

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340324076>

# Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Counties

Article · March 2020

DOI: 10.29333/ejgm/7869

CITATIONS

17

READS

5,564

3 authors, including:



Davit Gondauri

Business and Technology University

5 PUBLICATIONS 24 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Georgian Railways [View project](#)

# Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries

Davit Gondauroi<sup>1\*</sup>, Ekaterine Mikautadze<sup>2</sup>, Mikheil Batiashvili<sup>1</sup>

<sup>1</sup>Business and Technology University, GEORGIA

<sup>2</sup>The Shota Rustaveli National Science Foundation of Georgia, GEORGIA

\*Corresponding Author: [dgondauroi@gmail.com](mailto:dgondauroi@gmail.com)

**Citation:** Gondauroi D, Mikautadze E, Batiashvili M. Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries. Electron J Gen Med. 2020;17(4):em209. <https://doi.org/10.29333/ejgm/7869>

## ARTICLE INFO

Received: 30 Mar. 2020

Accepted: 30 Mar. 2020

## ABSTRACT

The novel coronavirus COVID-19 originally identified in December 2019, based on the data issued by March 30, 2020 daily report, the epidemic of SARS-CoV-2 so far has caused 693224 cases and resulted in 33106 deaths in more than 200 countries. Referring to the data reported, World Health Organization declared the outbreak a pandemic. We considered the chain-binomial type of the model which involves short stages of high infectivity and approximately constant incubation periods. This research paper is to study and analyze the COVID-19 Virus spreading statistics on the examples of the cases from the different countries. High correlation coefficients (91.64%) and determinants (83.98%) between the total volumes of virus spread and recovery are considered to be high and indicate the correctness of the Bailey model. Thus, as of March 30, with the results of statistical and mathematical data processing, it is difficult to predict the future spread-reduction variables of the pandemic.

**Keywords:** COVID-19, BAILEY's Model, regression slope, pearson correlation method

## INTRODUCTION

The novel coronavirus COVID-19 originally identified in December 2019 as a severe case of pneumonia in Wuhan province of China and since that has become a global pandemic, affecting greatest nations around the whole world (1-5). Following the few days after diagnosing the first case of this previously unknown pneumonia, a novel coronavirus and its contributing agents have been identified by the several independent laboratories (6-8). For the time being the causative virus has been named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the relevant infected disease has been named as coronavirus disease 2019 (COVID-19) by the World Health Organization respectively. According to the daily report of the World Health Organization, the epidemic of SARS-CoV-2 so far registered 118 319 cases and 4292 deaths to 113 countries that reported by March 11, 2020 and the World Health Organization declared the outbreak a pandemic ([https://www.who.int/docs/default-source/coronavirus/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57\\_10](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10)). On March 30 reported 693 224 confirmed cases, including 33 106 deaths in more than 200 countries (9).

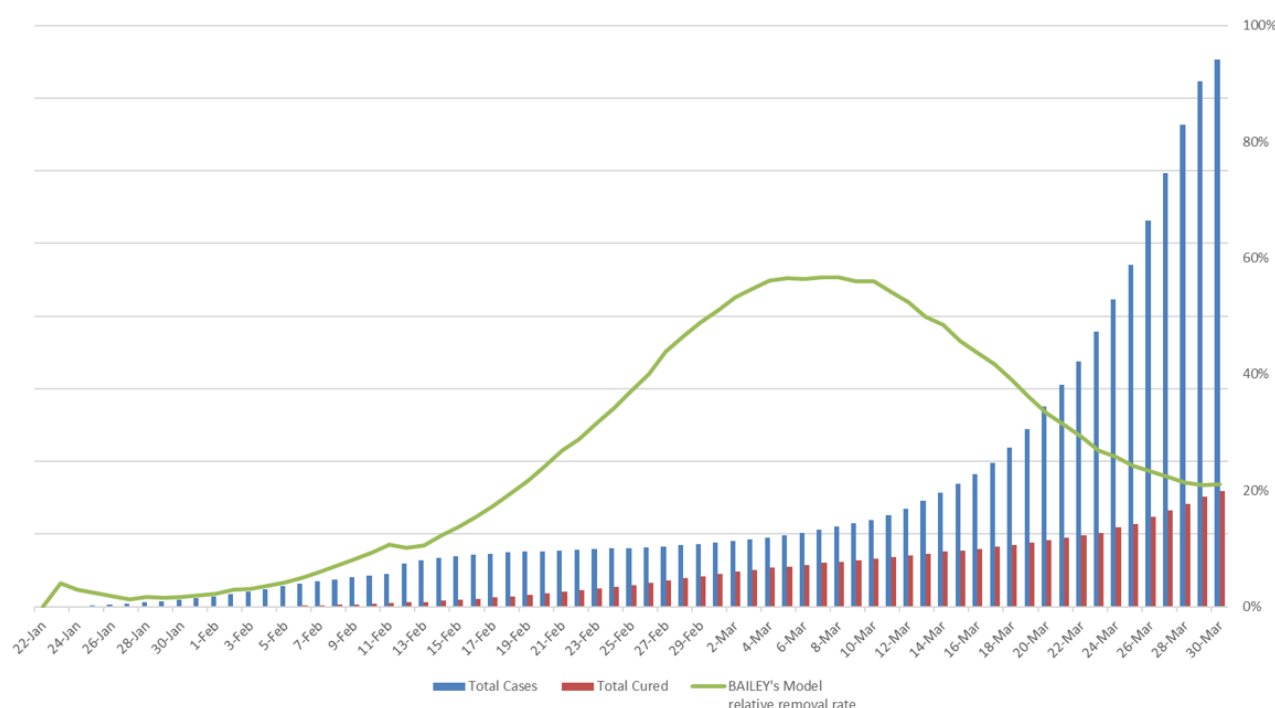
By the year 1957 there were about a hundred primarily mathematical references to the population theory of infectious disease. These information was broadly distributed in the literature, and no single work attempted to do a systematic treatment of the field. It was against the above described background when Norman Bailey formed his brilliant little book, The Mathematical Theory of Epidemics, which was the first edition of the present work (10). In 1953 Bailey discussed

the distribution of the total size of a stochastic epidemic, involving both infection and removal, in a given group of homogeneously mixing susceptible. The model employed was based of the 'continuous infection' type, rendering to which, the infected individuals continue as sources of infection until they are removed from population by recovery, death or isolation (11).

Statistical analysis, embracing modelling, parameter estimation, hypothesis testing and the design of studies, plays an essential role in connecting the gap between the mathematical theory and public health practice, and it is this aspect that motivates the present discussion. In other words, the world attempts to promote the use of statistical analyses that provide practical insight and guidance for the disease control, with emphasis on identifying issues that have not been addressed sufficiently (12).

## BAILEY'S MODEL

The chain-binomial type of the model which involves short stages of high infectivity and approximately constant incubation periods. The basic assumptions are that, with  $x$  susceptible and  $y$  infectious persons in circulation, the chance of one new infection taking place in time  $dt$  is  $\beta xy dt$ , while the chance of a removal is  $\gamma y dt$ , where,  $\beta$  and  $\gamma$  are the infection and removal rates, respectively. For a full discussion, the paper referred to (11) should be looked up. Particular attention was paid to the total size, i.e. when  $t \rightarrow \infty$ , of the epidemic occurring in small groups following the introduction of a single infectious case, the obvious application being to intra-household



**Figure 1.** Presents the number of cumulative cases infected from January 22 to March 30 in blue columns; Number of recovered cumulative cases in red columns; Change in BAILEY's Model relative removal rate

epidemics. It is important to note that from this ultimate distribution of epidemic size we cannot estimate  $\beta$  and  $\gamma$  separately, though we can estimate the *relative removal rate*,  $\rho = \gamma/\beta$ .

The purpose of this research paper is to study and analyze the COVID-19 Virus spreading statistics on the examples of the cases from the different countries.

### Research Design

On the first phase of the study, we have collected the data from the World Health Organization's official daily reports. The data covers the number of people infected with COVID-19 virus from January 22 to March 30, 2020. As of January 22, there were 580 cases of infections, and as of March 30, the number reached 752687. As of January 23, 2020, there have been 34 recovered patients. And as of March 30, 158700 people out of 752687 infected were cured of the virus. The death toll from the virus is quite high and it is 36276 worldwide as of March 30.

On the second phase of the study, we processed COVID-19 virus distribution statistics by 204 countries of distribution. We monitored the number of infected and recovered patients from January 22 to March 30. According to BAILEY's Model, we used the relative removal rate, for which we calculated the percentage of recovered patients in the infected population. The article presents only countries selected from 200 outbreaks where the prevalence of COVID-19 exceeds 50 and the BAILEY's coefficient is greater than 0%.

In the next step of the study, we calculated the results, and analyzed the data using regression slope and Pearson correlation methods. The regression slope method implies how much of a second variable changes by 1% of one variable. We looked at the effect of the 1% change in the number of infected and the percentage change in the number of infected. The Pearson correlation method showed correlations between the total infection rate and the total recovery rate.

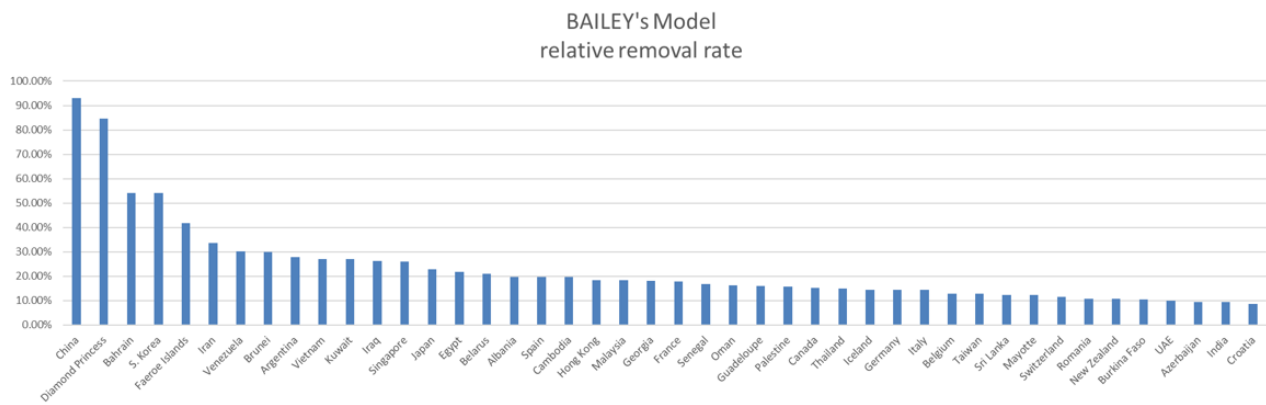
## RESEARCH FINDINGS

Looking at the statistics globally, since January 22, 2020, the prevalence of COVID-19 and the number of confirmed infections have been rising, increasing at a rapid pace for almost two months. As it can be seen from the graph (**Figure 1**), the spread of the virus has slowed down, which has a positive effect on the BAILEY coefficient. From March 4 to March 10, BAILEY's coefficient was at the highest level at 56%. On January 31, a confirmed case of COVID-19 virus was reported in Italy, where the prevalence was quite high, reaching 10149 on March 11. This led to a drop in the BAILEY's ratio.

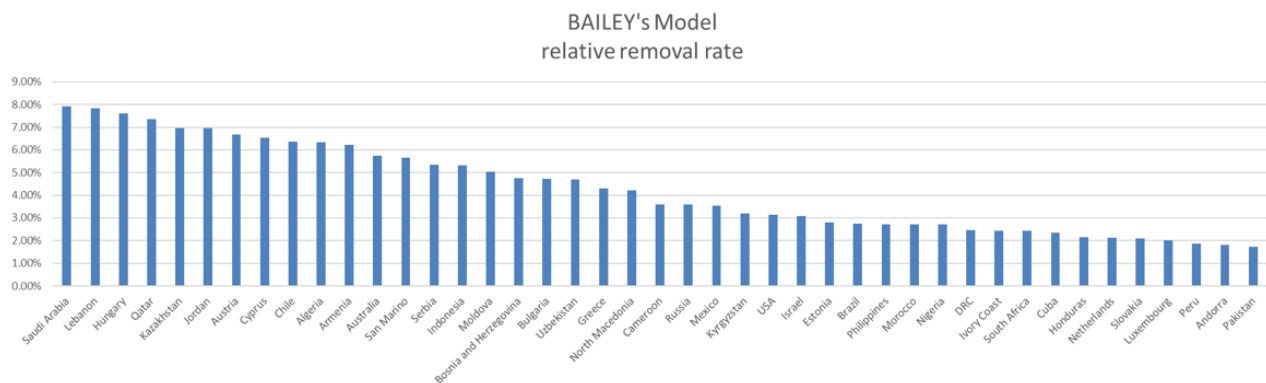
**Figures 2-4** show the relative recovery margin in different areas. Countries are selected based on the prevalence of COVID-19 virus - over 50. According to the BAILEY's coefficient, the following are the paths to stop and overwrite the virus:

- China with 93%,
- Cruise ship "Diamond Princess" - 84.7%,
- Bahrain - 54.17%,
- South Korea - 54.11%,
- Faeroe Islands - 41.67%,
- Iran - 33.52%,

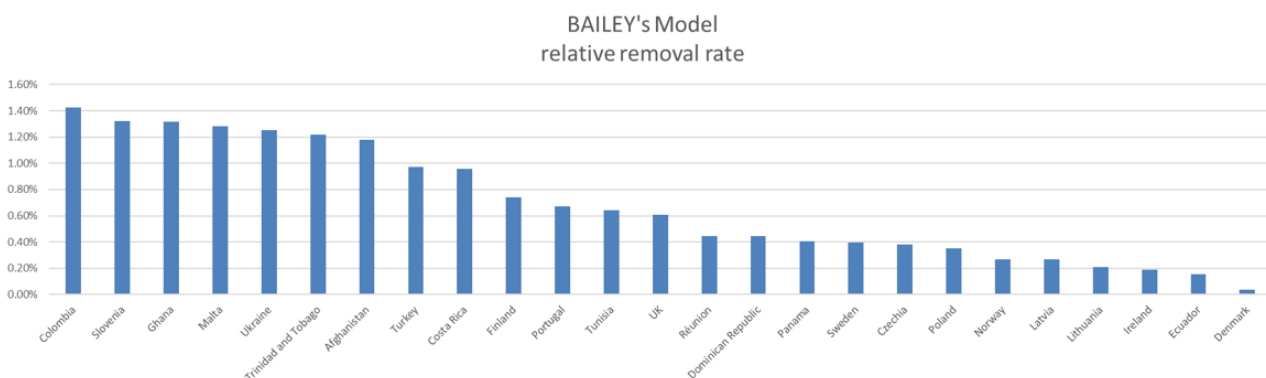
It should be noted that the prevalence of COVID-19 virus is high at this stage, and in some countries the epidemic is at a new stage. And those countries, including the cruise ship that came in the top five with the resources available, can be said that the epidemic is defeated. It should also be noted that there is no COVID-19 virus vaccine and single treatment at this stage, and if infected with COVID-19, the time from symptom onset to recovery ranged from 12 to 32 days (13), which is reflected in the cure rates. BAILEY's coefficient decreases as the virus spreads and the probability of recovery increases. BAILEY's ratio as of March 30 is 21.08%.



**Figure 2.** BAILEY's Model relative removal rate as of March 30 in different countries with  $p > 8\%$



**Figure 3.** BAILEY's Model relative removal rate as of March 30 in different countries, where  $p > 1.5\%$



**Figure 4.** BAILEY's Model relative removal rate as of March 30 in different countries, where  $p < 1.5\%$

The modified version of BAILEY's model we provide involves a percentage of recoveries in the number of confirmed cases without fatal cases. Which is 15.94% as of March 30th. According to our mathematical calculations, when the number of infected is equal to the number of recovered patients, the coefficient will reach 100% threshold and the epidemic will be extinguished.

As of March 30, Georgia is ranked 22th in the world according to BAILEY's model, which is the result of a quick study of government decision-making and action in the virus containment phase, epidemic situation management and spreading in the country, according of recommendations from World Health Organization. A detailed review of other countries' interventions was not included in the study.

A worldwide regression analysis of the number of the cases and cure regression results shows 53.38%, which means that the total spread of the virus, when varied by 1% between January 22

and March 30 (69 reporting periods) by the regression slope model over this period, recovery rate was growing at 53.38%.

As for the Pearson correlation model, between January 22 and March 30 (69 reporting periods) the correlation coefficient between total virus spread and total recoveries was 91.64%, while its coefficient of determination ( $R^2$ ) was 83.98%, which is When modeling, they are considered high values and indicate on the correctness of the Bailey's model.

Results should account in some measure for the frequency of occurrence of epidemics in populations whose density increase by the importation of unaffected individuals. The longer the epidemic is withheld the greater will be the catastrophe, provided that the population continues to increase, and the threshold density remains unchanged. Such a prolonged delay may lead to almost complete extinction of the population. Similar results, though of a somewhat more complicated form, hold for epidemics transmitted through an intermediate host. In

this case, in place of the threshold density we have to consider the threshold product (14).

## CONCLUSION

High correlation coefficients (91.64%) and determinants (83.98%) between the total volumes of virus spread and recovery are considered to be high and indicate the correctness of the Bailey model.

Thus, as of March 30, with the results of statistical and mathematical data processing, it is difficult to predict the future spread-reduction variables of the pandemic.

## REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020. <https://doi.org/10.1056/nejmoa2001017> PMID: 31978945
2. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. 2020;1-3. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
3. Hopkins Johns. University Center for Systems and Science Engineering. Coronavirus COVID-19 Global Cases. Available at: <https://coronavirus.jhu.edu/map.html> (Accessed March 25, 2020).
4. He F, Deng Y, Li W. Coronavirus Disease 2019 (COVID-19): What we know? *J Med Virol*. 2020;2019:0-2. <https://doi.org/10.1002/jmv.25766> PMID:32170865
5. Wu Z, McGoogan JM. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *Jama*. 2020;2019:3-6. <https://doi.org/10.1001/jama.2020.2648>
6. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020; 395(10224):565-74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
7. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020. <https://doi.org/10.1038/s41586-020-2012-7> PMID:32015507 PMCID:PMC7095418
8. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020; 382(8):727-33. <https://doi.org/10.1056/NEJMoa2001017> PMID:31978945
9. Organization WH. Coronavirus disease 2019 (COVID-19) Situation Report - 70. 30 March 2020.
10. Wood PHN. The Mathematical Theory of Infectious Diseases and its applications. *Immunology*. 1978;34(5):955-6.
11. Bailey NTJ. The total size of a general stochastic epidemic. *Biometrika*, 1953a;40:177. <https://doi.org/10.1093/biomet/40.1-2.177>
12. Becker NG, Britton T. Statistical studies of infectious disease incidence. *J. R. Statist. Soc. B* 1999;61(2):287-307. <https://doi.org/10.1111/1467-9868.00177>
13. Lan L, Xu D, Ye G, et al. Positive RT-PCR Test Results in Patients Recovered From COVID-19. *JAMA*. Published online February 27, 2020. <https://doi.org/10.1001/jama.2020.2783> PMID:32105304
14. Kermack WO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 1927;115(772):700-21. <https://doi.org/10.1098/rspa.1927.0118>
15. Bailey NTJ. Maximum-likelihood estimation of the relative removal rate from the distribution of the total size of an intrahousehold epidemic. *J Hyg (Lond)*. 1954;52(3):400-2. <https://doi.org/10.1017/s0022172400027595> PMID:13212043 PMCID:PMC2217790