International Journal of Mathematical, Engineering and Management Sciences
Vol. 5, No. 5, 812-819, 2020
https://doi.org/10.33889/IJMEMS.2020.5.5.063

COVID-19 Highest Incidence Forecast in Russia Based on Regression Model

Iosif Z. Aronov
Department of Commerce and Trade Regulation,
MGIMO (Moscow State Institute of International Relations) University, Moscow, Russia.
E-mail: izaronov@itandi.ru

Olga V. Maksimova
Department of Global Climate Stabilization Research,
Yu. A. Izrael Institute of Global Climate and Ecology, Moscow, Russia.
E-mail: o-maximova@yandex.ru

Nataliia M. Galkina
Department of Trade Barriers Analysis,
International Trade and Integration (ITI) Research Center, Moscow, Russia.
Corresponding author: nmgalkina@itandi.ru

(Received May 8, 2020; Accepted June 12, 2020)

Abstract
The authors suggest a simple regression model of COVID-19 highest incidence prognosis in Russia on the basis of the revealed correlation between the duration of coronavirus peak (plateau) and air traffic volume. The study base included 37 countries in Europe, South America and Asia. Cluster analysis on the basis of the Euclidean metric for these countries showed the necessity of classifying the USA and China into a separate group, which gave grounds to exclude these countries from the analysis. In addition, Ireland was excluded from the analysis due to its special geographical location. For the remaining countries, the correlation coefficient between the number of airline passengers and the duration of the epidemic before reaching its peak was 0,87, which shows a high level of linear relationship between these indicators. Point forecast for the highest incidence in Russia by regression line falls on the 4th of May. The forecast interval with confidence levelγ=0.9 is ±14 days from the calculated date. The one-way analysis of variance showed that from April 22 to May 2, there was a slowdown in the growth rates of the diseased, which indicates an exit to the plateau.

Keywords- COVID-19, Regression model, Forecast, Peak (plateau).

1. Introduction and Statement of the Problem

In December 2019, a new type of coronavirus (SARS-CoV-2) was discovered, which causes acute pneumonia in humans. An outbreak of this disease was registered in mainland China (Wuhan city). According to the World Health Organization (WHO), as of April 26, COVID-19, which is caused by this virus, has been exported to 210 countries and territories, including Russia.

The significant impact on the economies of these countries caused by the SARS-CoV-2 epidemic stimulated coronavirus research related to different aspects of virus communication. For example, Chen et al. (2020) suggested a virus extension model from its source (bats) to humans, called Bats-Hosts-Reservoir-People. The model was identified from early data on coronavirus extension in China.

Krantz and Rao (2020) obtained preliminary retrospective results based on wavelets and deterministic modeling in relation to the peak of COVID-19 in China, USA, South Korea, etc.

Koo et al. (2020) justified activities to mitigate the early spread of SARS-CoV-2 in Singapore, based on an adapted simulated influenza epidemic model. Measures included isolation for infected persons and quarantine of family members; quarantine plus school closures; quarantine plus social distance; quarantine plus school closures plus social distance at the workplace.

Kucharski et al. (2020) examined the dynamics of coronavirus transmission based on a stochastic model in 20 countries involved with Wuhan.

The results of a large study on the spread of the virus in China, taking into account undetected cases, are presented in the article of B. Ivorra (Ivorra et al., 2019).

Rabajante (2020) provides an overview of some mathematical probabilistic models of the new virus communication, which is in the author's opinion, admissible for predicting the disease development in the Philippines.

Finally, Sethy et al. (2020) discuss the coronavirus diagnostic radiography method.

It is important to note that all presented models, firstly, refer to the early coronavirus period, and secondly, assume a large number of input data for calculation.

Meanwhile, as the coronavirus spreads in the country (e.g. Russia), it is important to estimate the t_p exit time of the number of infected patients to the "plateau" in order to plan measures for the country's recovery from the crisis, which makes it possible to hope for a relatively soon end of the epidemic, the abolition of the regime of self-isolation (quarantine) and the restart of economies, i.e. to improve the quality of life.

Herein under "plateau" shall be understood the stage of virus communication, when the highest incidence is reached, and then this index remains approximately on the same level. Entering the plateau means when the number of new cases during some period remains on average constant.

2. Correlation Model

Taking into account that in Russia in comparison with other countries (China, Singapore, European Union countries, etc.) faced the spread of the infection a little later (only on 2 March 2020 the first patient infected with coronavirus was identified), one can use the information about the time of reaching the plateau (peak) of coronavirus disease in those countries where the epidemic has decreased.

This fact was the reason for the formation of a rather simple statistical analysis model of the t_p period before the plateau in Russia.

It should be noted that a statistical analysis of coronavirus development dynamic in the country currently considers such a characteristic as the number of infected per million inhabitants. Unfortunately, this characterization is full of uncertainties associated, for example, with the vagueness of coronavirus diagnosis and other medical factors.

The authors decided to move away from quantitative estimation of the maximum number of detected diseased prognosis in different countries. The dynamics of the diseased number and its' increase is due to a multitude of factors, among which both the measures taken by Governments to prevent morbidity, which is crucial, and the response of the population to the measures taken. Indeed, this affects the future scenario of the epidemic in a country after reaching the plateau, but does not change the overall picture of the increase in disease itself, which is exponential.

A study conducted according to primary data among 30 countries in mid-March by Aronov and Maksimova (2020) allowed to make an assumption that in the countries with high activity of citizens' movement the epidemic has a longer - term character. In part, the high activity of the country's citizens reflects its connection with the high quality of life and partly with the economic situation of the country among other countries. Therefore, the authors decided to study not the number of the diseased in the country by a certain day, but, as it was said earlier, the duration of the epidemic from the moment the first diseased in the country was discovered to the highest incidence t_p.

At the time of the forecast as of the 2nd of May, many countries have already passed the epidemic peak, that broadens database for the study. The Johns Hopkins University website collected data on the variable *tp* for 37 countries and data on the activity of citizens' movement (air, rail and road transport volumes of citizens during the year), as well as GDP per capita for the same countries. The data contained in the international passenger transport databases are for 2018. Also, Aronov and Maksimova (2020) revealed that a weak correlation between the chosen *tp* variable is observed with the number of population and the area occupied by the country, which we see as crucial in interpreting. It is also noted that the coefficient of correlation of the *tp* variable "volume of road and rail conveyance of passengers, expressed in the value of million passenger-kilometers)", was small, which confirms the expediency of using in the calculation of the indicator "number of passengers carried by air transport". This fact confirms the assumption that the duration of the epidemic before reaching the peak is related to the quality of life of the country's citizens, expressed in part in the ability of people to travel by air transport.

The air transport industry is an important engine of global socio-economic growth. It is vital for economic development, the creation of direct and indirect employment, support for tourism and local business, and the promotion of foreign investment and international trade. The data used for air transport represent the total (international and domestic) planned number of transportations by air carriers registered in each country. For statistical research, departures are equal to the number of landings or flight stages performed. Countries submit data on air transport to the International Civil Aviation Organization (ICAO) based on standard instructions and definitions issued by ICAO.

Due to the fact that the highest incidence in Russia has not yet passed (at the time of conducting research), it was not included in the general study. Indicators for Ireland were also not taken into account due to its special geographical location and the use of the Dublin airport (a disproportionate flow of air passengers with a small population) and some other countries (e.g. Africa, Ukraine, Belarus) for which there is no correct data on the selected indicators.

Cluster analysis for these countries on the basis of the Euclidean metric by the method of finding the nearest neighbor (dendrogram build on two factors: the duration of the disease development to a peak and the number of passengers carried by the air transport for a year in the country), showed the need to include the U.S. and China in a separate group, which allowed to exclude them from further analysis as Figure 1 shows.

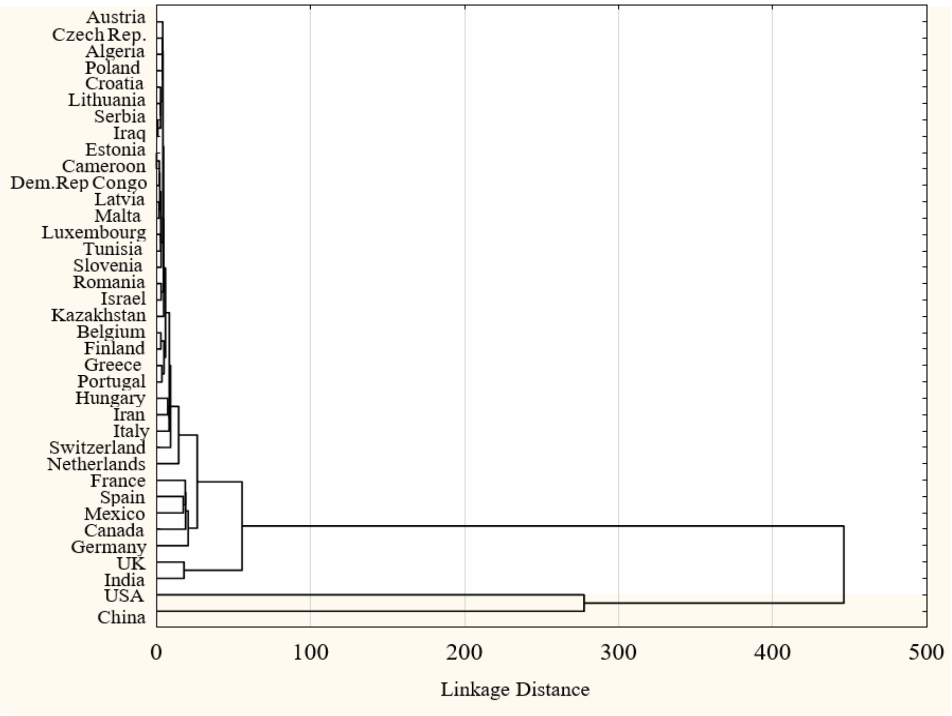

Figure 1. Dendrogram for 37 countries regarding the indicators: number of passengers carried by air transport across the country, and the epidemic duration to its peak

For the remaining countries, the correlation coefficient between the number of air passengers and the duration of the epidemic before reaching its peak was 0,87, indicating a high level of linear relationship between these indicators. Thus, the determination coefficient was R^2=0.75, i.e. 75% of the data variability is due to the linear relation and the remaining 25% is unaddressed factor.

3. Time Forecast to Reach the Peak (Plateau) of the Disease Based on Regression

Figure 2 shows the obtained regression between the selected indicators: the number of transported passengers by air (horizontal axis, OX) and the epidemic duration to a peak in days (vertical axis, OY), the quality of which is confirmed by checking the importance of its coefficients. It should be noted that for countries with low passenger activity, there is a greater spread of points (condensation of points on the left) compared to the number of countries with high activity. For the latter countries, the proximity of points to the regression line is high as Figure 2 shows.

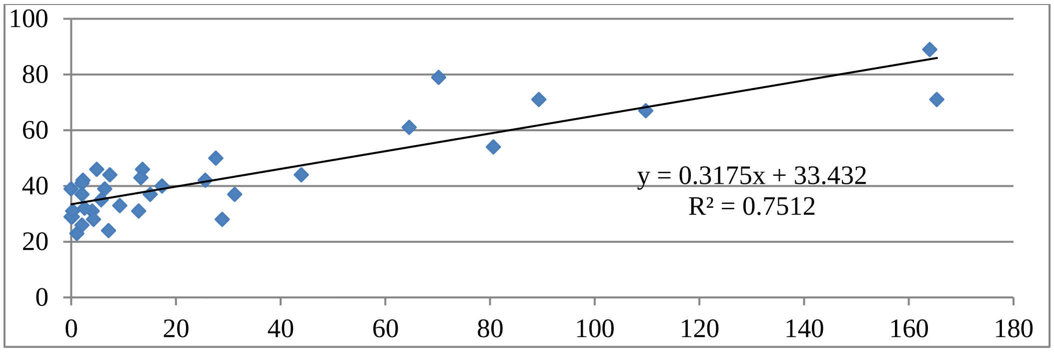

Figure 2. Linear regression between the number of passengers (OX-axis) carried by air in 2018 and the epidemic duration (OY-axis) to peak (countries excluding China, USA and Russia are included)

The list of countries with high activity in the regression analysis is the following: India, France, Germany, Great Britain, Spain, Mexico, Canada, and Netherlands. It should be noted that Russia is in the upper quartile by the number of air transportation (99.3 million passengers carried by air). This gives grounds to assume a rather high quality of possible forecast.

Point prognosis of highest incidence in Russia by regression line is on the 64th day from the key mark of 2nd of March, which leads to the date of the 4th of May.

To improve the quality of prognosis, a forecast on highest incidence interval has been built. With a probability belief $\gamma=0.9$ the calculations give a deviation of ±14 days from the calculated date. The resulting forecast displays the passage of peak in Russia at an interval from April 20 to May 18 as Figure 3 shows.

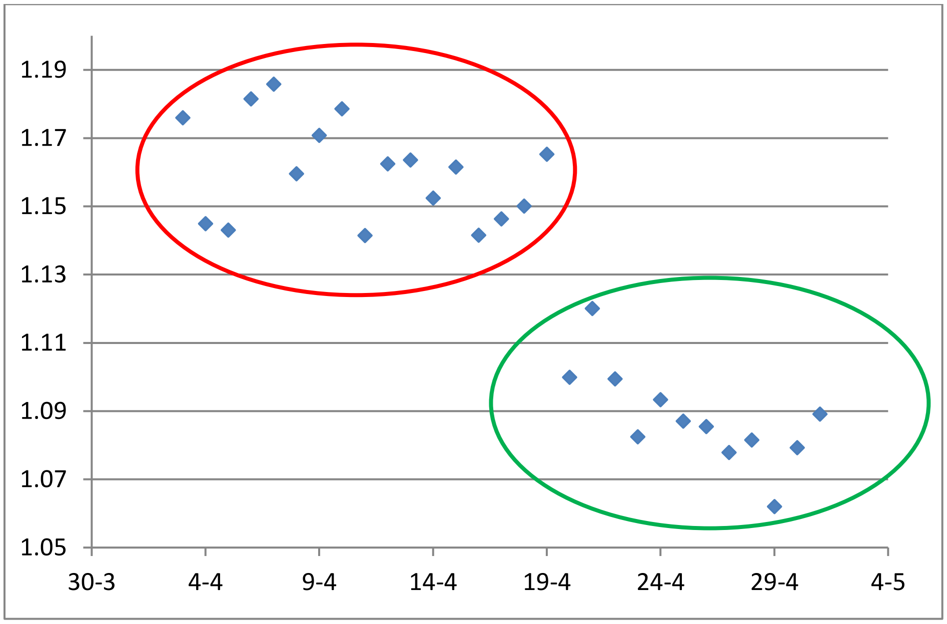

Figure 3. Rate of growth correspondence (OY axis) in the number of COVID-19 cases in Russia by days

The relatively wide forecast interval is due to a significant deviation in the number of passengers carried by air transport in Russia compared to the average performance for all countries included in the research.

4. Dispersion Analysis

It was noted that the increase in the number of cases was exponential, to describe which it was convenient to apply the rate of increase in the number of cases. At the moment of writing this paper the given forecast fixes a peak of morbidity which from the mathematical point of view should be characterized by a slowdown of growth rates in number of diseased.

Primary analysis of the growth rate

$$T_i = \frac{S_i}{S_{i-1}}, \tag{1}$$

where, S_i −number of cases in the country by the time of the i −day, shows their reduction since April 22 (see the selected groups in Figure 3). Breaking down all the growth rates of the cases number from the beginning of statistics on morbidity in Russia until May 2, 2020 to determine the differences, the authors use done-way analysis of variance with the selected level of significance α = 0.05 represented in Table 1.

Table 1. Analysis of variance of groups distributions

Variable	Analysis of Variance Marked effects are significant at *p* < 0.05000							
	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	*F*	*p*
Ratesofgrowth	0.188	1	0.188	0.770	41	0.019	10.022	0.003

For Fisher *F* statistics based on comparison of dispersion caused by intergroup MSEffect and dispersion caused by intergroup MSError, the probability p=0.003 is calculated. The obtained value is much lower than the accepted value level, which may indicate statistically significant differences among the groups. Graphic interpretation of the performed analysis is presented in Figure 4, the analysis of which shows the confirmation of the assumption about the significant slowdown of the rate of growth in the number of the diseased that testifies to the fact that Russia is coming to the plateau by the number of diseased.

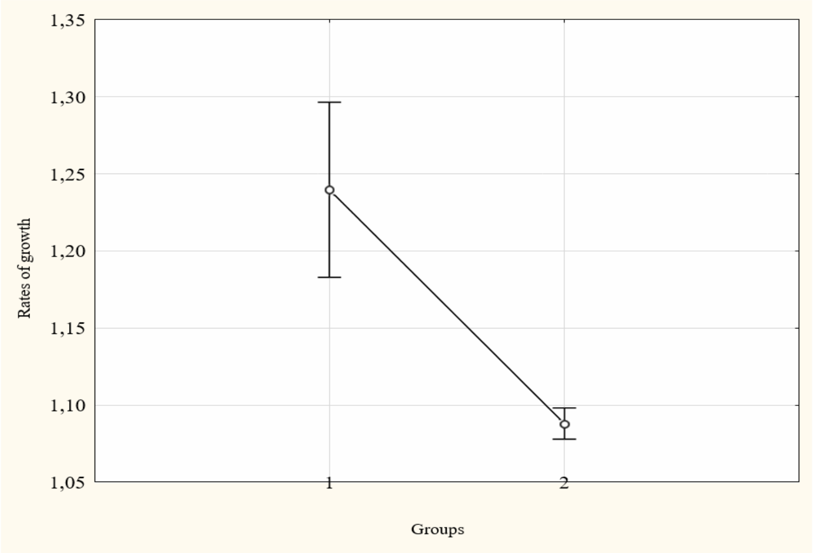

Figure 4. Graphs of average values and their 95% confidence intervals for each group of growth rates of the number of diseased in Russia by the date of May 2, 2020 (1 - data group of growth rates up to 22.04, 2 - data group of growth rates from 22.04 to 2.05.2020)

5. Conclusion

The obtained preliminary forecast of the coronavirus peak duration in Russia might be considered as one of the possible. In addition, the presence of several peaks (plateaus) is not excluded, which is due to both inconsistent measures of the authorities and the behavior of individuals, effecting the epidemic in the country.

The forecast does not take into account socio-epidemiological and medical factors (e.g. type of virus, spread of the virus over a large territory of Russia, etc.) and also does not show the epidemic picture in the country's regions as Table 2 shows.

Table 2. Number of the revealed individuals, diseased COVID-19 in Russia during the period from May 3 to May 18.

Date	3.05	4.05	5.05	6.05	7.05	8.05	9.05	10.05
Number of diseased	9623	10581	10102	10559	10533	10699	10817	11012
Date	11.05	12.05	13.05	14.05	15.05	16.05	17.05	18.05
Number of diseased	11656	10899	10028	9974	10598	9200	9709	8926

It should be noted that between the date of writing the article (May 2) and the date of its revision (May 18) additional data became available. Table 2 contains data on the number of individuals who were exposed to COVID-19 in Russia from May 3 to May 18 inclusive. These data indicate a high quality of the forecast: a plateau has been observed in the period from May 3 to May 18.

Scientists from the Singapore University of Technology and Design have published a study of the development cycles of COVID-19 in 131 countries using the SIR-model which also predicted one of the possible dates when Russia will peak disease (April 26-27). How the epidemic will decline is a question and monitoring of the near future that awaits us. Apparently, the forecast presented in this article has been more effective.

Conflict of Interests
The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgements
The authors sincerely thank anonymous reviewers and the Editor-in-Chief for providing critical comments and suggestions for improving the quality of the paper.

References

Aronov, I., & Maksimova, O. (2020). Life quality and prognosis of COVID-19 peak morbidity. Available at: https://ria-stk.ru/stq/adetail.php?ID=187663.

Chen, T., Rui, J., Wang, Q.P., Zhao, Z.Y., Cui, J.A., & Yin, L. (2020). A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infectious Diseases of Poverty*, *9*, 24. https://doi.org/10.1186/s40249-020-00640-3.

Ivorra, B., Ferrández, M.R., Vela-Pérez, M., & Ramos, A.M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, *88*. https://doi.org/10.1016/j.cnsns.2020.105303.

Koo, R.J., Cook, A.R., Park, M., Sun, Y., Sun, H., & Lim, J.T. (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Available at: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30162-6/fulltext.

Krantz, S.G., & Rao, A.S.S. (2020). Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: preliminary retrospective results based on wavelets and deterministic modeling. *Infection Control & Hospital Epidemiology*, 1-3. https://doi.org/10.1017/ice.2020.116.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, *20*(5), 553-558.

Rabajante, J.F. (2020). Insights from early mathematical models of 2019-nCoV acute respiratory disease (COVID-19) dynamics. *ArXiv Preprint ArXiv:2002.05296*.

Sethy, P.K., Behera, S.K., Ratha, P.K., & Biswas, P. (2020). Detection of coronavirus disease (Covid-19) based on deep features and support vector machine. *International Journal of Mathematical, Engineering and Management Sciences*, *5*(4), 643-651.