A method for estimating the case fatality rate during the ongoing epidemic: A Case Study of COVID-19

Wanling Hu, Xiaoyun Liu, Tao Wang, Changlong Zhou, Xiaofei Zeng, Chen Dong, Yuanming Zhang, Xuan She, Zhongli Hu, Dingfu Liu, Ying Diao

1 Hubei Asset Management Co. Ltd., Wuhan, 430062, P.R. China
2 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
3 Wuhan Windoor Information Technology Co. Ltd., Wuhan, 430040, P.R. China
4 Yongchuan Hospital, Chongqing Medical University, Chongqing, 402160, P.R. China
5 School of Medicine, Southern University of Science and Technology, Shenzhen, 518055, PR China
6 College of Biological Engineering, Henan University of Technology, Zhengzhou 450001, P.R. China
7 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
8 College of Life Science, Wuhan University, Wuhan 430072, P.R. China
9 Hubei Academy of Agricultural Sciences, Wuhan, 430064, P.R. China
10 Hubei Asset Management Co. Ltd., Wuhan, 430062, P.R. China
11 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
12 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
13 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
14 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
15 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
16 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
17 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
18 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
19 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
20 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
21 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences, Chongqing, 402160, P.R. China
Abstract

Human beings are constantly struggling with various epidemics. Although we gained a lot of experience and success, in the face of the new epidemic, we still inevitably face pressure from public health, politics, and the economy. Case fatality ratio (CFR) received widespread attention as one of the indicators describing the severity of the epidemic and evaluating treatment options. However, due to the ongoing epidemic situation and the constant changes in the death and diagnosis data, no scientific method for this situation to calculate the CFR exists. This study proposes a method for estimating CFR in the continuation of the epidemic. CFR is estimated by "ratio of the cumulative number of deaths before j days from a given day to the sum of the number of patients discharged from a given day and the cumulative number of deaths before j days from a given day ". Take the ongoing outbreak of COVID-19 in December 2019 as an example. The results show that, regardless of the size of the estimated value or its changing trend, the estimated CFR given by the new method shows better stability.
and better reflects the true situation of the case fatality rate; additionally, the improvement of medical conditions
can also be clearly reflected in the change in valuation. When $j = 10$, according to the data of March 10, the CFR
of COVID-19 in Wuhan, China and China (excluding Hubei) is 6.23%, 4.46%, and 0.87%, respectively. This
method of estimating CFR can be used in time to evaluate the therapeutic effect of different medical schemes
and different regions, which is of great value and significance for the decision-making in the epidemic
prevention and control.

**Key words:** epidemic, case fatality ratio, COVID-19, estimation

**Introduction**

Looking back on the history of civilization for thousands of years is also a review of the epic history of the
"war against plague ". The history of human civilization is accompanied by the pace of the plague and spiraled
upward in history. With the development of civilization and the continuous optimization of transportation,
infectious diseases also developed at the same or even faster speed. Since 2000, three coronavirus epidemic
attacks worldwide occurred. SAR S in 2003 caused 774 deaths worldwide, with a CFR of 9.6%. MERS in 2012
caused 858 deaths worldwide, with a CFR of 34.4%. In December 2019, Wuhan, China suffered a new crown
pneumonia (COVID-19) epidemic caused by new coronavirus SARS CoV-2. The number of infected people, the
scope of transmission, and the number of deaths are still increasing. On March 8, 2020, more than 80,000 people
were infected in China, and a total of 3,123 people died. There have been confirmed cases in 105 countries
overseas, with more than 30,000 people diagnosed and 760 people dead. However, due to the difficulty in
identifying and accounting for mild and asymptomatic infections, the number of COVID-19 cases may actually
be higher. There is no doubt that the outbreak of the COVID-19 is a severe test of public health safety in China and the world.

The case fatality ratio (CFR) indicates the severity of an epidemic disease, and it also reflects the ability of medical diagnosis and treatment to a certain extent. Generally, the popular method is used to calculate the CFR, which is the ratio of the number of accumulated deaths to the number of diagnoses over a period of time. The traditional method of calculating the CFR is certain and accurate for the epidemic situation that ended. However, for the novel and ongoing epidemic, a large number of patients are still being treated in the hospital, and the final treatment effect of these patients cannot be defined at the time of calculation. The data of a large number of patients treated in hospital are mixed in it, which interferes with the accuracy of the calculation results of CFR. This defect makes the obtained results deviate greatly from the real situation, may mislead the public, and even cause decision makers to misjudge the future situation. In the Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), as of February 20, the CFR of the COVID-19 in China was 2.9%1, and according to the data on March 8, the CRF reached about 3.9%. Although more than 60,000 people have been cured and discharged, nearly 16,000 patients still remain in hospitals.

Therefore, we propose a method to estimate the CFR during the course of the epidemic and estimate the CFR of COVID-19 in order to provide important reference for the diagnosis, treatment, and containment of epidemic disease.

82 Results

83 Determine the j value of COVID-19

As a result of the February 12 diagnostic criteria change, resulting in a surge in cases, only the data before
February 11 were selected. Taking the national data as an example, starting from February 11 and backtracking to January 20, in Equation (4), \( j = 1, 2, 3, \ldots, 15 \) was calculated, and the CFR was calculated daily. When \( j = 10 \), the value of CFR was the most stable among the days, and the coefficient of variation was the smallest.

According to the published data, in addition to calculating the national CFR, the same calculations were performed according to the division data of Wuhan, Hubei, Hubei except Wuhan, and China except Hubei (Fig.1). The coefficient of variation of the estimated CFR of Wuhan, Hubei, China, Hubei excluding Wuhan, and China excluding Hubei reached the minimum value at \( j = 8, 10, 10, 10, 12 \), respectively, indicating that this is the smallest difference in the estimated value of each day.

**CFR**

Dividing data into five categories as Wuhan, Hubei, Hubei excluding Wuhan, China and China excluding Hubei, \( j = 8, 10, 12 \) were selected to calculate the day-to-day CFR tracking back from March 5 (Fig. 2 Table S2–4). The results showed that the difference is large in the early stage and small in the near future in either different \( j \) values or different data categories (Fig.2a-c). According to Table 1 and Fig.2, when \( j = 10 \), if no specific drugs and better treatments emerge later and no significant variation in SARS-CoV-2 pathogenicity, the CFR of COVID-19 is approximately 4.46%, which is lower to the CFR of SARS (6.6%) in Mainland China. In Wuhan, the CFR of COVID-19 is approximately 6.23%, which is close to the CFR of SARS in Mainland China.

Because of the large number of cases and low cure rate in Wuhan, the CFR in the Hubei province is higher than the national average. The CFR of China (excluding Hubei) is the lowest, with an estimate of 0.87%. This may be due to factors such as the large number of imported cases in different provinces outside Hubei, the different age structure of imported cases, relatively few patients with better treatment conditions, and earlier hospital admissions.
Comparison of CFRs obtaining from three methods

Estimates of CFR $q_1$, $q_2$, $q_j$ vary over the duration of the epidemic (Fig.3). The difference among the three is large at the beginning of the epidemic, but over time, the difference between the three becomes smaller and smaller, and when the epidemic is over and treatment is over, the three are bound to present the same value.

It can be seen from Figure 3 and Table 2 that the estimated value of $q_j$ is between the estimated values of $q_1$ and $q_2$. Over time, the estimated values of $q_1$ and $q_2$ move closer to the middle, regardless of the size of the estimated value or its changing trend. The estimated value of $q_j$ shows better stability, indicating that the $q_j$ value can better reflect the true situation of mortality. By comparing the current $q_1$, $q_2$, and $q_j$ estimates, the $q_1$ and $q_2$ estimates exactly give an interval estimate of CFR (Table 2).

Detection of CFR stability and effects of medical events

It is known from figure 2 and figure 3 that the CFR estimates vary with time. In the course of treatment, two important time nodes occur: on February 12, 13,797 new cases were confirmed in Hubei Province, and on February 23, the number of patients died in Hubei Province suddenly dropped sharply from 149 on February 23 to 68 on the 24th and then continued to decline. If the stability of the CFR estimates and whether the above two important time nodes are reflected in the fatality rate estimate, it can be tested by formula (5). At this time, two dummy variables, $D_1$ and $D_2$ are introduced.

The regression analysis results are showed in Table 3. First, the estimated value of $q_j$ is basically stable. In the analysis of the five data categories, the correlation between the value of $q_j$ and time $t$ is not significant. This result is consistent with its definition. The estimated value of $q_1$ and $q_2$ is highly correlated with time $t$, in which the $q_1$ estimate is rapidly decreasing, while the $q_2$ estimate is significantly increasing. Second, on February 12, as a result of changes in diagnostic criteria, a large number of suspected cases were converted to diagnosis cases,
but it exhibits a relatively small impact on the CFR estimates.

The significant reduction in deaths after February 23 is due to the large number of square cabin hospitals in use, the improvement of related medical conditions, and the large-scale entry of traditional Chinese medicine into the anti-epidemic battlefield, which effectively avoided mild to severe illness, severe to critical illness, and reduced CFR in severe cases; therefore, these improvements of medical treatment have a relatively large impact on CFR estimates.

Methods

Data sources

Data from January 20 to March 10 released by Wuhan Municipal Health Commission, Hubei Provincial Health Commission, and National Health Commission, including cumulative diagnoses, cumulative deaths, and cumulative number of patients discharged (Table S1).

The definition, model and correction of case fatality rate

"Cure" means the patient was recovered and discharged from the hospital at the end of treatment, and "death from illness" is the end of treatment due to death.

If the epidemic situation ended and all patients also completed treatment, if the cumulative number of discharged patients is $\Sigma x_i$ and the cumulative number of deaths is $\Sigma y_i$, then the cumulative number of confirmed diagnoses is $(\Sigma x_i + \Sigma y_i)$.

$$CFR(q) = \frac{\langle \Sigma y_i \rangle}{\Sigma x_i + \Sigma y_i} \quad (1)$$

Here, $i=1,2,..., n$, when the epidemic is over.

On this basis, several different CFR can be estimated using different data:
Bring the cumulative number of patients cured and discharged on the day (n) ($\sum x_i$) and the cumulative number of death ($\sum y_i$) on the day into formula (1) to calculate the CFR, that is:

$$q_1 = \frac{\sum y_i}{\sum x_i + \sum y_i} \quad (2)$$

Here, $i=1, 2, ..., n$.

In fact, this estimate has not been used in reality. The possible reason is that the method highly overestimates CFR.

The widely accepted estimate of CFR in real life is calculated using the "ratio of the cumulative number of dead patients on the day ($\sum y_i$) to the cumulative number of patients diagnosed on that day ($\sum T_i$)". which is called the crude fatality rate, and is recorded as:

$$q_2 = \frac{\sum y_i}{\sum T_i} \quad (3)$$

Here, $i = 1, 2, ..., n$.

As the epidemic is ongoing, diagnosed patients present as the following three types: as cured and discharged, died of illness, or hospitalized. It is unclear whether the diagnosed patients treated in the hospital will eventually be cured or die. Therefore, the results are inaccurate by using the above formula to calculate the CFR directly during the outbreak, especially since a large number of patients were treated in the hospital. The results are inappropriate to infer the trend of the epidemic. Because the disease itself demonstrates a certain time course, a period of time from diagnosis to treatment in the hospital to discharge or death must occur. In other words, the daily published patients who are dead in the hospital or discharged from the hospital are not the same patients admitted to the hospital at the same day. For the diagnosed patients in the same day, a time difference (j days) between the cured patients and the deaths occurs. Therefore, we propose that the CFR estimate should be
Here, \( i = 1, 2, \ldots, n \).

**Selection of j value**

According to the results of clinical research, COVID-19 is a self-limiting disease with milder early symptoms. After one week, the condition may worsen, and the patient can gradually recover as long as the patient passes the most dangerous period of time. In addition, according to the treatment plan, the cured patients must wait for all symptoms to disappear and be negative for two SARS-CoV-2 nucleic acid tests before they can reach the discharge standard for to be considered cured.

According to statistical theory, if the COVID-19 epidemic event satisfies the statistical random event hypothesis, and the virus is not mutated without considering the discovery of future specific medicines and special medical treatments, the estimated daily CFR should be theoretically stable.

According to this idea, formula (4) can be used to perform simulation calculations on actual data. Two screening criteria were used to select the value of \( j \): (1) the minimum variance or coefficient of variation of the day-to-day estimated CFR; and (2) the average value of the daily CFR is not lower than the average value of the daily value of the crude CFR at that time. The best \( j \) is determined by referring to clinical observation statistics value.

**Vent study**

The definitions of \( q_1 \) and \( q_2 \) are not required to meet the premise that "the CFR is theoretically stable."

Although the definition of \( q_j \) meets the premise that "assuming that the medical conditions and medical care technology level are unchanged, and the virus has not mutated, then the daily CFR is theoretically stable", but in fact, the value of the three may change over time due to the occurrence of the epidemic and the medical events
in the development process. Therefore, in order to detect the stability of CFR estimates and the effects of medical events, this study used a statistical model:

\[ q = \beta_0 + \sum \beta_k D_k + \beta_{k+1} t + \varepsilon \]  

Here, \( t \) is time, and the \( D \) is a dummy variable.

Discussion

About j value

The CFR estimate is negatively correlated with the \( j \) value size. The larger the \( j \) value, the smaller the CFR estimates; the smaller the \( j \) value, the larger CFR estimates. In relation to the estimated value of CFR, CFR will vary with the value of \( j \). Therefore, to obtain accurate estimates of CR and CFR, we need to choose a \( j \) value, which is also worthy of further study. However, using this strategy and method is certainly more reasonable than directly “calculating the CFR for the COVID-19 in the country with the ratio of cumulative death to cumulative confirmed cases.”

COVID-19 happens suddenly, with little information known about this disease. Assuming that the COVID-19 epidemic event satisfies the statistical randomness, the virus demonstrates no variation, and the daily CFR should be basically stable, with approximately the same medical means and measures in a certain period of time and the estimated value of \( j = 10 \) able to meet this condition.

According to clinical experience, the median time from onset to clinical recovery is approximately 2 weeks for mild and approximately 3–6 weeks for severe or critical illness. The time from onset to progression to severe illness such as hypoxia is 1 week, and among patients who die, the time from onset to death ranges from 2–8 weeks\(^1\). In this article, \( j = 10 \) was consistent with the development of the disease course.
Some researchers also noticed that when estimating the CRF, the selected data ignored a large number of hospitalized patients. Recently, the Imperial College London predicted the CFR of Hubei Province and outside China. However, their model is based on the time difference between the onset of the disease and death or cure, but in fact the symptoms of the disease are very similar to that of the common flu, and the time difference is difficult to determine accurately.

About CFR

Because the epidemic outbreak in Wuhan, the number of patients has grown too fast, inadequate treatment conditions, delayed diagnosis and delayed treatment, and insufficient early understanding of the virus affected the treatment effect, leading to a higher clinical fatality. Using the original case-fatality estimation method, its estimated value is over 10%. For examples, in Wuhan, the CFR of early 41 cases and 99 cases of COVID-19 was 15% and 11%, respectively. These results put tremendous pressure on both the public and government decision makers. This study corrects the current problems in calculating the CFR of COVID-19. The estimated CFR of COVID-19 in China is about 4.5%, which is lower than SARS. The estimated value of Wuhan is about 6.2%, which is equivalent to SARS. On the basis of the estimated CFR in Wuhan on March 10, the number of deaths in Wuhan caused by COVID-19 in the current medical conditions is around 3,000. Although the epidemic situation continues and the number of confirmed patients continues to increase by a small amount every day, due to the tremendous progress in knowledge, methods and tools of scientific research related to COVID-19, the number of deaths caused should further decrease. Of course, all undiagnosed deaths due to COVID-19 are not included.

The strong transmission of SARS-CoV-2, the uncertainty that causes death in high-risk groups, and the destructive power of socio-economic order are extraordinary among human coronaviruses. Because China
resolvedly and timely adopted the main strategies of controlling the source of infection, blocking transmission, and preventing spread, the virus failed to erupt throughout the country. The case fatality rate in China (excluding Hubei) is less than 1%. At present, the COVID-19 epidemic in China entered a relatively gradual phase, while other countries, such as Italy, Iran, and South Korea, exhibit a trend of outbreak. Due to different genetic backgrounds of different ethnic groups, inconsistent levels of containment, diagnosis and treatment in each country, and possible mutations in the virus, the severity of the COVID-19 epidemic abroad cannot be estimated. However, using the method in this article, selecting the appropriate value of $j$, and based on the cumulative cure and cumulative death toll in the epidemic country, the estimated CFR can still be used to judge the severity of the outbreak. Scientifically predicting the case fatality rate can provide an important reference for local decision makers in the formulation of prevention and control policies and measures, and this can also help the macro evaluation of the effectiveness of diagnosis and treatment. This novel coronavirus pneumonia will exhibit a profound impact on political, economic, cultural, and other aspects at home and abroad. We believe that the use of modern medicine and government governance methods and legal systems can minimize the panic and the mortality of the disease and finally win the victory against the epidemic.

Contributors

WH, XL and TW had roles in data analysis, and data interpretation. XZ, CD, YZ, CZ, XS and ZH had roles in data collection and edit manuscript. YD and DL had roles in the study design, and write manuscript.

Conflicts of Interest

The authors declare no conflict of interest.
**References**

1. Mission, J. (WHO, 2020).
2. Gu, X., Cao, B. & Wang, J. Full spectrum of COVID-19 severity still being depicted—Authors’ reply. *The Lancet* (2020).
3. Dorigatti, I. *et al.* (Imperial College London, London, 2020).
4. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395, 497-506 (2020).
5. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 395, 507-513 (2020).

**Figure legends**

Fig. 1 Coefficient of variation of estimated case fatality rate (CFR) ($q_j$).

Fig. 2 Estimated CFR at $j=8,10,12$.

Fig. 3 Trends of three types of CFR.

**Table 1** Estimated CFR at different $j$ value

| $j$ | Wuhan | Hubei | China | Hubei excluding Wuhan | China excluding Hubei |
|-----|--------|-------|-------|------------------------|-----------------------|
|     |        |       |       |                        |                       |
| 8   | 6.38%  | 5.46% | 4.57% | 3.51%                  | 0.87%                 |
| 10  | 6.23%  | 5.33% | 4.46% | 3.41%                  | 0.87%                 |
| 12  | 6.06%  | 5.18% | 4.34% | 3.32%                  | 0.85%                 |

Data as of March 10, 2020

**Table 2** Estimated CFR of three methods

| Method | Wuhan | Hubei | China | Hubei excluding Wuhan | China excluding Hubei |
|--------|-------|-------|-------|------------------------|-----------------------|
| $q_1$  | 6.83% | 5.85% | 4.89% | 3.74%                  | 0.89%                 |
| $q_j$  | 6.23% | 5.33% | 4.46% | 3.41%                  | 0.87%                 |
| $q_2$  | 4.85% | 4.49% | 3.91% | 3.50%                  | 0.86%                 |

Data as of March 10, 2020
Table 3: Regression analysis of estimated CFR (j=10)

|        | $\beta_0$       | $\beta_1 D_1$     | $\beta_2 D_2$     | $\beta_3 t$     | $R^2$  | $F$          | $n$ |
|--------|-----------------|-------------------|-------------------|-----------------|--------|--------------|-----|
| Wuhan  |                 |                   |                   |                 |        |              |     |
| $q_1$  | 0.569481 ***    | -0.06278 **       | 0.018152 ns       | -0.01237 ***    | 0.971429 | 430.6764 *** | 42  |
| $q_j$  | 0.105325 ***    | 0.008931 ns       | -0.04112 **       | 4.69E-05 ns     | 0.507853 | 12.72692 *** | 41  |
| $q_2$  | 0.048146 ***    | -0.01139 **       | 0.01343           | -0.00013 ns     | 0.552524 | 15.64023 *** | 42  |
|        |                 |                   |                   |                 |        |              |     |
| Hubei  |                 |                   |                   |                 |        |              |     |
| $q_1$  | 0.571495 ***    | -0.09764 **       | 0.08331           | -0.01389 ***    | 0.934823 | 181.6764 *** | 42  |
| $q_j$  | 0.088106 ***    | -0.00633 ns       | -0.02348          | 5.5E-05 ns      | 0.509997 | 12.83657 *** | 41  |
| $q_2$  | 0.028796 ***    | -0.00352 ns       | 0.007245 ***      | 0.000285 **     | 0.817685 | 56.81016 *** | 42  |
|        |                 |                   |                   |                 |        |              |     |
| China  |                 |                   |                   |                 |        |              |     |
| $q_1$  | 0.458278 ***    | -0.05948 ns       | 0.113951 **       | -0.01306 ***    | 0.86653 | 82.23591 *** | 42  |
| $q_j$  | 0.054454 ***    | -0.00225 ns       | -0.00866          | 0.00013         | 0.262221 | 4.383498 *** | 41  |
| $q_2$  | 0.0185          | -0.00038 ns       | 0.004074 ***      | 0.000425 ***    | 0.946167 | 222.6296 *** | 42  |
|        |                 |                   |                   |                 |        |              |     |
| Hubei excluding Wuhan | |                   |                   |                 |        |              |     |
| $q_1$  | 0.633681 ***    | -0.12128 ns       | 0.224608 **       | -0.02009 ***    | 0.806271 | 52.71684 *** | 42  |
| $q_j$  | 0.054779 ***    | -0.00682 ns       | -0.00426 ns       | -0.00024 ns     | 0.40583  | 7.740883 *** | 38  |
|      | q_2  | 0.011317 *** | 0.003341 *** | 0.002912 ** | 0.000438 *** | 0.97218 | 442.6448 | 42 |
|------|------|-------------|-------------|------------|-------------|---------|----------|---|
| China excluding Hubei | q_1  | 0.087139 *** | 0.006119 ns  | 0.045973 *  | -0.00375 *** | 0.524476 | 13.97059 | 42 |
|      | q_j  | 0.006268 *** | -0.00263 **  | 0.001765 *  | 8.44E-05 ns  | 0.55965 | 14.40379 | 38 |
|      | q_2  | 0.001547 *** | 0.002403 *** | 0.000284 ns  | 0.00012***   | 0.964824 | 347.4229 | 42 |

Data as of March 10, 2020.