

1 Estimation of COVID-19 case fatality ratio based on a bi-directional 2 correction method

3 Wanling Hu^{2#}, Xiaoyun Liu^{3#}, Tao Wang^{7#}, Changlong Zhou⁵, Dingfu Liu⁴,
4 Yuanming Zhang⁶, Zhongli Hu⁷, Ying Diao^{1*}

5
6 1 College of Landscape Architecture and Life Sciences, Chongqing University of Arts and Sciences,
7 Chongqing, 402160, P.R. China

8 2 Hubei Asset Management Co. Ltd., Wuhan, 430062, P.R. China

9 3 Wuhan Windoor Information Technology Co. Ltd., Wuhan, 430040, P.R. China

10 4 Hubei Academy of Agricultural Sciences, Wuhan, 430064, P.R. China

11 5 Yongchuan Hospital, Chongqing Medical University, Chongqing, 402160, P.R.China

12 6 College of Plant Science and Technology, Huazhong Agricultural University,

13 Wuhan 430070, China

14 7 College of Life Science, Wuhan University, Wuhan 430072, P.R. China

15

16 #Contribute equally.

17 *Correspondence: ydiao@whu.edu.cn; [Tel:+86-23-4968-2201](tel:+86-23-4968-2201)

18

19 **Abstract:** The case fatality rate (CFR) can be used to predict the number of potential
20 deaths in the epidemic and thus can reflect the appropriateness and quality of medical
21 measures developed by public health. When a new disease breaks out, it is
22 particularly important to accurately estimate the CFR. However, while the epidemic is
23 still developing, the crude CFR is often lower than the true value and the hospital
24 CFR is often higher than the true value due to differences in occurrence time, patient
25 number, and treatment plans. Therefore, this study proposes a bi-directional correction
26 method to estimate the CFR. COVID-19 data from China were used to evaluate this
27 method. The results show that this method provides more accurate results than both
28 the crude CFR and hospital CFR. Additionally, this method was used to estimate the
29 CFR of COVID-19 in other countries, with an aim to provide a reference for
30 prevention and control decisions for the COVID-19 epidemic and for the evaluation
31 of medical efforts.

32 **Keywords:** COVID-19; case fatality ratio; bi-directional correction; estimate; crude
33 CFR

34

35 1. Introduction

36 At the end of December 2019, COVID-19, an epidemic disease caused by a novel
37 coronavirus called SARS-CoV-2, broke out in Wuhan, China [1]. European and
38 American countries generally failed to pay enough attention at the beginning of the
39 outbreak. However, as the outbreak spread worldwide, the World Health Organization
40 (WHO) characterized it as a pandemic on March 11, 2020. Similar to the SARS and
41 MERS outbreaks, this serious epidemic disease was caused by a coronavirus. As of
42 April 22, 2020, the number of people diagnosed with SARS-CoV-2 worldwide has

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

43 exceeded 2.5 million, and the cumulative deaths have exceeded 170,000. Conversely,
44 8,098 were confirmed to have been infected by SARS worldwide, with a cumulative
45 death toll of 774 and a case fatality rate (CFR) of 9.6%. As of January 2020, MERS
46 had reported 2,519 confirmed infections worldwide and 866 deaths with a CFR of
47 34.3%. Obviously, the number of confirmed infections and deaths from COVID-19 is
48 far higher than that of SARS and MERS. The published crude CFRs of COVID-19
49 are reported as 6.1% globally, with the highest CFR (8.3%) in Europe and the lowest
50 (3.4–3.5%) in the Pacific and America. The CFR varies greatly in different countries,
51 with the highest (14.3%) in France, CFRs above 12% in Italy and the United
52 Kingdom, and 3.4% in the United States.

53 As the epidemic develops, there will be large differences in the calculated crude CFRs
54 because of the inconsistent time that the epidemic has existed in various regions. For a
55 certain region, the calculated crude CFR will display large deviations over time. That
56 is, the CFR will be low at the beginning, but will gradually increase during the
57 epidemic process, which may lead to incorrect judgments and decisions and may also
58 affect the public's confidence and trust in leaders [2]. As the severity of the
59 COVID-19 epidemic has increased, and in the light of the strict government control
60 measures, frequent medical resource problems in the news, and the increasing number
61 of confirmed cases and deaths, people began to realize that COVID-19 is no longer an
62 ordinary flu. Will COVID-19 exceed the CFR of SARS? Or even reach the CFR of
63 MERS? At present, COVID-19 has been essentially curbed in China. Currently, more
64 than 84,000 people have been diagnosed, nearly 93% have been cured and discharged,
65 and more than 4,600 have died. This provides the basis and conditions for more
66 objectively assessing the mortality of patients with COVID-19. Therefore, to improve
67 the deviations in the estimates of the crude fatality rate caused by the continuously
68 changing data during the epidemic, this study proposed a bi-directional correction
69 method to estimate the CFR. This method was evaluated using the COVID-19 data
70 from China. It further estimates the CFR of COVID-19 in other countries. The
71 research results not only provide a reference for the evaluation of medical treatment
72 and decision-making for epidemic prevention and control, but they also have
73 important significance for stabilizing society and people's feelings regarding the
74 epidemic.

75 **2. Methods and data sources**

76 2.1 Definition, model, and correction of CFR

77 The CFR is defined as the number of people or animals who die from a disease during
78 a certain period of time out of the total number of patients or animals infected. For
79 human diseases, "death" is the end of medical treatment due to illness, and "cure" is
80 defined as discharge from the hospital after the end of treatment. Consider the
81 following scenario: if the epidemic situation of a certain disease has ended and all
82 patients have finished treatment, suppose the cumulative number of people cured is
83 $\sum x_i$ and the cumulative number of deaths $\sum y_i$. Then, the cumulative number of
84 confirmed cases is $(\sum T_i) = (\sum x_i + \sum y_i)$, as follows:

$$85 \quad q = (\sum y_i) / (\sum T_i) \quad (1)$$

86 or,

87 $q = (\sum y_i) / (\sum y_i + \sum x_i)$ (2)

88 Here, $i = 1, 2, \dots, n$, but the epidemic has ended. After the epidemic is over, the crude
89 method can be a good estimate of the CFR. Additionally, formulas (1) and (2) give
90 exactly the same value.

91 However, while the epidemic is still occurring, calculating the CFR involves many
92 complicated factors. The WHO used two methods to estimate the SARS CFR using
93 accumulative case data [3]. One is the widely adopted estimator of the CFR,
94 commonly known as the crude CFR, which underestimates the true CFR as the
95 epidemic continues. The crude CFR is calculated by formula (1) using the cumulative
96 number of deaths $(\sum y_i)$ on day (n) and the cumulative number of diagnosed cases
97 $(\sum T_i)$ on the same day (n) , as follows:

98 $q_c = \sum y_i / \sum T_i$ (3)

99 Here, $i = 1, 2, \dots, n$.

100 The second method involves using only the final number of deaths and the number of
101 discharges at the time of the epidemic, that is, it substitutes the cumulative number of
102 discharges $(\sum x_i)$ and cumulative deaths $(\sum y_i)$ of day (n) into formula (2). This results
103 in an overestimation of the CFR during the ongoing epidemic. For convenience, this
104 method is called the hospital CFR, and it is calculated as:

105 $q_a = \sum y_i / (\sum y_i + \sum x_i)$ (4)

106 Here, $i = 1, 2, \dots, n$.

107 In general, the natural procedures for diseases such as SARS and COVID-19 follow
108 the general pattern of viral infection-onset-treatment/hospitalization-diagnosis-death
109 or rescue-discharge, but different diseases have different courses [4]. To utilize
110 easy-to-obtain statistical data, special consideration is given to the "confirmed-death
111 or rescue-discharged" time period, which involves only three time points: confirmed,
112 death or rescue, and discharged (Figure 1). For example, the period from diagnosis of
113 SARS cases to death is approximately 10 days [5], which indicates that the death
114 cases observed until a certain day are generally patients who were diagnosed 10 days
115 ago, and the cumulative number of cases includes the number of new cases in the last
116 10 days. In fact, most of these patients have not reached the time point of "death or
117 rescue" caused by the disease. In formula (3), this part of the population should not be
118 included in the denominator of the CFR calculation. The same is true for COVID-19
119 case data. Considering its generality, that is, because is a time difference $(k$ days)
120 between when the patients confirmed and when they died (Figure 1), formula (3) is
121 corrected as:

122 $q_{ck} = \sum y_i / \sum T_{i-k}$ (5)

123 Here, $i = 1, 2, \dots, n$.

124 When evaluating this from the perspective of COVID-19, according to clinical
125 research, the early symptoms are mild. After a week, the disease may worsen or the
126 patient may even die; however, as long as the patient passes the most dangerous
127 period of time, he can gradually recover. Moreover, in China, according to the
128 treatment plan, cured patients must wait for all symptoms to disappear and then test
129 negative for SARS-CoV-2 nucleic acid twice before classified as cured and being
130 discharged. In other words, the patients with COVID-19 who died and those who

131 were discharged on a particular day were not admitted to the hospital at the same time.
132 For patients who were confirmed at the same time, there was a time difference (j days)
133 between when those patients were either discharged or died. The same was true for
134 SARS case data. Without a loss of generality, to correct formula (4), we propose an
135 estimator of CFR:

$$136 \quad q_{aj} = \sum y_i / (\sum y_i + \sum x_{i+j}) \quad (6)$$

137 Here, $i = 1, 2, \dots, n$.

138 Obviously, formulas (3) and (4) are special cases of formulas (5) and (6) when $k = 0$
139 and $j = 0$, respectively.

140 2.2 Estimated CFR with the bi-directional correction

141 According to the structure of formulas (5) and (6), the estimated values of q_{ck} and q_{aj}
142 calculated using formulas (5) and (6), respectively, can be combined with the
143 cumulative death toll of a certain day and observed in the "Golden Cross", that is, the
144 interval where $q_{ck} = q_{aj}$ appears should contain the true CFR. It is clear that the three
145 cumulative data points corresponding to the three time points involved in the
146 "diagnosis-death or rescue-discharge" periods have the following logical relationship:

$$147 \quad \sum T_{i-k} = \sum y_i + \sum x_{i+j} \quad (7)$$

148 It is easy to derive $q_{ck} = q_{aj}$ from equation (7).

149 Since k and j are both variable, some conditions must be set for estimating the true
150 mortality rate. These conditions are easy to understand and are in line with statistical
151 principles and clinical laws, for example: $q_{ck} < q_{a0}$, $q_{aj} > q_{c0}$; $k < j$; etc. Therefore, we
152 propose the following principles and steps of the bi-directional correction method to
153 estimate the mortality rate:

154 1. Take $k = 0, 1, 2, \dots, 15$ and $j = 0, 1, 2, \dots, 20$, and enter them into formulas (5) and
155 (6), respectively, calculate the values of q_{ck} and q_{aj} , according to $q_{ck} < q_{a0}$ and $q_{aj} >$
156 q_{c0} , and determine the maximum values of k and j (k_{\max} and j_{\max} , respectively), which
157 constitute the interval of $q_{ck} = q_{aj}$.

158 2. According to the $q_{ck} = q_{aj}$ interval composed of k_{\max} and j_{\max} , calculate the mortality
159 using the values of the mortality (time series) corresponding to $k = 0, 1, 2, \dots, k_{\max}$ and
160 $j = 0, 1, 2, \dots, j_{\max}$. Find the Manhattan distance or Euclidean distance between the
161 sequences, and determine the two fatality sequences with the smallest distance (they
162 are derived from q_{ck} and q_{aj} , respectively).

163 3. Average the two fatality series with the smallest distance to obtain an estimated
164 value series. Two types of averages can be considered: 1) arithmetic average and 2)
165 weighted average, "weight" is taken as $k / (j + k)$ and $j / (j + k)$, respectively, in which
166 k and j correspond to the two CFR sequence with the smallest distance.

167 For convenience, we can call the method of determining k and j values as the
168 real-time CFR sequence distance minimum determination method".

169 2.3 Data source and description

170 China data sources: (1) The data for mainland China (except for Hong Kong, Macao,
171 and Taiwan) were derived from the data released by the Hubei Provincial Health
172 Commission and the National Health Commission from January 20 to April 22, 2020,
173 including cumulative and new diagnoses, deaths, and discharges. (2) However, the
174 Wuhan Municipal Government revised the data on April 17 because the revised data

175 (an increase of 1290 cases) in the cumulative death toll was not included in the
176 previously published cumulative number of confirmed cases, so it was not included in
177 the calculation in this study (TableS1).

178 Foreign data sources: (1) The cumulative number of diagnoses and the number of
179 deaths come from the data published by the WHO and span from the date of the
180 announcement of the cumulative deaths (February 3, 2020) to April 9. (2) The
181 cumulative number of cured people in eight countries comes from the Tencent News
182 Network.(3) Because Tencent News Online does not have data on the cumulative
183 number of cured people in "Global except China", data from Phoenix News Online
184 were used. The cumulative number of cured people has not been updated, but the
185 corresponding Phoenix News online data were updated for the same time period; thus,
186 the data from Phoenix News Online were used (TableS2).

187 ChinaCOVID-19 data were used to evaluate and verify the bi-directional correction
188 method for CFR proposed in this paper, and this method was also used to estimate the
189 global CFR and the CFR in eight countries with the most severe epidemics.

190

191 **3 Results**

192 3.1 Estimation and comparison of the COVID-19 CFR in China

193 3.1.1 Bi-directional correction estimates

194 According to geographical divisions, China's COVID-19 case data can be divided into
195 three groups: Wuhan, Hubei except Wuhan, and national except Hubei. The data do
196 not overlap with each other. In all regions, the characteristics of the cases are
197 essentially the same, and the medical treatment methods are essentially the same, so
198 the course of the disease is similar. The three sets of data and the national data were
199 analyzed according to the bi-directional correction method to estimate the CFR. The
200 intermediate results are shown in Table 1. Based on these results, the average of the
201 mortality rate series corresponding to k and j in Table 1 was calculated. The time
202 series of the estimated value of the CFR of the COVID-19 in China is shown in
203 Figure 2, in which the relationship between the time series of various CFR estimates
204 clearly displays. The CFR (arithmetic mean) estimated by the bi-directional correction
205 is essentially in the middle, which is slightly higher than the crude CFR. The CFR of
206 the bi-directional correction method makes up the deficiencies of the lower estimates
207 for the crude CFR so as to produce a more accurate estimate. Of course, all the
208 estimates tend to reach the same value in the end, that is, the "final CFR".

209 3.1.2 Comparison of the estimated CFR with the final CFR

210 To compare the differences in the arithmetic mean CFR of the bi-directional
211 correction method, the weighted average CFR, the crude CFR, and the final CFR (q_z),
212 two evaluation indicators were used: (1) the Manhattan distance, i.e., absolute
213 difference $= \sum |q - q_z| / n$; (2) relative difference $= \sum (q - q_z) / n$. In addition, the
214 weighted average CFR from the bi-directional correction is divided into two
215 conditions: the calculated CFR is recorded as the weighted average CFR 1 using the
216 weight of q_{ck} is $k / (j + k)$ and the weight of q_{aj} is $j / (j + k)$; conversely, the calculated
217 CFR recorded as the weighted average CFR 2 using the weight of q_{ck} is $j / (j + k)$ and
218 the weight of q_{aj} is $k / (j + k)$. The calculation results are listed in Table 2.

219 Overall, for the Wuhan data, the difference between the crude CFR and the final CFR
220 was the smallest; conversely, for the Hubei except Wuhan, China except Hubei, and
221 China data, the difference between the crude CFR and the final CFR was the largest,
222 which indicates that the CFR estimated by the bi-directional correction is closer to the
223 true CFR in most cases. Figure 3 shows the differences and trends of the weighted
224 average CFR 1, the weighted average CFR 2, the arithmetic mean CFR, the crude
225 CFR, and the final CFR of COVID-19 in China. In the early stage of the epidemic in
226 Wuhan (Fig.3 a), when several estimates deviated from one another, the crude CFR
227 appeared closer to the final CFR; but 20 days after the epidemic had begun, the
228 bi-directional correction estimate was closer to the true CFR than the crude CFR.

229 3.1.3 The relationship between the statistical distribution of the new daily data of
230 COVID-19 in China and the k and j values

231 Using the daily data of newly diagnosed, discharged, and dead patients in China, the
232 trend of newly diagnosed, discharged, and dead cases of COVID-19 in China from
233 January 20 to April 9 can be evaluated (Figure 4).

234 In Figure 4, the first peak in data occurs for the number of diagnosed cases, followed
235 by peaks for death and discharge. This is a good verification of the logical relationship
236 given by formula (7). First, the peaks appeared earlier in areas with mild epidemic and
237 the peaks appeared later in areas with heavy epidemics. Second, for the interval time
238 from the diagnosis peak to the death peak was 6, 9, 9, and 6 days in Wuhan, Hubei
239 except Wuhan, China except Hubei, and China data, respectively; and the interval
240 time from the death peak to the discharge peak was 10, 8, 8, and 10 days, in Wuhan,
241 Hubei except Wuhan, China except Hubei, and China data, respectively. Therefore,
242 the time from diagnosis to discharge may be approximately 17 days.

243 Here, another statistical method can be used to determine the k and j values, which is
244 tentatively called "the method for determining the k and j values by the peak interval
245 time". However, this method may only be applied in the middle and later periods,
246 even after various peaks of the epidemic situation data appear. Of course, this
247 statistical distribution result given by China's COVID-19 data is definitely of reference
248 value for verifying the theoretical method of estimating the CFR and analyzing
249 foreign COVID-19 data. Compared with the results in Table 1, the peak interval time
250 given here corresponds to a larger k value and a smaller j value. If the CFR is
251 estimated based on this method, the estimated CFR is higher.

252 3.2 Estimated global mortality rate of COVID-19

253 According to the bi-directional correction method and procedure for estimating the
254 CFR, the statistical analysis of the data from eight countries around the world is
255 shown in Table 3. Since the k and j values calculated by foreign data are too diverse,
256 the additional filter $k \geq 3$ must be added. For comparison, the table also lists the current
257 crude CFR. Because there is a time lag of j days, the estimated mortality rate of the
258 bi-directional correction is actually the mortality rate of j days ago, so Table 4 gives
259 the estimated results for March 20 and March 30, respectively, while the estimation
260 result for April 9 is only given by formula (5) with a one-directional correction, that
261 makes a slightly worse accuracy for the estimate value than bi-directional correction,
262 but more accurate than the crude CFR. In addition, the trend of the CFR in the eight

263 countries is shown in Figure 5 (TableS3).

264 As shown in Table 3, the time lags of K are relatively similar and range from three to
265 five days in the eight countries, but the time lag of j is higher and ranges from 7 to 15
266 days. Since j depends on the cure data, the difference of j is related to the number of
267 discharges/recoveries announced by each country and the timeliness of the
268 announcement. The disease termination rate or case withdrawal rate reflects the
269 overall progress of epidemic treatment and the epidemic situation. For countries with
270 a disease termination rate of close to or greater than 50%, such as South Korea and
271 Iran, the crude CFR is close to our estimates, which also shows that these two
272 estimates are approaching the final CFR. For countries with a low disease termination
273 rate and where the epidemic situation is more serious, the CFR estimated by the
274 bi-directional correlation method has risen and fallen in time to reflect the epidemic
275 situation at that time. For example, the estimated CFR of Italy was lower on April 9
276 than that reported on March 20, indicating that the epidemic is improving. The crude
277 CFR cannot show this trend because the crude CFR increases from low and high over time,
278 which is one of its characteristics. From the perspective of change overtime, the change
279 in our estimates is smaller than the change in the crude CFR. For example, our
280 estimate of the CFR in Germany has a maximum difference of approximately
281 seven-fold, and the difference in the crude CFR is more than ten-fold; moreover, our
282 estimate of the worldwide CFR except for China has been relatively stable and has
283 only increased by 0.6% whereasthe crude CFR has increased by approximately 2%.
284 The increase in CFR estimates in most countries indicates that the epidemic continues
285 to worsen, and the range of the increase indicates the intensity of the outbreak.

286

287 **4 Discussion**

288 4.1 The estimated CFR of COVID-19 in China using the bi-directional correction
289 method

290 The COVID-19 epidemic is essentially over in China, and the CFR has almost reached
291 its final value. By using this complete epidemic data, we can evaluate or compare
292 different CFR estimation methods. The CFR estimated by the bi-directional correction
293 method is closer to the final CFR than the crude CFR. In the early stages of the
294 Wuhan epidemic (approximately the first 20 days of the epidemic), those estimates
295 deviated largely from one another. The crude CFR appeared closer to the final CFR,
296 but in fact, the estimated CFR of the bi-directional correction may be closer to the true
297 CFR than the crude CFR. This occurs because for an unknown novel epidemic, with
298 the occurrence and development of the epidemic, the true CFR should not be a
299 horizontal line, nor a curve from bottom to top (low to high); instead, the dynamic
300 CFR curve should be an “ Ω ” curve with steep fronts and gradual backs [6]. We think
301 this may be more consistent with actual clinical conditions, that is, with the
302 characteristics of a steep front and gradual back. Therefore, the estimates of the
303 bi-directional correction are closer to such a curve, which unfortunately cannot be
304 verified because it is almost impossible to give the exact number of infections, deaths
305 and so on at each time point of the epidemic to calculate true CFR.

306 In theory, the bi-directional correction method takes the average value of the corrected

307 confirmed CFR q_{ck} and the corrected in-patient CFR q_{aj} to estimate the CFR. q_{ck} is
308 known to be underestimated and q_{aj} is overestimated. On average, the two exactly
309 hedge the deviations in their independent estimates. Therefore, the estimated CFR of
310 the bi-directional correction method proposed in this article may be able to predict the
311 severity of the outbreak early for confirmed hospitalizations. Of course, in situations
312 when less than 100% of confirmed cases are hospitalized, the hospitalization rate
313 must be considered when calculating the average values of q_{ck} and q_{aj} to determine the
314 estimated CFR.

315 4.2 The time delay (k, j) in the bi-directional correction estimation method

316 When estimating the CFR using accumulative case data, if the time lag (k, j) is taken
317 into account for correction calculations, statistical methods or clinical experience, or a
318 combination of both, may be used to determine the k and j values. Based on clinical
319 observations, Lei (2003) believed that the time from diagnosis to death in SARS cases
320 was approximately 10 days (5). Thus, the determination of $k = 10$ was based on
321 clinical experience. According to statistical theory, Hu et al. (2020) used “the
322 minimum coefficient of variation of the daily estimated CFR” as the selection metrics,
323 that is, they used the real-time CFR minimum coefficient of variation to determine the
324 best values of j, which were 8, 10, 10, and 12 in Wuhan, Hubei except Wuhan, China
325 except Hubei, and China, respectively [7]. According to the bi-directional correction
326 method and program proposed in this paper, the analysis of the method and procedure
327 is based on the real-time CFR sequence distance minimum determination method, in
328 which $k = 3$ or 4 and $j = 11$ were determined, which is essentially consistent with
329 actual clinical conditions. China adopts unified medical treatment standards,
330 particularly the principles that all suspected and confirmed patients should be
331 admitted to the hospital and all confirmed patients should be treated, that is, once the
332 diagnosis is confirmed and regardless of whether it is mild, the patient must be
333 isolated at a designated medical place. Thus, because all confirmed patients had
334 complete records of hospitalization to recovery and discharge, these data can yield
335 more consistent k and j values.

336 Different countries have different attitudes, prevention and control strategies, and
337 medical treatment methods for COVID-19, and the nature or quality of published
338 statistical data also differ; these differences will affect data analysis and results and
339 cause deviations. For example, due to the different diagnosis and treatment standards
340 abroad, most countries with mild patients require home isolation. After asymptomatic
341 isolation for several days, they can be regarded as cured, without the need for viral
342 nucleic acid testing again; thus, the course data of these COVID-19 patients are
343 incomplete. In some countries, the case data are incomplete or only record the entry of
344 patients and ignore the exit; thus, the withdrawal rate is very low, or the number of
345 confirmed cases does not equal the sum of the number of deaths, cures, and hospital
346 treatment. Therefore, the estimated k value and j value are diversified.

347 Further research is required to determine whether or how the third method mentioned
348 in Section 2.1.3 of this paper, that is, the method for determining the k and j values by
349 the peak interval time, or referring to the information given by the cumulative data
350 curve of the cases, can be used to estimate the actual CFR.

351 4.3 The CFR of COVID-19 in foreign countries

352 The crude CFR was low at the beginning of the epidemic, which may have created
353 serious misjudgment by decision-makers. For example, the United Kingdom first
354 proposed the strategy of herd immunity; the US President also believes that the
355 COVI-19 CFR will not exceed 1%, which is lower than the ordinary flu. Unfortunately,
356 these original judgments were wrong, and those leaders missed the effective time for
357 epidemic containment. Our estimate was relatively high from the beginning and may
358 be closer to the true CFR. This information may remind policy makers to be vigilant
359 and take appropriate epidemic prevention measures to reduce the impact of the
360 epidemic on society.

361 From the perspective of the valuation curve, in the early stage of the epidemic, most
362 countries were in line with China's trend, and there was a large fluctuation in the CFR.
363 This may be related to inadequate medical resources during the early stages and the
364 intensity of the outbreak. With the use of strict prevention and control measures, the
365 estimation of the CFR will stabilize. It is worth noting that regardless of the crude
366 CFR and our estimates, the current epidemic situation in European countries may be
367 becoming serious. The existing medical methods and prevention and control measures
368 have not been able to adequately stop the epidemic.

369 The latest research shows that SARS-CoV-2 has been differentiated into three types
370 labeled A, B, and C, of which type C is the predominant type in Europe [8]. At present,
371 the rapid increase in the number of confirmed cases in Europe and the high CFR, in
372 addition to failing to effectively block the spread of the virus, may be related to the
373 change in the pathogenicity caused by the differentiation of the virus.

374

375 **Supplementary Materials:**

376 Table S1: Case number of COVID-19 in China

377 Table S2: Case number of COVID-19 in some countries

378 Table S3: The CFR estimated by bi-directional correction method for some countries

379

380 **Author Contributions:** Y.D. conceived the study, W.H., X.L. and T.W. analyzed
381 using models. C.Z., Y.Z. and D.L. collected the data. All authors participated in the
382 study design. Y.D. and Z.H. drafted the manuscript. All authors gave comments on the
383 earlier versions of the manuscript. All authors edited the manuscript and approved the
384 final version.

385

386 **Funding:** None.

387

388 **Conflicts of Interest:** The authors declare no conflict of interest.

389

390 **Reference**

391 1 World Health Organization. Report of the WHO-China Joint Mission on coronavirus
392 disease 2019 (COVID-19).

393 [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-oncoronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-oncoronavirus-disease-2019-(covid-19)) [Accessed Feb28,2020.]

394

- 395 2 Lipsitch, M et al., Potential biases in estimating absolute and relative case-fatality
396 risks during outbreaks. *PLoS Negl Trop Dis* (2015), 9: e0003846.
- 397 3 World Health Organization. Update 49-SARS case fatality ratio, incubation period.
398 https://www.who.int/csr/sars/archive/2003_05_07a/en/ [Accessed May 7, 2003]
- 399 4 Nishiura,H. Case fatality ratio of pandemic influenza. *Lancet Infect Dis* 2010,
400 10:443–444. DOI:[https://doi.org/10.1016/S1473-3099\(10\)70120-1](https://doi.org/10.1016/S1473-3099(10)70120-1)
- 401 5 Lei, H et al. Research on the estimation of the fatality rate of SARS. *Chin J Hosp*
402 *Admin* (2003) 19:541-544
- 403 6 Chen, Q. A Case Study of Mathematical Model of SARS case fatality rate, *J Acad*
404 *Equip Com & Tech* (2003) 14: 98-101
- 405 7 Hu, W et al. A method for estimating the case fatality rate during the ongoing
406 epidemic-A Case Study of COVID-19 (2020). DOI: 10.21203/rs.3.rs-21381/v1
- 407 8Froster P et al. Phylogenetic network analysis of SARS-CoV-2
408 genomes. *PNAS* (2020) DOI: <https://doi.org/10.1073/pnas.2004999117>

Table 1 The k and j values of China COVID-19 data

k/j	Wuhan	Hubei except Wuhan	China except Hubei	China	
k_{\max}	8	4	7	8	
j_{\max}	11	11	11	11	
Manhattan Distance	k	3	4	6	3
	j	11	11	11	11
Euclidean Distance	K	3	4	3	3
	j	11	11	11	10

Table 2 Comparison of CFR estimates and final CFR of COVID-19 in China

Area	Final CFR (q_z)	Evaluation index	Crude CFR	Arithmetic mean CFR	Weighted average CFR1	Weighted average CFR2
Wuhan	5.16%	Relative difference	-0.0047	0.0133	0.0157	0.0108
		Absolute difference	0.0072	0.0136	0.0159	0.0136
Hubei except	3.61%	Relative	-0.0077	-0.0015	-0.0008	-0.0021

Wuhan		difference				
		Absolute	0.0077	0.0033	0.0025	0.0041
		difference				
China except	0.82%	Relative	-0.0013	-0.0006	-0.0005	-0.0007
Hubei		difference				
		Absolute	0.0018	0.0012	0.0011	0.0014
		difference				
China	4.05%	Relative	-0.0075	0.0021	0.0026	0.0012
		difference				
		Absolute	0.0076	0.0031	0.0026	0.0053
		difference				

Note: The "final CFR (q_z)" listed in the table is also an estimate. Based on April 9, the number of deaths in before 10 days of this date is d , the number of patients cured is c , and the number of patients currently in hospital is y , then the number of patients who may die in hospital at the end is $y * d / (c + d)$; and this number of deaths plus the current cumulative number of deaths, divided by the cumulative number of confirmed cases, is called the final CFR (q_z).

Table 3 Comparison of COVID-19 CFR in several countries

Country	Time lag		March 23			April 2nd			April 12			April 22		
	k	j	Crude CFR	Arithmetic mean CFR	Withdrawal rate	Crude CFR	Arithmetic mean CFR	Withdrawal rate	Crude CFR	Arithmetic mean CFR	Withdrawal rate	Crude CFR	Arithmetic mean CFR	Withdrawal rate
Italy	5	7	9.26%	23.49%	21.14%	11.90%	24.20%	27.13%	12.79%	22.46%	34.15%	13.40%	14.59%	41.45%
Spain	4	7	6.02%	12.41%	15.03%	8.86%	14.20%	31.04%	10.10%	14.80%	46.62%	10.42%	11.32%	50.84%
Iran	3	7	7.79%	10.56%	44.36%	6.38%	9.01%	38.89%	6.22%	6.98%	66.12%	6.25%	6.55%	78.14%
South Korea	4	14	1.17%	1.48%	33.87%	1.69%	1.95%	60.11%	2.04%	—	72.13%	2.23%	2.23%	79.62%
Japan	6	15	3.68%	5.85%	28.68%	2.39%	4.75%	22.19%	1.45%	—	12.74%	2.41%	2.30%	14.20%
Germany	4	9	0.38%	0.92%	2.08%	1.19%	1.65%	26.84%	2.22%	2.71%	46.97%	3.35%	3.55%	68.69%

y		%												
USA	3	1.27 2 %	3.99%	1.84%	2.05%	5.76%	5.87%	3.76%	—	9.68%	4.84%	5.41%	14.30%	
France	4	1.426 0 %	6.60%	18.17%	7.14%	11.96%	26.79%	14.89 %	21.81%	43.64%	17.88 %	19.20%	52.18%	
Global except China	4	1.447 0 %	8.58%	15.50%	5.19%	9.52%	20.20%	6.36%	10.94%	26.70%	6.89%	7.92%	33.75%	

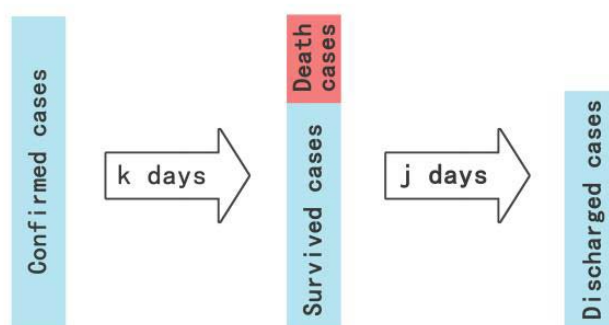


Figure 1 A course schematic diagram of COVID-19

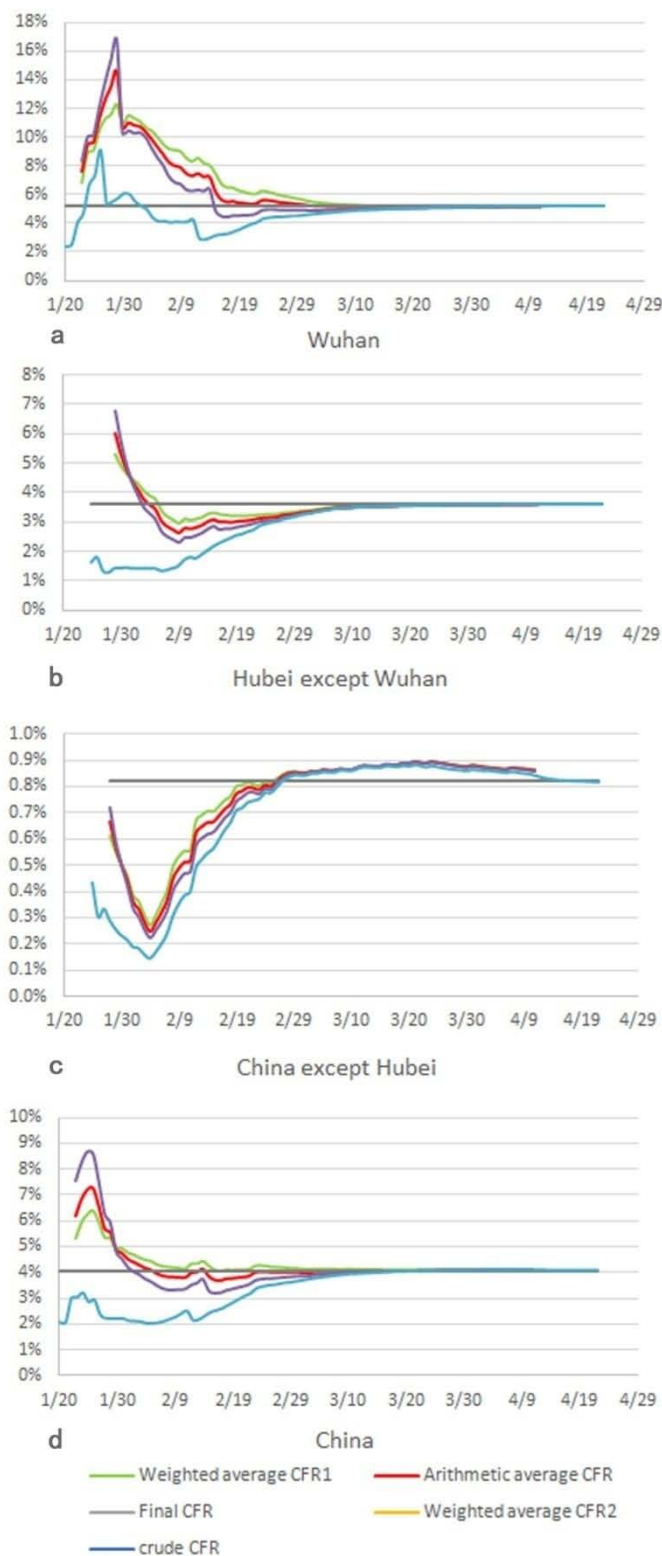


Figure 2 Time series of CFR estimates for COVID-19 in China

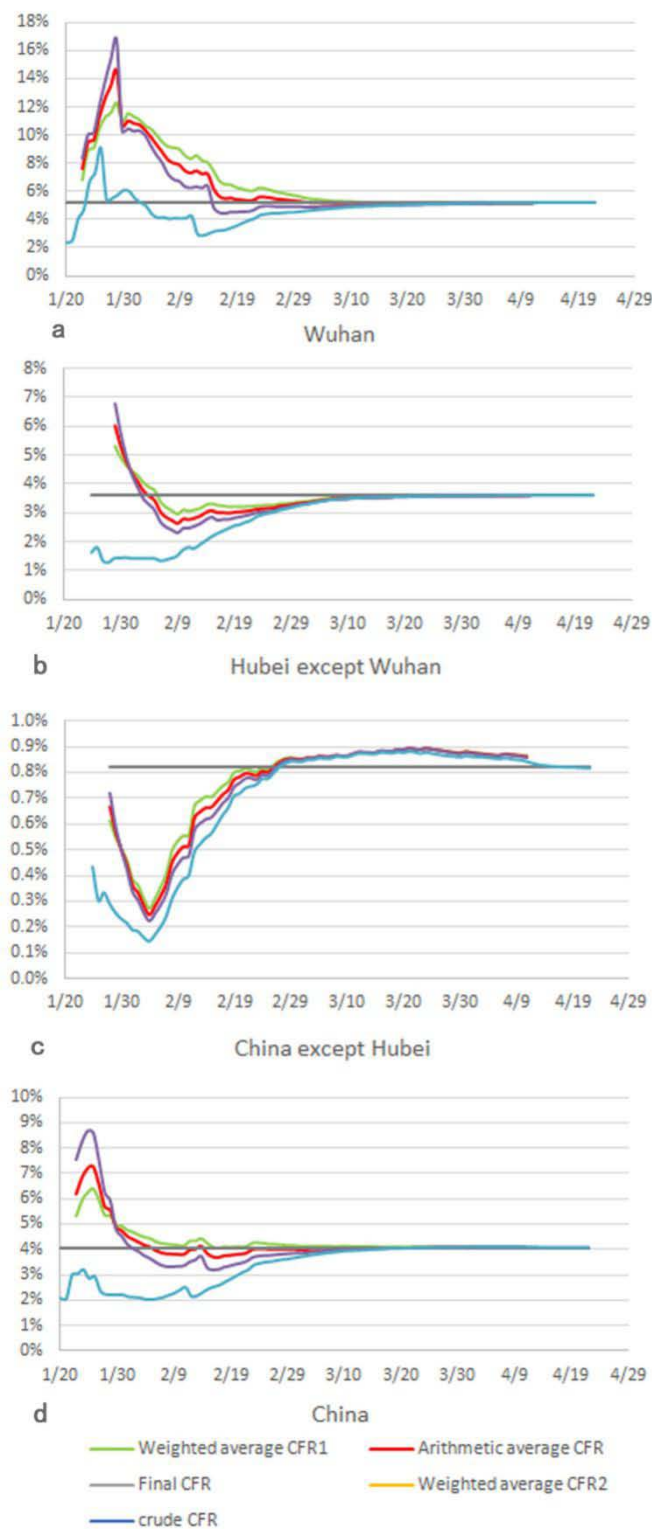


Figure 3 Trends of the CFR estimates calculated by four methods for COVID-19 in China

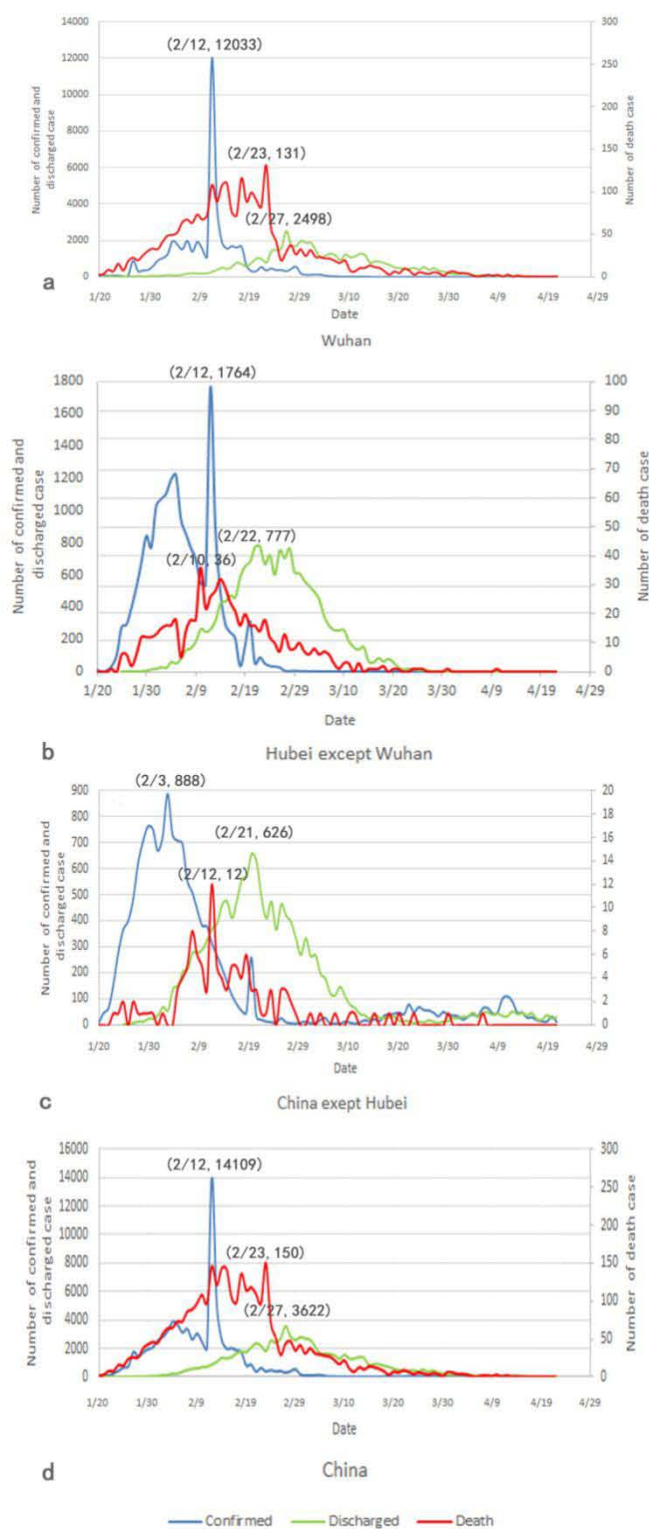


Figure 4 Trends of the confirmed, discharged and dead Cases of COVID-19 in China

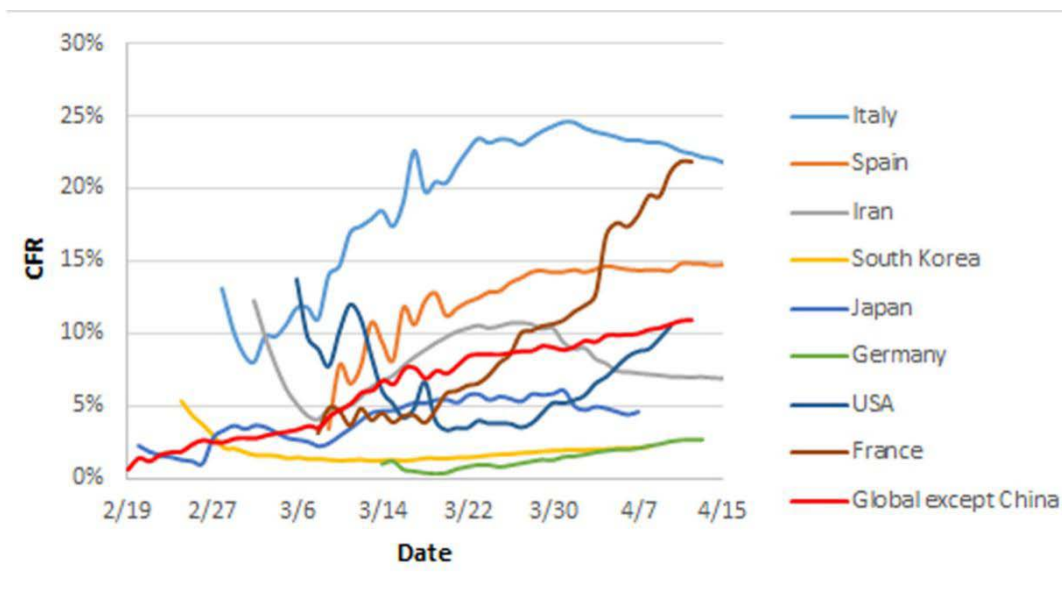


Figure 5 Trends of the CFR estimates of COVID-19 worldwide