Agent-based modeling in epidemiology of airborne infections

Tatiana E. Sannikova[1,*]
Marchuk Institute of Numerical Mathematics, Russian Academy of Science, Moscow, Russia[1]
SannikovaTE@yandex.ru

Abstract: *Agent-based modeling proved to be a powerful tool for studying the complex multifactorial processes that take place in human population. In this paper we describe how this approach can be used to study the spread of airborne infections in a big city and the ways to control them. Agent-based modeling includes three main stages: a creating a synthetic population, simulation of disease spread in synthetic population during a fixed time period, and an analysis of the results. We created the population of 10 million agents and united them in a complex network according to their individual characteristics such as age, sex, marital status and occupation. In addition, each agent has a property that characterizes its state of health: susceptible, infected, and recovered. A susceptible agent can become infected if has an infected agent among its contacts and if an event of disease transmission occurs. Disease transmission is simulated as a random event with probability p which calculated for every pair susceptible-infected and depends on their individual characteristics and on the length of the contact. In so doing, the heterogeneity in a number of contacts and in resistance can be modelled. This approach was applied to model a dynamic of COVID-19 in Moscow during the period between October 2020 and December 2021.*

Keywords*: ABM, SYNTHETIC POPULATION, CONTACT NETWORK, DISEASE TRANSMISSION*

1. Introduction

When a new virus or a new strain appears, the most important question that should be answered is: "Will there be an epidemic or will the incidence remain at a low level?" Mathematical epidemiology developed intensively in the 20th century, and a row of models based on systems of differential equations were elaborated and thoroughly analyzed [1-3]. It was shown that the dynamics of morbidity depends on the value of the basic reproductive number of the disease, R_0. If R_0 is greater than one, then there will be an epidemic, if less, then the incidence will not grow, and the disease will eventually disappear from the population.

Formulas for the basic reproductive number were derived and it was proved that it was directly proportional to the rate of virus transmission, β, and to the length of the time interval during which an infected person stayed infectious [4]. To simplify the form of the first mathematical models, the rate of transmission was supposed to be a product of three probabilities – the probability of an infected person meeting with a susceptible, the probability of susceptible person to be frail enough to be infected and the probability of infectious person to shed virus. It was assumed that there was an equal probability of meeting everyone with everyone, the same susceptibility in uninfected and the same infectivity in patients.

Such models provide good fit to the data at the early stage of epidemics when the exponential growth is observed but afterward their application turned out to be impossible. The parameter estimates obtained at the initial phase of the rapid increase in morbidity failed to give the correct dynamics after a month [5]. The explanation of the limited applicability of the models lies in basic assumption of population homogeneity both in terms of the number of contacts, vulnerability, and the ability to spread the disease. Agent-based modeling allows us to consider the heterogeneity of the population, and also to model their behavior, which can also affect the dynamics of the disease.

In this paper we present agent-oriented models in which we construct complex networks of contacts and simulate the spread of the disease on the networks. Agents have such characteristics as gender, age and social status and are united in families, educational and working groups. Agent-based models can be used to analyze the current epidemiological situation and make forecasts, as well as to assess the effectiveness of anti-epidemic measures.

2. Method

An agent-based model is a computational model that simulates the interactions between entities called agents [6]. Agents are supposed to have constant and variable attributes. In our model each agent represents an inhabitant of big city with assigned age, sex, marital status, and occupation as constant properties and an attribute reflecting its state of health as a variable.

Agent-based modeling includes three main stages:

(1) creating a synthetic population of agents and assigning values to agents’ attributes;

(2) accomplishing *k* circles of the model simulating ‘everyday’ interactions between agents and related changes in values of their variable attributes;

(3) an analysis of individual trajectories and generalized curves.

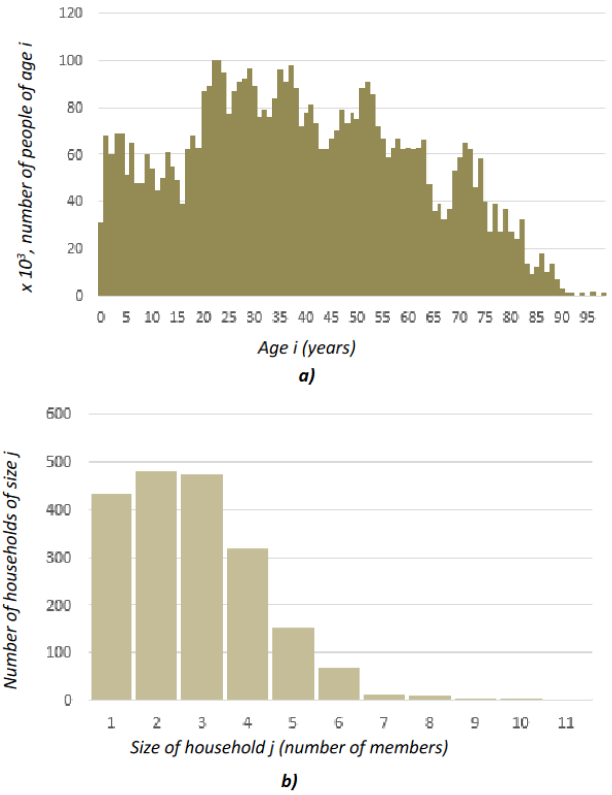

Fig.1 a) *Age distribution of synthetic population of Moscow constructed on the base of the 2010 All-Russian Population Census. The size of population is 10 million.* ***b)*** *Size distribution of households. Total number of households is 4 million*

3. Creating a synthetic population

When building agent models for studying the processes occurring in the human population, the problem of creating a model population arises. As a general rule, detailed data on real households are either unavailable or confidential. An alternative way to create a model population is to generate synthetic populations with structural characteristics (such as sex and age distribution, a number of households and their size distribution, population density and a number of educational organizations and workplaces) corresponding to the data collections of the real population.

In this paper we describe the construction of the synthetic population with general structural characteristics corresponding to the real population of the city of Moscow within the Moscow Ring Road. In so doing we use data from the statistical collections of the 2010 All-Russian Population Census for Moscow. The information used includes demographic and socio-economic information of the city's residents in 107 municipalities. In addition, geographical coordinates are used for 989 schools, 1,967 kindergartens and 138 higher educational institutions, taken from the Open Data Portal of the Moscow Government.

The synthetic population of the model is assumed to consist of 10 million agents and 4 million households, without loss of generality. The creation of the population is started with generating households and assigning them a number of members and geographical coordinates. Then, the created households are filled with agents. First, an adult agent is created with age and sex taken from age distribution. Then, we create the remaining household members with age depending on the age of the original agent and the size of the household in such a way as to have age difference in marriage and age difference between children and parents in accordance with studies [7-9]. The obtained age distribution of the synthetic population is presented in Fig. 1a. It has the same waves as the age distribution for the real Moscow population but can have slight variations unlike the size distribution for households (Fig1b) which is absolutely the same since we started modeling with a household.

Thus, the synthetic population of 10 million agents is constructed. Each agent is assigned age, sex and a list of family or household members. Further, we will assume that members of a household have a close contact every day for a period of time sufficient to inhale a effective dose of the infectious agent if one of them is infected. To model the spread of infections outside the occupational property is to be defined. This property can have one of the following values: 'preschooler', 'student', 'worker' or 'inactive'. These values are assigned basing on the statistical information about numbers of preschooler, students and workers in Moscow. In next section the way of network construction for different types of occupation is described.

4. Complex network of contacts

As we aim to study transmission of air-born infection, the interactions at the distance less then 1.5 meter and longer then 5 minutes are of interest. It is quite difficult to identify all interactions in population when people have physical contact or talk face-to-face for sufficient time interval although certain researches on the number of such contacts have been done [10-13].

Basing these studies and on the data available we constructed network of interactions during which air-born infection could be passed. The fragment of such a network is represented in Fig.2. The vertices of the network represent agents, a links between two vertices corresponds to a contact. The color of a vertex describes the type of occupation an agent has. It is easy to see that members of one family or household form a complete subgraph. As well, groups of preschoolers formed from 15-25 children also can be represented by a complete subgraph.

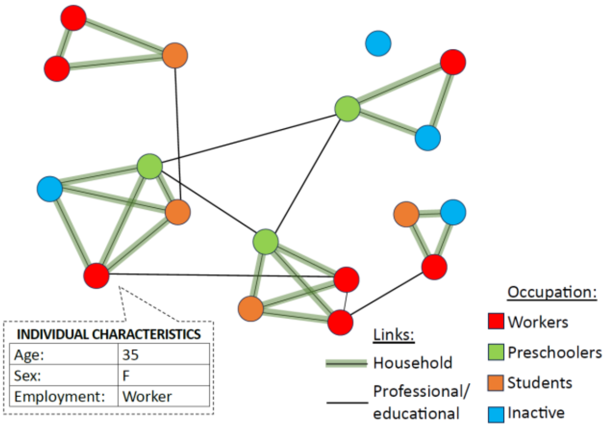

Fig.2 *Fragment of the contact network for different occupations. Households are represented as a complete subgraph.*

Groups of colleagues or working groups are composed of agents aged 16 and above, using data on employment in different age groups from Federal State Statistics Service (2015). Working collectives, unlike preschooler groups and classes, do not have geographical coordinates. We use Federal State Statistics on pre-school education and on the level of education of Moscow population to allocate agents to educational institutions. Agents are united in classes and enterprises according to their age. The number of agents in a group of students depends on the year of study, and age difference in the same group may not exceed three years. In addition, for each group of students, an adult working agent is chosen to be a teacher. Social connections within groups of colleagues and educational institutions are presented either as complete graphs or as Barabashi-Albert graphs [14] depending on the size of the group. The parameter for Barabashi-Albert network corresponding to the minimum number of social connections is 5 for groups of colleagues and 10 for educational institutions.

Further, each link is assigned a weight corresponding to the time interval during which the disease transmission can occur. Thus, almost all population was united in one complex network with weighted links. In order to use the constructed network in modelling of disease transmission it is necessary to validate it and to show that it reproduces the structure of contacts obtained in other studies.

To validate the structure of contacts per day in the synthetic population we use the structure of contacts built by Prem et al. in [13]. Frequencies of contacts between different age groups received by use of our model and by Prem's analysis are represented by coloured matrices in Fig.3. The main diagonal for the household contact matrix is fully reproduced since it corresponds to contacts of close-aged couples living together, at the same time diagonals corresponding to children's contact with their parents can be distinguished. We have also managed to reproduce quite well the structure of contacts in the workforce, taking into account certain number of working pensioners (65 years and older). The number of contacts between students is slightly overestimated.

Thus, the comparison of the main characteristics of the constructed synthetic population with census data and data from the literature confirms the reliability of the proposed algorithms. We can conclude that created synthetic population can be used as the basis for the further study of processes occurring in the human population.

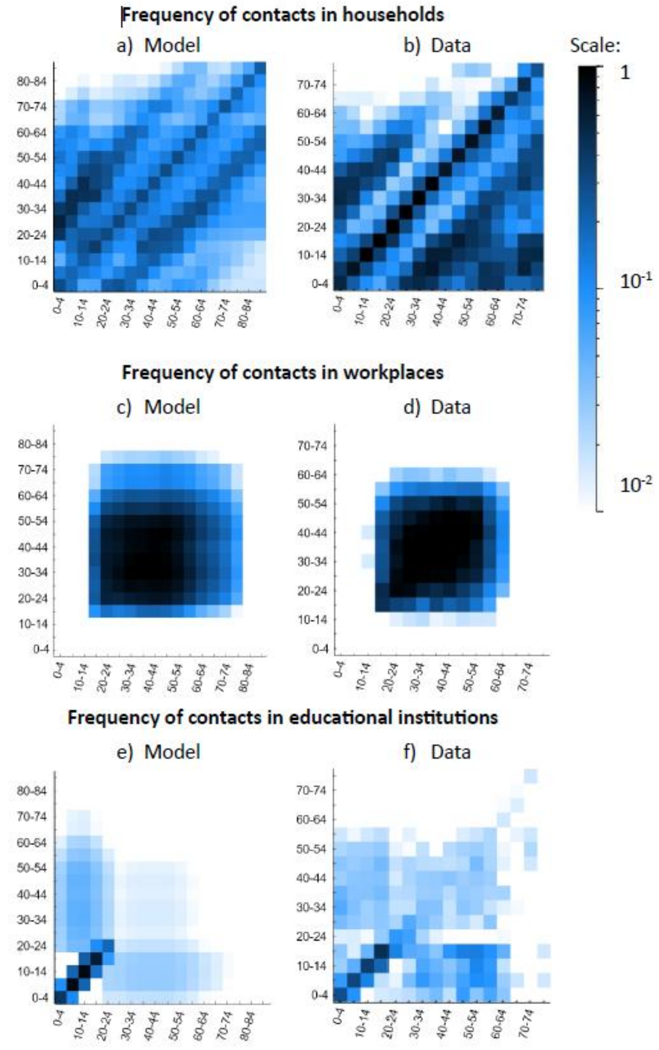

Fig.3 Colored matrices of the frequencies of contacts per day between different age groups at different location a), c), e) in synthetic population and b), d), f)in real population according to [13].

5. Epidemiology of airborne infections

In order to simulate the spread of airborne infections it is necessary to defined some agents' attributes describing their state of health. The states of health referring to the classical epidemiological models are 'susceptible', 'infected' and 'recovered'. For infected agents, the day of the contact is stored in order to determine the level of infectivity and the moment of symptom manifestation, in other words, the first day of disease. The infected agents recover with a constant rate and acquire an immune memory that declines and completely disappears after 6-12 months. Then, the agent becomes susceptible. In general, the length of disease is quite short in comparison with lifespan, for this reason deaths and births of agents are not modelled. For some severe infections, disease-caused mortality can be added.

The advantage of agent-based models is their capacity to take into account population heterogeneity. For the sake of simplicity, the only process that depends on individual characteristics is virus transmission from an infected agent to susceptible. For each link connected an infected agent to susceptible one the probability of getting infection is calculated. For air-born infections, the probability depends on the level of vulnerability of the susceptible agent, on infectivity of the infected one, on the length of the contact and possibly on other factors. It is argued that some environmental conditions as air-temperature, humidity or air pollution can influence the probability of getting infected. For example, some respiratory viruses propagate faster in epithelial cells of upper respiratory tract at low temperatures that causes seasonal increase in respiratory morbidity [15]. At the same time, strains of SARS-Cov2 did not demonstrate dependence on air temperature [16].

The probability of *i*th susceptible agent to be infected during the contact with *j*th infected individual can be calculated as follows

$$P_{ij}=p_s\cdot p_{inf}\cdot p_t\cdot p_{env}$$

Here p_s corresponds to the level of vulnerability of *i*th susceptible individual. As a rule, susceptibility is low when an individual has a high level of immune memory cells and specific antibodies, and high for new infections when there is no immune memory. Additionally, some chronic states and disease can make an individual more susceptible to air-born infections. p_{inf} is depicted the level of infectivity of *j*th infected agent. This parameter depends on the time that passed form the moment when the contact with infected person occurs. For many respiratory viruses including SARS-Cov2, infection transmission is possible before first symptoms manifest.

The third parameter defining the probability to become infected is the length of time interval during which the contact takes place p_t. Contact between students are assumed to be 4-6 hours, contacts between preschoolers and between adults at work places are 7-9 hours, and contacts between family members 10-12 hours. Thus, the probability to be infected at home once somebody in family is infected is maximum.

The fourth parameter in the formula for probability to get infected is assumed to be optional. It could describe the environmental factors or influence of some preventive measures as wearing masks.

6. Simulation

The main stage of agent-based modelling is implementation of *k* circles simulating every day activities of agents that change values of agents' attributes. First of all, pseudo-random events of disease transmission are simulated, resulting in certain number of new infected individuals. Then, recovering and loss of immunity are modelled. At each circle, the numbers of susceptible, infected and recovered agents are calculated as well as location of the disease transmission.

While simulating *k* days it is crucial to remember that the sequence of procedure for each circle is the same but not the contact networks. On weekend, on state holidays, during school and university holidays the contact networks transform, the links between classmates should be ignored on these days as well as links between colleagues during weekend and state holidays. For these days number of links decrease significantly and it can stop the spread of infection. But, it is not the case, so some random links can be added to simulate some unscheduled meetings.

As a rule, the curves received during of *k* circles of simulation should be fitted to some observed data. Since the structure of population and frequencies of contacts were validated, it is reasonable to change parameters that influence disease transmission.

The daily dynamics of COVID-19 in Moscow from October,1, 2020 to December,15, 2021 was model. In Fig 4 comparison of model results with statistical data is represented. In COVID model we have to split the group of infected agents to infected detected and infected undetected since not all infected were diagnosed. So, we also have the rate of detection which also was changed to get best fit. Besides transmission parameters, some other parameters have strong influence on disease dynamics. One of the most important parameters is the duration of immunity. Now it is assumed to be the same for all agents as well as the rate of decline.

But efficacy of immunity depends on sex, age, stress and other physiological conditions.

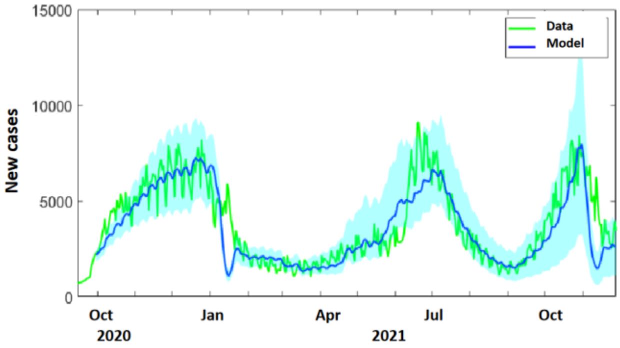

Fig. 4 *Comparison of model results with statistical data for the number of new cases per day. Light blue color corresponds to 95% confidence intervals due to 5% variation of parameters*

8. Conclusion

To summarize it should be noticed that there are many various approaches to construct and validate agent-based models: from abstract, simple models designed to study the role of several factors or general principles of agent interaction, to models that reproduce detailed reality, including cartographic, transport, real-life, behavioral data.

The objective of our study is to create a tool that allows to simulate the actual epidemic processes and estimated losses and benefits of control measures that could restrain the spread of infections. The dynamics of the epidemic are determined by processes of different nature: socio-economic, demographic, behavioral, physical and immune. The strength of agent-based modeling is that we can incorporate as many processes as we want. But, at the same time the complexity of the model grows and validation requires more datasets that are not used on the stage of construction. In our model we tried to keep balance between complexity of the model and clarity of the validation procedures.

References

1. R.M. Anderson, R.M. May, *Infectious diseases of humans: dynamics and control*. (Oxford university press, 1991)
2. O. Diekmann, J.A.P. Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Vol. 5*. (John Wiley & Sons, 2000).
3. H. W. Hethcote, SIAM review, **42**(4), 599(2000).
4. K. Dietz, Stat. Methods Med. Res. **2**, 23 (1993).
5. I. Cooper, A. Mondal, C. G. Antonopoulos, Chaos, Solitons & Fractals, **139**, 110057 (2020).
6. E. Hunter, B. Mac Namee, J. D. Kelleher, JASSS, **20**(3), 2 (2017)
7. A. Avdeev, A. Monnier, Population: an English selection, **12** (1), 7 (2000).
8. S. Scherbov, H. Van Vianen, Dem. Res., **10**, 27 (2004).
9. D. Philipov, A. Jasilioniene, Dem. Res., **19**, 2057 (2008).
10. J. Wallinga, P. Teunis, M. Kretzschmar, Am J Epidemiol, **164**(10), 936(2006).
11. S. Y. Del Valle, J.M. Hyman, H.W. Hethcote, S.G. Eubank Social Networks, **29**(4), 539 (2007).
12. J Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, ..., W. J. Edmunds, PLoS medicine, **5**(3), e74 (2008).
13. K. Prem, A.R. Cook, M. Jit. PLOS Comput. Biol. **13**, e1005697 (2017).
14. R. Albert, A.L. Barabasi, Rev. Mod. Phys. **74**, 47–97 (2002)
15. A. I. Vlad, A. A. Romanyukha, T.E. Sannikova, Bull. Math. Biol. **85**(10), 100, (2023).
16. A. A. Romanyukha, K. A. Novikov, K. K. Avilov, T. A. Nestik, T. A., T. E. Sannikova, Inf. Dis. Mod., **8**(2), 403 (2023).