Case fatality rate of novel coronavirus disease 2019 in China

Rui Qi, Chao Ye, Xiang-rong Qin, Xue-Jie Yu

1. Wuhan University School of Health Sciences, State Key Laboratory of Virology, Wuhan University, Wuhan, Hubei Province, China. (Rui Qi, Xue-Jie Yu, Xiang-rong)

2. Department of Medical Record Statistics, Shandong Maternal and Child Health Care Hospital, Jinan, Shandong Province, China. (Chao Ye)

Correspondence to:

Xue-Jie Yu, M.D., Ph.D

Professor and Dean

Wuhan University School of Health Sciences

No. 115 Donghu Road, Wuhan, 430071

Hubei Province, China

E-mail: yuxuejie@whu.edu.cn.

Tel: +86 27 68759399

Adjunct Professor

Departments of Pathology

University of Texas Medical Branch

Galveston, Texas 77555-0609, USA.

xuyu@utmb.edu

Abstract

Background: A pandemic of coronavirus disease 2019 (COVID-19) which have caused more than 75 thousand persons infected globally is still ongoing. This study aims to calculate its case fatality rate (CFR).

Methods: The method was based on the formula of dividing the number of known deaths by the number of confirmed cases T days before, where T was an average time period from case confirmation to death. It was found that supposing a T, if it was smaller (bigger) than the true T, calculated CFRs would gradually increase (decrease) to infinitely near the true CFR with time went on. According to the law, the true T value could be determined by trends of daily CFRs calculated with different assumed T values (left of true T is decreasing, right is increasing). Then the CFR could be calculated.

Results: CFR of COVID-19 in China except Hubei Province was 0.8%. So far, the CFR and T value (T was 9) had accurately predicted the death numbers more than two weeks continuously. CFR in Hubei of China was 5% by which (and T = 5) the calculated death number corresponded with the reported number in the last week.

Conclusions: The method could be used for diseases to calculate CFR without requiring the end of pandemics. Dynamic monitor of the assumed CFRs trends could help outbreak-controller have a clear vision of the case-confirmation situation**.**

Background

An outbreak of pneumonia caused by a novel coronavirus occurred in Wuhan, Hubei Province, China at the end of 2019 [1]. On Feb 11, 2020 the World Health Organization (WHO) announced an official name for the disease as coronavirus disease 2019 (COVID-19) and the International Committee on Taxonomy of Viruses named the novel coronavirus as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak of the disease was linked to a live animal market firstly and then was reported person-to-person transmission [2, 3]. The disease has rapidly spread from Wuhan to other areas. As of Feb 19, 2020, more than 74 thousand cases in China alone have been confirmed. Cases also have been reported in more than 25 countries of 5 continents. The case fatality rate (CFR) represents the proportion of people who eventually die from a specified disease. CFR typically is used as a measure of disease severity and is often used for prognosis where comparatively high rates are indicative of relatively poor outcomes [4]. It also can be used to compare the effect of treatments among different areas. In general, when a pandemic has ended, CFR can be calculated by dividing the number of known deaths by the number of confirmed cases. A major difficulty in estimating case fatality rate is ensuring the accuracy of the numerator and the denominator. At present, it is tempting to estimate the case fatality rate by dividing the number of known deaths by the number of confirmed cases. The resulting number, however, does not represent the true case fatality rate because this calculation does not account for the delay between case confirmation and disease outcome [5]. In that case, the CFR will be underestimated. To estimate the CFR while a pandemic is still ongoing, the denominator should be corrected as cases at T days before, where T is an average time period from case confirmation to death. This study

aims to calculate the CFR of the COVID-19 in China by estimating the average time period from case confirmation to death based on population level data.

Methods

Data: population level data in this study included daily accumulative numbers of cases and deaths of COVID-19 in China from Jan 21 to Feb 19, 2020. Data was collected from WHO, China CDC, and provincial level health authorities.

Estimation of T (average time period from case confirmation to death): To calculate CFR, it should be realized that deaths at day X are averagely from cases at day X-T rather than day X. Given a T value, CFR at day X can be calculated by case number at day X-T dividing deaths number at day X. Thus, a group of CFRs (daily CFRs) can be obtained from different X days. As known that death number at day X should be less than case number at day X-T (if more than day X-T, CFR would be greater than 100% which is illogical). Based on this point, the range of T can be narrowed. More importantly, no matter what T value is assumed, even it is far away from the true T value, the daily CFRs would theoretically converge towards (infinitely approach to but never be over) the true CFR with time (X) increases. The following example will illustrate this principle (Table 1). Assuming CFR = 10%, T = 4 for a disease, the cases number was from 100 to 10000 at day X (X=1 to 100), then the deaths number would be 10 (10, 20 and so on) at day X+4 (5, 6 and so on). When calculating daily CFRs based on case and death numbers with formula deaths (X) divided by cases (X-T), law 1: if assumed T was equal to the true T value (4 in the example), calculated daily CFRs at different day X would constantly be the true CFR (0.1); if assumed T was greater than the true T (5 and 6), daily CFRs would be greater than the true CFR (0.1) and infinitely reduce to near it with the

time (X) increased; if assumed T was smaller than the true T (1 to 3), daily CFRs would be smaller than the true CFR and infinitely increase to near it. Besides, it could be found that, law 2: if assumed T was more far away (bigger absolute difference) from the true T, daily CFRs would be more far away from the true CFR and they would need more times to converge towards it. In this example, case numbers were given from 100 to 10000 by 100 increments per day, however, cases growth every day would not be evenly for an infectious disease. Then case numbers in this example were replaced by real case numbers of COVID-19 and the convergence tendency still remained except for individual data points. Based on convergence laws, we used exhaustive method to calculate daily CFRs of COVID-19 by different T values. If an assumed T resulted in relatively constant daily CFRs, and T+1 resulted in decreasing daily CFRs and T-1 in increasing, it could be determined as the true T.

Table 1. Convergence tendency at different T values for a CFR=10% assumed disease

Daily CFR when			Days (X)	Cases (at X)	Deaths (at X+4)	Daily CFR when		
T=6	T=5	T=4	Days (X)	Nc	(Nc*CFR)	T=3	T=2	T=1
			1	100				
			2	200				
			3	300				
			4	400				
		0.100	5	500	10	0.050	0.033	0.025
	0.200	0.100	6	600	20	0.067	0.050	0.040

0.300	0.150	0.100	7	700	30	0.075	0.060	0.050
0.200	0.133	0.100	8	800	40	0.080	0.067	0.057
0.167	0.125	0.100	9	900	50	0.083	0.071	0.063
0.150	0.120	0.100	10	1000	60	0.086	0.075	0.067
0.140	0.117	0.100	11	1100	70	0.088	0.078	0.070
……	……	……	……	……	……	……	……	……
0.102	0.101			10000	960	0.099	0.098	0.097

Results

CFR in China except Hubei Province (non-Hubei regions)

Since a total of 85 deaths were reported by Feb 19 in non-Hubei regions, the number was bigger than case numbers in Jan 21 (Table 2), we could narrowed the range of time T to 29 days. A number of T values less than 29 were selected for screening based on convergence laws. After different T values were tried, as Figure 1 showed, when assumed T was 11, the daily CFRs were decreasing and had no pronounced increase, when it was 0 to 7; the daily CFRs had pronounced increase after early time. CFRs increased as expected according to laws at later stage when in some assumed T values (e.g. T=0), but it decreased at early stage which seemed not satisfy the convergence laws. Actually, it was normal. Convergence laws happened due to the force of the true CFR drawing daily CFRs towards its direction by truly dominating accumulated number. At early stage, the outcome of death had not yet occurred resulting in daily CFRs decreasing with the growth of case number. Thus, T value exploration by convergence laws should depend on period of death growth.

Results of Figure 1 indicated the true T should be in the range of 8 to 10. As differences

between CFRs were too small at converging stage to compare and scales of y axis in different plots of Figure 1 varied greatly, Figure 1 was only used for preliminary tendency exploration. Converging stage CFRs had been cut out to plot with same y axis scales for the true T and CFR estimation (Figure 2).

Table 2. Death and case numbers of COVID-19 in China from Jan 21 to Feb 19

Date	Cumulative cases	Cumulative deaths	Cumulative cases	Cumulative death
	non-Hubei		Hubei	
J_21	56	0	270	0
J_22	105	0	444	0
J_23	321	2	549	24
J_24	601	2	729	32
J_25	937	3	1052	52
J_26	1332	4	1423	76
J_27	1714	6	2714	100
J_28	2439	7	3554	125
J_29	3117	8	4586	162
J_30	3247	9	4903	162
J_31	4006	9	5806	204
F_1	4737	10	7153	249
F_2	5283	10	9074	294
F_3	5970	11	11177	350

F_4	6911	11	13522	414
F_5	7755	14	16678	479
F_6	8473	15	19665	549
F_7	9154	19	22112	618
F_8	9720	25	24953	699
F_9	10189	33	27100	780
F_10	10631	38	29631	871
F_11	11019	43	31728	974
F_12	11399	48	33366	1068
F_13	11676	58	35699*	1310
F_14	11946	63	39479	1318
F_15	12170	67	41899	1457
F_16	12335	70	43742	1596
F_17	12453	76	45675	1696
F_18	12539	81	47482	1789
F_19	12597	85	49175	1921

* 13,332 clinical diagnosed cases were reported together on Feb 13; in our calculation the number had been excluded from confirmed case numbers since then.

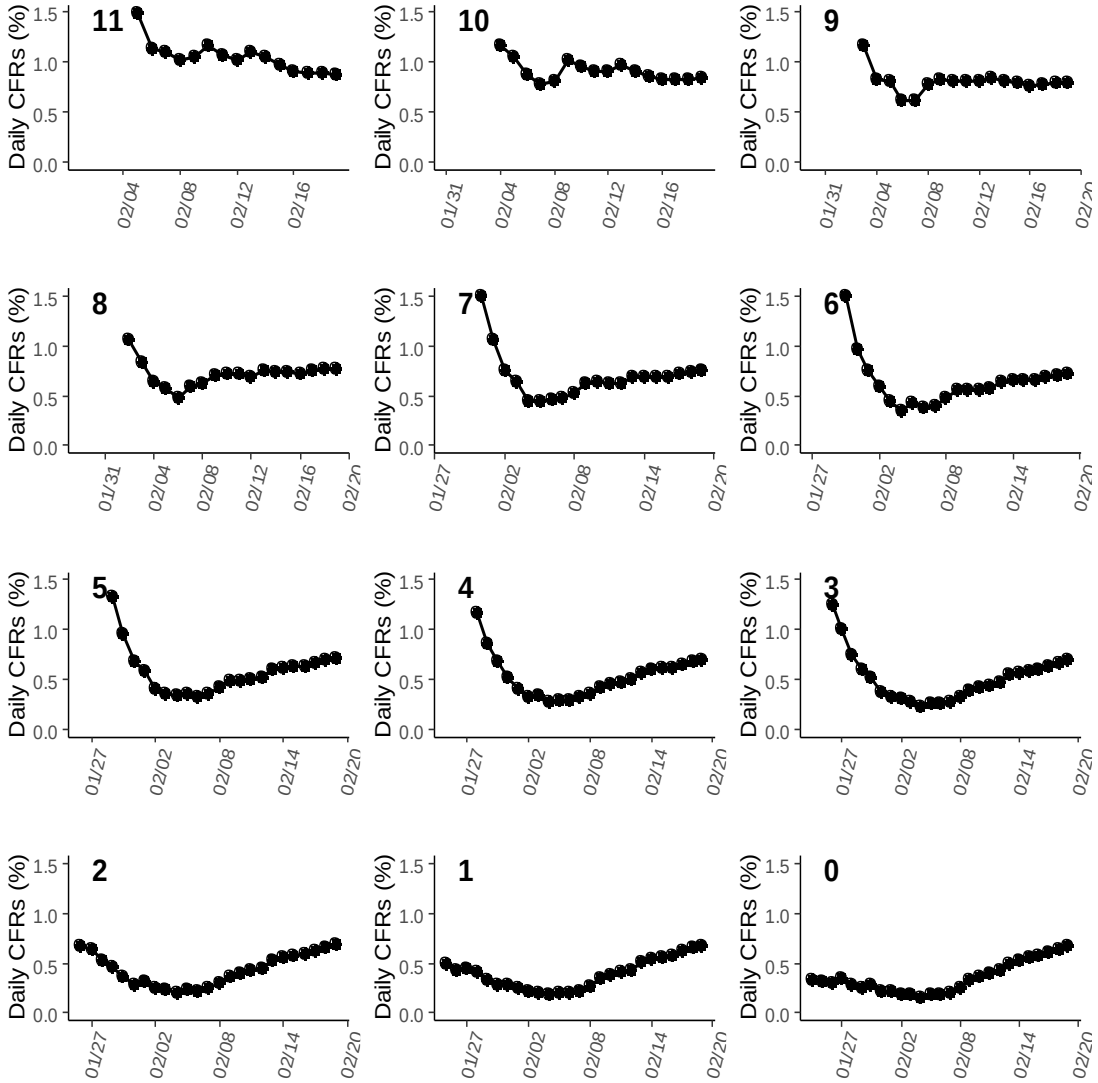

Figure 1. Calculated daily CFRs of non-Hubei regions from Feb 02 to Feb 19 when T was assumed from 0 to 1.

As mentioned in Methods, with the time increased, even under a false T in calculation, the daily CFR could converge towards the true CFR though more times needed. As Figure 2 showed, for T= 11 and 10, CFRs had pronounced decreasing trends, according to the laws 1 in the Methods, the true T should be bigger than 8 and less than 11. For T = 8, CFRs were slightly increasing, as mentioned, an assumed T was the closer to the true T value, the earlier convergence happened. When T = 9, the CFRs were almost staying in one line (red dotted line

in Figure 2). Linear models (blue lines) were generated for analysis of variances and linear trends of theses CFR points in each plot. The slopes of models became flatter and approached towards to be 0 when T was from 8 to 9 and 11 to 9. The results could accord with laws in Methods accurately. Then, T = 9 was determined for true CFR calculation of non-Hubei regions. There were two outliers which were thought to be random occurrence. The mean value of data in plot 9 of Figure 2 was 0.80% which could be determined as the true CFR of COVID-19 in China excluding Hubei Province. The CFR was higher than reported by the NHC of China (0.2%).

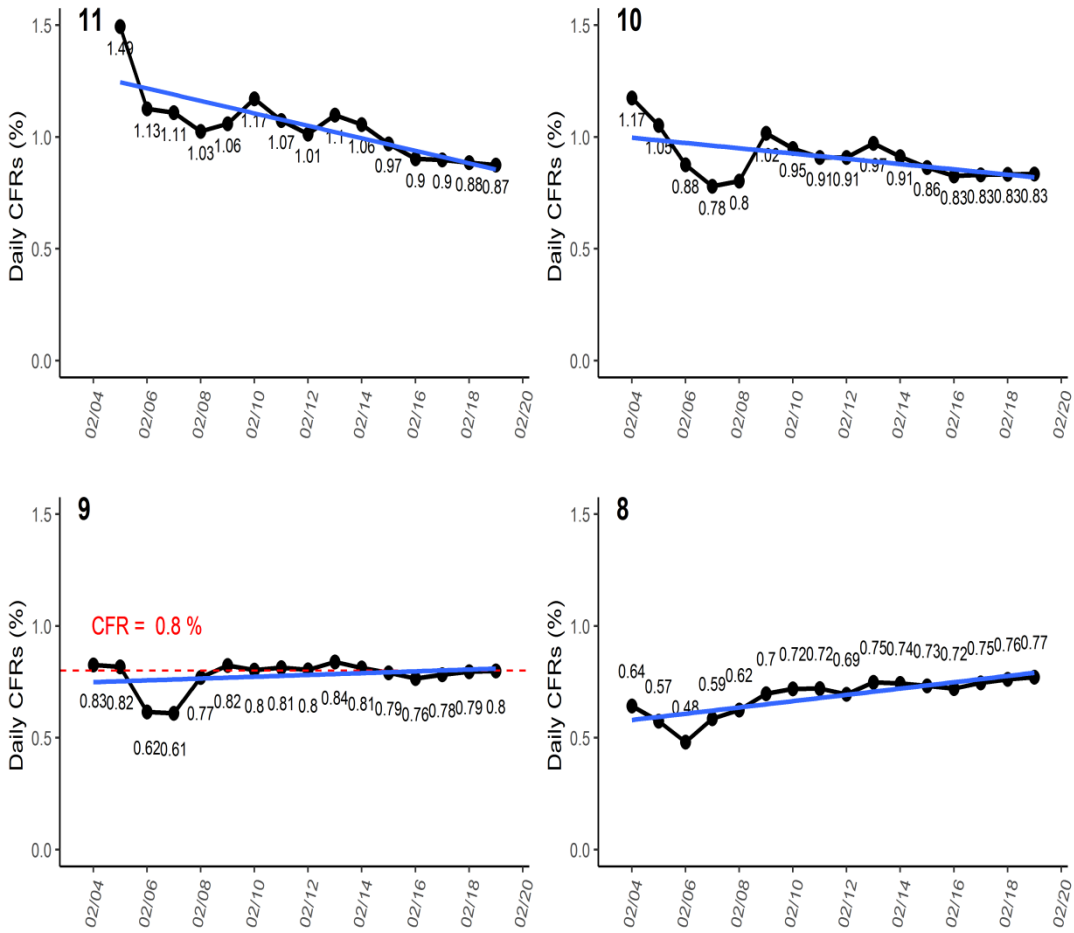

Figure 2. Converging stage daily CFRs of non-Hubei regions when T was from 8 to 11

Blue lines represented linear models generated for analysis of variances and linear trends of

these data points in plots. Red dotted line and number were the true CFR.

CFR in Hubei Province

As shown in Figure 3, after Feb 3, death number (350) were more than Jan 21 case number (270), if the T was 12 (Feb 3 minus Jan 1), the CFR would be illogically greater than 100%. In another words, death numbers only when before day 12 were less than case number at day 1. So the time T should be less than or equal to 11 days (12-1). The death number when was firstly more than the case number at day 2 was Feb 5 (day 15), so the T should be less than or equal to 12 (14-2). The rest could be done in the same manner. Finally, the smallest T value (T = 11) was selected as the upper limit for convergence screening.

Figure 3 was the calculation of daily CFRs with assumed T values (0 to 11), when assumed T was 7 to 11, daily CFRs were continuously decreasing. When T = 0 and 2, there were increase trends at later which meant they were smaller than the true T value. Thus, converging stage CFRs data when T = 3 to 6 was selected for plotting with the same y axis scales (Figure 4). As it showed, For T= 4 to 3, CFRs had increase trends and when T = 5, the CFRs had not pronounced increase or decrease trend though these points slightly fluctuated around. When T was 6, plateau stage appeared lather than T = 5.The slopes of linear models became flatter and approached towards to be 0 when T approached to 5. Then T =5 was selected as the true T value for the true CFR calculation. The true CFR of COVID-19 in Hubei calculated by mean value of the daily CFRs of plot 5 in Figure 4 was 5%.

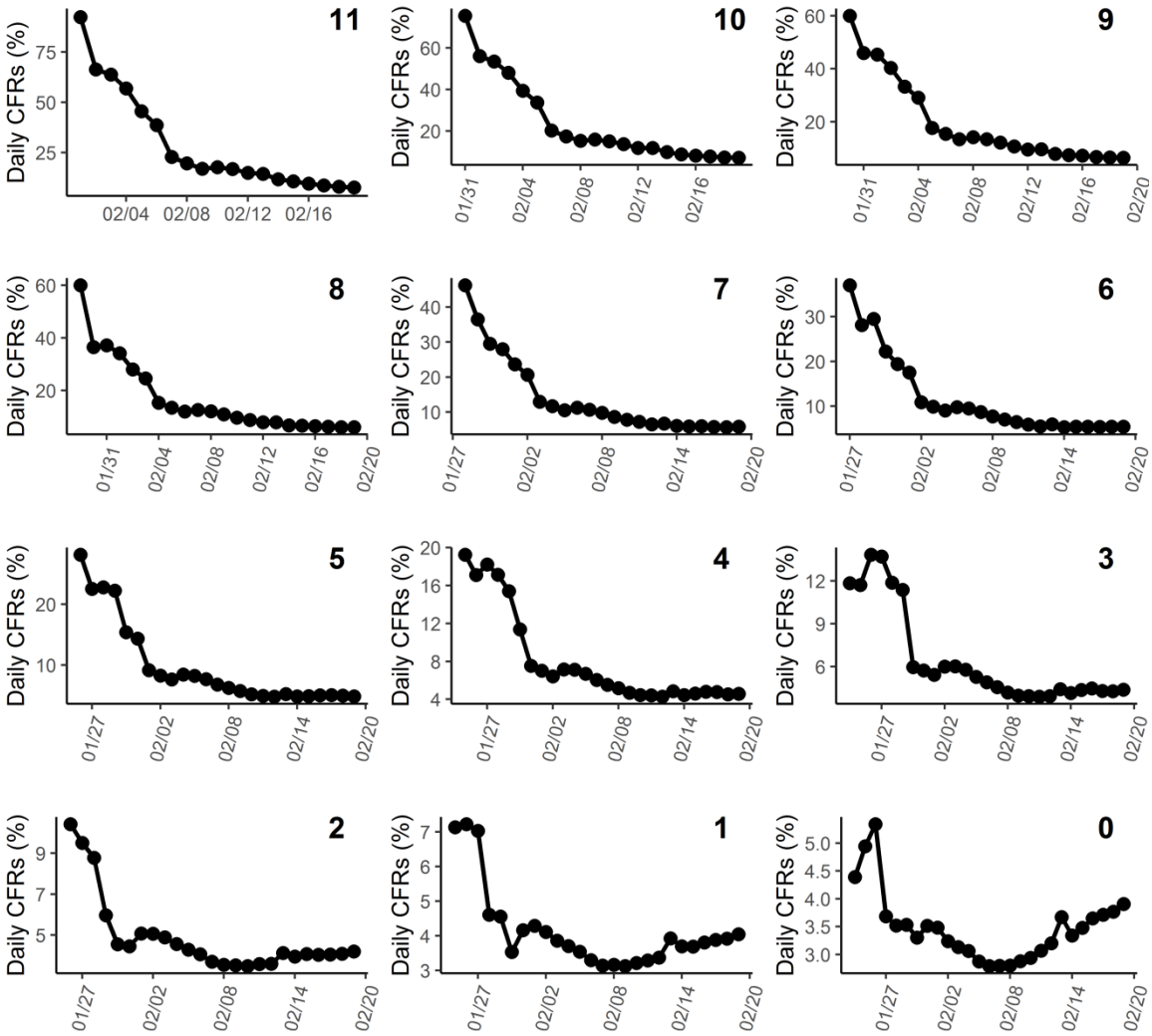

Figure 3. Calculated daily CFRs of Hubei Province from Feb 02 to Feb 19 when T was from 0 to 5.

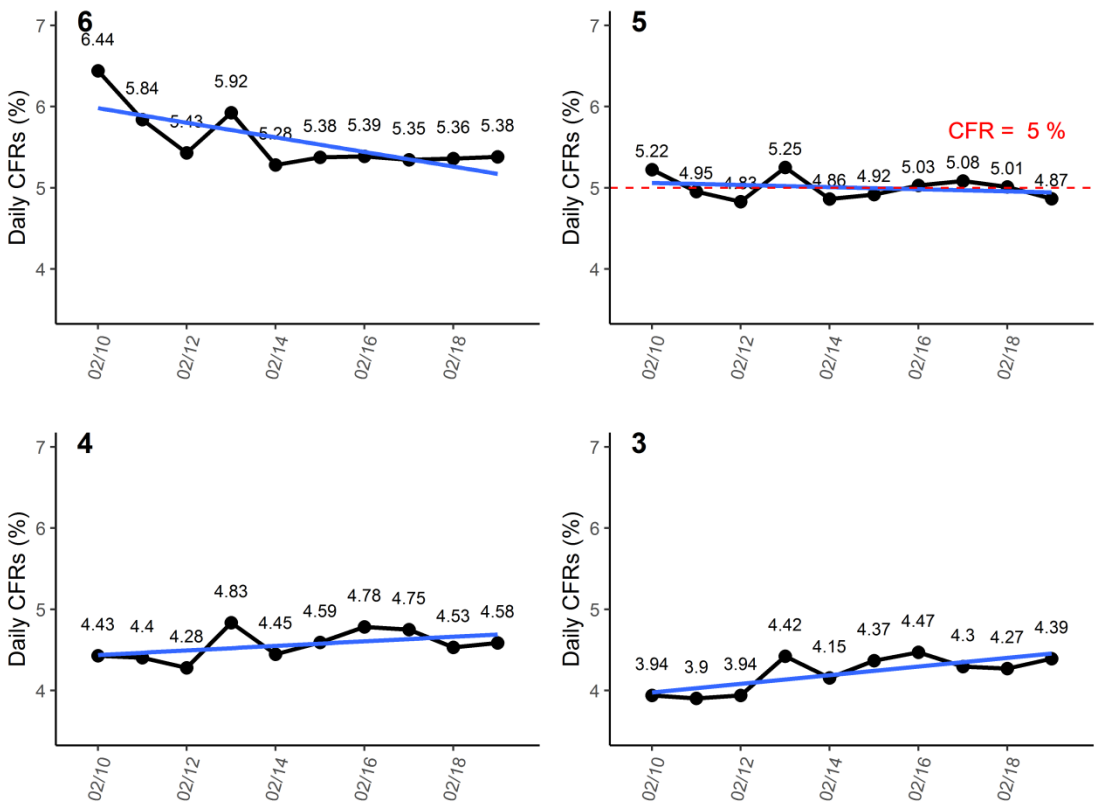

Figure 4. Converging stage daily CFRs of Hubei Province when T was from 3 to 6

Blue lines represented linear models generated for analysis of variances and linear trends of these data points in plots. Red dotted line and number were the true CFR.

Validation of calculation

True numbers of death were compared with numbers estimated by the calculated T and CFR to validate the accuracy of our method. The cumulative cases at day X multiplied calculated CFR should be approximately equal to true death number at day X + T theoretically. As shown in Figure 5-non-Hubei, since Feb 4, calculated death numbers had the best fit to the true death data. The two curves came closest to coinciding in shape. For Hubei (Figure 5-Hubei), the predictive curve was similar in shape with true death line, however, from Jan 23 to Feb 10, predicted death numbers were smaller than the true numbers. It was important to note that CFR calculation in our study was based on T value exploration, CFR of

a disease could be assumed fixed in a short period, but timeliness of case confirmation could be likely changing, especially in an uncontrollable outbreak like COVID-19 in Wuhan City of Hubei. A subset data from Jan 21 to Feb 12 was selected to recalculate the T and CFR, and results were the same in non-Hubei regions and could accurately predict the upcoming death numbers of Feb 12 to Feb 19. But in Hubei, T was 2 days and the CFR was 3.6%. It could been found in Figure 3 that before Feb 12, when assumed T was smaller than 2, daily CFRs had decrease trends, and was slightly increasing when T was 1. Without later data, it could result in a misleading false CFR. It was not surprising as it seemed that time of case confirmation to death was 2 days. Previously in Wuhan City of Hubei Province, many patients had not been confirmed and reported timely due to overwhelmed medical services and lack of testing kits. The death number (from confirmed and unconfirmed population) could prefer to "select" forward case pools with bigger population. Thus, to obtain an accurate CFR, averagely conformation time to death should be as fixed as possible. On the other hand, outbreak-controllers could indirectly have information about timeliness of case confirmation by monitoring daily CFRs. Stable CFRs trends meant the denominator for CFR calculation, case number, was accurate enough statistic. As extant cases were in quarantine, combining with transmission potential of diseases, it could provide policy-makers information about the risk of second infection, which could help them with evaluation of when people in regions could go back on production. In summary, as daily CFRs calculated by T = 9 had kept stable around 0.8% as long as two weeks, it could be considered as the true CFR of COVID-19 in China except Hubei Province. For Hubei, the possibility could not be rule out that the short converging stage of daily CFRs before Feb 19 were transient due to difficult situation.

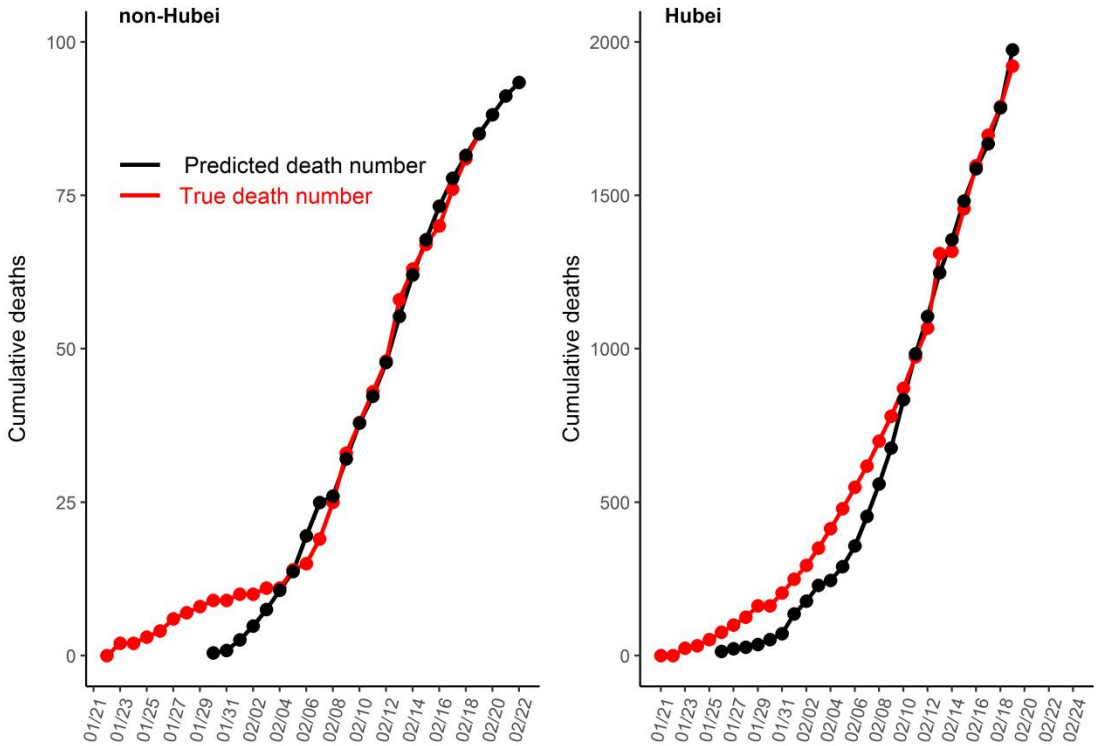

Figure 5 Comparison of true numbers of death and estimated numbers by the calculated T and CFR.

Discussion

CFR was calculated by dividing the number of deaths from a specified disease. For a infectious disease, the outcome of death were determined by virulence of causative pathogens, immunity and health status of those infected, medical conditions, received treatment and so on. Whether all infected cases had been completely included into the denominator also affected the CFR. That meant, for the same disease, CFR were not always constant and could vary between populations [6]. COVID-19 firstly occurred in Wuhan City, Hubei Province, China and quickly went into a big outbreak and overwhelmed local medical facilities. Then it extended to the whole Hubei Province and other regions in China during the heavy-travel Chinese Spring Festival holidays. The Chinese government rapidly isolated Wuhan and took emergency measure nationwide to prevent and control disease. So other non-Hubei regions

response to COVID-19 could be regarded as timely. The situations of outbreak in Hubei and non-Hubei regions were quite different. So CFR were calculated separately. Diagnose and confirmation towards patients presenting with more severe disease had priority in Hubei, especially Wuhan as the limited healthcare-facilities and testing capacities. Thus, the calculated CFR for Hubei was higher due to the under detection of mild or asymptomatic cases. Conversely, other regions in China had token completely epidemical investigation of diagnosed cases under the nationwide strict quarantine and screening policy. Close contactor investigation by CDC could help find mild or asymptomatic cases. Thus, CFR calculated from these regions could be regarded as accurate values in the situation of medical services were not overwhelmed.

So far, only a few studies reported CFR of COVID-19. Study of Wuhan's earliest 41 cases gave a 15% death rate [7]. However, regardless of the sample size, these cases were highly biased towards the more severe cases for CFR calculation. Another study reported the CFR was 4.3% which also had a biased study population (Wuhan hospitalized patients) [8]. A newly epidemiological study estimated the CFR was 3.06% (95% CI 2.02-4.59%) from 4,021 cases [9]. This study included data from non-Hubei regions, so the CFR should be smaller than that of Wuhan. When epidemic was still ongoing, CFR could be estimated by following a cohort, however, it was time-consuming and difficult to included size-enough and representative patients from unbiased population. Considering the features of daily CFRs convergence, true CFR estimation based on population-level big data might be a good way.

In our study, calculated T values were different, T = 9 in non-Hubei but = 5 in Hubei. The time in Hubei from confirmation to death was shorter comparing with non-Hubei. On Feb

13, more than 10 thousands cases were reported one day including clinical diagnosed cases without laboratory confirmation. It indicated there would be a lag in case confirmation in Hubei. Thus, cases 5 days ago were supposed should have been confirmed in earlier days theoretically. When calculating the true CFR of Hubei by subset data before Feb 12, the CFR was 3.6% and T was 2 days which were different with results from the entire data set. In Figure 3, when T was 0 to 4, daily CFRs kept continuously non-decreasing (keep or increase) which indicated they must be smaller than true T even when time went on. Because if an assumed T was bigger than true T value, when converging, it could be very hard to continuously keep non-decreasing unless increased denominator (case number) was bigger and bigger each day to counteract the force of the true CFR pulling daily CFRs into its direction. Therefore, the CFR in Hubei in the current situation was no less than 5%. If the T = 5 could not present the true time from case confirmation to death in Hubei, it should move towards bigger and the daily CFR would go up to approach the true CFR and at least stay some time until the factors making the true CFR higher were controlled (true CFR decreased).

CFR calculated in our study was dynamic, which could be used to real-time monitor the case confirmation tendency. If daily CFRs kept on a horizontal line and the confirmed cases were continuously decreasing, it meant the control measures had worked well. Not only infectious diseases, but also other diseases which were difficult to follow cohorts can be monitored to calculate the CFR. But a limitation should be taken into consideration that daily CFR would approach to true CFR only when deaths started to appear. When calculated T was too small, it might give outbreak-controllers information that if there remained a lot of infected persons unconfirmed. And if calculated T started to move backwards with time, it

meant confirmation of patients had become timely at then.

In conclusion, by convergence laws of daily CFRs, the true CFR of COVID-19 in China except Hubei Province was 0.8%. This calculated CFR could accurately predict the death numbers for more than two weeks continuously. The CFR in Hubei was not lower than 5% in the current situation. The method in this study can be used for CFR calculation when a pandemic is still ongoing and monitoring the case confirmation tendency.

Declaration of interests

We declare no competing interests.

Fundings

This study was partially supported by a grant from National Natural Science Funds of China (No. 81971939).

Reference

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382**(8): 727-33.

2. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 2020; **395**(10223): 507-13.

3. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020; **395**(10223): 497-506.

4. Lipsitch M, Donnelly CA, Fraser C, et al. Potential biases in estimating absolute and relative case-fatality risks during outbreaks. *PLoS Negl Trop Dis* 2015; **9**(7).

5. Kucharski AJ, Edmunds WJ. Case fatality rate for Ebola virus disease in west Africa. *The*

Lancet 2014; **384**(9950): 1260.

6. Garske T, Cori A, Ariyarajah A, et al. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016. *Proc. Biol. Sci* 2017; **372**(1721): 20160308.

7. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *The Lancet* 2020; **395**(10223): 470-3.

8. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama* February 07, 2020. doi:10.1001/jama.2020.1585 2020.

9. Yang Y, Lu Q, Liu M, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *medRxiv* 2020: 2020.02.10.20021675.