower for Sweden than for Russia almost throughout the whole epidemic (in the last section, $kN_{max} = 0.079$ and 0.164 (day)⁻¹, respectively). As regards the constant k, for Russia as a whole it is $4.1 \cdot 10^{-7}$ (Yakutia is an exception, $8.7 \cdot 10^{-5}$), while for Sweden it is approximately 50 times higher: $2.2 \cdot 10^{-6}$. It is for virologists and politicians to decide which value is "better" and which is "worse".

Thus, in this paper, the elementary logistic function is proposed for analysis of the spread of coronavirus infection. A procedure is described for determining the maxi-



Fig. 11. Semilogarithmic transformation of the coronavirus spread in the beginning of the epidemic in Sweden (March 5 to March 13, 2020) in the $\ln(N_t)$ —time (*t*) coordinates.

mum number of people $N_{\rm max}$ that can be infected with coronavirus during a specified epidemic outbreak in a specified region and in a country as a whole. This function has a predictive power for describing the number of people infected with coronavirus. The deviations of the real situation from the prediction can be used to draw conclusions about the effectiveness of measures on limiting the epidemic spread and events of breaking the restrictions; they also provide quick information about sharp changes in the rate of growth of infected cases.

References

- 1. https://ddi.sutd.edu.sg/.
- S. D. Varfolomeev, K. G. Gurevich, *Biokinetika (Prakticheskii kurs)* (Biokinetics. Practical Course), FAIR-PRESS, Moscow, 1999, 557 pp. (in Russian).
- 3. S. D. Varfolomeyev, K. G. Gurevich, *J. Theoretical Biology*, 2001, **212**, 367–372.

Received June 5, 2020; in revised form July 27, 2020; accepted July 29, 2020