

Original Paper

A Recursive Model of the Spread of COVID-19: Modelling Study

Sergey O Ilyin, PhD

AV Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences, Moscow, Russian Federation

Corresponding Author:

Sergey O Ilyin, PhD

AV Topchiev Institute of Petrochemical Synthesis, Russian Academy of Sciences

29 Leninsky prospekt

Moscow, 119991

Russian Federation

Phone: 7 9168276852

Email: s.o.ilyin@gmail.com

Abstract

Background: The major medical and social challenge of the 21st century is COVID-19, caused by the novel coronavirus SARS-CoV-2. Critical issues include the rate at which the coronavirus spreads and the effect of quarantine measures and population vaccination on this rate. Knowledge of the laws of the spread of COVID-19 will enable assessment of the effectiveness and reasonableness of the quarantine measures used, as well as determination of the necessary level of vaccination needed to overcome this crisis.

Objective: This study aims to establish the laws of the spread of COVID-19 and to use them to develop a mathematical model to predict changes in the number of active cases over time, possible human losses, and the rate of recovery of patients, to make informed decisions about the number of necessary beds in hospitals, the introduction and type of quarantine measures, and the required threshold of vaccination of the population.

Methods: This study analyzed the onset of COVID-19 spread in countries such as China, Italy, Spain, the United States, the United Kingdom, Japan, France, and Germany based on publicly available statistical data. The change in the number of COVID-19 cases, deaths, and recovered persons over time was examined, considering the possible introduction of quarantine measures and isolation of infected people in these countries. Based on the data, the virus transmissibility and the average duration of the disease at different stages were evaluated, and a model based on the principle of recursion was developed. Its key features are the separation of active (nonisolated) infected persons into a distinct category and the prediction of their number based on the average duration of the disease in the inactive phase and the concentration of these persons in the population in the preceding days.

Results: Specific values for SARS-CoV-2 transmissibility and COVID-19 duration were estimated for different countries. In China, the viral transmissibility was 3.12 before quarantine measures were implemented and 0.36 after these measures were lifted. For the other countries, the viral transmissibility was 2.28-2.76 initially, and it then decreased to 0.87-1.29 as a result of quarantine measures. Therefore, it can be expected that the spread of SARS-CoV-2 will be suppressed if 56%-64% of the total population becomes vaccinated or survives COVID-19.

Conclusions: The quarantine measures adopted in most countries are too weak compared to those previously used in China. Therefore, it is not expected that the spread of COVID-19 will stop and the disease will cease to exist naturally or owing to quarantine measures. Active vaccination of the population is needed to prevent the spread of COVID-19. Furthermore, the required specific percentage of vaccinated individuals depends on the magnitude of viral transmissibility, which can be evaluated using the proposed model and statistical data for the country of interest.

(*JMIR Public Health Surveill* 2021;7(4):e21468) doi: [10.2196/21468](https://doi.org/10.2196/21468)

KEYWORDS

epidemiology; COVID-19; model; modelling; prediction; spread; infection; effective; contagious; transmission

Introduction

The first mathematical models to predict the development of infectious diseases were used in the early 20th century [1,2]. In

1927, Kermack and McKendrick [3] proposed the use of differential equations for calculations, dividing the human population into people susceptible to disease (S) and those who had already recovered (R). The susceptible persons became

infected (I) at some rate of transmission and then recovered at a different rate. Their model became known by the acronym SIR, which means that the model simultaneously calculates the number of susceptible, infected, and recovered persons. This model served as a basis for the development of subsequent models—by modifying the equations and adding to the calculation other persons not belonging to the three specified basic categories, which allowed consideration of the features of particular diseases. Since then, various models have been created that consider the possibility of re-infection (SIS model) [4] and death (SIRD model) [5], the existence of an incubation period (SEIR model) [6], and temporary immunity of infants (MSIR model) [7], among others.

When a new infection appears, neither the set of population categories to be considered in the model nor the rate of transition of people from one category to another is known. Current information about the features of the COVID-19 infection caused by the novel coronavirus (SARS-CoV-2) and the manner in which people perceive it and act should serve as a basis for building a model to describe the spread of this virus. These features can be described as follows: first, the presence of a long incubation period, during which the infected persons are contagious to others, and second, the isolation of discovered infected persons, which as a result become conditionally noncontagious. The combination of these two factors makes this novel coronavirus infection unique. In general, the opposite is true—infected people are not dangerous to others during the incubation period and become contagious after its expiry. For this reason, a new model that considers these circumstances is needed to predict the spread of COVID-19. However, the duration of the immunity produced after recovery from COVID-19 is currently unknown. In addition, there is also very little information available to accurately calculate the rate of recovery among patients with COVID-19: a small percentage of the population recovers within just a week after contracting infection, whereas the majority of people experience the illness for a long time. Therefore, the proposed model cannot be final, but it is necessary for forecasting and management decisions.

Methods

The model for COVID-19 spread is based on a set of parameters whose values are unique for each country due to differences in population density, human behavior, date of virus penetration, and government actions. The set includes the following parameters:

- d_0 is the date of the initiation of the epidemic; it is not the date of detection of the first infected person but the date of appearance of the first undetected (or detected too late) person.
- d_1 , d_2 , and d_3 are dates of change in the behavior of the population, for example, due to the awareness of the reality of what is happening and the introduction of quarantine and its tightening.
- t_D is the average time from infection to isolation of the infected person, which is equal to the incubation period assumed to be 6 days (ranging from 5.2 to 6.4 days according to different sources [8,9]); theoretically, this

parameter can be reduced by testing of the entire population, but it is feasible only for small communities.

- R_0 , R_1 , R_2 , and R_3 are the viral transmissibilities that are equal to the average number of people who will be infected by one person before he or she is isolated and depend on the behavior of the population at different stages of the epidemic; when R is less than 1.0, the epidemic fades, and vice versa.
- r_0 , r_1 , r_2 , and r_3 are the reduced viral transmissibilities that are equal to the average number of people who will be infected by one person per day: $r = R/t_D$; to suppress the spread of COVID-19, r should be less than 0.167.

The evaluation of the spread of the virus is based on the calculation of the following data:

- $N_D(d_i)$ is the number of infected persons detected on d_i date, which equals the total number of infected persons 6 days earlier:

$$N_D(d_i) = N_T(d_i - t_D)$$

- $N_T(d_i)$ is the total number of infected persons on date d_i , which is the sum of the total number of infected persons the day before and the number of new infected persons that, in turn, is equal to the product of the reduced transmissibility and the number of active infected persons the day before (taking into account that those who have been previously infected cannot be reinfected):

$$N_T(d_i) = N_T(d_i - 1) + r_0 \times N_A(d_i - 1) \times [1 - N_T(d_i - 1)/N_P],$$

where N_P is the total population.

In the case of vaccination of the population and considering the temporary nature of the immunity received due to SARS-CoV-2 infection or vaccination, the above expression will be as follows:

$$N_T(d_i) = N_T(d_i - 1) + r_0 \times N_A(d_i - 1) \times [1 - [N_T(d_i - 1) + N_V(d_i - 1) - N_T(d_i - 1 - t_{im}) - N_V(d_i - 1 - t_{im})]/N_P],$$

where t_{im} is the average duration of preserving full immunity against the virus after vaccination or disease, whereas $N_V(d_i)$ is the total number of vaccinated persons on date d_i who have not had COVID-19 in the last t_{im} days;

- $N_A(d_i)$ is the total number of active (undetected) infected persons on date d_i , which equals the difference between the total number of infected persons and the number of infected persons detected on the same day:

$$N_A(d_i) = N_T(d_i) - N_D(d_i).$$

At the start of the epidemic (date d_0), $N_A(d_0) = 1$, $N_T(d_0) = 1$, and $N_D(d_0) = 0$.

Thus, in order to calculate the virus spread dynamics, it is necessary to know the values of only two parameters— d_0 and r_0 . In the case of changing the behavior of the population from the date d_1 , parameter r_0 changes its value from this date to become r_1 . If the behavior changes again, a pair of d_2 and r_2 will appear, and so on.

However, it is more difficult to model human losses correctly. Two more parameters need to be considered:

- L is the apparent lethality rate that is equal to the ratio of the number of deaths to the sum of those who died or recovered;
- t_L is the average time from infection to death.

These two parameters depend on the efficacy of treatment and may vary as physicians gain experience and as hospitals overflow. The number of deaths on date d_i equals the total number of people infected t_L days earlier multiplied by the lethality rate:

$$N_L(d_i) = N_T(d_i - t_L) \times L$$

Because of the presence of two parameters (t_L and L) in the equation, which have the same effect on the resulting value, the precision of their evaluation is lower than that for viral transmissibility. It should be understood that the fewer the number of asymptomatic and mild cases of the disease have been detected, the more the lethality rate is overestimated. The average time from infection to death was found to be about 8 days, and this duration will be used to make calculations for all countries.

The situation with predicting the number of recovered persons is even worse due to the appearance of an even greater number of independent parameters:

$$N_R(d_i) = N_T(d_i - t_M) \times k_M + N_T(d_i - t_S) \times k_S$$

where k_M and k_S are the shares of mildly and seriously ill patients ($k_M + k_S + L = 1$), and t_M and t_S are the corresponding times from infection to healing:

$$k_M + k_S + L = 1$$

The model equations are presented in the discrete form (instead of differential one), so that the model can be easily reproduced for calculations in any spreadsheet editor. At first glance, it seems that the model does not take into account the existence of asymptomatic carriers of infection, but this is not true: since

the share of asymptomatic carriers in the population does not change over time, their presence is taken into account implicitly by the value of the transmissibility. This model can be denoted by the abbreviation SILRD, which means that it takes into account Susceptible, Infected, Isolated, Recovered, and Dead persons.

Results

Based on historical data on disease development in eight countries (China, Italy, Spain, the United States, the United Kingdom, Japan, France, and Germany [10]), the model was tested (Figure 1) and most of its parameters were found (Table 1). For all the countries, the viral transmissibility at the start of the epidemic was between 2.28 and 3.12. The highest viral transmissibility was found in China, wherein one person infected three others, probably because of higher population density. The introduction and progressive strengthening of quarantine measures resulted in a decrease in the viral transmissibility, which was noticeable 6 days later in the decline in the rate of new cases. All the countries had introduced quarantine measures gradually. The initial restrictions reduced the viral transmissibility to 1.20-1.74, which was not adequate (it was necessary to achieve a transmissibility of less than 1.0), and the virus continued to spread with acceleration. As a result, all the countries, with the exception of Japan, initiated stricter measures, thus reducing the transmissibility to 0.87-1.14. Japan focused on timely detection and isolation of infected persons. It can be concluded that this strategy does not work, as can be observed from the curve of the total number of cases in Japan that alternately slows down and then accelerates again. This is the result of the fact that Japan has been successfully isolating most of the infected persons, but a few infected people remain nonisolated and they can cause another outbreak to occur. By contrast, China further strengthened its containment measures, which resulted in a reduction of the transmissibility to 0.36 and a quick win over the epidemic (in 6 weeks according to the model).

Figure 1. Time dependences of the total number of COVID-19 cases, deaths, and recovered cases. Dots show actual data, whereas lines represent the result of calculations using the model.

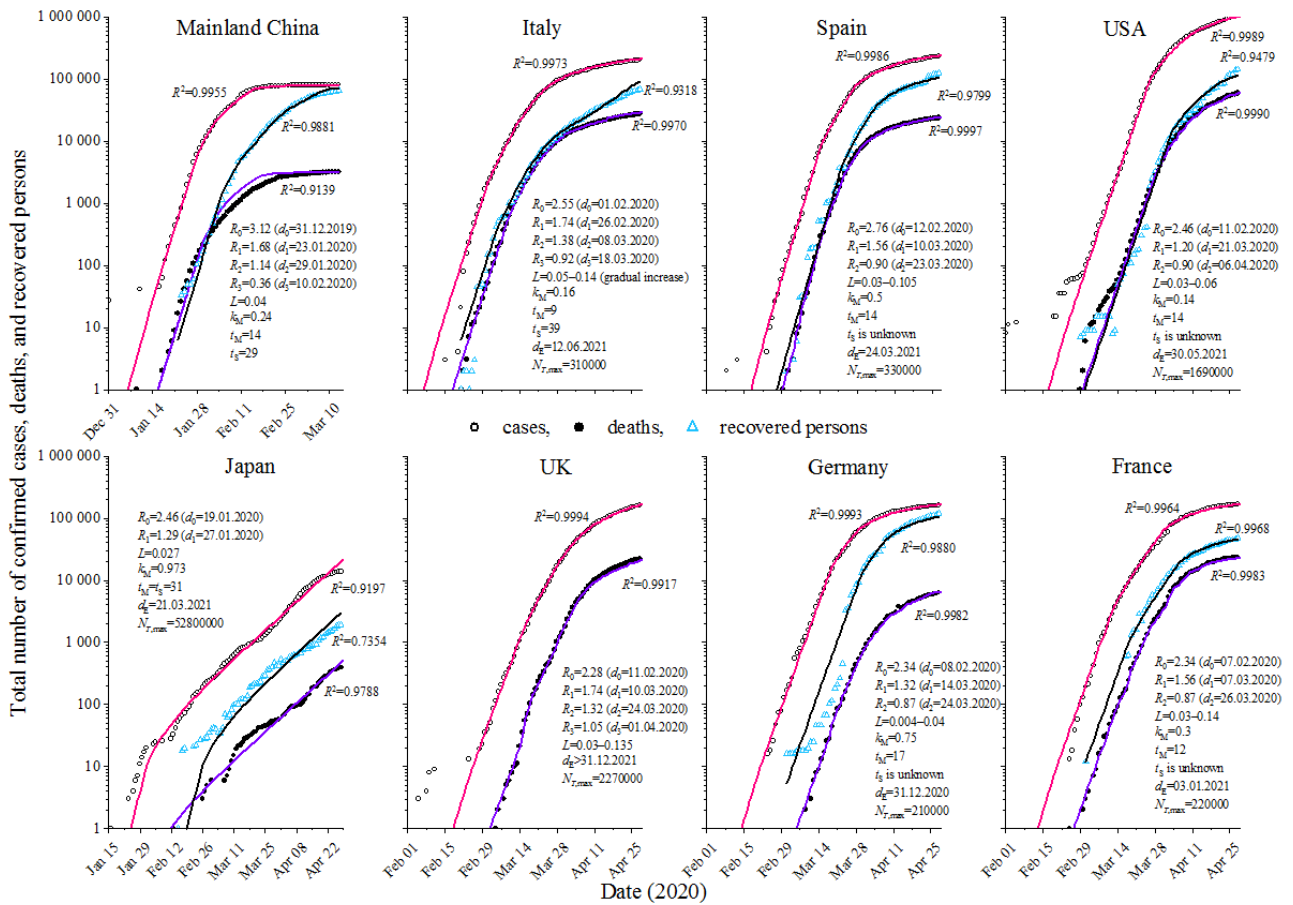


Table 1. Parameters identified from the models used in different countries.

Parameter	China	Italy	Spain	USA	Japan	UK	Germany	France
R_0	3.12	2.55	2.76	2.46	2.46	2.28	2.34	2.34
d_0	31.12.19	01.02.20	12.02.20	11.02.20	19.01.20	11.02.20	08.02.20	07.02.20
R_1	1.68	1.74	1.56	1.20	1.29	1.74	1.32	1.56
d_1	23.01.20	26.02.20	10.03.20	21.03.20	27.01.20	10.03.20	14.03.20	07.03.20
R_2	1.14	1.38	0.90	0.90	N/A ^a	1.32	0.87	0.87
d_2	29.01.20	08.03.20	23.03.20	06.04.20	N/A	24.03.20	24.03.20	26.03.20
R_3	0.36	0.92	N/A	N/A	N/A	1.05	N/A	N/A
d_3	10.02.20	18.03.20	N/A	N/A	N/A	01.04.20	N/A	N/A
L^b	0.04	0.05–0.14	0.03–0.105	0.03–0.06	0.027	0.03–0.135	0.004–0.04	0.03–0.14
k_M	0.24	0.16	0.5	0.14	0.973	N/A	0.75	0.3
t_M	14	9	14	14	31	N/A	17	12
t_S	29	39	unknown	unknown	31	N/A	unknown	unknown
d_E	N/A	12.06.21	24.03.21	30.05.21	21.03.21	>31.12.21	31.12.20	03.01.21
$N_{T,max}$	N/A	310000	330000	1690000	52800000	2270000	210000	220000

^aNot applicable.

^bIf an interval is indicated, it means gradual growth.

Discussion

On the date on which the analyzed data set ends, all European countries (except the United Kingdom) only managed to reduce the transmissibility slightly below 1.0. From a practical point of view, this means that the number of people falling ill on a daily basis in these countries was gradually decreasing, but it was at such a slow pace that the end date of the epidemic in these countries could not have been before, at best, the end of 2020. In reality, these countries have partially canceled quarantine measures, causing an increase in viral transmissibility and, consequently, a new rise in the number of infected persons and a shift in the date of a possible end of the epidemic to the future. It should be understood that any alleviation of quarantine measures would lead to increased transmissibility and resumption of an accelerated spread of the virus. To prevent this from happening after the quarantine restrictions have been removed, the viral transmissibility must remain below 1.0. By way of example, the original transmissibility was 2.55 in the case of Italy; therefore, it is necessary that 61% of the Italian population be either infected and then recovered (provided the immunity produced is durable and strong) or vaccinated against SARS-CoV-2 so that when quarantine measures are lifted, the transmissibility remains less than 1.0. At the time of writing this manuscript, 0.33% of the Italian population had been infected according to official statistics [10]. Statistics may not

take into account asymptomatic and mild cases of the disease, numbers of which may be 4-50 times higher (for the time being, only by rumor) than that of the officially recorded cases. Even if this were true, the percentage of infected and recovered persons is still significantly lower than necessary, and the removal of quarantine measures will inevitably lead to a return of the growth rate of the number of infected people to almost their original level. In other words, the rapid development, production, and subsequent application of a vaccine are vital to overcoming the COVID-19 crisis in the near future.

Thus, the model allows forecasting of the situation development and concluding about the effectiveness of quarantine measures. By way of example, it helps determine the current number of active infected persons ($N_A(d_i)$), the approximate date of isolation of the last infected person (d_E), and the number of people that could eventually be infected under the current quarantine ($N_{T,max}$). According to the calculations, the efforts made by many European countries, the United States, and Japan to stop the spread of COVID-19 are not as effective as those implemented previously in China. Most countries have been able to achieve a daily reduction in the number of infected people, but even in these cases, the viral transmissibility remains high enough, which does not allow the country to overcome the epidemic crisis within a reasonable time. At the same time, suppressing the epidemic, albeit slowly, allows time for vaccine development and launch into mass production.

Acknowledgments

This work was carried out within the State Program of A.V. Topchiev Institute of Petrochemical Synthesis.

Conflicts of Interest

None declared.

References

1. Brauer F, Castillo-Chávez C. *Mathematical Models in Population Biology and Epidemiology*. New York: Springer; 2001.
2. Daley DJ, Gani J. *Epidemic Modelling: An Introduction*. New York: Cambridge University Press; 2005.
3. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A* 1927 Jan;115(772):700-721. [doi: [10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118)]
4. Näsell I. The quasi-stationary distribution of the closed endemic SIS model. *Adv Appl Probab* 2016 Jul 1;28(03):895-932. [doi: [10.1017/s0001867800046541](https://doi.org/10.1017/s0001867800046541)]
5. Wang P, Jia J. Stationary distribution of a stochastic SIRD epidemic model of Ebola with double saturated incidence rates and vaccination. *Adv Differ Equ* 2019 Oct 15;2019(1):1-16. [doi: [10.1186/s13662-019-2352-5](https://doi.org/10.1186/s13662-019-2352-5)]
6. Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. *Math Biosci* 1995 Feb;125(2):155-164. [doi: [10.1016/0025-5564\(95\)92756-5](https://doi.org/10.1016/0025-5564(95)92756-5)] [Medline: [7881192](https://pubmed.ncbi.nlm.nih.gov/7881192/)]
7. Bichara D, Iggidr A, Sallet G. Global analysis of multi-strains SIS, SIR and MSIR epidemic models. *J Appl Math Comput* 2013 Jun 7;44(1-2):273-292. [doi: [10.1007/s12190-013-0693-x](https://doi.org/10.1007/s12190-013-0693-x)]
8. Backer J, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* 2020 Feb;25(5):2000062 [FREE Full text] [doi: [10.2807/1560-7917.ES.2020.25.5.2000062](https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062)] [Medline: [32046819](https://pubmed.ncbi.nlm.nih.gov/32046819/)]
9. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316)] [Medline: [31995857](https://pubmed.ncbi.nlm.nih.gov/31995857/)]
10. Coronavirus Update (Live). Worldometer. URL: <https://www.worldometers.info/coronavirus> [accessed 2020-10-12]

Edited by G Eysenbach; submitted 07.08.20; peer-reviewed by H Akram, J Opoku; comments to author 22.09.20; accepted 28.10.20; published 19.04.21

Please cite as:

Ilyin SO

A Recursive Model of the Spread of COVID-19: Modelling Study

JMIR Public Health Surveill 2021;7(4):e21468

URL: <https://publichealth.jmir.org/2021/4/e21468>

doi: [10.2196/21468](https://doi.org/10.2196/21468)

PMID:

©Sergey O Ilyin. Originally published in JMIR Public Health and Surveillance (<https://publichealth.jmir.org>), 19.04.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.