Agent-Based Simulation of the COVID-19 Epidemic in Russia

G. N. Rykovannov, S. N. Lebedev, O. V. Zatsepin, G. D. Kaminskii, E. V. Karamov, A. A. Romanyukha, A. M. Feigin, and B. N. Chetverushkin

A consortium of scientists (RFNC‒VNIITF, RAS IAP, RAS KIAM, RAS INM, NRCEM, NMRC PPI) is working on the creation of new mathematical models and a set of programs for modeling the development of epidemics of viral infections. This will make it possible to obtain a forecast of the effectiveness of regulatory measures for informed decision making on the part of government bodies to control the burden on medical institutions. The multidisciplinary team includes physicists, mathematicians, physicians,

Received February 8, 2022; revised February 20, 2022; accepted March 10, 2022

Abstract—The COVID-19 pandemic has created a public health emergency in Russia and across the world. The wavelike spread of the new coronavirus infection, caused by newly emerging variants of the coronavirus, has led to a high incidence rate in all subjects of the Russian Federation. It is becoming extremely topical to get the opportunity to manage the development of the epidemic and assess the impact of certain regulatory measures on this process. This will help government agencies make informed decisions to control the burden on healthcare organizations. It is often impossible to obtain such assessments without using modern mathematical models.

Keywords: COVID-19 pandemic, mathematical modeling, forecast for the development of the COVID-19 epidemic

DOI: 10.1134/S1019331622040219
virologists, and epidemiologists. They work within the framework of the project of the Russian Ministry of Education and Science “Modeling of Epidemics of Viral Infections” under the subprogram Basic Research for the Long-Term Development and Competitiveness of Society and the State of the state program Scientific and Technological Development of the Russian Federation. The topicality of this work is due to the lack of systems applicable in practice and adapted to Russian conditions for predicting the spread of viral infections and the response of the epidemic process to various control measures.

The development of an epidemic is, in essence, a chain reaction. Chain reactions in chemistry and nuclear physics are well studied and are described by equations close to those characterizing the development of an epidemic. However, there is a significant difference between these processes. As for nuclear chain reactions that occur during the operation of systems traditionally handled by the Federal Nuclear Center, everything is strictly determined. The neutron interaction cross sections and the geometry of the system are well known. This makes it possible to describe accurately the quantitative characteristics and effectiveness of nuclear system control measures. There are no such data for viral infections. As a rule, the characteristics of the infection and the parameters of the disease course are known with some degree of uncertainty and can change during the development of the epidemic. For new infections that humanity encounters for the first time, research and characterization of the disease take time. For some infections, the immune state of the population is unknown; in addition, the effectiveness of the fight against infection depends on the behavior of individuals. Therefore, a long-term accurate description of the development of an epidemic is hardly possible.

The question may arise: Is mathematical modeling of epidemics necessary in this case? Indeed, in two limiting cases (low and high lethality of the disease), the benefit of mathematical modeling is rather low. In the former case, the epidemic ends when the required value of herd immunity is achieved; in the latter, the methods of control are well developed—strict localization of possible foci of infection. In this way, in about 70 days, the outbreak of COVID-19 in Wuhan (China, Hubei Province) was eliminated [1]. Another example of a successful fight against a particularly dangerous disease is the liquidation of the outbreak of smallpox in Moscow in 1960 [2], when the infection was stopped within 19 days from the moment of its identification. Note that mass vaccination of almost ten million residents of Moscow and Moscow oblast was carried out in the shortest possible time.

The role of mathematical modeling increases if it is supposed to control the process of epidemic development. This requires a tool to predict the effectiveness of planned restrictive measures and the burden on the healthcare system. It is this, and not the accuracy of the description of actual data on morbidity, that is the main task of mathematical modeling.

Forecasting is necessary to determine, in particular, the effectiveness of the influence of the detection and isolation of sources of infection, as well as that of treatment, vaccination, and prevention. Note that a mathematical model should show results close to statistical data throughout the entire period of the disease to provide the most reliable forecast for one scenario or another of the introduction of regulatory measures. Modeling should take into account not only the very fact of possible infection of the susceptible part of the population but also the loss of health of the uninfected part of the population due to the current epidemic. In addition to the consequences of the infectious disease itself, it is necessary to consider the loss of the population due to both exacerbations of chronic diseases caused by violations of psychological well-being in society (anxiety, depression) and untimely provision of medical care because of the overload of the healthcare system.

Traditionally, epidemics are described using models derived from the SIR model proposed by Kermack and McKendrick in 1927 [3]. The solution of the problem obtained using these models can be compared with the solution of the kinetic equations for a homogeneous medium, for example, a homogeneous nuclear reactor. To develop effective measures to counteract viral infections, this approach is not enough. Having extensive experience in numerical methods for solving the neutron transport equation, the authors tried to apply these methods to describe the development of an epidemic. One of them, the Monte Carlo method [4], can easily be adapted to simulate a viral infection. Using this approach, an agent-based model of the spread of a viral infection was developed [5–7]. Note that agent-based models are currently widely used to analyze various aspects of the spread of COVID-19 and predict the development of the pandemic [8–12].

The agent-based model can be viewed as a kind of analogue of the method for solving the neutron transport equation for a heterogeneous medium in the multigroup approximation. The population is divided into characteristic groups according to the type of daily behavior—employees of enterprises, offices, and the life support sphere; school students; preschoolers; university students; and pensioners. Each group has its own daily life cycle—the sequence of places of stay and the time interval in which they are there. The model belongs to the class of individual-based models [13, 14] and makes it possible to account for the complex demographic structure, the characteristics of the spread and course of diseases, medical statistics, and the system of contacts of urban residents.

The city model is based on the following:

- description of the city-specific distribution of residents by social status (social groups, age, gender) and by visiting public places (cells)—centers of infec-
tion, including houses (apartments), kindergartens, schools, workplaces, places of recreation, transport, etc.;

• the breakdown of society into a set of social groups united by visiting cells that form identifiable clusters through which infection is possible;

• description of a set of random trajectories assigned to each individual (agent) in accordance with a given scenario of the day; there can be several scenarios for visiting cells, which allows one to build various modes of counteracting the spread of the epidemic (self-isolation mode, quarantine, complete isolation for a certain population group);

• describing epidemic characteristics such as the degree of contagiousness, the radius of infection, the probabilistic characteristics of the transition from one state of an individual to another, the phases of the disease, and the time distributions of being in one phase or another;

• the statistical nature of population changes during the development of an epidemic, which makes it possible to evaluate not only the actual parameters of the population over time but also the statistical characteristics: the probabilities of different scenarios for the development of the epidemic, the average number of patients over the ensemble of populations depending on time, the average statistical spread, and other parameters that allow statistical analysis.

The urban environment is approximately described in the form of cells in which contact is possible, leading to infection of the agent: an apartment, three types of transport (metro, bus, fixed-route taxi), a place of work or study, public places, or three types of stores (hypermarkets, supermarkets, convenience stores). The characteristics of the cells (primarily the area per person), established according to statistical data for Moscow, turned out to be the same for other large cities of Russia. The probability of infection in a cell is a function of the area of the cell, the number of infected people, and the total number of people in the cell.

For each agent, a time budget for one day is defined—the time during which the agent interacts with other agents in a given set of cells. Each cell describes an apartment (house), a place of work or study, transport, shops, and shopping centers. Groups of residents differ in ways of interaction, which are characterized by different algorithms for moving from one cell to another. For example, the dynamics of the behavior of employees of enterprises and offices is described by the algorithm “home—transport—work—transport—shop—home.” The dynamics of the behavior of pensioners corresponds to the algorithm “home—transport—shop—transport—home.” School students and preschoolers do not use the “transport” and “shop” cells.

The pattern of the course of the COVID-19 disease in the model is shown in Fig. 1. The duration of the disease phases changed repeatedly as factual data accumulated. One of the modeling parameters is the percentage of “hidden” patients. This parameter has not been explicitly determined by any of the researchers; the data on it are radically different. Thus, at the beginning of our work in March—April 2020, the proportion of latent patients varied from 1%, according to the Russian Ministry of Defense [15], to 25–40%, according to the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor) [16]. With such a difference, it simply does not make sense to talk about the accuracy of predicting the burden on the healthcare system and the development of the epidemic. This parameter was estimated based on data from testing for coronavirus in New York and Moscow conducted in May 2020 and subsequently corrected due to the reprocessing of statistical data. The duration of the COVID-19 phases was taken from the results of medical studies. The model assumes that the average latent period (when a person is infected but not contagious) lasts 3 days; the subclinical one (when the patient is contagious but does not show symptoms of the disease), 2 days; for latent patients, the period when they can infect others, is 4 days; for outpatients who pose a danger to their families, the isolation period is 10 days; hospitalized patients stay in the hospital for an average of 14 days. The duration of immunity to the same strain in case of a previous disease is 180 days. When the strain changes, the average duration of immunity is reduced to 120 days. In vaccinated individuals, the duration of immunity for all strains is 180 days.

To obtain a reliable forecast of the development of an epidemic situation, a mathematical model should show results close to statistical data. Any approximate model needs to be calibrated against factual data. Such
A calibration was carried out in two megacities—Wuhan and New York, which are close in terms of population structure to Moscow. Incidence data for these cities were taken from the website of the Johns Hopkins University [1].

Most residents of Wuhan were isolated in their places of residence, identified patients were placed in hospitals, and residents with suspected illness were isolated in so-called health centers. Note that the actions of the authorities were the same for all groups of the population; for modeling, such an urban population is homogeneous. In this case, the influence of hidden patients on the development of the epidemic is small, and the modeling relies largely on the duration of the disease phases. The development of the epidemic in Wuhan was described in a wide range of changes in the proportion of hidden patients—from 50 to 90%. Figure 2 shows the simulation results. There is a leap in the factual incidence data, which indicates both the possibility of an error in the current information on morbidity and mortality, including in other countries, and a possible change in the approach to registering the number of cases.

The development of the epidemic in New York in the spring of 2020, where limited quarantine measures were introduced, could only be described assuming a proportion of hidden patients in the range of 70—90%. This value significantly exceeded the Rospotrebnadzor estimate, 25—40%, which raised doubts about the correctness of the model. In mid-May 2020, factual data appeared on the presence of antibodies in New Yorkers in a sample of 15000 people (Table 1). Comparison with the factual incidence rate made it possible to estimate the proportion of hidden patients at 90%. Considering possible errors in the factual data, an average value of the proportion of hidden patients of 80% was chosen for further modeling. Figure 3 compares simulations of early New York epidemic development with factual data.

**Table 1.** Hidden case share data in New York as of April 30—May 10, 2020

| Borough       | Positive antibody tests, % | Infected per 100 000 people | Registered cases per 100 000 people | Share of hidden carriers, % |
|---------------|-----------------------------|------------------------------|------------------------------------|-----------------------------|
| Bronx         | 27.6                        | 27600                        | 2624                               | 90.5                        |
| Staten Island | 19.2                        | 19200                        | 2407                               | 87.5                        |
| Queens        | 18.4                        | 18400                        | 2093                               | 88.6                        |
| Brooklyn      | 19.2                        | 19200                        | 1654                               | 91.4                        |
| Manhattan     | 17.3                        | 17300                        | 1111                               | 93.6                        |
| Average       | 20                          | 20000                        | 1943                               | 90.3                        |
Why can the results of mathematical modeling be useful? The agent-based model makes it possible to reveal the probability of infection through various channels and the change in this probability over time. The simulation showed that the main channels of infection at the beginning of the epidemic were transport (35% of the total) and places of work (25–30%). After about two months of the epidemic, the probability of infection at work decreased, and the probability of infection at home increased to the level of infection at work (~22%). This means that the restrictions themselves, that is, the measures to control the epidemic, may depend on time.

Simulation results make it possible to quantify the impact of the measures used both in aggregate and across various channels. With the introduction of restrictions on visiting public places for various groups of the population, the pace of development of the epidemic changes (Table 2). Thus, restrictions for pensioners change the burden on the healthcare system by about 10%. At the same time, the transfer of the working population, except for workers in the life support sphere of the city, to a remote mode reduces the maximum value of the number of patients in the hospital by almost three times. The modeling of restrictions on visiting public places (Table 3) showed the weak effect of store closures on the reduction in incidence (20%) but confirmed the significant role of public transport in the spread of the infection. Such a hypothetical measure as the widespread closure of public transport would lead to a twofold reduction in the burden on medical institutions.

The model allows for a more detailed description of the urban infrastructure. You can enter “special points” with a high probability of infection, for example, polyclinics, and evaluate their role in the spread of infection. As mentioned above, the main task of mathematical modeling of an epidemic is to quantify the function of the system’s response to external influences (restrictions). However, the degree of confidence in the results still depends on the accuracy of simulation of development of the epidemic. Let us illustrate the accuracy of the description of the development of the epidemic with two examples.

The first one is the replacement of the dominant alpha strain (B.1.1.7) with the delta strain (B.1.617.2, AY.1, AY.2). According to UK Department of Health and Social Care Public Health Technical Briefing

![Fig. 3. Simulation of COVID-19 epidemic in New York in the spring of 2020.](image_url)

| Table 2. Impact of the introduction of quarantine for certain categories of the population on the burden of the medical system |
|----------------------------------|----------------|-------------------|-------------------|
| Category of the population      | Pensioners | Pensioners and school students | Pensioners, school students, and university students | Working population (except life support) |
| Coefficient of the reduction in the burden on the medical system | 1.1 | 1.2 | 1.4 | 3 |
no. 17 of June 25, 2021, sequencing results indicate that the replacement of the alpha strain with the delta strain in Britain occurred in approximately 70 days. Upon introducing into the model the possibility of the simultaneous presence of two different strains in the population, it was shown that, in the calculations, the replacement of the strain (B1) prevailing in Moscow occurred approximately during the same period (Fig. 4). Later the duration of the replacement of the original strain with the delta strain in Moscow was confirmed by studies [17]. The contagiousness of the delta strain was assumed according to the information provided by the Gamaleya NRCEM and NMRC PPI virologists and was 2 times higher than the original strain.

The second example is the duration of the forecast. Since the model approximately describes the heterogeneous structure of Moscow, and the behavior of people during the epidemic changes and may differ from the lifestyle assumed by the model, the model constants for describing the factual data have to be periodically recalibrated. Figure 4 shows that the forecast calculated after calibration for August 1, 2021, rather accurately describes the factual data for three months. The introduction of quarantine from the end of October 2021 was not yet expected at that time and, hence, was not considered in the calculation. In October 2021, Russian President V.V. Putin announced the introduction of restrictive measures in connection with the complicated epidemic situation. There was a need to obtain a forecast that would account for such measures. The calculations not only considered the quarantine in the period from October 28 to November 7, 2021, but also assumed that a similar quarantine would be in effect in the upcoming New Year (December 31, 2021—January 9, 2022) and May (April 30, 2022—May 8, 2022) holidays.

Figure 5 shows the forecast for the incidence in Moscow with factual data on the vaccination rate as of November 17, 2021. The same graph shows the calculation with the vaccination rate increased to 50000 people per day, indicating the suppression of the development of the epidemic. The emergence of a new strain can drastically change the incidence pattern, as already happened with the emergence of the delta strain.

Before switching to simulating the development of the epidemic in the country, it was necessary to generalize the model for the metropolis to the region considering the impact of commuting labor migration. As an example, the commuting migration for Moscow oblast was analyzed. According to a study by Higher School of Economics [18], the agglomeration of Moscow and Moscow oblast amounts to about 16.7 million people. Daily labor migration amounts to 1.3 million. The modeling considered cities with a population of more than 100000 inhabitants, in which a total of 3.7 million people live. The factual data are described with a sufficient degree of accuracy. The morbidity maxima correlate with the data obtained in Moscow, the lag of the peaks being no more than 2–3 days.

Figure 4. Change of strains of the SARS-CoV-2 virus in Moscow. Calculation of the forecast of August 1, 2021.

| Restrictions (closed) | Shops | Transport |
|-----------------------|-------|-----------|
| Coefficient of the reduction in the burden on the medical system | 1.2   | 2         |

Table 3. Impact of the closure of certain types of public places on the burden of the medical system.
A model has been constructed to obtain numerical estimates of the spread of infection throughout Russia. The main idea of this model is to present the country as a set of settlements with local epidemics, interconnected by traffic flows. The country model is a combination of a set of city models and a matrix of population flows defined for each day. Each city model from the set describes a locality in the Russian Federation and has its own independent parameters. To simulate the development of the epidemic in the Russian Federation, the model was supplemented with transport links; three communication schemes were implemented.

The first scheme considered air and rail links between major transport hubs. Passenger traffic was modeled with account for the actual flight and trip schedule. The amount of passenger traffic and occupancy were taken in accordance with the annual reports of Russian Railways and the Aeroflot Group. Seasonal fluctuations in passenger traffic were taken into account. For example, the minimum number of passengers transported by air, according to statistics, is observed from mid-January to mid-March, and the maximum traffic occurs from July to the first half of September [19]. The second scheme includes transportation that ensures interactions of regional centers with their nearest neighbors. These include transportation by passenger, rail, and road transport and small aircraft. The third scheme covers suburban communication between the settlements of the agglomeration and its center. This is mainly personal transport, suburban bus service, and suburban electric trains.

The use of the schemes individually or in combination makes it possible to model the development of the epidemic situation in the country at various levels of detail. Figure 6 presents the first results of modeling the development of the epidemic in Russia. All calculations considered the actual quarantine restrictions introduced in different regions of the country. The first calculation covers cities with a population of more than one million people. It takes into account only large traffic flows. In the second calculation, large regional centers with a population of more than 500,000 people are added. This increased the coverage of the population from 33 to 46 million people. Finally, the third calculation considers cities with a population of at least 100,000 people, in which a total of 73 million people live. The simulation results shown in Fig. 6 are normalized to the total population of the country, 146 million people. Thus far, the results do not fully describe the factual data, but they are approaching them as the accuracy of describing the urban structure of regions and the country increases.

In conclusion, let us note that within the framework of the project of the Russian Ministry of Education and Science, a model and a set of programs are being developed to predict various scenarios for the development of epidemics of viral infections on a national scale. A more correct description of the development of epidemics requires considering the impact of restrictive measures on the complex state of society in the economic and social spheres. It is necessary to learn how to predict population losses caused by both exacerbations of chronic diseases and untimely medical care. Predicting the realistic vari-
ability of public sentiment during an epidemic based on artificial intelligence methods and a more accurate description of society through the development of databases in the field of virology, sociology, medicine, aviation, and rail and road transport can improve the predictive capabilities of modeling.

By the Presidential Decree of October 12, 2020, the Interdepartmental Commission of the Security Council of the Russian Federation was established on the creation of a national system of protection against new infections. Forecasting the pace of development of epidemics and the effectiveness of countermeasures taken should be an integral part of the national system of protection against new infections. The developed model and computational complex can become a prototype of a system for predicting the development of epidemics in Russia.

ACKNOWLEDGMENTS

The authors are grateful to the RFNC–VNIITF associates A.A. Bragin, V.V. Vlasov, A.M. Deryabin, A.L. Karmanov, S.I. Samarín, A.V. Sokolov, N.A. Solomin, N.A. Teplykh, M.S. Urakov, and K.E. Khatuntsev; the RAS IAP associates D.N. Mukhin and E.M. Loskutov; and the RAS KIAM associate V.P. Osipov for discussion of the work and the materials provided.

FUNDING

This work was supported by the Ministry of Science and Education of Russia within Agreement no. 075-11-2020-011 of October 19, 2020 (IGK 0000000007520RHT0002).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19. Cited May 5, 2020.
2. O. V. Baroyan and A. F. Serenko, “Smallpox outbreak in Moscow in 1959–1960,” Zh. Mikrobiol. Epidemiol. Immunol., No. 4, 72–79 (1961).
3. W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” Proc. R. Stat. Soc. 115, 700–721 (1927).
4. O. V. Zatsepin, Ya. Z. Kandiev, E. A. Kashaeva, et al., “Calculations for the VVER-1000 core by the Monte Carlo method implemented in the prizma code,” VANT, Ser. Fiz. Yadern. Reakt., No. 4, 64–73 (2011).
5. V. A. Adarchenko, S. A. Baban’; A. A. Bragin, et al., Modeling the Development of the Coronavirus Epidemic Using Differential and Statistical Models (RFYaTs–VNIITF, Snezhinsk, 2020) [in Russian].
6. O. V. Zatsepin, A. A. Bragin, V. V. Vlasov, et al., “Agent-based model of the development of the COVID-19 epidemic,” in Zababakhin’s Scientific Readings: Collection of Report Abstracts at the XV International Conference, September 27–October 1, 2021 (RFYaTs–VNIITF, Snezhinsk, 2021) [in Russian].
7. A. M. Deryabin, A. A. Bragin, V. V. Vlasov, et al., “Accounting for the movement of the population in the agent model of the development of the epidemic of a viral infection,” in Zababakhin’s Scientific Readings: Col-
8. C. C. Kerr, R. M. Stuart, D. Mistry, et al., “Covasim: An agent-based model of COVID-19 dynamics and interventions,” PLOS Comput. Biol. 17 (7), e1009149 (2021).
9. J. Li and P. Giabbanelli, “Returning to a normal life via COVID-19 vaccines in the United States: A large-scale agent-based simulation study,” JMIR Med. Inf. 9 (4), e27419 (2021).
10. R. Hinch, W. J. Probert, A. Nurtay, et al., “Open ABM-COVID19: An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing,” PLoS Comput. Biol. 17 (7), e1009146 (2021).
11. S. Howick, D. McLafferty, G. H. Anderson, et al., “Evaluating intervention strategies in controlling coronavirus disease 2019 (COVID-19) spread in care homes: An agent-based model,” Infect. Control Hosp. Epidemiol. 42 (9), 1060–1070 (2021).
12. F. Lorig, E. Johansson, and P. Davidsson, “Agent-based social simulation of the covid-19 pandemic: A systematic review,” J. Artif. Soc. Soc. Simul. 24 (3), 5 (2021). http://jasss.soc.surrey.ac.uk/24/3/5.html. Cited July 18, 2021.
13. V. Grimm and S. F. Railsback, Individual-Based Modeling and Ecology (Princeton Univ. Press, Princeton, 2005).
14. M. A. Kornilina and V. D. Perminov, “Individual-based model for simulation of urban epidemic,” Mat. Model. 19 (5), 116–127 (2007).
15. A. A. Sergoventsev, I. I. Azarov, M. B. Patsenko, et al., “Diagnosis, treatment, and prevention of novel coronavirus infection (COVID-19),” in Kirov Military Medical Academy: Materials on Coronavirus Infection. https://www.vmeda.org/wp-content/uploads/2020/04/covid-190406.pdf
16. Report of the Head of Rospotrebnadzor Anna Popova at a meeting with Putin. https://iz.ru/1005598/video/doklad-glavy-rospotrebndzora-anny-popovo-nasoveshchani-s-putinym. Cited April 28, 2020.
17. Overview of COVID-19 variants in countries. https://covariants.org/per-country. Cited July 6, 2021.
18. A. Makhrova and A. Bochkarev, “Commuting in Moscow region: New data.” http://demoscope.ru/weekly/2017/0727/tema01.php. Cited September 6, 2021.
19. PJSC Aeroflot 9M 2021 IFRS Financial Results. https://www.aviastat.ru/reports/116-finansovye-rezultaty-pao-aeroflot-za-9-mesyacev-2021-goda-po-msfo. Cited November 10, 2021.

Translated by B. Alekseev