2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)

A Comparison of the Infectiousness of SARS-COV-2 N501Y Mutation and Pre-existing Variants: Based on SIR Model

Ziyi Huang*
School of Life Sciences;
Zhejiang University;
Hangzhou; Zhejiang Province; China; 310000
* Corresponding author: hzyspace@163.com

2021 International Conference on Electronic Information Engineering and Computer Science (EIECS) | 978-1-6654-1674-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/EIECS53707.2021.9587904

***Abstract*—Severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) is causing a world-wide pandemic. A variant of SARS-COV-2 (20I, Alpha) recently discovered in the United Kingdom has a single mutation from N501 to Y501 within the receptor binding domain (Y501-RBD), of the Spike protein of the virus. This variant is much more contagious than the original version (N501-RBD).**

With the virus' mutation, the variant distribution of SARS-COV-2 is changing. This article selected cases of two countries in different periods and cases of three states in the United States which are infected by different distributed virus group. The patient data is modeled by the Susceptible, Infected and Recover (SIR) model and the basic reproduction number R0 is used to compare the infectiousness.

It can be found that in Poland's and Slovenia's cases, the R0 of virus group with a larger volume of N501Y mutation variant is statistically higher than the R0 of group without N501Y mutation variant. And in three states of United States where N501Y mutation variants were distributed in gradient, R0 was increasing when the volume of N501Y mutation variants increased. The infectivity of N501Y mutant B.1.1.7 is stronger than that of variants without N501Y mutation.

Keywords- SARS-CoV-2; SIR; Basic reproduction number; Alpha variant; infectiousness

I. INTRODUCTION

The global spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a widespread threat to public health, and the medical treatment meet increased challenge due to the virus's variants. Mutations in the spike protein provide virus with enhanced infection and resistance to antibody such as D614G, N439K[1,2]. Among these mutations in the spike protein, N501 mutation site which is located at the 501st amino acid residue of the spike protein has appeared independently multiple times and each can accompany with different mutations. UK SARS-CoV-2 variant (20I(Alpha,V1) or B.1.1.7) which is found in September 2020 has the specific mutation N501Y(N501's amino-acid change) is transmitted at a much faster rate than pre-existing SARS-CoV-2. N501Y mutation site is also reported that it may increase ACE2 binding. [3]

The Susceptible, Infected and Recover (SIR) model is a basic communication model to estimate the epidemic dynamics of virus. [9] In this model, the basic reproduction number R0 is one of the commonly used standards to estimate the virus' infectiousness which means how fast the virus is propagating, several researches in China have used R0 to evaluate the virus's growth. [4-8] With the development of the epidemic, the composition of variants within the virus population in a country is changing. By observing the patient sample data updating on the selected countries in different periods, it directly reflects current distribution of virus's infection situation and can be compared through R0. [10,11]

N501Y mutation variants cause global infection cases with different transmit rate. This article chose infection cases in two different periods of two countries and cases of three states with different U.S. virus populations at the same time and place. We use the data of confirmed patient and epidemiological model SIR model to discuss the N501Y mutation variant B.1.1.7's infection characteristics and its difference with the pre-existing SARS-CoV-2 variants.

II. MATERIAL AND METHODS

A. Infectiousness Measure

1) SIR model

SIR model is the most classic and basic communication model among infectious disease models. S represents susceptible, I represents infective and R represents removed. People are divided into three groups in SIR model. S(t) represents the number of the susceptible, including the number of people who were not infected at time t but may be infected; I(t) represents the number of the infective people, referring to the number of people who have been infected and become patients at time t; R(t) represents the recovered people. In this article, because the number of Covid-19 patient is much lower than the number of total population, we assume that the total population is consistent.

The differential equation of the model is as follows:

$$\frac{dS(t)}{dt} = \beta S(t)I(t) \quad (1)$$

$$\frac{dR(t)}{dt} = \gamma I(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \quad (3)$$

2021 International Conference on Electronic Information Engineering and Computer Science (EIECS) | 978-1-6654-1674-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/EIECS53707.2021.9587904

β is the infection coefficient, which means the number of susceptible persons that a patient can infect is proportional to the total number of susceptible persons in the environment S(t); γ represents the recovery rate, which means the number of people removed from infected persons per unit time is proportional to the number of patients.

2) Basic Reproduction Number

R0 refers to the average number of people who can be infected by a patient in an environment where all the susceptible people are without intervention which means the larger the value of R0, the stronger the infectiousness of the epidemic and the more difficult it is to control the disease. R0 is calculated by β/α in the basic SIR model.

B. Patient samples

We choose two districts' (Poland, Slovenia) patient data to fit the model. According to the variant frequencies available at nextstrain.org, in the previous period (in 2020), patients in these areas were all infected with non-N501Y-mutant variants, while in the latter period (in 2021), most patients in these areas were infected with N501Y-mutant B.1.1.7(assuming 80% as the threshold). So, we used 3 months during the original infection period and 3 months during the mutant infection period to fit the model. Poland: non-N501Y-mutant:2020 Oct.7 to 2021 Jan.7; N501Y-mutant: 2021 Mar.7 to 2021 Jun.7. Slovenia: non-N501Y-mutant:2020 Oct.7 to 2021 Jan.7; N501Y-mutant: 2021 Mar.7 to 2021 Jun.7.

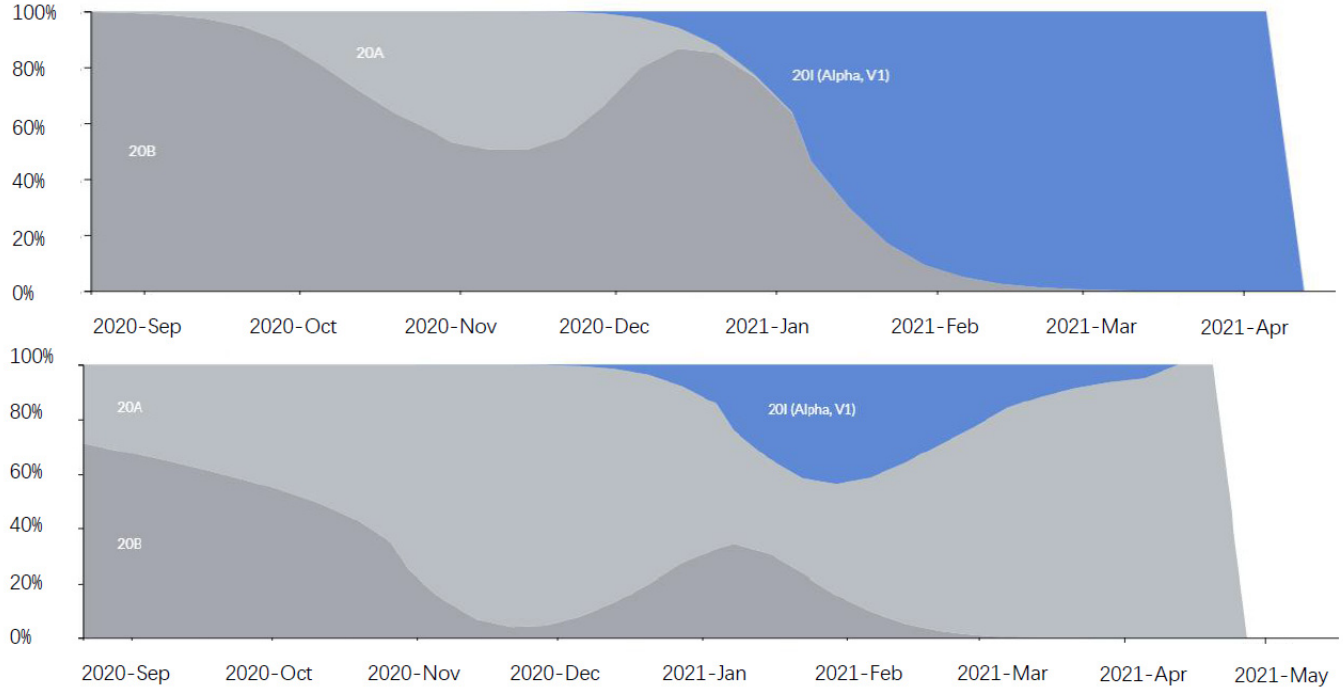


Figure 1. SARS-CoV-2 variant frequencies in Poland(above), Slovenia(below) [12]

Besides that, we try to study the change in the SARS-CoV-2 infection caused by the mutation in the same time period and the same region. Therefore, we located three neighboring states(Colorado, Texas and Louisiana) in the United States with different distribution frequencies(1 to 150, 301 to 450 and 751+) of the variant B.1.1.7 according to the Covid Data Tracker in CDC at 2021 April. Then we use the SIR model respectively to fit patient data in these states at the period from 2021 April.7 to 2021 July.7.

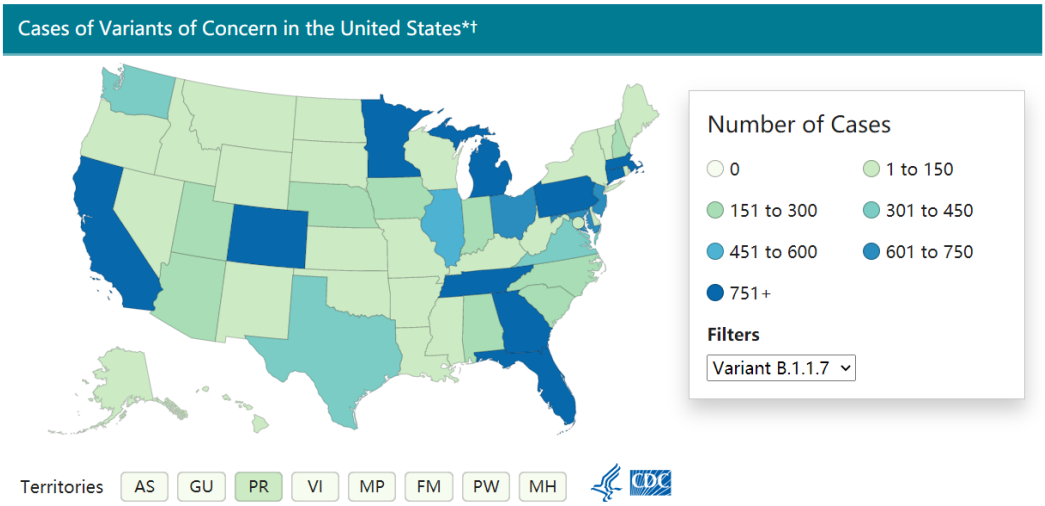


Figure 2. Cases of Variants of Concern in the US[13]

We use Student's t-test calculated for each parameter to evaluate the effects of parameter fitting, and use R0 to compare the difference in infectivity.

III. RESULT

In Poland's cases, after fitting to SIR model, the peak of N501Y mutant cases' data appears earlier and steeper than the peak of non-N501Y mutant cases' data. According to the t-test, β's and γ's Pr(>|t|) are both less than 2e-6, so the fitting effect is convincing. In the non-N501Y variant infection period, the R0 is 7.05e-4 and in the N501Y variant infection period the R0 is 9.71e-4. So, the infectiousness of N501Y mutant variant is larger.

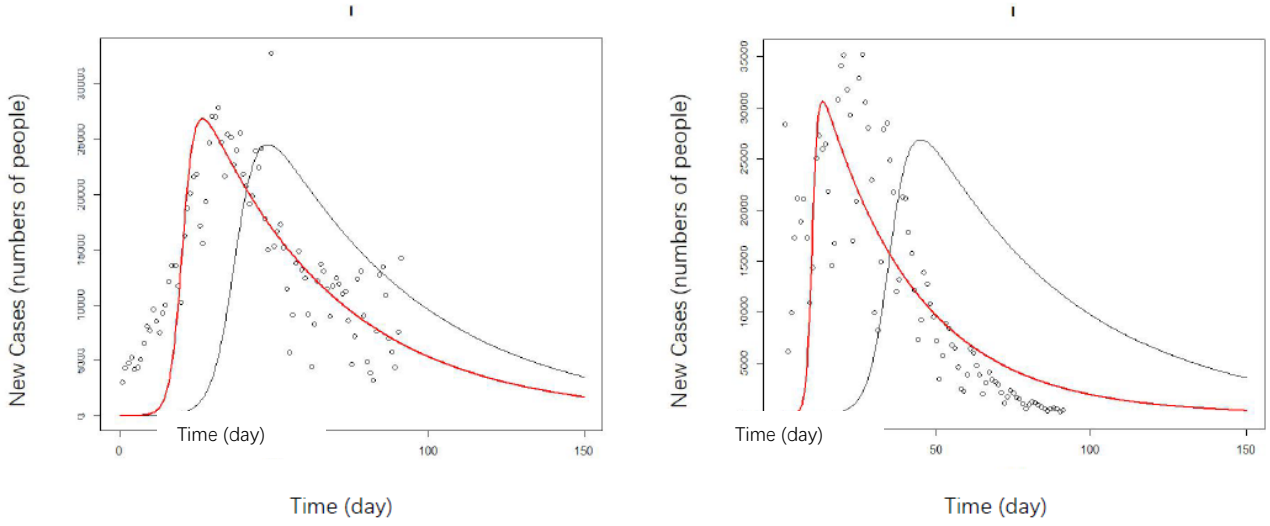


Figure 3. data fitting in the non-N501Y variant infection period(left) and the N501Y variant infection period(right)

TABLE I. T-TEST OF PARAMETERS(NON-N501Y IS ABOVE; N501Y IS BELOW)

	Estimate	Std.Error	t value	Pr(>|t|)
β	1.618e-05	4.175e-07	38.75	<2e-16 ***
γ	2.295e-02	1.328e-03	17.28	<2e-16 ***

	Estimate	Std.Error	t value	Pr(>|t|)
β	3.129e-05	1.647e-06	19.00	<2e-16 ***
γ	3.224e-02	2.631e-03	12.26	<2e-16 ***

In Slovenia's cases, after fitting to SIR model, the peak of N501Y mutant cases' data appears earlier and steeper than the peak of non-N501Y mutant cases' data. According to the t-test, β's and γ's Pr(>|t|) are both less than 2e-6, so the fitting effect is convincing. In the non-N501Y variant infection period, the R0 is 1.67e-2 and in the N501Y variant infection period the R0 is 7.98e-2. So, the infectiousness of N501Y mutant variant is larger.

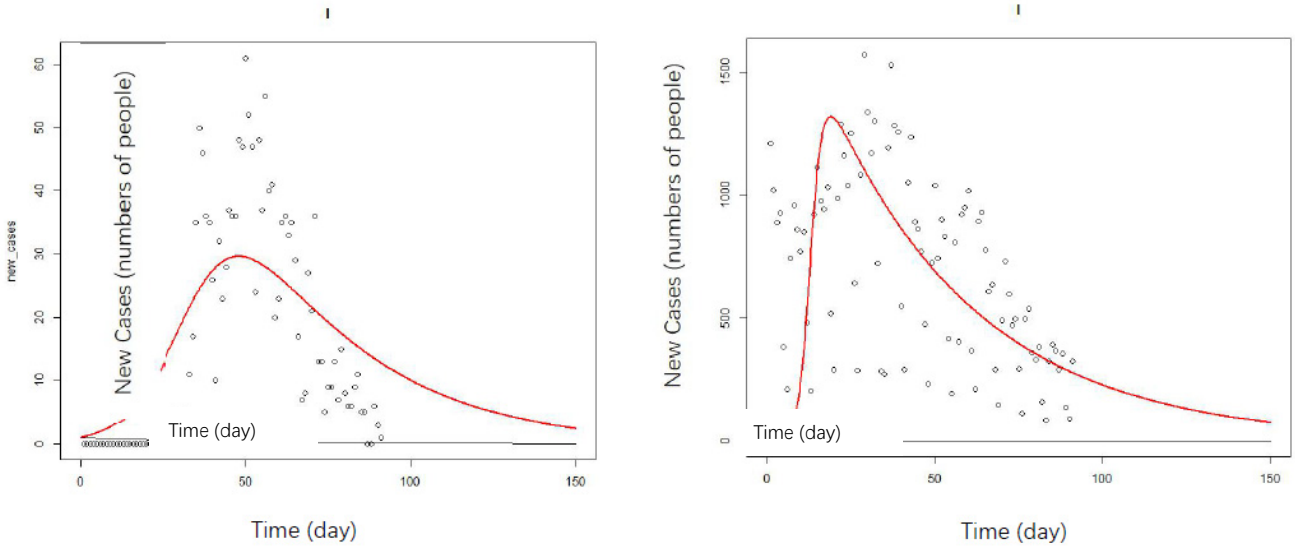


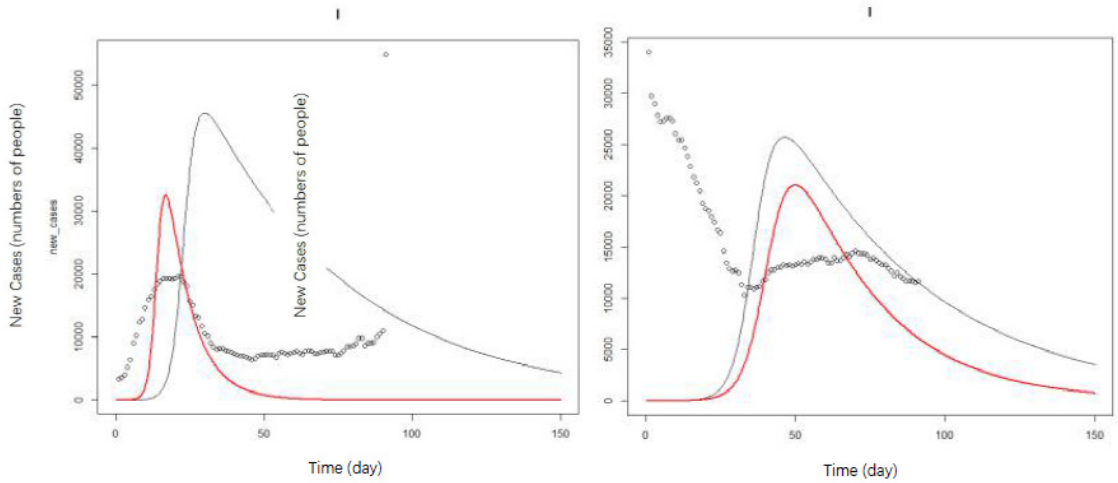
Figure 4. data fitting in the non-N501Y variant infection period(left) and the N501Y variant infection period(right)

TABLE II. T-TEST OF PARAMETERS(NON-N501Y IS ABOVE; N501Y IS BELOW)

	Estimate	Std.Error	t value	Pr(>|t|)
β	3.732e-04	2.551e-05	14.63	<2e-16 ***
γ	2.232e-02	2.187e-03	10.20	<2e-16 ***

	Estimate	Std.Error	t value	Pr(>|t|)
β	0.0023765	0.0001232	19.29	<2e-16 ***
γ	0.0297808	0.0024490	12.16	<2e-16 ***

In the data of three states in the US, the t-test of parameters indicate that the fitting results are all convincing. For Louisiana's data(1 to 150 cases), R0 is 1.33e-4; for Texas's data(301 to 450 cases), R0 is 2.39e-4; for Colorado's data(751+ cases), R0 is 1.44e-2. According to the data, we can see in the state where has higher B.1.1.7 variant frequency, the overall infectiousness of the virus is stronger.



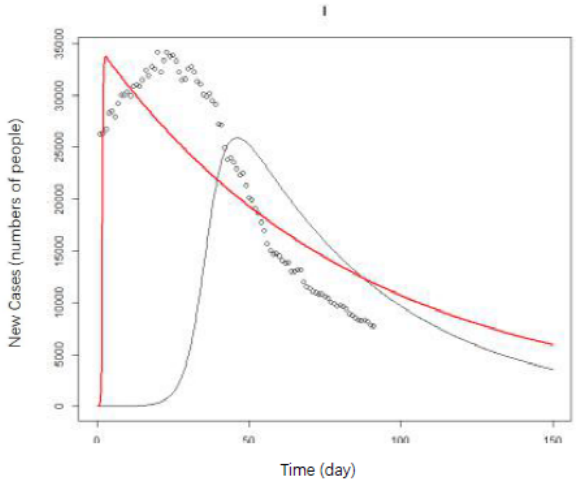


Figure 5. data fitting of three states (Louisiana(left) ,Texas(mid) and Colorado(right))

TABLE III. T-TEST OF PARAMETERS(LOUISIANA(ABOVE) ,TEXAS(MID) AND COLORADO(BELOW))

	Estimate	Std.Error	t value	Pr(>|t|)
β	1.624e-05	6.282e-07	25.853	<2e-16 ***
γ	1.217e-01	1.559e-02	7.805	1.09e-11***

	Estimate	Std.Error	t value	Pr(>|t|)
β	8.461e-06	4.658e-07	18.164	<2e-16 ***
γ	3.547e-02	6.965e-03	5.093	1.95e-06 ***

	Estimate	Std.Error	t value	Pr(>|t|)
β	1.708e-04	1.315e-05	12.99	<2e-16 ***
γ	1.186e-02	6.758e-04	17.55	<2e-16 ***

IV. CONCLUSION

According to the data of Poland and Slovenia in the same area at different times, as well as the data of the three states of the gradient distribution of the variants in the United States, the variants of the N501Y mutation site has increased the infectivity of the virus to a certain extent (represented by R0). The infectivity of N501Y mutant B.1.1.7 is different in different regions, but the infectivity is stronger than that of variants without N501Y mutation.

The N501Y mutation located on the spike protein not only appeared on B.1.1.7, but also appeared on a variety of variants. The increase in virus infectivity will apparently affect the prevention and treatment of SARS-CoV-2.

However, this article has some limitations. First, SARS-CoV-2 has an incubation period which indicates that the traditional SIR model may not be suit for fitting the case data. Second, because the infectiousness of the virus is also influenced by the temperature and climate, the comparison between different period's data may not completely reflect the N501Y mutation's effect.

ACKNOWLEDGMENT

I would like to thank all of you for helping me in writing this paper.

My deepest gratitude is first to professor Otto, for his constant encouragement and guidance. He taught me important knowledge about epidemiological analysis models, microbial population growth models and other important knowledge, which inspired me to find a way to complete this paper. Without his usual and enlightening guidance， this paper would not have been in its present form.

Secondly, I would like to express my heartfelt thanks to my mentor Shiqi Wen, who has helped me review the professor's knowledge in time and put forward important opinions on the topic of my thesis. I am also very grateful to professors and teachers who provided help and guidance on my way to complete the thesis and solved my puzzles in the experiment.

Finally, I would like to thank my dear family for their love and great faith in me for many years. I also want to thank my friends and classmates who have given me help and time to listen to my opinions and help me solve my problems in the difficult process of my paper.

REFERENCES

[1] Y.J. Hou, S. Chiba, P. Halfmann, C. Ehre, M. Kuroda, K.H. Dinnon 3rd, S.R. Leist, A. Schäfer, N. Nakajima, K. Takahashi, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo Science, 370 (2020), pp. 1464-1468

[2] E.C. Thomson, L.E. Rosen, J.G. Shepherd, R. Spreafico, A. da Silva Filipe, J.A. Wojcechowskyj, C. Davis, L. Piccoli, D.J. Pascall, J. Dillen, et al. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunitybioRxiv (2020)

[3] Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant Gard Nelson, Oleksandr Buzko, Patricia Spilman, Kayvan Niazi, Shahrooz Rabizadeh, Patrick Soon-Shiong bioRxiv 2021.01.13.

[4] World Health Organization (WHO). Naming the coronavirus disease (COVID-19) and the virus that causes it, Archived from the original on 28 February 2020. Retrieved 28 February; 2020.

[5] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China; 2020.

[6] S. Zhao, S.S. Musa, Q. Lin, J. Ran, G. Yang, W. Wang, et al. Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020, a data-driven Modelling analysis of the early outbreak J Clin Med (2020)

[7] I. Nesteruk Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China MedRxiv (2020)

[8] Ming WK, Huang J, Zhang CJ. Breaking down of healthcare system: Mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China. bioRxiv; 2020.

[9] Bacaër, N.; Dads, E.H.A. On the biological interpretation of a definition for the parameter R 0 in periodic population models. J. Math. Biol. 2011, 65, 601–621.

[10] Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 2020, 20, 533–534.

[11] Altizer, S.; Dobson, A.; Hosseini, P.; Hudson, P.; Pascual, M.; Rohani, P. Seasonality and the dynamics of infectious diseases. Ecol. Lett. 2006, 9, 467–484.

[12] https://nextstrain.org/

[13] https://covid.cdc.gov/covid-data-tracker/#datatracker-home