Monitoring and predicting viral dynamics in SARS-CoV-2-infected Patients

Shaoqing Wen[1,2]†*, Jianxue Xiong[1], Chang Sun[1], Barnaby Edward Young[3], David Chien Lye[3], Yee Sin Leo[3], Li Jin[1,4], Yi Wang[1,4]†*

1 Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China
2 Institute of Archaeological Science, Fudan University, Shanghai 200433, China
3 National Centre for Infectious Diseases, 16 Jln Tan Tock Seng 308442, Singapore
4 Human Phenome Institute, Fudan University, Shanghai 200433, China

† These authors contributed equally to this work

*Corresponding authors: wenshaoqing@fudan.edu.cn; godspeed.wang@gmail.com

Abstract

This study is based on the a simple but robust model we developed urgently to accurately monitor and predict viral dynamics for each SARS-CoV-2-infected patient, given the limited number of RT-PCR tests and the complexity of each individual's physical health situation. We used the mathematical model to monitor and predict the changes of viral loads from different nasal and throat swab of clinical specimens collected from diagnosed patients. We also tested this real-time model by using the data from the SARS-CoV-2-infected patients with different severity. By using this model (http://58.87.113.187:8080/), we can predict the viral dynamics of patients, minimize false-negative test results, and screen the patients who are at risk of testing positive again after recovery. We sincerely thank those who are on the front lines battling SARS-CoV-2 virus. We hope this model will be useful for SARS-CoV-2-infected patients.

Due to the principle of early identification, early isolation, early diagnosis and early treatment, the 2019 novel coronavirus (SARS-CoV-2) epidemic in China is under control. Viral dynamics studies are an important source of information on SARS-CoV-2 pathogenesis. The latest studies about SARS-CoV-2 viral dynamics revealed that:1) the viral load in lower respiratory tract (LRT; such as sputum) is generally higher than upper respiratory tract (URT; such as nasopharyngeal and oropharyngeal) [1] and, in the upper respiratory tract, higher viral loads are detected in the nose rather than in the throat [2]; 2) the viral loads in SARS-CoV-2-infected patients peaked soon after the onset of symptoms, resembling influenza-infected patients and distinct from SARS-CoV-infected patients [1-4]; 3) viral loads detected in asymptomatic or minimally symptomatic patients indicate the transmission potential in the early phase of infection [2].

However, there are many details that remain puzzled. For example, 1) some patients have presented positive chest CT findings showing multifocal ground-glass changes, but they had negative RT-PCR results at that time [5]; 2) some patients have rapid changes in viral loads during the window of infection, with the RT-PCR results sometimes positive and sometimes negative [2-4]; 3) few recovered patients (absence of clinical symptoms and radiological abnormalities and

2 consecutively negative RT-PCR test results separated by at least 1 day) had positive RT-PCR test results again 1 or 2 weeks after discharge from hospital [6]. Importantly, it is impossible to continuously collect samples from each SARS-CoV-2-infected patient over the course of the treatment and rehabilitation.

In this context, a simple but robust model is expected to accurately monitor and predict viral dynamics for each SARS-CoV-2-infected patient, given the limited number of RT-PCR tests and the complexity of each individual's physical health situation.

First of all, we reformat the original sparse matrix drawn from ref [2] (viral loads of 18 Chinese infected patients with many missing data points) into a dense matrix (training set) as follow:

Ct	Sex	Age	URT	Day	Day^2
38	1	48	0	8	64
39	1	56	1	3	9

…

We apply a linear regression to the above data format. To build a personal model with few data points, we borrow information from the population. We replicate the data points of the query person N times, where N is the number of population data points. We then concatenate the N folds of personal data points with the population data points to form a combined dataset. In this way, population data points are relatively down weights and act as a prior. The combined dataset is fed into a linear regression model to predict the response value: Cycle threshold (Ct). Ct values of SARS-CoV-2-specific gene on RT-PCR assay are inversely related to viral RNA copy number. Note that nasal data and throat data are also combined with an indicator URT in order to maximize the number of data points.

$$Ct = aSex + bAge + cURT + dDay + eDay^2 + f$$

Subsequently, we tested the personal model with another two sets of published data [3,4]. The test set included 13 mild patients, 5 moderate patients (requiring supplemental oxygen), and 2 severe patients (admitted to ICU), of which 7 patients were treated with lopinavir-ritonavir. The detailed epidemiologic features and clinical course of these patients are described in Supplementary Table 1. The Ct values of 20 confirmed patients were logged in the website, http://58.87.113.187:8080/, according to the timeline from symptom onset. For each SARS-CoV-2-infected patient, we predicted the Ct values in the near future based on the tested Ct values (Figure 1 and Supplementary Table 1). Figure 1 presented the Ct value curves of the 20 confirmed cases within 21 days since onset of symptoms, because of the time restriction of training set [2]. The absolute value of errors between predicted and tested Ct values is also shown in Supplementary Table 1. Over all, the mean and median of absolute value of errors is 2.57 and 2.31, respectively. The mild, moderate and severe patient group gave the average absolute value of errors for each patient, ranging from 0.99-5.51 (mean 2.75, median 2.43), 1.47-3.27 (mean 2.21, median 2.05), and 2.11-2.27 (mean 2.19, median 2.15), respectively. There was no significant difference between the drug group (mean 2.88, median 2.22) and non-drug group (mean 2.41, median 2.36). In figure 1, we observed that most samples' predicted Ct value peak at around 5-6 days after symptom onset, which is consistent with previous studies [1]. However, for each SARS-CoV-2-infected patient, the case ending time is quiet different. Within 21 days after symptom onset, we predicted that only 6 patients' viral loads would become undetectable. Notably, we did observe the drastic fluctuations of predicted Ct value curves. Some patients (such as patient 1, 2, 7, 8, 10, 11, 14, 16, and 20) have the positive and negative RT-PCR results, alternately. Combined with our predicted Ct value

curves, that could be because the treatment with lopinavir-ritonavir, or, more likely, because the occurrence of false-negative test results. Moreover, we should pay special attention to those who have negative RT-PCR results but lower predicted Ct values on day 21 after symptom onset, such as patient 2 and 7, which implies they may turn positive again.

In this study, we used a mathematical model to monitor and predict the changes of viral loads from different nasal and throat swab of clinical specimens collected from diagnosed patients. We also tested our real time model by using the data from the SARS-CoV-2-infected patients with different severity. By using this personal model, we can predict the viral dynamics of patients, minimize false-negative test results, and screen the patients who are at risk of testing positive again after recovery. The following factors can optimize our model:

1) more data regarding viral loads of the infected patients;
2) detailed information about clinical features and laboratory results;
3) more samples covering the clinical course, especially key time points (exposure, symptom onset, 5-6 days after symptom onset, recovery, and two weeks after recovery);
4) more attention to the early phase between exposure and symptom onset, which is currently a knowledge gap for SARS-CoV-2 viral dynamics studies.

Finally, we hope this model will be useful for SARS-CoV-2-infected patients.

Supporting Information

Table S1. Clinical course and viral loads (tested Ct values and predicted Ct values) of 20 SARS-CoV-2-infected patients according to the timeline from symptom onset.

Acknowledgments

We sincerely thank those who are on the front lines battling SARS-CoV-2 virus.

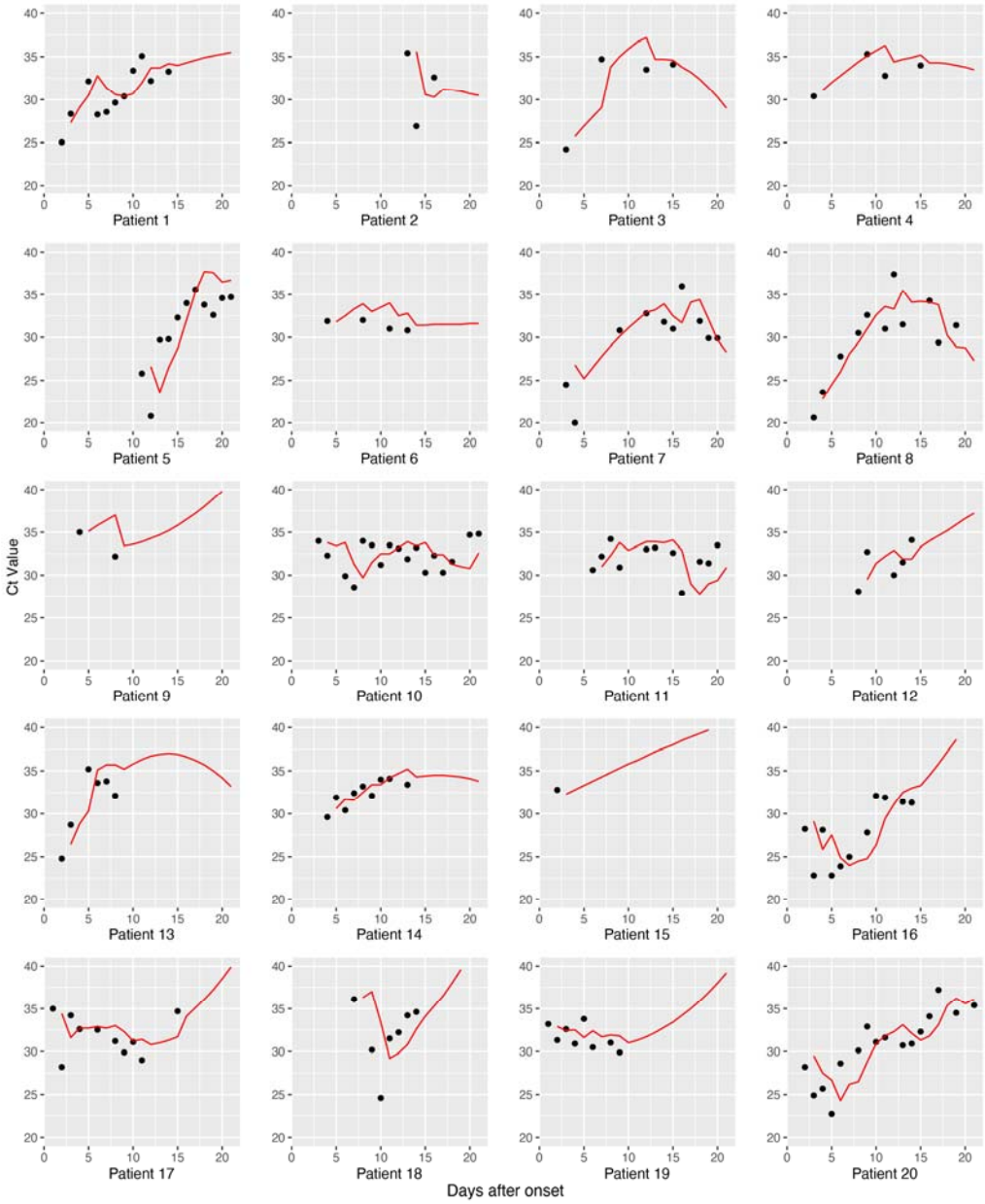
Conflict of interest

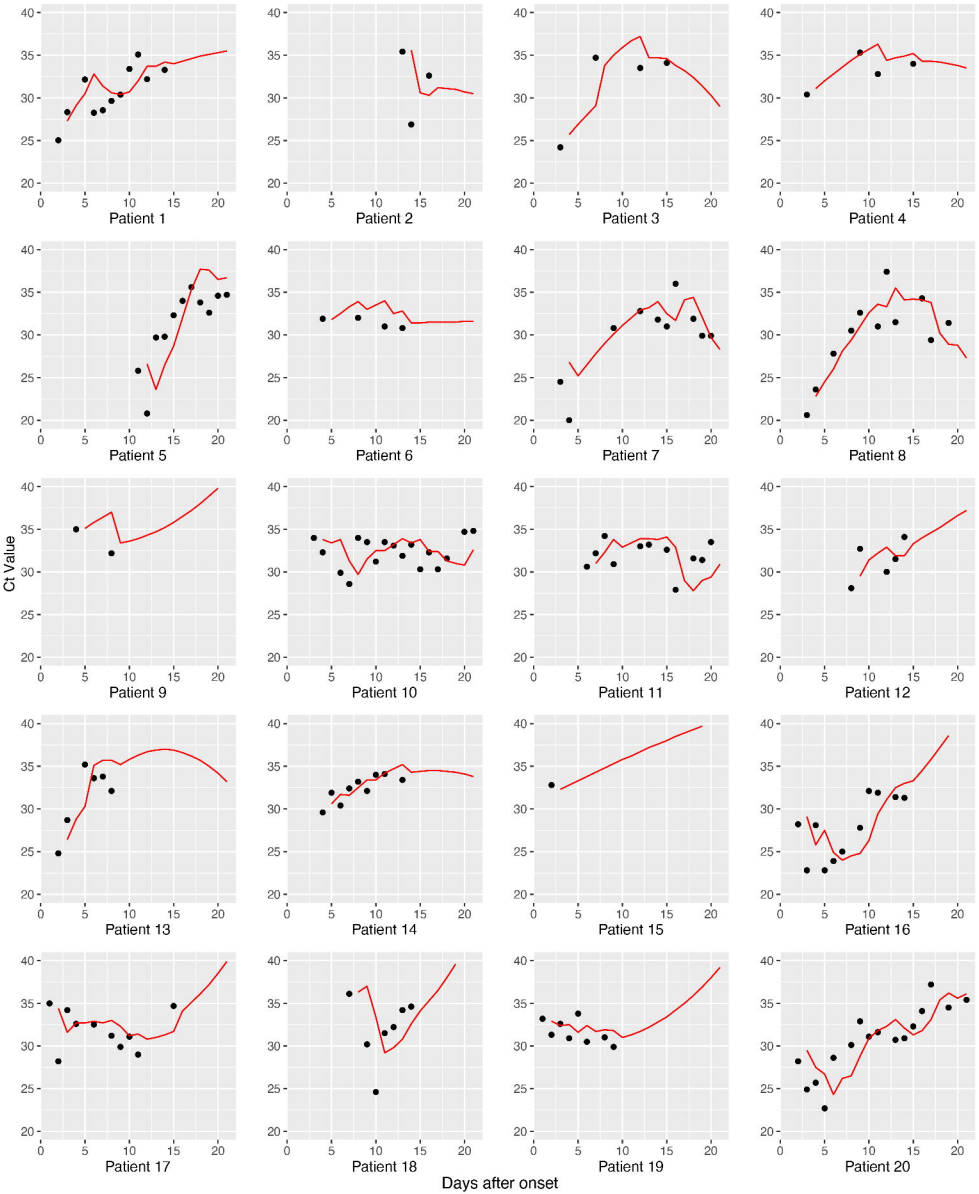
The authors declare that they have no conflict of interest.

References

1. Pan Y, Zhang D, Yang P et al (2020) Viral load of SARS-CoV-2 in clinical samples. Lancet Infect Dis. pii: S1473-3099(20)30113-4. doi: 10.1016/S1473-3099(20)30113-4.
2. Zou L, Ruan F, Huang M et al (2020) SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med. doi: 10.1056/NEJMc2001737.
3. Kim JY, Ko JH, Kim Y et al (2020) Viral Load Kinetics of SARS-CoV-2 Infection in First Two Patients in Korea. J Korean Med Sci. 35(7):e86. doi: 10.3346/jkms.2020.35.e86.
4. Young BE, Ong SWX, Kalimuddin S et al (2020) Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. JAMA. doi: 10.1001/jama.2020.3204.
5. Xie X, Zhong Z, Zhao W et al (2020) Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing. Radiology. 12:200343. doi: 10.1148/radiol.2020200343.
6. Lan L, Xu D, Ye G et al (2020) Positive RT-PCR Test Results in Patients Recovered From COVID-19. JAMA. doi: 10.1001/jama.2020.2783.

Figure 1. The tested Ct values of 20 SARS-CoV-2-infected patients and their predicted Ct value curves.


Ct Value
Days after onset
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient 10
Patient 11
Patient 12
Patient 13
Patient 14
Patient 15
Patient 16
Patient 17
Patient 18
Patient 19
Patient 20


Ct Value
Days after onset
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient 10
Patient 11
Patient 12
Patient 13
Patient 14
Patient 15
Patient 16
Patient 17
Patient 18
Patient 19
Patient 20