Title Page

Forecasting the Cumulative Number of COVID-19 Deaths in China: Can More Lives Be Saved?

Zuqin Zhang [1, #], Cheng Long [2, #], Wei Yao [1], Qi Ying [3, *] and Xinmiao Fu [1, *]

[1] Provincial University Key Laboratory of Cellular Stress Response and Metabolic Regulation, College of Life Sciences, Fujian Normal University, Fuzhou City, Fujian Province 350117, China
[2] West China Hospital of Sichuan University, Chengdu City, Sichuan Province, China
[3] Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843, USA.

Running title: Forecasting COVID-19 Deaths in China

[#] These authors contributed to this work equally.
* To whom correspondence should be addressed to Professors Xinmiao Fu (xmfu@fjnu.edu.cn) and Qi Ying (qying@civil.tamu.edu)

Keywords: COVID-19, SARS-CoV-2, 2019-nCoV, Boltzmann function, coronavirus, SARS, epidemic

Abstract

Background: An outbreak of 2019 novel coronavirus diseases (COVID-19) caused by SARS-CoV-2 is on-going in China and appears to approach late phase. It is highly demanding to estimate how many COVID-19 patients will die eventually. In this study, an estimate of the potential total number of COVID-19 deaths in mainland China, Hubei Province, Wuhan City, and other provinces is provided. The results may help to eluate the severity of the epidemic and facilitate mental health care.

Methods: Data of the cumulative number of COVID-19 deaths from January 21 to February 29, 2020, released daily by the National Health Commission of China and Hubei Provincial Health Commission, were used. The Boltzmann function was explored to simulate the data in each region. By using these established functions, the potential total number of deaths were forecasted. In addition, data of the cumulative number of 2003 SARS deaths in mainland China, Hong Kong and worldwide were collected from the WHO official website and analyzed in a similar manner. A Monte Carlo technique was applied to analyze the uncertainty of the estimates of the cumulative confirmed cases, and the results are presented using the resulting mean, median, and a 95% confidence interval (CI). For comparison, the potential total numbers of deaths were estimated by Richards function-based regression analyses.

Findings: The data of cumulative number of COVID-19 deaths with respect to each region were all well-fitted to the Boltzmann function (R^2 for all the regression analyses being close to 0.999). Consistently, the data for the cumulative numbers of 2003 SARS in mainland China, Hong Kong and worldwide were also well fitted to the Boltzmann function. The potential total number of COVID-19 deaths in mainland China, other provinces, Hubei Province, Wuhan City, and other cities in Hubei were estimated to be 3260 (95% CI 3187, 3394), 110 (109, 112), 3174 (3095, 3270), 2550 (2494, 2621) and 617 (607, 632), respectively. Similar results were obtained by Richards function-based regression analysis.

Interpretation: The observation that the data of the cumulative numbers of deaths for both the on-going COVID-19 outbreak and the 2003 SARS epidemic in each geographic region were well fitted with the Boltzmann function strongly suggests that it is suitable for the simulation of deaths associated with coronaviruses-induced diseases. The estimation of COVID-19 deaths may help governments to evaluate the severity of the outbreak and also facilitate timely mental health care for the families of dead patients.

An outbreak of 2019 novel coronavirus diseases (COVID-19) caused by SARS-CoV-2 is on-going in China and has spread worldwide [1-3]. As of Feb 29, 2020, there have been 79824 confirmed COVID-19 patients and 2870 deaths in China, and the epicenter of the outbreak, Wuhan city and related regions in Hubei province of China have reported 66906 confirmed patients and 2761 deaths. Although the number of new confirmed cases has substantially decreased since Feb 13, 2020 and the outbreak appears to approach late phase in China, people have raised grave concerns about the severity of the outbreak, especially questioning how many patients will die eventually. Here we estimated the potential total number of COVID-19 deaths by applying Boltzmann function-based regression analysis, an approach we recently developed for estimating the potential total numbers of confirmed cases for both the ongoing SARS-CoV-2 outbreak and the gone 2003 SARS epidemic [4].

We collected data for analysis on the officially released cumulative numbers of deaths in mainland China, other provinces than Hubei, Hubei Province, Wuhan City, and other cities in Hubei (from Jan 21 to Feb 29, 2020). We first verified that the cumulative numbers of confirmed cases with respect to each region were all well fitted to the Boltzmann function (R^2 all being close to 0.999);

Fig. 1A), consistent with our earlier report using the data from Jan 21 to Feb 14, 2020 [4]. Assuming that the number of deaths is proportional to the number of confirmed cases for the outbreak under specific circumstances, we speculated that the cumulative number of COVID-19 deaths would also obey the Boltzmann function. In support of this speculation, the cumulative numbers of COVID-19 deaths in the above regions were all well-fitted to the Boltzmann function (R^2 all being close to 0.999; **Figs. 1B**, **1C** and **Table 1**), with the potential total numbers of deaths being estimated as 3200±40, 108±1, 3100±40, 2500±40 and 604±6 respectively (**Table 1**). This result, in conjunction with our earlier observation that the cumulative numbers of confirmed cases of 2003 SARS in mainland China and worldwide were well fitted with the Boltzmann function, prompted us to analyze the cumulative numbers of 2003 SARS deaths in the same way. Consistently, we observed that the cumulative numbers of 2003 SARS deaths in mainland China, Hong Kong and worldwide were all well fitted to the Boltzmann function (**Fig. 1D**), strongly suggesting that the Boltzmann function is suitable to simulate the course of deaths associated with coronavirus-caused diseases.

One issue regarding our analyses is that some COVDI-19 deaths might be miss-reported such that the reported death numbers represent a lower limit. For instance, 134 new deaths were suddenly counted from more than 13000 clinically diagnosed patients in Hubei Province on Feb 12, 2020 (as clearly indicated by a sudden jump of deaths in **Fig. 1B**). Another uncertainty might result from those unidentified COVID-19 deaths at the early phase of the outbreak. We applied the Monte Carlo method (for detail, refer to the Methods section in SI file) to estimate such uncertainty assuming that the relative uncertainty of the reported numbers of deaths follows a single-sided normal distribution with a mean of 1.0 and a standard deviation of 2.5%. The potential total numbers of COVID-19 deaths in the above regions were estimated to be 3260 (95% CI 3187, 3394), 110 (109, 112), 3174 (3095, 3270), 2550 (2494, 2621) and 617 (607, 632), respectively (**Figs. 1E** and **S1**), which are slightly higher than those estimated without uncertainty (refer to **Table 1**).

To verify our Boltzmann function-based estimations, we calculated the potential total numbers of deaths in the above regions by applying Richards function-based regression analyses, which had been explored to simulate the cumulative numbers of confirmed cases of 2003 SARS in different regions [5]. The potential total numbers of COVID-19 deaths in mainland China, other provinces, Hubei Province, Wuhan City and other cities were estimated to be 3342 (3214, 3527), 111 (109, 114), 3245 (3100, 3423), 2613 (2498, 2767) and 627 (603, 654), respectively (**Figs. 1F** and **S2**), which are close to what are estimated by our Boltzmann function-based analyses (**Table 1**).

Collectively, we observed that all sets of data from both the COVID-19 deaths and the 2003 SARS deaths were well fitted to the Boltzmann function. We propose that the Boltzmann function is suitable for analyzing not only the cumulative number of confirmed COVID-19 cases, as reported by us recently [4] (also refer to **Fig. 1A**), but also those of deaths as reported here. We noticed that the COVID-19 deaths have been estimated by other groups using different models. Li et al recently reported in this Journal [6] using the data from Jan 20 to Feb 11 that a total of deaths in Hubei would be 2250, a number much lower than the currently observed (2761 as of Feb 29). Using the Susceptible-Infected-Recovered-Dead model Anastassopoulou et al forecasted that the total death might exceed 7,000 by February 29 [7], a number apparently much higher than the real one.

Since the case fatality and mortality rates in the epicenter of the outbreak are still much higher than that in other provinces of mainland China, there is a great potential for government to optimize preparedness and medical resource supplies therein, by which hundreds of lives of COVID-19 patients, particularly those severe and critically ill patients [3, 8], might be saved. This potential is reflected the consecutive decrease of the case fatality rate in Hubei and Wuhan since

Chinese government provided a great amount of medical supplies therein [2, 3]. In addition, our estimates on the course of COVID-19 deaths (refer to **Table S2**) may benefit the mental health service that needs to be timely provided to the families of dead patients [9, 10].

Acknowledgments

This work is support by the National Natural Science Foundation of China (No. 31972918 and 31770830 to XF).

We declare no competing interests.

References

1. Tang, J.W., P.A. Tambyah, and D.S.C. Hui, Emergence of a novel coronavirus causing respiratory illness from Wuhan, China. Journal of Infection, 2020; https://doi.org/10.1016/j.jinf.2020.01.014.
2. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 16-24 February 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.
3. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. Vital surveillances: the epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020, in China CDC Weekly. 2020. p. http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51.
4. Fu, X., et al., Simulating and Forecasting the Cumulative Confirmed Cases of SARS-CoV-2 in China by Boltzmann Function-based Regression Analyses. Journal of Infection, 2020; https://doi.org/10.1016/j.jinf.2020.02.019.
5. Hsieh, Y.H., J.Y. Lee, and H.L. Chang, SARS epidemiology modeling. Emerging Infect Dis, 2004. 10(6:) 1165–1167.
6. Li, Q., W. Feng, and Y.H. Quan, Trend and forecasting of the COVID-19 outbreak in China. Journal of Infection, 2020; https://doi.org/10.1016/j.jinf.2020.02.014.
7. Anastassopoulou, C.e.a., DATA-BASED ANALYSIS, MODELLING AND FORECASTING OF THE NOVEL CORONAVIRUS (2019-NCOV) OUTBREAK. https://doi.org/10.1101/2020.02.11.20022186, 2020.
8. Yang X., e.a., Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med, 2020; https://doi.org/10.1016/S2213-2600(20)30079-5.
9. Liu, S., et al., Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry, 2020; https://doi.org/10.1016/S2215-0366(20)30077-8.
10. Xiang, Y.T., et al., Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiatry, 2020. 7(3) 228-229.

Table 1 Summary of the estimated total numbers of COVID-19 deaths in China

Regions	**Boltzmann** [a]		**Boltzmann** [a] **(mean, 95% CI)**	**Richards** [a] **(mean, 95% CI)**
	Mean±SD	R^2		
mainland China	3200±40	0.999	3260 (3187, 3394)	3342 (3214, 3527)
other provinces	108±1	0.996	110 (109, 112)	111 (109, 114)
Hubei Province	3100±40	0.999	3174 (3095, 3270)	3245 (3100, 3423)
Wuhan City	2500±40	0.998	2550 (2494, 2621)	2613 (2498, 2767)
other cities in Hubei	604±6	0.999	617 (607, 632)	627 (603, 654)

[a] Boltzmann function-based regression analysis results assuming the reported cumulative number of deaths are precise and have uncertainty (a standard deviation of 2.5%), respectively (for detail, refer to the Methods section in the supporting information file and **Figs. 1E, S1**). The same uncertainty was set for Richards function-based regression analysis (for detail, refer to the Methods section and **Figs. 1F, S2**)

Figure 1

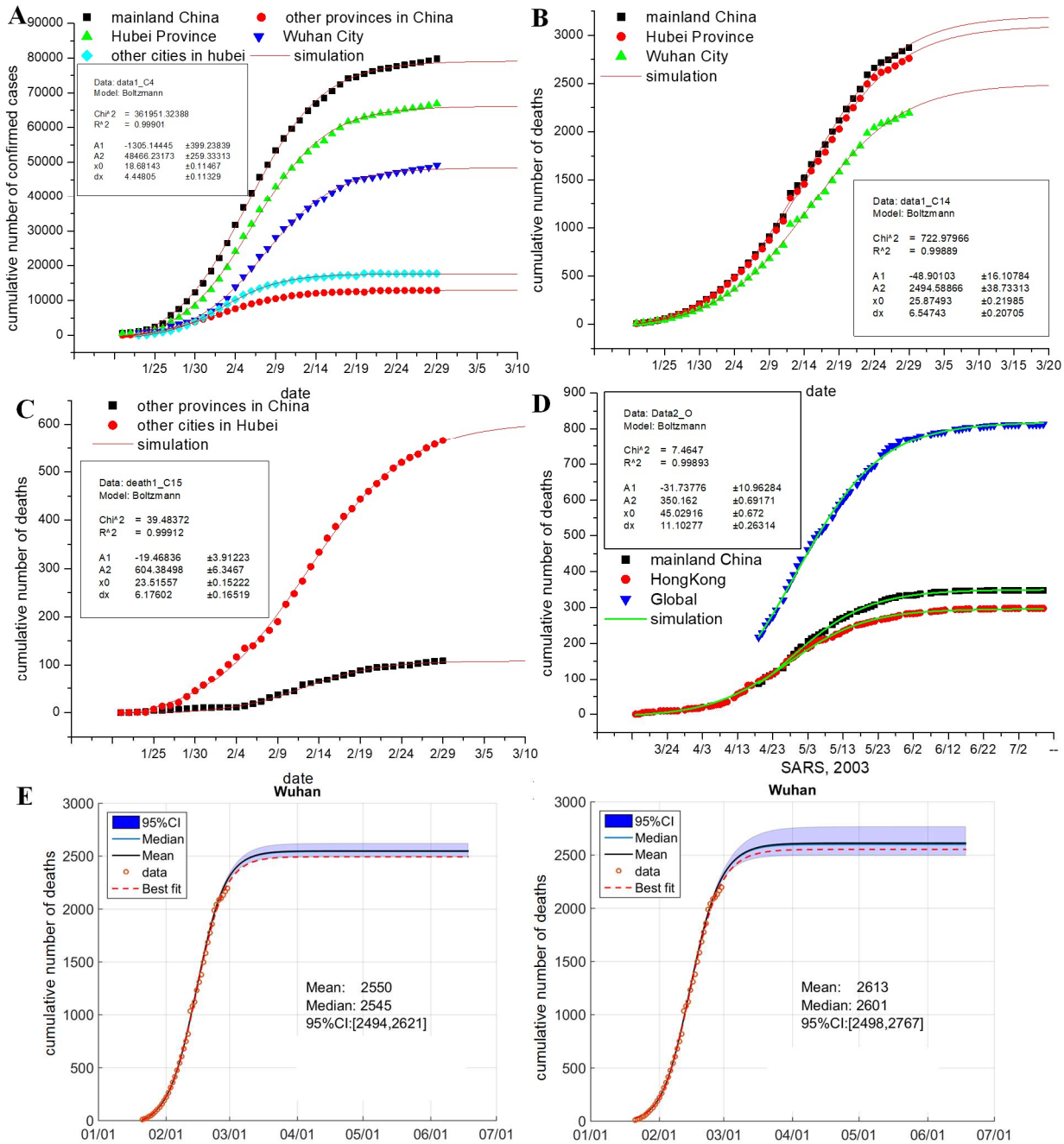

Figure 1. Fitting the cumulative number of COVID-19 deaths to Boltzmann function

(**A**, **B**, **C**) Boltzmann function-based regressions analysis results on the cumulative numbers of confirmed COVID-19 cases (panel A) and deaths (panels B and C) in the indicated geographic regions. Parameters of the established functions for Wuhan City (panels A and B) and for other cities in Hubei (panel C) are shown in insets. Note: the reported cumulative number of confirmed cases of Hubei Province and Wuhan City were re-adjusted for data fitting due to the suddenly added cases determined using clinical features (for details, refer to **Table S1**). (**D**) Boltzmann function-based analysis results on the cumulative numbers of 2003 SARS deaths in the indicated regions. Parameters of the established function for mainland China are shown in insets. (**E**, **F**). Regressions analysis results on the cumulative numbers of COVID-19 deaths in Wuhan City by Boltzmann (panel E) and Richards (panel F) functions, assuming that the relative uncertainty of the data follows a single-sided normal distribution with a mean of 1.0 and a standard deviation of 2.5%. Original data are shown as circles; simulated results are presented as colored lines as indicated. Inserts show key statistics. Results for other regions are presented in **Figs. S1** and **S2**.