

A Note on COVID-19 Diagnosis Number Prediction Model in China

Yi Li

Fudan University School of Life Sciences

Xianhong Yin

Fudan University <https://orcid.org/0000-0002-4021-1617>

Yi Wang (✉ wangyi_fudan@fudan.edu.cn)

Fudan University School of Life Sciences <https://orcid.org/0000-0002-3937-8268>

Meng Liang

Fudan University School of Life Sciences

Xiaoyu Liu

Fudan University School of Life Sciences

Meng Hao

Fudan University School of Life Sciences

Research

Keywords: COVID-19, coronavirus, Diagnosis

Posted Date: November 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-97682/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A Note on COVID-19 Diagnosis Number Prediction Model in China

Yi Li^{1,2,3}✉, Xianhong Yin^{1,2}✉, Meng Liang^{1,2}, Xiaoyu Liu^{1,2}, Meng Hao^{1,2}, Yi Wang^{1,2}*

¹Human Phenome Institute, Fudan University, Shanghai, China

²Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai, China

³Institute for Six-sector Economy, Fudan University, Shanghai, China

✉The authors contributed equally to the work

*Corresponding authors

Yi Wang

wangyi_fudan@fudan.edu.cn

Declarations:

1. Ethics approval and consent to participate

Our data were downloaded from the public. There is no any information which were involved in patients or animals.

2. Consent for publication

Our manuscript had no ethic problems and consent to participate.

3. Availability of data and material

Our data were downloaded from Internet and this data is show to the public.

4. Competing interests

Our research is pure scientific researcher, which is no conflict to interest.

5. Funding

This is no funding to support our research now.

6. Authors' contributions

Pro. Wang is responsible for the design of the project and revise the manuscript.

Yi Li and Xianhong Yin are responsible for data analysis and write the manuscript.

Meng Liang, Xiaoyu Liu and Meng Hao are responsible for retrieval of literature and revise the manuscript.

Abstract

In December 2019, pneumonia infected with the novel coronavirus burst in Wuhan, China. We aimed to use a mathematical model to predict number of diagnosed patients in future to ease anxiety on the emergent situation. Our model was based on the epidemic situation in China, which could provide referential significance for disease prediction in other countries, and provide clues for prevention and intervention of relevant health authorities. In this retrospective, all diagnosis number from Jan 21 to Feb 10, 2020 reported from China was included and downloaded from WHO website. We develop a simple but accurate formula to predict the next day

diagnosis number: $N_i/N_{(i-1)} = \left[\left(\frac{N_{(i-1)}}{N_{(i-2)}} \right) \right]^{\alpha}$, where N_i is the total

diagnosed patient till the i th day, and α was estimated as 0.904 at Feb 10. Based on

this model, it is predicted that the rate of disease infection will decrease exponentially. The total number of infected people is limited; thus, the disease will have limited impact. However, new diagnosis will last to end of March. Through the establishment

Significance of this study

What is already known about this subject?

- A novel coronavirus was verified and identified as the seventh member of the enveloped RNA coronaviruses as the cause of the disease, which is referred to as COVID-19.
- The spread of COVID-19 is expeditious.
- A simple but robust model is expected to work better than sophisticated epidemiology models.

What are the new findings?

- Our model indicates that the COVID-19 disease infection rate decreases exponentially.
- This model converges to a maximum number as time increases, indicating a limited impact of COVID-19.
- The model also predict that the new diagnosis number will last to March, indicating that the COVID-19 disease will be well controlled.

How might these results change the focus of research or clinical practice?

- The prediction is expected to provide practical significance on social and economic application.

of our model, we can better predict the trend of the epidemic in China.

Introduction

In December of 2019, the increasing number of patients with pneumonia of unknown cause emerged in Wuhan, the capital of Hubei province, Central China. Most of the early patients had directly or indirectly related to wild animal source in the seafood and wet animal wholesale market through epidemiological survey (1-5). Then laboratory tests found that the infection was caused by a novel coronavirus (now named as COVID-19 (6,7)).

The spread of COVID-19 is expeditious. According to WHO statistics, the total number of diagnosed people in China was 309 in Jan 21(8), but rapidly grew to 4537 in Jan 27(9). Considering a simple constant exponential growth rate model (10,11), the diagnosis number will be unimaginable soon. This triggered a major concern and motivated this study (12-14). A mathematical model is required to predict the spread trend and emergent situation.

Over the last few decades, mathematical models of disease transmission have been useful to gain insights into the transmission dynamics of infectious diseases and the potential role of intervention strategies (15). Epidemic transmission models focused on the spread of an infectious disease through a population. But this spread can be modeled in two fundamental ways: by looking at the entire population as a single group or set of linked subgroups (population-based) or by considering everyone that makes up the population and how interactions between unique individuals facilitate disease spread (individual-based). Common models of infectious disease included meta-population models, individual-based network models, and simple SIR-type models that incorporate the effects of reactive behavior changes or inhomogeneous mixing (16).

Considering the limited number of data points and the complexity of the real situation, a simple but robust model is expected to work better than sophisticated epidemiology models. In this paper, we propose a robust model for next day diagnosis number predict. This model had excellent performance on prediction of infected patients in past few days. The prediction is expected to provide practical significance on social and economic application.

Materials and methods

Data Download

Daily diagnosis number of China is download from WHO situation reports (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>). The data starts from Jan 21 and currently it ends at Feb 10.

Model

The following model is proposed based on our observations:

$$\frac{N_i}{N_{i-1}} = \left(\frac{N_{i-1}}{N_{i-2}}\right)^\alpha,$$

where N_i is the number of diagnosed patients in China according to WHO at the i th day, α is a parameter modeling the decay of diagnosis growth rate.

A simple estimation of α at the i th day would be

$$\hat{\alpha}_i = \frac{\ln\left(\frac{N_i}{N_{i-1}}\right)}{\ln\left(\frac{N_{i-1}}{N_{i-2}}\right)}$$

As we have multiple days of observations, historical $\hat{\alpha}_i$ with the median statistic, which is well known for its robustness, can be aggregated.

$$\hat{\alpha} = \text{median}(\hat{\alpha}_i)$$

The source code of model estimation is at:

https://github.com/wangyi-fudan/ncp_model

Results

Estimation of α

$\hat{\alpha}_i$ and its median aggregates were calculated (Figure 1). It was observed that $\hat{\alpha}_i$ manifests a random fluctuation around a constant center. It is reasonable to assume that α is a constant. The estimation of α currently (Feb 10) is 0.904. A future diagnosis number formula is obtained based on this estimation:

$$\frac{N_i}{N_{i-1}} = \left(\frac{N_{i-1}}{N_{i-2}}\right)^{0.904}$$

Next Day Prediction Table

Previous numbers of diagnoses in China are used to predict the number for the next day. The prediction is started from Jan 29 as several data points are needed to build the model (Table 1). We obtain nice concordance (<2.5% error) with the real numbers next day.

Table 1. Observed and Predicted number of diagnosis in China.

Date	WHO	Predict	error%
21-Jan	309		
22-Jan	571		
23-Jan	830		
24-Jan	1297		
25-Jan	1985		
26-Jan	2741		
27-Jan	4537		
28-Jan	5997		
29-Jan	7736	7602	-1.73%
30-Jan	9720	9760	0.41%
31-Jan	11821	11950	1.09%
1-Feb	14411	14088	-2.24%
2-Feb	17238	17240	0.01%
3-Feb	20438	20269	-0.83%
4-Feb	24363	23857	-2.08%
5-Feb	28086	28600	1.83%
6-Feb	31211	31959	2.40%
7-Feb	34598	34335	-0.76%

8-Feb	37251	37992	1.99%
9-Feb	40235	39824	-1.02%
10-Feb	42708#	43152	1.04%
11-Feb		45075	

latest number from <http://www.nhc.gov.cn/>

New Diagnosis Number Prediction

We predict the new diagnosis number in China according to our model as Fig.2. The model predicts that new diagnosis will last to end of March.

Conclusion

Our model indicates that the COVID-19 disease infection rate decreases exponentially. This model converges to a maximum number as time increases, indicating a limited impact of COVID-19. The model also predict that the new diagnosis number will last to March, indicating that the COVID-19 disease will be well controlled.

Discussion

We only analyzed the total number of diagnoses in China. The province scale data is available but is not as complete as total number. Here, we did not show these details analyses.

Our model is attractive as it not only predicts data accurately but also is minimal. The minimal number of parameters and robust statistics make our model robust in the real-world scenario. We develop a formula to predict the next day diagnosis number:

$$\frac{N_i}{N_{i-1}} = \left(\frac{N_{i-1}}{N_{i-2}}\right)^\alpha, \text{ and estimate that } \alpha \text{ is } 0.904 (\alpha < 1) \text{ which shows that the diagnosis}$$

number of COVID-19 patients will decrease. It is very effective in controlling the spread of infectious diseases that the Chinese health authorities have taken appropriate health care actions. If the outbreak is no longer controlled, the value of α will be around 1 indicating that the outbreak will wreak havoc.

We obtain this elegant formula with both curve plot visualization and mathematical intuition. We plot $\ln(N_i/N_{i-1})$ against time (Fig.3). First few observations showed a linear decay of spread rate. However, later data points realized us at Feb 4 that it is an exponential decay. Later data points further confirm our model.

Abbreviations and Definitions

COVID-19: Corona Virus Disease 2019

WHO: World Health Organization

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

YW conceived the idea and wrote the source code. YL, XY and YW contributed data analysis, generating tables and figures, and manuscript writing. YL, XY, ML, MH, LX and YW contributed the theoretical analysis and manuscript revision. All authors contributed to final revision of the manuscript.

Funding

Data in this study is publicly available and was downloaded from the WHO Website. Our research was supported by the Postdoctoral Science Foundation of China (2018M640333) and the Science and Technology Committee of Shanghai Municipality (2017SHZDZX01).

Acknowledgments

We thank the Fudan University High-End Computing Center for supporting computations involved in this study.

References

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*(2020); 395(10223), 497-506.
2. Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. *JAMA*(2020);323(8), 707-708.
3. Nishiura H, Jung SM, Linton NM, et al. The extent of transmission of novel coronavirus in Wuhan, China. *Journal of Clinical Medicine*(2020).
4. Liu T, Hu J, Kang M, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv*(2020).
5. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*(2020).

6. World Health Organization
website:<https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020>.
7. Chua MSQ, Lee JCS, Sulaiman S, Tan HK. From the frontlines of COVID-19—How prepared are we as obstetricians: a commentary. *BJOG: An International Journal of Obstetrics & Gynaecology*2020.
8. World Health Organization
website:https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200122-sitrep-2-2019-ncov.pdf?sfvrsn=4d5bcbca_2.
9. World Health Organization
website:https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200128-sitrep-8-ncov-cleared.pdf?sfvrsn=8b671ce5_2.
10. Zhao S, Lin Q, Ran J, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*(2020);92, 214-217.
11. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*(2020).
12. Shen M, Peng Z, Xiao Y, Zhang L. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China. *bioRxiv*(2020).
13. Chen T, Rui J, Wang Q, Zhao Z, Cui JA, Yin L. A mathematical model for simulating the transmission of Wuhan novel Coronavirus. *bioRxiv*(2020).
14. Zhao S, Lin Q, Ran J, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*(2020);92, 214-217.
15. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics IV. Analysis of experimental epidemics of the virus disease mouse ectromelia. *Epidemiology & Infection*(1937);37(2), 172-187.
16. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. *Physics of life reviews*(2016);18, 66-97.

Figures legends

Figure 1 . Estimation of Alpha Parameter.

Figure 2. Observed (Blue) and Predicted (Orange) number of daily new diagnosis.

Figure 3. $\ln (N_i/N_{i-1})$ against time.

Figures

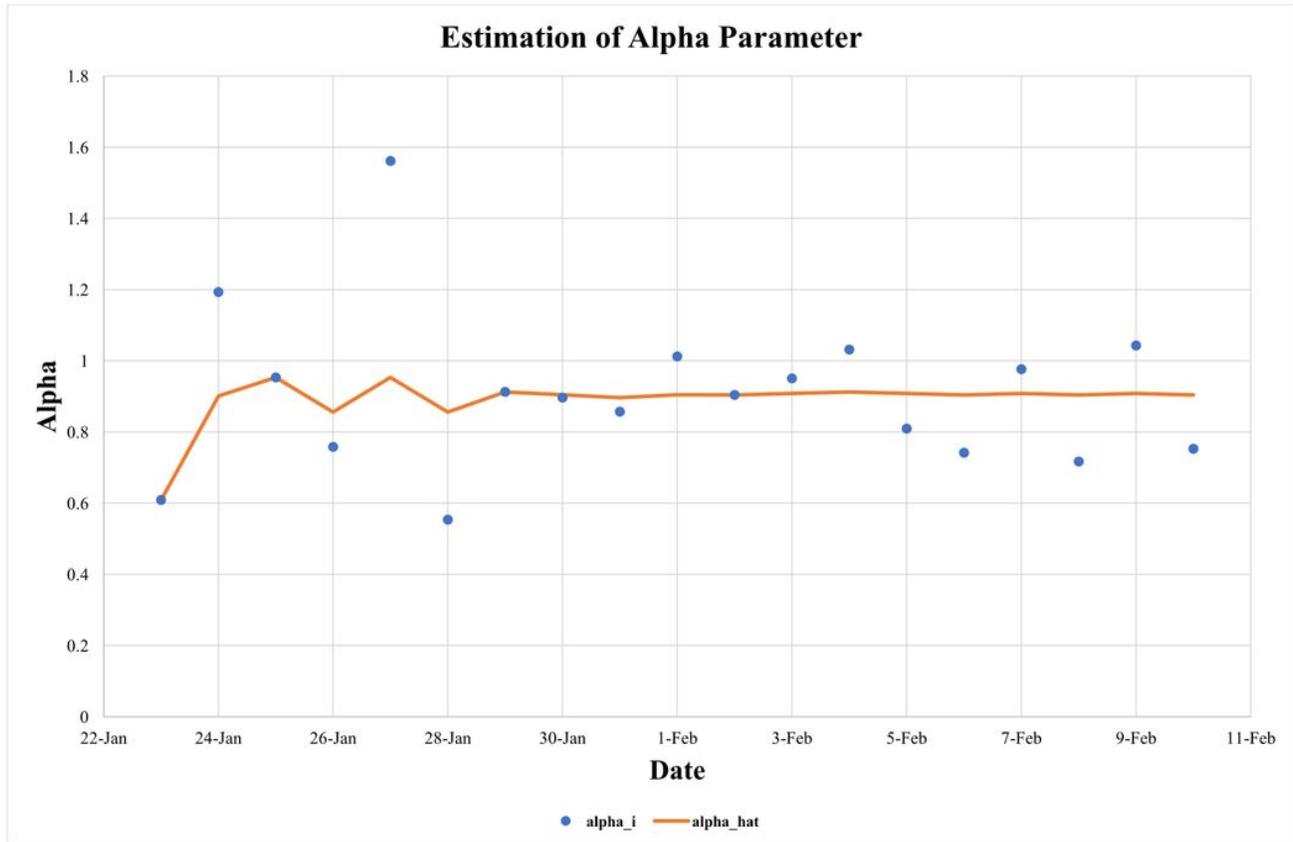


Figure 1

Estimation of Alpha Parameter.

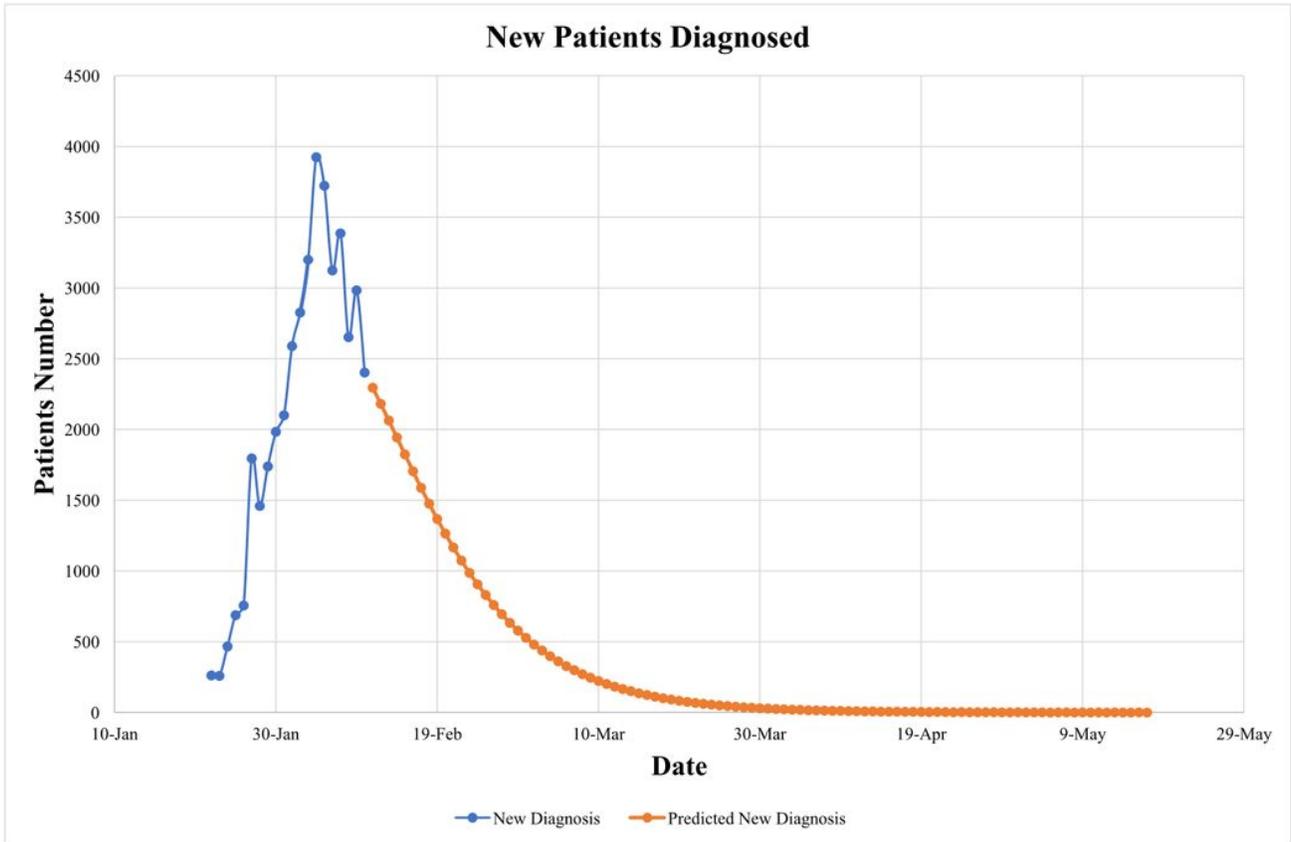


Figure 2

Observed (Blue) and Predicted (Orange) number of daily new diagnosis.

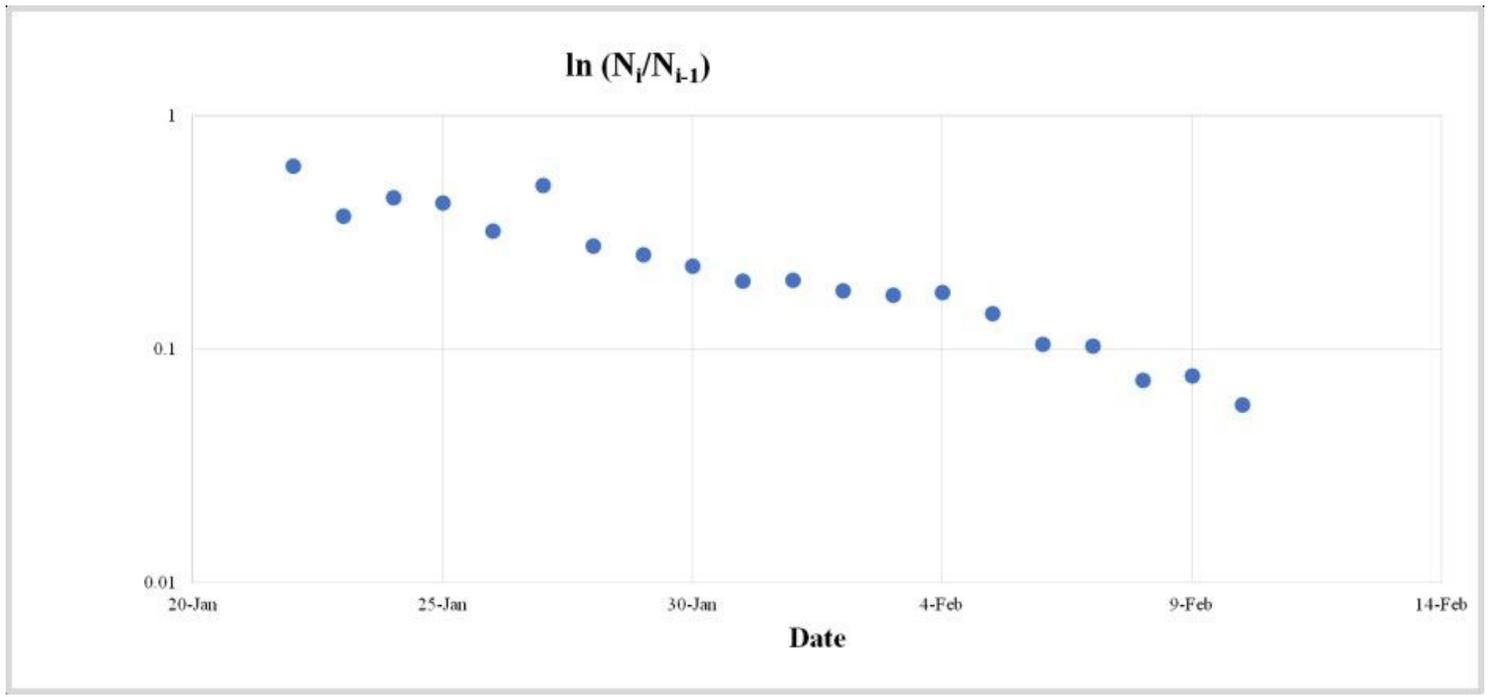


Figure 3

$\ln(N_t/N_{t-1})$ against time.