Research article

DEPENDENCE BETWEEN MORTALITY IN REGIONS AND PREVALENCE OF ACTIVE SARS-CoV-2 CARRIERS AND RESOURCES AVAILABLE TO PUBLIC HEALTHCARE ORGANIZATIONS

V.S. Stepanov

The Central Economics and Mathematics Institute of the Russian Academy of Sciences, 47 Nakhimovskii Ave., Moscow, 117418, Russian Federation

The paper dwells on certain mathematical models showing how epidemics develop, namely, logistic ones, SIR-model, and some others. There is also a review of articles that focus on such models showing dynamics of incidence with COVID-19 infection. These models are often successfully applied for data collected in a whole country but on a regional level there are difficulties due to peculiarities of calculating mortality figures in Russia. In this case regression models can be useful with their obvious advantage at the initial stage in an epidemic process. They also include exogenous variables that influence mortality, for example, a number of doctors and nurses per a hospital, how well hospitals are equipped with ALV devices, and a number of available beds in them.

Our research goal was to build up a linear regression model that could be used as a basis for estimating regional mortality caused by COVID-19 as well as for more efficient distribution of all the resources mentioned above.

The model is built as per a set of resource parameters including data on «active cases». Preliminary three variables that showed data on resources available to communicable diseases departments in hospitals were transformed into a new single one via linear transformation. Then the model was tested on a training sample containing an endogenous variable on mortality and four factor ones including prevalence of active virus carriers. Regions were included into training data with different lags; they were included into such daily samples when death cases were registered rarely. Then the estimated model was applied with other values. It turned out to be quite efficient in estimating COVID-induced mortality for regions from trainings samples as well as for several others (for certain intervals).

As a result, we built a regression model and estimated its precision; the model showed a relation between mortality in a region and prevalence of active SARS-CoV-2 carriers and availability of resources to hospitals in it. It can be useful when these resources are distributed. It can also be used to build SIRD, SEIR, and SEIRF models at a regional level when choosing parameters in them related to mortality. A methodology itself that can be similarly applied for other epidemic processes also deserves certain attention.

Key words: regression model, mortality estimation, COVID-19, coronavirus infection, logistic equitation, SEIR, SIR, ALV.

There are a lot of scientific works on issues related to SARS-CoV-2 pandemic; they can be found on *elibrary* web-site and some others. For example, K.V. Zhmerenetskii and E.N. Sazonova concentrate on epidemiologic and other peculiarities typical for COVID-19 in their review; other authors examine and discuss issues related to collecting data on morbidity and mortality as well as validity of statistics [1–7]. O.M. Drapkina and I.V. Samorodskaya declared in their work that it was necessary to build prognostic mathematical models; in particular, the present work focuses exactly on the matter [3]. S. Kozlovskii, apart from morbidity and mortality accounting, considers modeling question in his work as well [5].

Let us turn our attention to issues related to mathematic modeling of an epidemiological process (EP) as importance of building up a sufficiently precise model for COVID-19 has been widely recognized [3, 5, 8]. Such a model can give grounds for more efficient decision-making and it will allow reducing mortality, easing off negative consequences of the epidemic, and saving resources. Different approaches to building models showing EP dynamics are described in a review by M.A. Kondratyev in which he considers artificial neural networks, simulation and statistical modeling, etc. If we simplify a classification of approaches proposed in his work, we can spot out two types of models, deterministic and stochastic ones [9].

Vladimir S. Stepanov – Candidate of Physical and Mathematical Sciences, Senior researcher (e-mail: _stepanov@cemi.rssi.ru, vladstep0355@gmail.com; tel.: +7 (916) 872-52-07; ORCID: https://orcid.org/0000-0002-4478-376X).

First of all, we are going to give a brief description of models belonging to the first type and how to apply them to simulate COVID-19. When building up a model taking into account peculiarities in EP development, experts divide population living on a certain territory (a country or a region within it) into several sectors such as «susceptible to virus», «infected», etc. First, a block diagram showing EP development over time is built; it describes the possible transitions of persons from one sector to another with different intensities (model parameters). The latter are estimated in a different way; some – according to data taken form literature, others – as per statistic data collected for a given territory. Transfers between sectors in a block diagram are formalized basing on a system of differential equations. Thus, Baroyan-Rvachyov's model that was later implemented in *EpidMod* computer system used a system of non-linear integro-differential equations in partial derivatives with relevant boundary and initial conditions; it was taken as per analogic models existing in hydrodynamics. Models based on equations in partial derivatives of a function of two variables (time and a patient's age) exist at the moment [10]. However, most models that describe EP development over time are based on a system of ordinary differential first-order equations. SIR model was created quite a long time ago and its abbreviation comes from English words «susceptible», «infected», and «recovered» [5, 8, 9, 11]. Later its certain modifications were developed; they were as SEIR, SEIRF, SIRD models, etc.; the last two obviously take into account a decrease in a number of infected people due to mortality: D letter means death caused by an infection and F means death caused by complications. Starting from late 1990-ties, *elibrary* has been accumulating works with such models and now there are more than a hundred of them stored in it; most of them are performed by foreign researchers. SIR and SEIR models are widely used now to analyze dynamics of processes going on in social networks when experts examine how specific information is being spread through them [12]. SIRD model is used in a work by P.V. Khrapov, A.A. Loginova to analyze data on COVID-19 for the RF and some other countries [11]. A similar model, but in its

finite-difference form, was applied in a research work by O.V. Drugova, E.A. Pavlov et al. when they examined COVID-19 dynamics in Nizhniy Novgorod region [13]. Here, the number of people divided into four sectors (or compartments) was chosen as the main variables of the model, namely «susceptible» (S); «infected» (I); «recovered» (R); and «deceased» (D); it was assumed that up to 30 % in sectors I and R cannot be detected. After all the parameters were fitted, the model was adjusted to actual data on initial part of COVID-19 growth in the region; and then it was used as a basis for predicting further dynamics in the epidemic.

Baroyan-Rvachyov's model belongs to SEIR type where E letter means another sector, namely «exposed» (people who are going through incubation stage). The sector is determined by a period of time starting from a moment when new coronavirus enters a human body in a dose sufficient for contagion and up to a moment when a person falls sick with COVID-19. M.V. Tamm applies an expanded SEIR model in his work taking certain peculiarities existing on some territory into account; they can be population age groups etc. [14]. After all the parameters are fitted, the model is used to analyze several scenarios of fighting against this new EP in Moscow. A.V. Matveev applied a SEIR-type model in his work that has been modified taking COVID-19 peculiarities into account; the model was implemented with *AnyLogic* simulation modeling tool [15]. As a result, a few scenario calculations were accomplished for two capital cities and recommendations were given on equipping hospitals with beds for contagious patients with coronavirus and artificial lung ventilation (AVL) devices; a forecast on a number of deceased was made.

Another model that applies a stochastic approach to estimating is described in a work by A. Godio and F. Pace et al. [16]. They also use a generalization made for EP by L. Peng and W. Yang et al. [17]. As a result, a SEIR-model was developed that uses a number of specific parameters. They were estimated as per research data obtained by A. Godio and F. Pace et al.; in particular, there was a mortality coefficient $k(t)$ that depended on t counting and efficiency of treating patients. Besides, S. Peng-

peng and Shengli et al. [18] suggested another SEIR-model taking into account reduced contagiousness of infected people who didn't have any COVID-19 symptoms as well as using variability in incubation duration. et al. [19] described a SIR-model and two SEIR-model modifications in their work. Unknown parameters in the research were found via fitting the suggested model to data obtained in People's Republic of China from January 23 to February 10. As a result, a short-term forecast was made and it was estimated when a peak in contagion number was expected in China; the work also contains a code for the program in *MATHLAB*.

A.V. Nikitina, I.A. Lyapunova C. et al. [19] described a SIR-model and two SEIR-model modifications in their work. They suggested certain changes for these modifications that took into account peculiarities related to how immunity was acquired as well as a lag occurring in detecting infected people; they also examined sensitivity of the described models. These results can be applied in a model building for COVID-19. Also, almost 50 years ago Russian researchers started to use a SEIRF model and N.F. Gamalei's National Research Center for Epidemiology and Microbiology created relevant software for it at the end of the 20th century. This model is applied by Z. Liu, P. Magal et al., and F. Ndaïrou, I. Area et al. in their works to process data on COVID-19, and in the latter one it is built already with 8 sectors; there is an expression for $D(t)$, a number of people who died on a day t, that includes three parameters [20, 21]. We should note that the model (1) which we suggest in the present work can be useful for selecting values of a parameter $k(t)$ mentioned by A. Godio and F. Pace et al. or selecting parameters in the expression $D(t)$ [16, 21].

Apart from an approach that involves building and finding solutions to a system of differential equations, there are simpler ways to do modeling. They are used in many works focusing on predicting what characteristics an examined EP might have basing on finding a solution to an ordinary differential equation created by P. Verhulst and its generalizations [22–25]. That mathematician applied it a very long time ago to analyze population growth in dynamics as per time t. The equation is widely spread in biophysics; at present it is applied in

modeling a growth in number $N(t)$ of persons infected with SARS-CoV-2 as per data collected in People's Republic of China and Sweden [22]. This work focuses on three models based on Verhulst equation, gives relevant solutions, and also dwells on predicting problem. After $N(t)$ value has been differentiated, we can estimate both a moment when an EP reaches its peak and a number of necessary beds in hospitals according to the research performed by A.A. Kurkin, O.E. Kurkina et al. [22]. One of the described generalized models leads to a differential-difference equation. The model takes into account incubation period that occurs in EP development as well as the fact that approximately 4 % people who fall sick with COVID-19 might die due to it.

Models belonging to this type are less sensitive to quality of available statistical data than sector-type models that have been described above. They are built on daily counts of a time series, in particular, for a series on a number of people who fell sick with COVID-19 or died due to it, and include certain parameters. For example, N. Fabiano and S.N. Radenović fitted Verhulst model to the data on a number of Italian citizens who fell sick with COVID-19 [23]. In another research work the model was modified taking into account specific features of the infection and it allowed obtaining a precise analytical solution when certain assumptions were introduced [24]. Parameters of the solution were estimated as per data collected in China and, as a result, there was good congruence between the model and EP dynamics, both for people who were sick and overall number of deaths in People's Republic of China [24].

Verhulst model is also used in its discrete variant which has two parameters. This logistic equation started to be widely used to examine various processes after it was introduced by M. Feigenbaum, an American scientist. Feigenbaum's discrete model was applied to examine a number of infected people in a dozen countries and two megacities in a research work by E.M. Koltsova, E.S. Kurkina et al., and its parameters were estimated as per an overall number of people who fell sick with the infection in China [25]. There were 4 variants of growth in overall number of infected people in Moscow and the mildest one was called «Israeli». As per

this variant, by the end of the EP there will have been totally 300 thousand people living in Moscow infected with SARS-CoV-2 and it is well in line with 248 thousand people detected on August 10 with daily growth equal to 700 cases.

Finally, V.V. Boyarintsev, R.S. Pal'min et al. [8] revealed certain regularities in EP development in their research basing on analysis of overall number of infected people *I(t)* living on a given territory. As a result, basing on a set of hypotheses-assumptions, they suggested a recurrent formula for calculating *I(t)* on a *t*-th day in EP. Basing on it, dependence as per *t* was obtained for a function showing how a share of infected people was distributed, and then forecasts were made for different countries and Moscow as well.

However, a lot of researchers who examine EP basing on SIR-models and their modifications note that obtained solutions are not stable and it prevents them from making reliable forecasts how EP is going to develop over a period longer than a week [5, 8]. A.V. Matveev also states that forecasts have only approximate results [15]. S. Kozlovskiy and O. Boldyrev note that many parameters applied in EP models are to be estimated as per regional statistics in case the models be applied to describe the process in a specific region [5]. And many specialists on statistic data believe regional statistical data to be of low quality [4, 5, 7]. As per experts' estimates, corruption in statistical data has been growing since late April 2020. M.V. Tamm also has no trust in figures on mortality in regions [5, 14]. Moscow is commonly recognized as an etalon region when it comes to transparency and authenticity of data on mortality caused by COVID-19. As for RF regions, more qualitative data are provided by the Rosstat and regional register offices. The latter provide monthly data on their official web-sites stating the overall number of deaths over a month but without reasons that caused them; it should be mentioned that such publications are issued with considerable delays [4, 6, 7]. Besides, certain parameters are unknown at initial stages in EP development, and it is due to this fact that maximum simplest models are applied at these stages as such models contain minimum number of parameters [22]. Such models include regression analysis models and time series models.

If we consider non-deterministic EP models that naturally take into account existing data noisiness, than we should examine a work by D.V. Melik-Guseinov, N.N. Karyakin et al. [26], where several regression models for analyzing mortality caused by COVID-19 in the RF as a whole were built up and estimated. Thus, Model 1 in its variant 2 is given as $ln(Y) = ln(X) \cdot \beta_1 + \beta_0$, where Y is a number of new death cases over the last day and X is a number of new incidence cases over the same day; β_0, β_1 are parameters in the model that are to be estimated. After substituting β_0, β_1 estimates taken from the research by D.V. Melik-Guseinov, N.N. Karyakin et al. as per least square procedure (LSP) the authors obtained a linear regression model with determination coefficient $R^2 = 70\%$ [26]. Besides, they suggested Model 2 with $R^2 = 80\%$; it was built basing on Model 1 after introducing one more variable X_2 or «a number of severe COVID-19 cases in *j*-th region».

I.A. Lakman, A.A. Agapitov et al. suggest ARIMA model that is rather hard to build as well as an adaptive Holt – Winters' ARIMA model or Box – Jenkins' «autoregressive moving average» [9, 27]. When they are used together with SIRD model, it allows making highly precise forecasts for short-term morbidity, mortality, and survivability among COVID-patients on a specific territory. However, it is assumed for a SIRD-model that coefficients applied to estimate probabilities of contagion, recovery, or death are stable over a time period t and it is not in line with Russian practice. It results in a necessity to re-adjust a model frequently in order to obtain reliable forecasts [27].

Due to all the above mentioned difficulties modeling is less frequently applied for regions in the RF than for countries, Russia as a whole, or Moscow. Besides, some statistical models have some advantages in comparison with sectional or logistic ones. Therefore, sometimes both approaches are combined as it was made in research works outlined above where an Erlang distribution analogue was used together with a SEIR-model [9, 27–29]. Linear multiple regression models seem quite promising among statistical ones [9]. Their advantages include a well-developed theory; for example, one can build confidence intervals, calculate elasticity coefficients, etc. Such

a model can be easily built provided that a user is well-qualified to work with *IBM SPSS*, *Statistica*, *EViews,* or *Deductor* software. It is also present in Excel, *MATLAB* computer statistic programs, *Wolfram Mathematica* etc., *Python* or *R* (*R-Studio*) environment. Finally, it is quite easy to take additional factors into account that can influence COVID-19 mortality. For example, we can add a variable that describes clinical process gravity for a patient, or how available medical aid is to him or her; there can be a variable showing whether a patient smokes tobacco or electronic cigarettes and if yes than how frequently; there can be other factors added into the model.

We should also note that at present well-known scientific centers and universities are trying to build relevant models for EP in general and COVID-19 EP in particular. Among them there are Center for Mathematical Modeling of Infectious Diseases (CMMID) at the London School of Hygiene and Tropical Medicine, Center for Hygiene and Epidemiology at the Imperial College (London), Geneva University, Basel University, a lot of universities in the USA and other countries [5, 14]. Thus, US Centers for Disease Control publish their weekly estimates on EP development obtained by 24–26 research teams, mostly Americans, for overall number of patients who died due to COVID-19 [30]. These estimates are given on a date t as point and interval predictions for the next 4 weeks. A wide set of models is used in the process; thus, 11 research groups apply various SEIR models (and its stochastic variants SEIS, SEIRD, or SEIRX); 5 groups, SIR-model and its SEIR modification. Agent models, artificial neural networks, time series models, regression models – both linear and non-linear one with the «ridge» estimator of covariance matrix (UM model) are not so frequent [9, 30]. Several groups examine a situation existing in a particular US state. Also there are several foreign researchers who are trying to model COVID-19 dynamics in Russia [5, 31].

Our research goal was to build a linear multiple regression model that could give grounds for estimating COVID-19 mortality in a specific region and give a point prediction for it at an initial stage in EP. It can also be used to provide better substantiation for how to allocate certain resources available to public healthcare organizations in RF regions.

Data and methods. The model was built as per data on prevalence of active SARS-CoV-2 carriers at an initial stage in EP development in a specific region as well as with taking into account how well hospitals in it were provided with resources or necessary equipment. First we created a table containing data on «objects and variables» where objects were multi-dimensional observations (or p-dimensional vectors) collected in $n = 40$ RF regions, and initial variables were p = 7 parameters. Two out of them change every day, namely COVID-19-related mortality rate among population and a ratio between active virus carriers and average population number in a region over 2019, in other words, their prevalence («factor» or «exogenous variable» X_1)[1] [32]. The rest five indicators were taken from Rosstat[2] data collected over 2018 and 2019, research works by Yu. Apukhtina and S. Zobova, and from an article by A.A. Sokolov [33, 34]. They described how well regional hospitals were equipped with: X_2 (artificial lung ventilation devices in reanimation, 2019) as per 10^5 people; X_3 (doctors with sanitary and anti-epidemics specialties, 2018) per 10^5 people; as well as infectious disease doctors (2019), nurses (2018), excluding obstetrician nurses; and beds for contagious patients with communicable diseases (2019).

To substantiate our selecting these variables, let us note that there were factors influencing lethality during EP listed in several research works [3]. They included a situation in public healthcare in a period prior to EP as well as quality and availability of medical aid. Variables X_2, X_3... reflect it partially. A lot of them are taken into account in decision-making in foreign countries [13]. Thus, specialists in a university in Germany developed *CoronaVIS* system for monitoring over resources available to hospitals in the country, namely, number of

---

[1] Russian statistical yearbook. 2019. *Rosstat*. M., 2019, 708 p. (in Russian).
[2] Public healthcare in the RF. Supplement to the collection (data as per RF regions). *Rosstat*, 2019. Available at: https://rosstat.gov.ru/folder/210/document/13218 (06.08.2020) (in Russian).

beds and devices for oxygen therapy [35]. After an online inquiry has been made, one can instantly see a map of a given territory with actual data on how many beds are occupied and how significant is load on artificial lung ventilation devices in a specific hospital (its coordinates also available); there are also diagrams showing loads on the nearest hospitals over a day. It allows avoiding «bottlenecks» in delivering a patient to a hospital as we'll as optimizing their distribution over different hospitals. The IHME institute at Washington University developed models for predicting frequency of patients' deaths as well as number of occupied hospitals beds and ALV devices [30, 36]. Experts from a college in New York together with scientists from three universities created a calculator working within Excel; it allows estimating loads occurring due to patients with COVID-19, and one can see availability of beds in hospitals and loads on ALV devices in them in an electronic table. As a result, there is a model showing loads on infectious and reanimation departments in hospitals in a specific region taking into account a number of occupied beds, both ordinary ones and those aimed at intensive care [37].

Preliminarily all the above mentioned variables that described availability of necessary resources to hospitals (infectious disease doctors, nurses, and beds for patients with communicable diseases) were turned into a new one as per method of principal components. And here a slight modification was used; it was previously suggested by S.A. Aivazian [38, 39]. An algorithm used to calculate this variable given here as $(X_4)^*$ included several steps. First, each variable (a parameter) maps into [0, 10] segment where 10 corresponded to the maximum value in the training data table, and 0 to the minimum one. Then estimator of S covariance matrix including three variables was calculated. Then all eigenvalues and eigenvectors of the matrix S were calculated; one of the vectors set the weights required to calculate $(X_4)^*$ variable.

The learning sample containing $n=40$ RF regions was obtained after excluding 3 Federal cities, Moscow Region, and three regions it had common boundaries with, as well as the next 11 Regions: Astrakhan, Bryansk, Volgograd, Voronezh, Kaliningrad, Kemerovo, Murmansk, Orel, Penza, Pskov, and Rostov;

all Autonomous Areas; Republic of Kalmykia, Komi Republic, Republic of Mari El , Chuvash Republic, several North Caucasian regions and some others, as well as Altay, Krasnodar, and Stavropol Territories; and some regions where data on mortality were not published in spring. Statistic data provided in many of them are being criticized [7].

As a result, we built a non-adaptive linear regression model

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_4 \cdot (X_4)^* + \varepsilon, \quad (1)$$

where Y is mortality caused by COVID-19;

$X_1, \ldots, (X_4)^*$ are four «factors» outlines above;

ε shows influence exerted by neglected variables on Y;

$\beta_0, \ldots, \beta_4$ are unknown parameters estimated by as per a sampling [9].

Value of ε in the formula (1) is determined as random; by definition its mathematical expectation $E(\varepsilon) = 0$ and variance $var(\varepsilon) = \sigma^2$, and it is unknown. Estimators for $\beta_0, \ldots, \beta_4$ and σ^2 are found as per LSP in Excel as per a training data table. Let us express estimators of β-coefficients via $b_0, \ldots, b_4$. The table contained value sets $\{Y, X_1, X_2, X_3, (X_4)^*\}$ for all n regions and those sets were included into the table with different counting as per t. It was assumed that a number of new deaths M published over a day t for a specific region was a random value that approximately had binomial distribution; and m, a number of active COVID-19 cases detected among N people living in a region was interpreted as m for independent Bernulli tests over a set made up of N objects and M out of them had a property that was significant to us – «there was with death » [32]. As m grows, M value distribution gets closer to Poisson regularity, however, probability of death due to COVID-19 P_t in a region over a day t is not known precisely. To estimate it, we considered a sequence over 7–10 with growths in a number of deaths per a day t. Then, basing on binomial distribution, we built up a 95 % confidence interval and its middle was taken as P_t estimate. Then we took average value P^* for these estimates as per a given sequence of days. A mixture of Poisson distributions is known to also

have Poisson distribution but with different intensity. As a result, when creating the training sampling, we selected a lag in a time series in such a way so that a value of $P*$ estimate of probability of a COVID-death was within $(3 \cdot 10^{-5}, 9 \cdot 10^{-5})$, that is, it was lower than one death case per 1 million people. For example, Ulyanovsk Region was included into the table with data on May 11, 2020, and its value $P*$ calculated over a period from May 06 to May 17, 2020, is equal to 0.000049. Follow there are two comments.

Note 1. As 3 factor variables X_2, X_3 and $(X_4)^*$ that describe availability of resources to hospitals do not change (to be exact, we do not have any data on their changes available to us), the expression (1) can be given for each j-th region as a sum of two summands. The first one is a constant $C(j)$ which is by definition equal to $b_0 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_4 \cdot (X_4)^*$ where X_2, X_3, $(X_4)^*$ are variables for j–th region, $j=1,...,$ n. $C(j)$ value shows an expected mortality level in case there is low prevalence of active cases in j–th region when $X_1 \approx 0$. Obviously, the better available and developed are resources required by hospitals in a region, the lower $C(j)$ value is going to be. $C(j)$ value does not depend on t since ultimately parameters are estimated for (1) only once. The second summand in (1) contains a factor $X_1(j)$ which depends on t. As a result, the model (1) is given as follows after unknowns $\beta_0,...,\beta_4$ have been estimated:

$$Y(j) = C(j) + b_1 \cdot X_1(j), \qquad (2)$$

where $Y(j)$ is COVID-related mortality per 1 million people;

$X_1 = X_1(j)$ is prevalence of active cases with SARS-CoV-2 calculated per 10^4 people in j-th region.

Index t is omitted are here in order to make the formulation simpler. The formula (2) can be easily used to daily estimate Y. To do that, we should take some data from a spreadsheet located on the official web-site on coronavirus for a day t; these data are overall number of infected (I), deceased (D), recovered (R); then we have to find $X_1 = 10 \cdot [I - (R+D)]/N$ where N is population number taken from Table 2; the next

step is to put X_1 into (2) with using LSP-estimator b_1 for β_1 from (1) [32].

Note 2. However, if there was a change in one of variables X_2, X_3 and $(X_4)^*$ in i-th region, and a user has such information, then he (or she) should use the formula (1) instead of (2) with estimates of $b_0,...,$ b_4 instead of $\beta_0,...,$ β_4 (Table 1).

As a result, we detected LSP-estimates $b_0,...,b_4$ for parameters (1) in early June as per the regions sampling given above; they were equal to 4.6721, 0.4494, −0.04774, −0.05668, −0.3308; σ estimate was equal to 0.03474. Significance for each factor and a free member was lower than 0.001 when calculated via «Data analysis package» in Excel menu. Determination coefficient R^2 adjusted for lack of bias turned out to be higher than 95 % and it was partially due to peculiarities related to creation of a training sampling. Regression residuals by formula (2) were distributed approximately-normally. Their normality was tested with d-statistics that has been described by L.N. Bolshev and N.V. Smirnov [40]. It is calculated as an average value of absolute deviation found as per n observations for the residuals which is taken in a ratio to a root from sample variance. This statistics is distributed approximately normally with the parameters $E(d)$, $var(d)$ according to formulas given in the book [40]. Figure values for statistics of d criterion and its parameters when calculated as per the regression residuals in (2) and the sampling $n = 40$ turned out to be the following: $d = 0.7954$, $E(d) = 0.8029$, $(var(d))^{0.5} = 0.0324$; therefore, a hypothesis on normality is not rejected since $d \approx E(d)$, and deviation from $E(d)$ is lower than ¼ of a standard for this statistics.

Results and discussion. First, let's adjust Table 1 to Note 2 outlined above. A decrease in Y value combined with growth in X_2 is given for 3 regions in it. We should note that «Aventa-M» ALV device which is considered to be rather cheap costs nearby 1.8 million rubles.

Instead of actual ALV devices number in regions available on a day t (Table 1), we took their analogue quantity required as per state standards and calculated X_2 for both cases taking data provided by Rosstat[1] into account [33, 34]. Then, having put X_2 into the formula (1) and

Changes in *Y* occurring due to better provisions of regions with ALV devices

RF region	ALV devices, number	Ministry standard	Improvement in X_2	Decrease in Y	t
Voronezh Region	348	860	15 → 37	2.43 → 1.38	24.04
Tver Region	118	527	9.24 → 41.3	4.7 → 3.84	26.04
Rostov Region	643	1521	15.3 → 35	4.5 → 3.57	05.05

Calculation of constant *C(j)* and mean absolute percentage error δ

Region	$C(j)$ from (2)	$N(j)$, thousand people[1]	Interval as per t	Error δ (%), from (3)	Note
Republic of Karelia	1.060	620.3	06.05–05.07	5.0 (2.9)	0... +1
Republic of Crimea	1.163	1912.8	27.05–02.06	6.2 (2.2)	0
Republic of Tatarstan	2.164	3896.4	30.05–08.07	4.5 (3.9)	0
Arkhangelsk Region	0.120	1149.6	02.06–09.06	3.3 (2.3)	0
Vologda Region	2.074	1172.2	15.05–19.05	5.2 (3.2)	+ 1
Kurgan Region	3.191	840.1	15.05–02.07	5.3 (3.0)	+ 1… +3
Lipetsk Region	1.333	1147.1	03.06–14.06	2.3 (1.6)	+ 1
Orenburg Region	0.901	1970.4	04.05–25.05	5.6 (2.7)	0
Ryazan Region	1.541	1117.8	04.06–16.06	3.5 (2.0)	+ 4
Tambov Region	1.970	1024.7	27.05–14.06	3.0 (2.3)	+ 2
Ulyanovsk Region	1.529	1242.5	07.05–01.06	3.5 (2.2)	+ 1
Chelyabinsk Region	2.10	3484.4	18.05–27.05	1.0 (0.6)	+7…0

having preliminarily replaced β₀,…, β₄ with $b_0,…, b_4$ estimators, we estimated a probable decrease in mortality *Y* at a moment *t* due to growth in X_2. As we can see from Column 5 and calculation made as per the model, a significant decrease in *Y* is expected, from 22 % to 75 %.

The model as per the formula (1) with $b_0, …, b_4$, estimators or the formula (2) can provide satisfactory precision in daily estimates of mortality caused by COVID-19 in certain regions, Table 2 contains data for 12 regions and precision achieved for them as per a sequence of days taken from a time interval as per *t*. Several regions such as Republic of Karelia, Tambov Region and Kurgan Region were not included into training sampling. All the calculations in Table 2 are made as per the formula (2) with a focus on a period of time starting from the second half of May and finishing in July. Column 5 in Table 2 contains mean percentage error (MRPE) was taken modulo:

$$\delta = \frac{100}{K} \cdot \sum_{t=1}^{K} \left| \frac{Y_t - Y_t^*}{Y_t^*} \right|, \qquad (3)$$

which is calculated when *Y* estimates obtained as per the formula (2) are compared with its smoothed true value Y_t^* (%). The calculation was made as per *K* days taken from an interval shown in Column 4, Table 2.

Sample standard deviation found out of the said sequence consisting of *K* days is given in Column 5, Table 2. At the *t*-step in this calculation Y_t estimate as per the formula (2) was compared with a smooth value Y_t^* (with its probable preliminary correction according to Column 6, Table 2; but if there was «0», then data [32] were not changed). «Upward» adjustment in COVID-19 mortality parameter is quite justified in certain regions; therefore, if Column «Note» contains a figure different from 0, it means a number of deaths added to initial data [4–7, 32, 41]. The series of graphs shows dynamics of this mortality in regions built up on data provided by the information center at the RF Public Healthcare Ministry; the graphs are given in comparison with similar ones taken from the official web-site on issues related to coronavirus [32, 41]. They are identical for 31 RF regions; 6 regions provide data with a certain lag; 19 regions sometimes understate certain data in comparison with an investigation performed by «Mediazona» Web-site; 29 regions provide data that make the graphs for

them to be significantly different in many time counts. Values for δ similar to those given in Table 2 can be obtained for the remaining almost 30 regions for a time period «April – the first half of May» (they were included into the training data with earlier counting as per t).

This model allows obtaining precise data in Column 5, Table 2, provided that estimate of a death probability $P*$ has values similar to those from the training data. For example, in Ulyanovsk Region estimate $P*$ was equal to 0.00049 for a time series May 06 – May 17 (taking into account data given in Column 6, Table 2); and it was equal to approximately 0.0004 for a time series May 18 – June 02 for which a prediction as per the formula (2) was performed.

Let us dwell on results obtained for certain regions in greater detail. They were obtained as per the formula (2) for different time series t; graphs are given in Figures 1–4.

Y-axis in each graph shows mortality in j-th region Y. This rates is calculates as a total number of deaths per 1 million people. Figure 1 shows graphs obtained for Republic of Tatarstan. A broken line is a graph built as per the formula (2) for a day t form a time series (May 27, July 07); a solid line shows actual mortality[1] [32].

Then, Figure 2 shows graphs for Arkhangelsk Region[1]. A solid line shows smooth actual data; a dotted line is an exponent adjusted to Y; a broken line is a graph built as per the formula (2) [32].

The next Figure (3) shows graphs for Lipetsk Region. A broken line with dots is a graph built as per the formula (2). Values for a stepwise graphs were calculated as follows: we took a number of deaths, then, starting from April 17, we added one more death to them that probably remained unaccounted, then the sum was multiplied by 10^3 and divided by $N(j) = 1,147.1$ (it is given as «+1» in Column 6, Table 2)[1] [32]. A smooth curve given in Figure 3 is MA or «moving average» and it is calculated as per points in the stepwise graph via a moving centered 7-day window.

Figure 4 shows graphs for Republic of Karelia for a time period April 16 – July 29. Dark-colored dots are obtained via putting X_k values into the formula (2). An exponential trend shown as a thin line is adjusted to these dots.
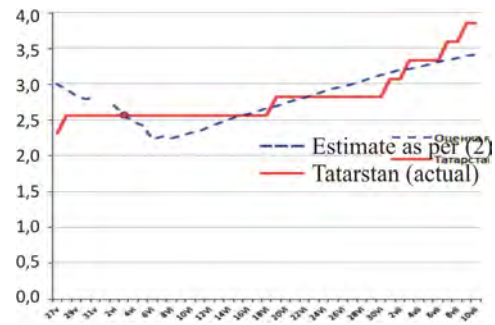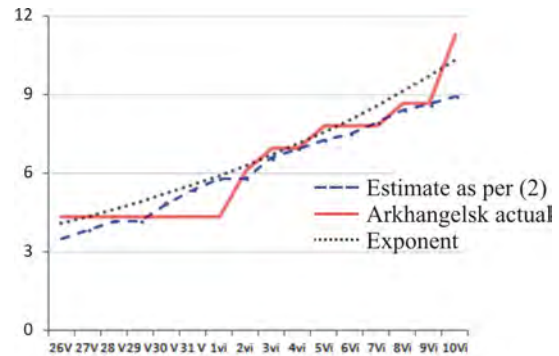

Figure 1. Dynamics in Y: Republic of Tatarstan

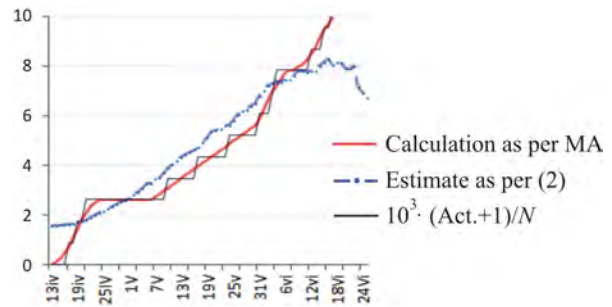Figure 2. Dynamics in Y: Arkhangelsk Region

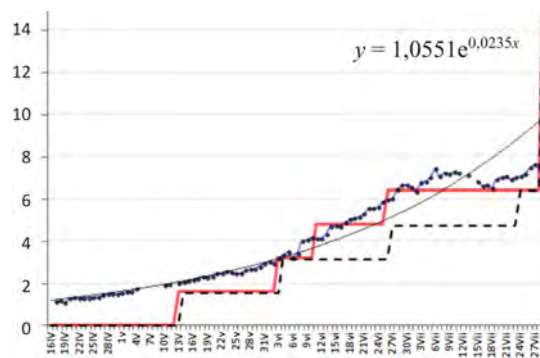Figure 3. Dynamics in Y: Lipetsk Region

Figure 4. Dynamics in mortality Y in Republic of Karelia and its estimate by (2)

The constant in the formula is 1.055 and it corresponds to almost zero prevalence of active COVID-19 cases when $C(j) = 1.06$ from

Table 2. A stepwise line (broken one) shows official mortality; a similar solid line shows Y adjusted as per Column 6 in Table 2. They are a bit shifted in order to avoid merging. Starting from June 11 we added one more death case when calculating an adjusted line which was probably accounted with a certain lag. This additional death case was reported by the regional authorities and it can be seen on «Mediazona» repository [7]. Shifts as per t in publications about death cases are also reported in other sources [6, 7, 26, 41]. We can try to logically substantiate this assumption via analyzing mortality in three regions that have a common border with Republic of Karelia, namely Leningrad Region, Vologda Region, and Arkhangelsk Region. And mortality here should be examined with close X_1, values that were detected in Republic of Karelia during the second decade of June (between values 6.8 and 9). As it can be seen from the graphs given in Figure 4, the model built with the formulas (1) and (2) gives quite reliable estimates of mortality in a time period June 06, … July 05 which is a bit adjusted «upward» (if we draw a line which is smoothed as per moving average instead of a stepwise one). The same was done for Kurgan Region. Approximately similar actions are performed by analytics in some organizations when they adjust data on mortality [5, 7], for example, in Moscow office of «Data Insight» company where specialists working at the research department are building up an index showing mortality underestimation [5].

Conclusion. Thus, we have built up a regression model and tested its precision. The model relates mortality in a given region with prevalence of active COVID-19 cases as well as with provision of hospitals with several necessary resources. It can be useful at initial stages in EP development when there are rare and insignificant growths in mortality in a region over 7–10 days and they have got Poisson distribution. Besides, the model can be useful for building up SEIR, SIRD, and some other models at a stage when mortality-related parameters are selected for them. It can also be applied when distributing resources available to public healthcare organizations.

Conflict of interests. The authors declare there is no any conflict of interests.

References

1. Zhmerenetskii K.V., Sazonova E.N., Voronina N.V., Tomilka G.S., Sen'kevich O.A., Gorokhovskii V.S., D'yachenko S.V., Kol'tsov I.P., Kutsyi M.B. COVID-19: scientific facts. *Dal'nevostochnyi meditsinskii zhurnal*, 2020, no. 1, pp. 5–22 (in Russian).

2. Danilova I.A. Morbidity and mortality from COVID-19. The problem of data comparability. *Demograficheskoe obozrenie*, 2020, vol. 7, no. 1, pp. 6–26 (in Russian).

3. Drapkina O.M., Samorodskaya I.V., Sivtseva M.G., Kakorina E.P., Briko N.I., Cherkasov S.N., Tsinzerling V.A., Mal'kov P.G. COVID-19: urgent questions for estimating morbidity, prevalence, case fatality rate and mortality rate. *Kardiovaskulyarnaya terapiya i profilaktika*, 2020, vol. 19, no. 3, pp. 302–309 (in Russian).

4. Ivanov S. Mortality from COVID-19 against the backdrop of other twentieth century mortality bursts. *Demograficheskoe obozrenie*, 2020, vol. 7, no. 2, pp. 143–151 (in Russian).

5. Kozlovskii S., Boldyrev O. Mozhno li predskazat' razvitie pandemii koronavirusa? Ob"yasnyaem na primere Rossii [Can we predict how coronavirus pandemic will develop? Let us explain using Russia as an example]. *BBC Russia*. Available at: https://www.bbc.com/russian/features-52762747 (30.07.2020).

6. Glava Rosstata rasskazal, kak schitayut zhertv COVID-19 [Head of Rosstat has told how victims of COVID-19 are calculated in the country]. *BBC Russia*. Available at: https://www.bbc.com/russian/features-53156041 (02.08.2020).

7. Koronavirus v Rossii: Infografika [Coronavirus in Russia: graphic data]. *Mediazona*, 2020. Available at: https://zona.media/coronagraph (30.07.2020).

8. Boyarintsev V.V., Pal'min R.S., Pal'min S.A., Pertsev S.F. Nique for predicting parameters of the epidemic process caused by COVID-19. *Kremlevskaya meditsina. Klinicheskii vestnik*, 2020, no. 2, pp. 14–21 (in Russian).

9. Kondrat'ev M.A. Forecasting methods and models of disease spread. *Komp'yuternye issledovaniya i modelirovanie*, 2013, vol. 5, no. 5, pp. 863–882 (in Russian).

10. Magpantay F.M.G., Kosovalić N., Wu J. An age-structured population model with state-dependent delay: derivation and numerical integration. *SIAM Journal on Numerical Analysis*, 2014, vol. 52, no. 2, pp. 735–756. DOI: 10.1137/120903622

11. Khrapov P.V., Loginova A.A. Comparative analysis of the mathematical models of the dynamics of the coronavirus COVID-19 epidemic development in the different countries. *International Journal of Open Information Technologies*, 2020, vol. 8, no. 5, pp. 17–22.

12. Minaev V.A., Sychev M.P., Vaits E.V., Bondar' K.M. System-dynamic modeling of network information operations. *Inzhenernye tekhnologii i sistemy*, 2019, vol. 29, no. 1, pp. 20–39 (in Russian).

13. Drugova O.V., Pavlov E.A., Bavrina A.P., Blagonravova A.S., Saperkin N.V., Kovalishena O.V. Statistic and dynamic aspects of the prediction of the COVID-19 spread in Nizhny Novgorod region. *Meditsinskii al'manakh*, 2020, no. 2 (63), pp. 27–36 (in Russian).

14. Tamm M.V. COVID-19 in Moscow: prognoses and scenarios. *Farmakoekonomika. Sovremennaya farmakoekonomika i farmakoepidemiologiya*, 2020, vol. 13, no. 1, pp. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51

15. Matveev A.V. The mathematical modeling of the effective measures against the COVID-19 spread. *Natsional'naya bezopasnost' i strategicheskoe planirovanie*, 2020, vol. 1, no. 29, pp. 23–39 (in Russian).

16. Godio A., Pace F., Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. & Public Health*, 2020, vol. 17, no. 10, pp. 3535. DOI: 10.3390/ijerph17103535

17. Peng L., Yang W., Zhang D., Zhuge C., Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *Med Rxiv Epidemiol*, 2020, 11 p. (in Russian).

18. Pengpeng S., Shengli C., Peihua F. SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration. *Med Rxiv*, 2020, no. 20, 5 p. DOI: 10.1101/2020.02.16.20023655

19. Nikitina A.V., Lyapunova I.A., Dudnikov E.A. Study of the spread of viral diseases based on modifications of the SIR model. *Computational mathematics and information technologies*, 2020, vol. 1, no. 1, pp. 19–30. DOI: 10.23947/2587-8999-2020-1-1-19-30

20. Liu Z., Magal P., Seydi O., G.F. Webb Predicting the Cumulative Number of Cases for the COVID-19 Epidemic in China From Early Data. *Populations and Evolution*, 2020, vol. 1, no. 10, 10 p. DOI: 10.20944/preprints202002.0365.v1

21. Ndaïrou F., Area I., Nieto J.J., Torres D.F.M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*, 2020, vol. 135, no. 6, pp. 109846. DOI: 10.1016/j.chaos.2020.109846

22. Kurkin A.A., Kurkina O.E., Pelinovskii E.N. Logistic models of epidemic growth. *Trudy NGTU im. R.E. Alekseeva*, 2020, no. 2 (129), pp. 9–18 (in Russian).

23. Fabiano N., Radenović S.N. On COVID-19 diffusion in Italy: data analysis and possible outcome. *Vojnotehnički glasnik*, 2020, vol. 68, no. 2, pp. 216–224. DOI: 10.5937/vojtehg68-25948

24. Cherniha R., Davydovych V. A mathematical model for the coronavirus COVID-19 outbreak. *ArXiv*, 2020. Available at: https://arxiv.org/abs/2004.01487v2 (25.07.2020).

25. Kol'tsova E.M., Kurkina E.S., Vasetskii A.M. Mathematical modeling of the spread of COVID-19 in Moscow. *Computational nanotechnology*, 2020, vol. 7, no. 1, pp. 99–105 (in Russian).

26. Melik-Guseinov D.V., Karyakin N.N., Blagonravov A.S., Klimko V.I., Bavrina A.P., Drugova O.V., Saperkin N.V., Kovalishena O.V. Regression models predicting the number of deaths from the new coronavirus infection. *Sovremennye tekhnologii v meditsine*, 2020, vol. 12, no. 2, pp. 6–13 (in Russian).

27. Lakman I.A., Agapitov A.A., Sadikova L.F., Chernenko O.V., Novikov S.V., Popov D.V., Pavlov V.N., Gareeva D.F. [et al.]. COVID-19 mathematical forecasting in the Russian Federation. *Arterial'naya gipertenziya*, 2020, vol. 26, no. 3, pp. 288–294 (in Russian).

28. Fontes E. Modelirovanie v COMSOL Multiphysics rasprostraneniya virusa COVID-19 [Modeling the spread of the COVID-19 virus in COMSOL Multiphysics]. *COMSOL*, 2020. Available at: https://www.comsol.ru/blogs/modeling-the-spread-of-covid-19-with-comsol-multiphysics/ (30.07.2020) (in Russian).

29. Getz W.M., Dougherty E.R. Discrete stochastic analog of Erlang epidemic models. *J. of Biological Dynamics*, 2018, vol. 12, no. 1, pp. 16–38. DOI: 10.1080/17513758.217.1401677

30. The CDC portal: Forecasts of total deaths at the USA. *Centers for Disease Control and Prevention*, 2020. Available at: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html (15.07.2020).

31. My izuchili 858 vashikh voprosov pro koronavirus [We have examined 858 questions you have about coronavirus]. *Meduza*, 2020. Available at: https://www.meduza.io/feature/2020/05/13 (15.07.2020) (in Russian).

32. Ofitsial'nyi internet-resurs po voprosam koronavirusa [The official web-site on issues related to coronavirus]. *Stopkoronavirus.rf*, 2020. Available at: https://стопкоронавирус.рф (17.07.2020) (in Russian).

33. Apukhtina Yu., Zobova S. Issledovanie o tom, skol'ko bol'nichnykh koek mozhet spasti rossiiskuyu meditsinu [Research on a number of available beds in hospitals that can save public healthcare in Russia]. *Proekt*, 2020. Available at: https://www.proekt.media/research/koronavirus-regiony (01.07.2020) (in Russian).

34. Sokolov A.A. Gotovo li rossiiskoe zdravookhranenie k bor'be s koronavirusom [Is public healthcare in Russia ready to fight coronavirus?]. *Vedomosti*, 2020. Available at: https://www.vedomosti.ru/society/articles/2020/04/09/827471-gotovo-rossiiskoe (02.12.2020) (in Russian).

35. Free and occupied beds for COVID-19 patients in Germany. *Coronavis*, 2020. Available at: https://coronavis.dbvis.de/en/ (01.07.2020).

36. Murray C.J.L. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*, 2020, no. 30, 26 p. DOI: 10.1101/2020.03.27.20043752

37. COVID caseload calculator C5V. *Weill Cornell Medicine*, 2020. Available at: https://phs.weill.cornell.edu/cornell-covid-caseload-calculator-c5v (01.08.2020).

38. Aivazian S.A. Quality of life and living standards analysis: an econometric approach. Berlin/Boston, De Gruyter Publ., 2016, 399 p. DOI: 10.1515/9783110316254

39. Stepanov V.S. integral indicator of the living conditions at the Crimea Republic and some other territories: dependence from factorial variables. *Vestnik TsEMI*, 2019, no. 2, pp. 8. DOI: 10.33276/S265838870004976-6

40. Bol'shev L.N., Smirnov N.V. Tablitsy matematicheskoi statistiki [Mathematical statistics tables]. 3-th edition. Moscow, Nauka Publ., 1983, 416 p. (in Russian).

41. S aprelya v koronavirusnykh bol'nitsakh umerlo ne men'she 74,9 tysyach chelovek. Kak region skryvayut eti dannye [Not less than 74.9 thousand people have died in coronavirus hospitals since April. How do regions manage to hide these data?]. *Mediazona*, 2020. Available at: https://zona.media/news/2020/11/30/75k (02.12.2020).