



Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis

Shu Yang^{1,2#}, Peihua Cao^{3#}, Peipei Du⁴, Ziting Wu^{5,6}, Zian Zhuang⁷, Lin Yang⁸, Xuan Yu⁹, Qi Zhou¹⁰, Xixi Feng⁴, Xiaohui Wang¹¹, Weiguo Li^{12,13}, Enmei Liu^{12,13}, Ju Chen¹, Yaolong Chen^{9,14,15,16}, Daihai He⁷; on behalf of COVID-19 evidence and recommendations working group

¹College of Medical Information Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China; ²Digital Institute of Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China; ³Clinical Research Center, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China; ⁴School of Public Health, Chengdu Medical College, Chengdu 610500, China; ⁵School of Public Health, Peking University, Beijing 100191, China; ⁶School of Public Health, Yale University, New Haven, CT, USA; ⁷Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China; ⁸School of Nursing, Hong Kong Polytechnic University, Hong Kong, China; ⁹Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China; ¹⁰The First Clinical Medical College of Lanzhou University, Lanzhou 730000, China; ¹¹School of Public Health, Lanzhou University, Lanzhou 730000, China; ¹²Department of Respiratory, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, China International Science and Technology Cooperation base of Child development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing 400014, China; ¹³Chongqing Key Laboratory of Pediatrics, Chongqing 400014, China; ¹⁴WHO Collaborating Centre for Guideline Implementation and Knowledge Translation, Lanzhou 730000, China; ¹⁵GIN (Guidelines International Network) Asia, Lanzhou 730000, China; ¹⁶Key Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province, Lanzhou University, Lanzhou 730000, China

Contributions: (I) Conception and design: S Yang, P Cao, Y Chen, D He; (II) Administrative support: W Li, E Liu, J Chen, Y Chen; (III) Provision of study materials or patients: P Du, Z Wu, Z Zhuang, X Yu, Q Zhou, X Feng; (IV) Collection and assembly of data: S Yang, P Du, Z Wu; (V) Data analysis and interpretation: S Yang, P Chao, P Du, L Yang, X Wang, Y Chen, D He; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yaolong Chen. Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China. Email: chenyaolong@lzu.edu.cn; Daihai He. Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China. Email: daihai.he@polyu.edu.hk.

Background: An ongoing outbreak of pneumonia caused by a novel coronavirus [severe acute respiratory syndrome coronavirus (SARS-CoV)-2], named COVID-19, hit a major city of China, Wuhan in December 2019 and subsequently spread to other provinces/regions of China and overseas. Several studies have been done to estimate the basic reproduction number in the early phase of this outbreak, yet there are no reliable estimates of case fatality rate (CFR) for COVID-19 to date.

Methods: In this study, we used a purely data-driven statistical method to estimate the CFR in the early phase of the COVID-19 outbreak. Daily numbers of laboratory-confirmed COVID-19 cases and deaths were collected from January 10 to February 3, 2020 and divided into three clusters: Wuhan city, other cities of Hubei province, and other provinces of mainland China. Simple linear regression model was applied to estimate the CFR from each cluster.

Results: We estimated that CFR during the first weeks of the epidemic ranges from 0.15% (95% CI: 0.12–0.18%) in mainland China excluding Hubei through 1.41% (95% CI: 1.38–1.45%) in Hubei province excluding the city of Wuhan to 5.25% (95% CI: 4.98–5.51%) in Wuhan.

Conclusions: Our early estimates suggest that the CFR of COVID-19 is lower than the previous coronavirus epidemics caused by SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV).

Keywords: 2019 novel coronavirus; severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2); COVID-19; case fatality rate (CFR)

Submitted Feb 09, 2020. Accepted for publication Feb 12, 2020.

doi: 10.21037/atm.2020.02.66

View this article at: <http://dx.doi.org/10.21037/atm.2020.02.66>

Introduction

On December 31, 2019, the discovery of a novel coronavirus (SARS-CoV-2)-infected pneumonia (COVID-19) in Wuhan, China, was reported to the World Health Organization (WHO) and has caused serious illness and death (1). The symptoms of COVID-19 included fever, cough, shortness of breath, and occasionally watery diarrhea (2). As of February 3rd, 2020 (8:00 PM, GMT+8), 17,238 cases of COVID-19 infections and 361 deaths have been confirmed in mainland China (3).

In the 21st century, two highly pathogenic human coronaviruses, the severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), have caused global epidemics with more than 10,000 cumulative cases in the past two decades (4,5). Of significant concern is the pandemic potential of the virus, as well as its transmission dynamics. The basic reproduction number (R_0) and best estimates of case-fatality rates (CFR) are two key parameters to understand the basic epidemiological features of the outbreak (6). Although several studies have been done to estimate the basic reproduction number in the early phase of this outbreak, there is still lack of reliable estimates of CFR for COVID-19 to date (7-9). The global epidemic of the 2002-2003 SARS outbreak was reported to have an estimated CFR of 9.6% globally and a considerably lower CFR of 6.4% in mainland China, while the MERS outbreak was estimated to have a much higher CFR of 34.5% worldwide and 37.1% in Saudi Arabia (10-13). Therefore, it is of great significance for clinicians and public health practitioners to implement efficient and effective disease control interventions.

In this study, we aim to estimate the CFR of COVID-19 during the first weeks of the epidemic by employing the cumulative number of confirmed cases and deaths through a simple linear regression model.

Methods

Sources of data

We collected the daily number of laboratory-confirmed COVID-19 cases and the daily number of COVID-19 deaths released by Wuhan Municipal Health Commission, China and National Health Commission of China from January 10, 2020 to February 3, 2020 to construct a real-time database. Because the virus detection rates and the healthcare resources may differ in different regions, we further divided the database into three clusters: Wuhan city, Hubei province (except Wuhan), and mainland China (except Hubei province). Overseas data were not included in our estimation of CFR due to the small number of confirmed cases and only one death occurred (14).

Statistical model

Simple linear regression model was applied to estimate the CFR during the observation period in each cluster. We used cumulative number of laboratory-confirmed cases as predictor variable and cumulative deaths as outcome variable. The slope of the fitted line can be used as an estimate of the CFR, and the confidence interval of CFR can be calculated from the standard error of the slope. The impacts of patients' incubation time, duration of disease, hospitalization time and other factors can be filtered by the intercept of the fitted line without changing the slope. We also modelled the epidemic curve obeying the exponential growth as comparison. In order to avoid the influence on the CFR estimation in the initial stage of outbreak when no illness or death occurred, the time of reporting the first case of death was used as the starting point for modeling in each cluster.

Results

The simple linear fitting results are shown in *Figure 1*.

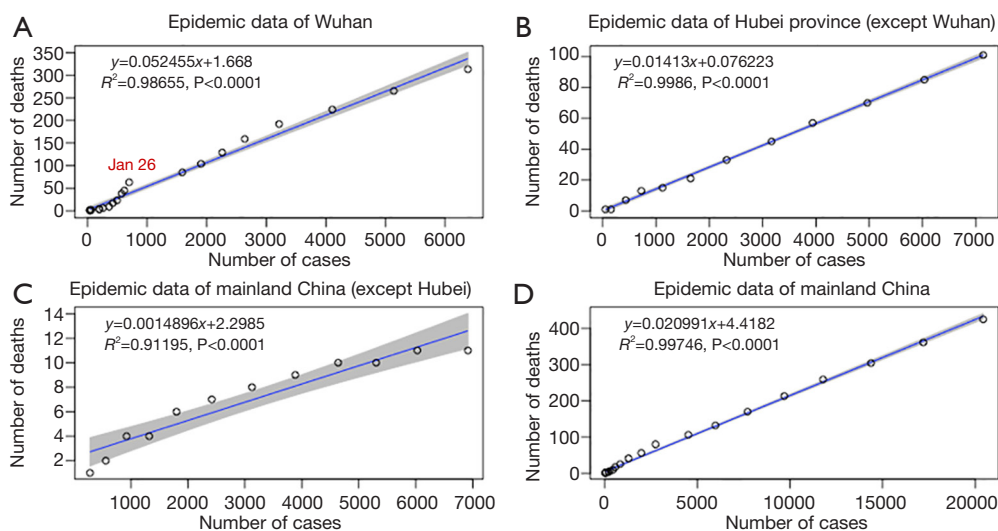


Figure 1 The simple linear model fitting for different clusters: (A) Wuhan, (B) Hubei province (except Wuhan), (C) mainland China (except Hubei), and (D) mainland China.

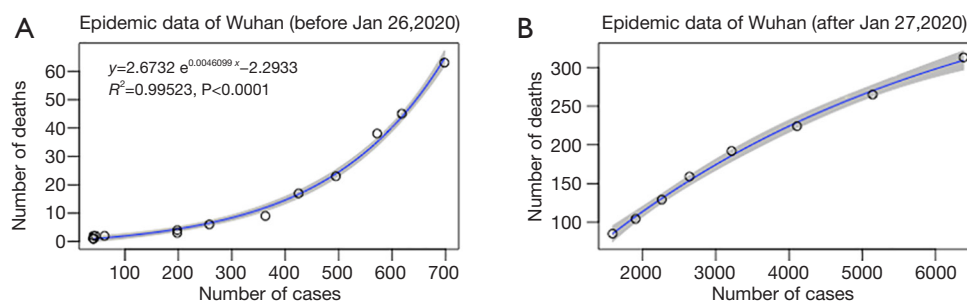


Figure 2 The exponential growth fitting for Wuhan before (A) and after (B) January 26, 2020.

The coefficient of determination, R-squared, ranged from 0.912 to 0.999 for all clusters, which implies that the early epidemic data of laboratory-confirmed cases and deaths were largely following the linear trend. As expected, the estimated CFR of COVID-19 in Wuhan was the highest, with an overall CFR of 5.25% (95% CI: 4.98–5.51%) (Figure 1A). As shown in the figure, two different trends appeared before and after January 26, 2020. Prior to January 26, 2020, the slope (CFR) gradually increased, while it slightly decreased afterwards. We further modelled the epidemic curve obeying the exponential growth for the time periods before and after January 26 separately. The goodness of fitting for both time periods improved compared with the linear model, with R-squared of 0.995 and 0.997, respectively (Figure 2).

Hubei province (except Wuhan) has the best goodness of linear fitting, with R-squared of 0.999 and CFR of 1.41%

(95% CI: 1.38–1.45%) (Figure 1B). The linear fitting result for mainland China (except Hubei province) is not as effective as previous two clusters, with R-squared of 0.912 (Figure 1C). The estimate of CFR was 0.15% (95% CI: 0.12–0.18%), which is the lowest among three clusters. An exponential curve fitting for this cluster generated a better R-squared of 0.992 (Figure 3). We also applied the linear model to fit the epidemic data of whole mainland China. The estimated CFR is 2.10% (95% CI: 2.05–2.14%) and R-squared is 0.997 (Figure 1D).

Discussion and Conclusions

In the early phase of the COVID-19 outbreak, the detection rate of mild patients was low in Wuhan, only patients with serious conditions could be confirmed due to hospitalization. Therefore, the case fatality rate (CFR)

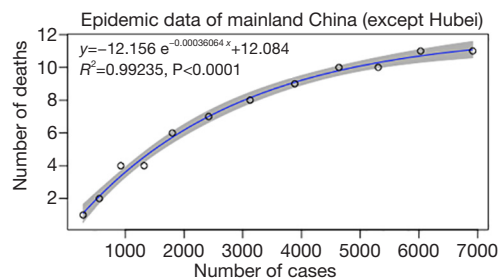


Figure 3 The exponential growth fitting for mainland China (except Hubei).

has been increasing over time, resulting in a high overall estimate of 5.25% (95% CI: 4.98–5.51%). Subsequently, a large number of new cases were detected from January 27 onwards due to the improved the efficiency of detection in Wuhan. Thereafter the CFR has gradually decreased, but it is still the highest in the country, indicating that the pressure of epidemic prevention and control and clinical treatment in Wuhan are still very high.

The Hubei province (except Wuhan) has an estimated CFR of 1.41% (95% CI: 1.38–1.45%). Based on our estimate of the confidence interval, the CFR in this cluster will be less than 1.5%. We predict that the CFR level in Hubei province (except Wuhan) will be maintained in the near future if there are no major changes in detection methods, healthcare resources, clinical treatment, and other factors. The cluster of mainland China (except Hubei province) still has a decreasing trend in CFR with an estimate of 0.15% (95% CI: 0.12–0.18%), clearly lower than Hubei. Since the epidemic situation in Wuhan has been further aggravated, the Chinese Government has implemented metropolitan-wide quarantine of Wuhan and several nearby cities from January 23, 2020 (15). Considering the initial spread of the epidemic outside Hubei province, the state and local governments at all levels have paid enough attention and adopted strong control measures. Numerous domestic airports and train stations have adopted temperature screening measures to detect passengers with fever. Individuals from Hubei province will be quarantined and further observed. Thus, the estimated CFR in regions other than Hubei province, where efficient detection and timely treatment can be guaranteed by sufficient healthcare resources, is more likely to be closer to the true value of clinical treatment prognosis estimates of COVID-19. Our estimate of CFR for whole mainland China, which is 2.10% (95% CI: 2.05–2.14%), is consistent with the tentative estimate of 2% for CFR in a preliminary

estimation (16).

In this study, we provide a purely data-driven statistical method to estimate the CFR in the early phase of the COVID-19 outbreak. Compared to dynamic transmission models, our model is simple and does not require complex parameter assumptions. The model fitting efficiency is high and the estimation of confidence interval is accurate in the early phase of the outbreak. Similar model was previously implemented to give future population projections and we employed it for the first time to estimate CFR (17). To our knowledge, this is by far the most comprehensive estimation for the CFR of the COVID-19 jointly completed by a multidisciplinary team, including respiratory physicians, statisticians, epidemiologists, and evidence-based medicine experts. Our findings could be used to inform guideline development, systematic review and knowledge translation programs (18–20).

The CFR of the same disease varies greatly in different countries or even different regions of the same country, and will be affected by numerous factors such as health control policies, medical standards, and detection efficiency. In view of the regional differences in distribution of factors that may affect the estimation of CFR, it is our strength to estimate the CFR in three separate clusters of China to provide more practical and referential importance. Taking into account that the vast majority of COVID-19 infections occur in China, we believe that the results of this study are more in line with the actual situation. Nonetheless, our study is still subject to limitations. First, the cumulative number of confirmed cases and the cumulative number of deaths we employed in our study are monotonic over time, resulting in generating generally higher R-squared in linear fitting models. Therefore, we also did exponential curve fitting and also obtained very high goodness of fitting. These curves can be used for studying the future trend of CFR. Second, our study is a statistical analysis of the data explaining the early epidemic of COVID-19, with only limited ability to predict what the CFR will finally be. Considering the ongoing national and global outbreaks of COVID-19, the CFR of the outbreak could change at the later stage. It is challenging but worth trying to estimate the CFR in the early phase to provide information for successful control on the COVID-19.

In summary, our estimates of CFR, ranging from 0.15% (95% CI: 0.12–0.18%) in mainland China excluding Hubei through 1.41% (95% CI: 1.38–1.45%) in other cities of Hubei province to 5.25% (95% CI: 4.98–5.51%) of Wuhan, indicate that the fatality caused by COVID-19 is lower than

the previous coronavirus epidemics caused by SARS-CoV (CFR: 9.6% globally and 6.4% in mainland China) (10,11) and MERS-CoV (CFR: 34.5% worldwide and 37.1% in Saudi Arabia) (12,13).

Acknowledgments

We are extremely thankful to Janne Estill for his comments and revisions on earlier drafts of our paper.

Funding: This study was supported by National Natural Science Foundation of China (Grant Number 81903406, 81804222), China Postdoctoral Science Foundation (Grant Number 2017M620378), 2020 Key R & D project of Gansu Province, China (Reference Number 2020001), and General Research Fund (Grant Number 15205119) of the Research Grants Council (RGC) of Hong Kong, China.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethics Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The ethical approval or individual consent was not applicable.

References

1. Wang C, Horby PW, Hayden FG, et al. A novel coronavirus outbreak of global health concern. *Lancet* (London, England) 2020.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497-506.
3. World Health Organization. Novel Coronavirus(2019-nCoV) Situation Report – 14. Available online: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200203-sitrep-14-ncov.pdf?sfvrsn=f7347413_4. February 3rd, 2020.
4. de Wit E, van Doremalen N, Falzarano D, et al. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016;14:523-34.
5. Song Z, Xu Y, Bao L, et al. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses* 2019. doi: 10.3390/v11010059.
6. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 2004;160:509-16.
7. 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. [Epub ahead of print].
8. Zhao S, Musa SS, Lin Q, 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. doi: 10.3390/jcm9020388.
9. Imperial College London MRC Centre for Global Infectious Disease Analysis. News/Wuhan Coronavirus, 2020. Available online: <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/>. January 29, 2020.
10. Donnelly CA, Ghani AC, Leung GM, et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 2003;361:1761-6.
11. Jia N, Feng D, Fang LQ, et al. Case fatality of SARS in mainland China and associated risk factors. *Trop Med Int Health* 2009;14 Suppl 1:21-7.
12. Majumder MS, Rivers C, Lofgren E, et al. Estimation of MERS-Coronavirus Reproductive Number and Case Fatality Rate for the Spring 2014 Saudi Arabia Outbreak: Insights from Publicly Available Data. *PLoS Curr* 2014. doi: 10.1371/currents.outbreaks.98d2f8f3382d84f390736cd5f5fe133c.
13. Lin Q, Chiu AP, Zhao S, et al. Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. *Stat Methods Med Res* 2018;27:1968-78.
14. World Health Organization. Novel Coronavirus(2019-nCoV) Situation Report – 16 Available online: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200205-sitrep-16-ncov.pdf?sfvrsn=23af287f_4. February 6, 2020.
15. Phelan AL, Katz R, Gostin LO. The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance. *JAMA* 2020. [Epub ahead of print].
16. World Health Organization. Update on the situation regarding the new coronavirus. Available online: https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-script-ncov-rresser-unog-29jan2020.pdf?sfvrsn=a7158807_4. January 29, 2020.
17. Haque M, Ahmed F, Anam S, et al. Future population projection of Bangladesh by growth rate modeling using logistic population model. *Annals of Pure and Applied*

- Mathematics 2012;1:192-202.
18. Tian J, Zhang J, Ge L, et al. The methodological and reporting quality of systematic reviews from China and the USA are similar. *J Clin Epidemiol* 2017;85:50-8.
 19. Yao L, Sun R, Chen YL, et al. The quality of evidence in Chinese meta-analyses needs to be improved. *J Clin Epidemiol* 2016;74:73-9.
 20. Yang K, Chen Y, Li Y, et al. Editorial: can China master the guideline challenge? *Health Res Policy Syst* 2013;11:1.

Cite this article as: Yang S, Cao P, Du P, Wu Z, Zhuang Z, Yang L, Yu X, Zhou Q, Feng X, Wang X, Li W, Liu E, Chen J, Chen Y, He D; on behalf of COVID-19 evidence and recommendations working group. Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis. *Ann Transl Med* 2020;8(4):128. doi: 10.21037/atm.2020.02.66