

# 1 Early epidemiological assessment of the transmission 2 potential and virulence of coronavirus disease 2019 3 (COVID-19) in Wuhan City: China, January-February, 4 2020

5 **Authors:** Kenji Mizumoto<sup>1,2,3</sup> §, Katsushi Kagaya<sup>4</sup>, Gerardo Chowell<sup>3</sup>

## 6 **Affiliations:**

7 <sup>1</sup> Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto  
8 University Yoshida-Nakaadachi-cho, Sakyo-ku, Kyoto, Japan

9 <sup>2</sup> Hakubi Center for Advanced Research, Kyoto University, Yoshidahonmachi,  
10 Sakyo-ku, Kyoto, Japan;

11 <sup>3</sup> Department of Population Health Sciences, School of Public Health, Georgia State  
12 University, Atlanta, Georgia, USA

13 <sup>4</sup> Center for Education and Research in Information Science and Technology (CERIST),  
14 Graduate School of Information Science and Technology, The University of Tokyo,  
15 Tokyo, Japan

16

17 §Corresponding author

18 Email addresses:

19 KM: mizumoto.kenji.5a@kyoto-u.ac.jp, KK: kagaya.katsushi.8e@kyoto-u.ac.jp,

20 GC: gchowell@gsu.edu

## 21 **Article type:**

22 Original Research

## 23 **Word count:**

24 Abstract: 347 (Max 350)

25 Main: 3103

## 26 **Abstract**

### 27 **Background:**

28 Since the first cluster of cases was identified in Wuhan City, China, in December, 2019,  
 29 coronavirus disease 2019 (COVID-19) rapidly spread around the world. Despite the  
 30 scarcity of publicly available data, scientists around the world have made strides in  
 31 estimating the magnitude of the epidemic, the basic reproduction number, and  
 32 transmission patterns. Accumulating evidence suggests that a substantial fraction of the  
 33 infected individuals with the novel coronavirus show little if any symptoms, which  
 34 highlights the need to reassess the transmission potential of this emerging disease. In  
 35 this study, we derive estimates of the transmissibility and virulence of COVID-19 in  
 36 Wuhan City, China, by reconstructing the underlying transmission dynamics using  
 37 multiple data sources.

### 38 **Methods:**

39 We employ statistical methods and publicly available epidemiological datasets to jointly  
 40 derive estimates of transmissibility and severity associated with the novel coronavirus.  
 41 For this purpose, the daily series of laboratory-confirmed COVID-19 cases and deaths  
 42 in Wuhan City together with epidemiological data of Japanese repatriated from Wuhan  
 43 City on board government-chartered flights were integrated into our analysis.

44 **Results:**

45 Our posterior estimates of basic reproduction number ( $R$ ) in Wuhan City, China in  
 46 2019–2020 reached values at 3.49 (95%CrI: 3.39–3.62) with a mean serial interval of  
 47 6.0 days, and the enhanced public health intervention after January 23<sup>rd</sup> in 2020 was  
 48 associated with a significantly reduced  $R$  at 0.84 (95%CrI: 0.81–0.88), with the total  
 49 number of infections (i.e. cumulative infections) estimated at 1906634 (95%CrI:  
 50 1373500– 2651124) in Wuhan City, elevating the overall proportion of infected  
 51 individuals to 19.1% (95%CrI: 13.5–26.6%). We also estimated the most recent crude  
 52 infection fatality ratio (IFR) and time–delay adjusted IFR at 0.04% (95% CrI:  
 53 0.03%–0.06%) and 0.12% (95%CrI: 0.08–0.17%), respectively, estimates that are  
 54 several orders of magnitude smaller than the crude CFR estimated at 4.06%

55 **Conclusions:**

56 We have estimated key epidemiological parameters of the transmissibility and virulence  
 57 of COVID-19 in Wuhan, China during January-February, 2020 using an ecological  
 58 modelling approach. The power of this approach lies in the ability to infer  
 59 epidemiological parameters with quantified uncertainty from partial observations  
 60 collected by surveillance systems.

61     **Keywords:** epidemic; transmissibility; mathematical model; COVID-19; China

62

## 63     **Background**

64             The novel coronavirus (Severe acute respiratory syndrome coronavirus 2;  
65     SARS-CoV-2) that erupted from China is a deadly respiratory pathogen that belongs to  
66     the same family as the coronavirus responsible for the 2002-2003 Severe Acute  
67     Respiratory Syndrome (SARS) outbreaks [1]. Since the first cluster of cases was  
68     identified in Wuhan City, China, in December, 2019, the novel coronavirus disease 2019  
69     (COVID-19) continues its relentless march around the world as of May 12<sup>nd</sup>, 2020 [2].  
70     Nevertheless, China was hit hard by this emerging infectious disease, especially the city  
71     of Wuhan in Hubei Province, where the first cluster of severe pneumonia caused by the  
72     novel virus was identified. Meanwhile, the cumulative number of laboratory and  
73     clinically confirmed cases and deaths in mainland China has reached 82918 and 4633,  
74     respectively, as of May 10<sup>th</sup>, 2020 [3].

75             Because the morbidity and mortality burden associated with the novel  
76     coronavirus has disproportionally affected the city of Wuhan, the center of the epidemic  
77     in China, the central government of the People's Republic of China imposed a lockdown  
78     and social distancing measures in this city and surrounding areas starting on January  
79     23<sup>rd</sup> 2020. Indeed, out of the 82918 COVID-19 cases reported in China, 50339 cases  
80     (60.7%) are from Wuhan City. In terms of the death count, a total of 3869 deaths  
81     (83.5%) have been recorded in Wuhan city out of the 4633 deaths reported throughout  
82     China. To guide the effectiveness of interventions, it is crucial to gauge the uncertainty  
83     relating to key epidemiological parameters characterizing the transmissibility and the  
84     severity of the disease. Despite the scarcity of publicly available data, scientists around  
85     the world have made strides in estimating the magnitude of the epidemic, the basic  
86     reproduction number, and transmission patterns [4-5]. Moreover, accumulating evidence

suggests that a substantial fraction of the infected individuals with the novel coronavirus show little if any symptoms, which suggest the need to reassess the transmission potential of this emerging disease [6]. For this purpose, in this study we employ statistical methods and publicly available epidemiological datasets to jointly derive estimates of transmissibility and severity associated with the novel coronavirus.

## Methods

### Epidemiological data

We linked our model to two different datasets. First, the daily series of laboratory-confirmed COVID-19 cases and deaths in Wuhan City were extracted according to date of symptoms onset or reporting date from several sources [3, 7-8]. Our analysis relies on epidemiological data reported prior to February 11<sup>th</sup>, 2020 because of the change in case definition that was announced on February 12th, 2020 [9]. As of February 11<sup>th</sup>, 2020, a total of 19559 confirmed cases including 820 deaths were reported in Wuhan City. Second, epidemiological data of Japanese evacuees from Wuhan City on board government-chartered flights were obtained from the Japanese government. After arriving in Japan, all of the Japanese evacuees were kept in isolation for about 14 days and examined for infection using polymerase chain reaction (PCR) tests [7]. As of February 11<sup>th</sup>, a total of four flights with the Japanese evacuees left Wuhan City. We collected information on the timing of the evacuee flights that left Wuhan City as well as the number of passengers that tested positive for COVID-19 in order to calibrate our model (Table S1).

## 110 Statistical analysis

111 Using the following integral equation model, we estimate the reproduction  
112 number of COVID-19. Here, infected and reported cases are denoted by  $i$  and  $c$ ,  
113 respectively.

114 We connected the daily incidence series with a discrete-time integral equation  
115 to describe the epidemic dynamics. Let  $g_s$  denote the probability mass function of the  
116 serial interval, e.g., the time from illness onset in a primary case to illness onset in the  
117 secondary case, of length  $s$  days, which is given by

$$g_s = G(s) - G(s - 1) ,$$

118 For  $s > 0$  where  $G(\cdot)$  represents the cumulative distribution function of the gamma  
119 distribution. Mathematically, we describe the expected number of new cases with day  $t$ ,  
120  $E[c(t)]$  as follows,

$$E[c(t)] = \sum_{s=1}^{\infty} E[c(t-s)]R,$$

121 where  $E[c(t)]$  represents the expected number of new cases with onset day  $t$ , where  $R$   
122 represents the average number of secondary cases per case.

123 Subsequently, we also employed the time-dependent variation in  $R$  to estimate  
124 the impact of enhanced interventions on the reproduction number. This time dependence  
125 was modelled by introducing a parameter  $\delta_t$ , which is given by

$$\delta_t = \begin{cases} 1 & \text{otherwise} \\ \beta_1 & \text{if } t = \text{period}_1 \\ \beta_2 & \text{if } t = \text{period}_2 \end{cases}$$

126 where  $\text{period}_1$  and  $\text{period}_2$  represent the corresponding period from January 23<sup>rd</sup> to  
127 February 2<sup>nd</sup> 2020 and from February 3<sup>rd</sup> to February 11<sup>th</sup>, 2020, respectively. January  
128 23<sup>rd</sup> 2020 is the date when the central government of the People's Republic of China

129 imposed a lockdown in Wuhan and other cities in Hubei in an effort to quarantine the  
130 epicentre of the coronavirus (COVID-19) to mitigate transmission. Furthermore, we  
131 evenly divide the interval into two periods to incorporate the time-dependent effects on  
132  $R$  using the parameters  $\beta_1$  and  $\beta_2$  which scale the effects of the intervention, taking  
133 values smaller than 1[10].

134 To account for the probability of occurrence,  $\theta$  [11], we assume that the number  
135 of observed cases on day  $t$ ,  $h(t)$ , occurred according to a Bernoulli sampling process,  
136 with the expected values  $E(c_i; H_{t-1})$ , where  $E(c_i; H_{t-1})$  denotes the conditional expected  
137 incidence on day  $t$ , given the history of observed data from day 1 to day  $(t-1)$ , denoted  
138 by  $H_{t-1}$ . Thus, the number of expected newly observed cases is written as follows:

$$E[h(t); H_{t-1}] = \begin{cases} (1 - \theta) + \theta E[c(t); H_{t-1}], & \text{if } h = 0, \\ \theta E[c; H_{t-1}], & \text{otherwise,} \end{cases}$$

139 Further, we model the time-dependent variation in the reporting probability.

140 This time dependence was modelled by introducing a parameter  $\delta_2$ , which is given by

$$\delta_2 = \begin{cases} \alpha_1, & \text{if } t = \text{period}_3, \\ \alpha_2, & \text{if } t = \text{period}_4, \\ 1, & \text{otherwise,} \end{cases}$$

141 where  $\text{period}_3$  and  $\text{period}_4$  represent the corresponding periods from the start of our  
142 study period to Jan 16 and from Jan 17 to Jan 22, respectively, while  $\alpha_1$  and  $\alpha_2$  scale the  
143 extent of the reporting probability (where  $\alpha_1$  and  $\alpha_2$  is expected to be smaller than 1).  
144 We evenly divide the time interval before the lockdown was put in place into two  
145 periods in order to incorporate the time dependency of the reporting probability. The  
146 number of expected newly observed cases should be updated as

$$E[h(t); H_{t-1}] = \begin{cases} (1 - \theta) + q\delta\theta E[c(t); H_{t-1}], & \text{if } h_a = 0, \\ q\delta\theta E[c(t); H_{t-1}], & \text{otherwise,} \end{cases}$$

147 We assume the incidence,  $h(t)$  is the result of the Binomial sampling process with the



148 expectation  $E[h]$ . The likelihood function for the time series of observed cases that we  
149 employ to estimate the effective reproduction number and other relevant parameters is  
150 given by:

$$L_1(U; c) = \prod_{t=1}^T \binom{E(h(t); H(t-1))}{c(t)} q^{c(t)} (1-q)^{E(h(t); H(t-1))-c(t)},$$

151 where  $U$  indicates parameter sets that are estimated from this likelihood.

152 Subsequently, the conditional probability of non-infection given residents in  
153 Wuhan City at the time point of  $t_i$ ,  $p_{ti}$ , was assumed to follow a binomial distribution,  
154 and the likelihood function is given by:

$$L_2(p_{t_i}; M_{t_i}, m_{t_i}) = \binom{M_{t_i}}{m_{t_i}} p_{t_i}^{m_{t_i}} (1-p_{t_i})^{M_{t_i}-m_{t_i}},$$

155 Where  $M_{ti}$  and  $m_{ti}$  is the number of government chartered flight passengers and  
156 non-infected passengers at the date of  $t_i$ , respectively, and  $p_{ti}$  is the proportion of the  
157 estimated non-infected population in Wuhan at the date of  $t_i$ , calculated from the  $h(t)$   
158 and catchment population in Wuhan City [3,13].

159 Serial interval estimates of COVID-19 were derived from previous studies of  
160 COVID-19, indicating that it follows a gamma distribution with the mean and SD at 6.0  
161 and 2.9 days, respectively, based on ref. [14,15]. The maximum value of the serial  
162 interval was fixed at 28 days as the cumulative probability distribution of the gamma  
163 distribution up to 28 days reaches 1.000.

164

## 165 **Infection fatality ratio**

166 Crude CFR and crude IFR is defined as the number of cumulative deaths  
167 divided by the number of cumulative cases or infections at a specific point in time

without adjusting the time delay from illness onset or hospitalization to death. Next, we employed an integral equation model in order to estimate the real-time IFR. First, we estimated the real-time CFR as described elsewhere [16-18]. For the estimation, we employ the delay from hospitalization to death,  $f_s$ , which is assumed to be given by  $f_s = F(s) - F(s-1)$  for  $s > 0$  where  $H(s)$  follows a gamma distribution with mean 10.1 days and SD 5.4 days, obtained from the available observed data [19].

$$L_3(\pi; c_t, \theta) = \prod_{t_i} \binom{\sum_{t=1}^{t_i} c_t}{D_{t_i}} \left( \pi \frac{\sum_{t=2}^{t_i} \sum_{s=1}^{t-1} c_{t-s} f_s}{\sum_{t=1}^{t_i} c_t} \right)^{D_{t_i}} \left( 1 - \pi \frac{\sum_{t=2}^{t_i} \sum_{s=1}^{t-1} c_{t-s} f_s}{\sum_{t=1}^{t_i} c_t} \right)^{\sum_{t=1}^{t_i} c_t - D_{t_i}}$$

where  $c_t$  represents the number of new cases with reported day  $t$ , and  $D_{t_i}$  is the number of new deaths with reported day  $t_i$  [16-18]. We assume that the cumulative number of observed deaths,  $D_t$  is the result of the binomial sampling process with probability  $\pi$ . Subsequently, crude IFR and time-delay adjusted IFR are calculated using the estimated  $\pi$  and  $h_t$ .

The total likelihood is calculated as  $L=L_1L_2L_3$  and model parameters were estimated using a Monte Carlo Markov Chain (MCMC) method in a Bayesian framework. Posterior distributions of the model parameters were estimated based on sampling from the three Markov chains. For each chain, we drew 100,000 samples from the posterior distribution after a burn-in of 20,000 iterations. Convergence of MCMC chains were evaluated using the potential scale reduction statistic [20-21]. Estimates and 95% credibility intervals for these estimates are based on the posterior probability

186 distribution of each parameter and based on the samples drawn from the posterior  
187 distributions. All statistical analyses were conducted in R version 3.5.2 (R Foundation  
188 for Statistical Computing, Vienna, Austria) using the ‘rstan’ package.

189

## 190 **Results**

191 The daily series of COVID-19 laboratory–confirmed incidence and cumulative  
192 incidence in Wuhan in 2019–2020 are displayed in Figure 1. Overall, our dynamical  
193 models yield a good fit to the temporal dynamics (i.e. incidence, cumulative incidence)  
194 including an early exponential growth pattern in Wuhan. In incidence data, a few  
195 fluctuations are evident, probably indicating that the surveillance system likely missed  
196 many cases during the early transmission phase (Figure 1).

197 Our posterior estimates of basic reproduction number ( $R$ ) in Wuhan City, China  
198 in 2019–2020 was estimated to be 3.49 (95%CrI: 3.39–3.62). The time–dependent  
199 scaling factors quantifying the extent of enhanced public health intervention on  $R$  is  
200 0.99 (95%CrI: 0.95–1.00), declining  $R$  to 3.44 (95%CrI: 3.36–3.52) from January 23<sup>rd</sup>  
201 to February 1<sup>st</sup> and 0.24 (95%CrI: 0.23–0.26), declining  $R$  to 0.84 (95%CrI: 0.81–0.88)  
202 from February 2<sup>nd</sup> to February 11<sup>th</sup>, 2020. Other parameter estimates for the probability  
203 of occurrence and reporting rate are 0.97 (95% CrI: 0.84–1.00) and 0.010 (95% CrI:  
204 0.007–0.014), respectively. Moreover, the time–dependent scaling factor quantifying the  
205 extent of reporting rate,  $\alpha$ , is estimated to be 0.07 (95% CrI: 0.03–0.18) before January  
206 16<sup>th</sup> and to be 0.99 (95% CrI: 0.96–1.00) from January 17<sup>th</sup> to January 22<sup>nd</sup>.

207 We conducted sensitivity analyses to examine how varying the mean serial  
208 interval between 5.0 and 7.0 days affects our  $R$  estimates.  $R$  estimates are sensitive to  
209 changes in the serial interval, ranging from 2.86 (95%CrI: 2.79–2.96) to 4.10

210 (95%CrI: 3.96–4.38).

211 The total number of estimated laboratory–confirmed cases (i.e. cumulative  
212 cases) is 18967 (95% CrI: 16428–19680) while the actual numbers of reported  
213 laboratory–confirmed cases during our study period is 19559 as of February 11th, 2020.  
214 Moreover, we inferred the total number of COVID-19 infections (Figure S1). Our  
215 results indicate that the total number of infections (i.e. cumulative infections) is  
216 1906634 (95%CrI: 1373500– 2651124).

217 The Observed and posterior estimates of the cumulative number of deaths from  
218 COVID-19 in Wuhan are displayed in Figure 2, and model–based posterior estimates of  
219 the cumulative number of deaths is 821 (95%CrI: 751–892), while actual number of  
220 reported deaths is 820. The estimated temporal variation in the death risk caused by  
221 COVID-19 in Wuhan, China, 2019–2020 is shown in Figure 3 and Figure S2. Observed  
222 and posterior estimates of the crude CFR in Wuhan City is presented in Figure 2A,  
223 while observed and posterior estimates of time–delay adjusted CFR is shown in Figure  
224 2B. Furthermore, Figure 3A and 3B illustrates time–delay no–adjusted IFR and  
225 time–delay adjusted IFR, respectively.

226 The latest estimate of the crude CFR and time–delay adjusted CFR in Wuhan  
227 appeared to be 4.3% (95% CrI: 3.9–5.0%) and 12.2% (95% CrI: 11.4–13.1%),  
228 respectively, whereas the latest model–based posterior estimates of time–delay not  
229 adjusted IFR and adjusted IFR, presented in Figure 3 C and D, are 0.04%(95% CrI:  
230 0.03%–0.06%) and 0.12% (95%CrI: 0.08–0.17%), respectively, while the observed  
231 crude CFR is calculated to be 4.06% (Table 1).

232

## 233 Discussion

234 In this study we derived estimates of the transmissibility and virulence of  
235 COVID-19 in Wuhan City, China, by reconstructing the underlying transmission  
236 dynamics using multiple data sources. Applying dynamic modeling, the reproduction  
237 number, death risks as well as probabilities of occurrence and reporting rate were  
238 estimated.

239 Our posterior estimates of basic reproduction number ( $R$ ) in Wuhan City, China  
240 in 2019–2020 is calculated to be 3.49 (95%CrI: 3.39–3.62). The time-dependent scaling  
241 factor quantifying the extent of enhanced public health intervention on  $R$  is 0.99  
242 (95%CrI: 0.95–1.00), declining  $R$  to 3.44 (95%CrI: 3.36–3.52) from January 23<sup>rd</sup> to  
243 February 1<sup>st</sup> and a scaling factor at 0.24 (95%CrI: 0.23–0.26), declining  $R$  to 0.84  
244 (95%CrI: 0.81–0.88) for February 2<sup>nd</sup> to February 11<sup>th</sup>, 2020. These  $R$  estimates  
245 capturing the underlying transmission dynamics modify the impact of COVID-19, with  
246 the total number of infections (i.e. cumulative infections) estimated at 1906634  
247 (95%CrI: 1373500–2651124) in Wuhan City, raising the proportion of infected  
248 individuals to 19.1% (95%CrI: 13.7–26.5%) with a catchment population in Wuhan  
249 City of 10 million people. Our estimates of mean reproduction number reached values  
250 of 3.44, an estimate consistent with previous mean estimates in the range 2.2–3.8  
251 derived by fitting epidemic models to the initial growth phase of the observed case  
252 incidence [14,22,23]. By comparison, the  $R$  estimate for the Diamond Princess cruise  
253 ship in Japan reached values as high as ~11 [24]. Further, these estimates are higher than  
254 recent mean  $R$  estimates derived from the growth rates of the COVID-19 outbreaks in  
255 Singapore ( $R \sim 1.1$ ) [25] and Korea ( $R \sim 1.5$ ) [26].

256

257 The sustained high  $R$  values in Wuhan City even after the lockdown and mobility  
 258 restrictions suggests that transmission continues inside the household or amplified in  
 259 healthcare settings [19], which is a landmark of past SARS and MERS outbreaks  
 260 [27-28]. Considering the potent transmissibility of COVID-19 in confined settings, as  
 261 illustrated by COVID-19 outbreaks aboard cruise ships, including the Diamond Princess  
 262 cruise ship, where the total number of secondary or tertiary infections reached 705  
 263 among more than 3,700 passengers as of February 28<sup>th</sup>, 2020 and also by the COVID-19  
 264 outbreak tied to the Shincheonji religious sect in South Korea where church members  
 265 appear to have infected from seven to 10 people [29-31], it is crucial to prevent  
 266 transmission in confined settings including hospital-based transmission by  
 267 strengthening infection control measures as well as transmission stemming from large  
 268 social gatherings.

269 Our most recent estimates of the crude CFR and time-delay adjusted CFR for  
 270 Wuhan city are at 4.3% (95% CrI: 3.9–5.0%) and 12.2% (95% CrI: 11.4–13.1%),  
 271 respectively. In contrast, our most recent crude IFR and time-delay adjusted IFR is  
 272 estimated to be 0.04% (95% CrI: 0.03%–0.06%) and 0.12% (95%CrI: 0.08–0.17%),  
 273 which is several orders of magnitude smaller than the crude CFR estimated at 4.06%  
 274 and another recent estimate of the infection fatality ratio at 0.66% (95%CrI: 0.39–1.33)  
 275 and 0.6% (95% CI: 0.2-1.3) in China [32, 33]. Several data and methodological  
 276 differences can explain these differences, which we list in Table S2. For instance, Verity  
 277 et al. conducts an age adjustment based on the data of age-stratified COVID-19 deaths  
 278 from mainland China, assumes a constant attack rate by age and adjusts for  
 279 demographic structure. Our IFR estimates will be compared with estimates emerging  
 280 from ongoing several mass serological studies in China (Wuhan City), Italy, Germany

281 the U.K., and New York. Yet, these serological studies should be carefully validated  
 282 since these are not exempt of limitations as discussed elsewhere [34, 35]. Also, these  
 283 findings indicate that the death risk in Wuhan is estimated to be much higher than those  
 284 in other areas, which is likely explained by hospital-based transmission [36]. Indeed,  
 285 past nosocomial outbreaks have been reported to elevate the CFR associated with  
 286 MERS and SARS outbreaks, where inpatients that tend to be older and affected by  
 287 underlying diseases have raised the CFR to values as high as 20% for a MERS outbreak  
 288 [37-38].

289 Public health authorities are interested in quantifying both  $R$  and CFR to  
 290 measure the transmission potential and virulence of an infectious disease, especially  
 291 when emerging/re-emerging epidemics occur in order to decide the intensity of the  
 292 public health response. In the context of a substantial fraction of unobserved infections  
 293 due to COVID-19,  $R$  estimates derived from the trajectory of infections and the IFR are  
 294 more realistic indicators compared to estimates derived from observed cases alone [18,  
 295 39-40].

296 Our analysis also revealed a high probability of occurrence and quite low  
 297 reporting probabilities in Wuhan City. High probability of occurrence in the above  
 298 equation suggests that zero observed cases at some point is not due to the absence of  
 299 those infected, but more likely due to a low reporting rate. A very low reporting  
 300 probability suggests that it is difficult to diagnose COVID-19 cases or a breakdown in  
 301 medical care delivery. Moreover, we also identified a remarkable change in the  
 302 reporting rate, estimated to be 14-fold lower in the 1<sup>st</sup> period (–Jan 16<sup>th</sup>, 2020) and  
 303 about the same during the 2<sup>nd</sup> period (January 17<sup>th</sup> – 22<sup>nd</sup>, 2020), relative to that  
 304 estimated after January 23<sup>rd</sup> 2020.

305

306 Our results are not free from the limitations. First, our methodology aims to capture the  
307 underlying transmission dynamics using multiple data sources. By implementing mass  
308 screening in certain populations is a useful approach to ascertain the real proportion of  
309 those infected and a way of adding credibility to the estimated values. Second, it is  
310 worth noting that the data of Japanese evacuees from Wuhan employed in our analysis  
311 is not a random sample from the Wuhan catchment population. Indeed, it also plausible  
312 that their risk of infection in this sample is not as high as that of local residents in  
313 Wuhan, underestimating the fatality risk. Third, given the likely under-ascertainment of  
314 cases, there may also exist unreported deaths, and this might underestimate the death  
315 risk. Fourth, case fatality ratio (CFR) varies with age, gender, presence or absence of  
316 comorbidities, race, whether the healthcare system is overloaded or not, and other  
317 factors such as poverty risk, infant mortality risk, and the cumulative morbidity ratio  
318 [41-45]. As CFR is influence by reporting rate and ascertainment bias, subgroup  
319 analysis of IFR based on individual-level data is essential to capture the overall disease  
320 burden of COVID-19.

321

## 322 **Conclusion**

323 In summary, we have estimated key epidemiological parameters of the  
324 transmissibility and virulence of COVID-19 in Wuhan, China, January-February, 2020  
325 using an ecological modelling approach and several epidemiological datasets. The  
326 power of our approach lies in the ability to infer epidemiological parameters with  
327 quantified uncertainty from partial observations collected by surveillance systems.



328

## 329 **List of abbreviations**

330 CFR: Case fatality ratio, IFR: Infection Fatality ratio, SARS: Severe Acute Respiratory

331 Syndrome, MERS: Middle East Respiratory Syndrome

332

## 333 **Additional files**

334 **Additional file 1: Table S1.** Information related to Japanese evacuees from Wuhan City

335 on board government-chartered flights. **Table S2.** Main differences between our study

336 and previous study.

337 **Additional file 2: Fig. S1. Observed daily new cases and posterior estimates of the**

338 **daily new infections of the COVID-19 in Wuhan, China, 2019–2020.**

339 Observed daily new cases and posterior estimates of infections of the COVID-19 are

340 presented. Observed data are presented in the dot, while dashed line indicates 50

341 percentile, and areas surrounded by light grey and deep grey indicates 95% and 50%

342 credible intervals (CrI) for posterior estimates, respectively. Epidemic day 1

343 corresponds to the day that starts at January 1<sup>st</sup>, 2020.

344

### 345 **Additional file 3: Fig. S2. Temporal variation of the case fatality risks caused by**

### 346 **COVID-19 in Wuhan, China, 2019–2020.**

347 (A) Observed and posterior estimates of crude case fatality ratio in Wuhan City, (B)  
 348 Observed crude case fatality ratio and posterior estimates of time–delay adjusted CFR  
 349 in Wuhan City. This figure is submitted to the ref [19]. The purpose of the study is to  
 350 compare the case fatality ration (CFR. Not IFR) in three different areas (Wuhan City, in  
 351 Hubei Province excluding Wuhan City and in China excluding Hubei Province) to  
 352 interpret the current severity of the epidemic in China, and the purpose is different from  
 353 this study.

354

355

## 356 **Declarations**

### 357 **Ethics approval and consent to participate**

358 Not applicable.

### 359 **Consent for publication**

360 Not applicable.

### 361 **Availability of data and materials**

362 The present study relies on published data and access information to essential  
 363 components of the data are available from the corresponding author.

### 364 **Competing interests**

365 The authors declare that they have no competing interests.

366

## 367 **Funding**

368 KM acknowledges support from the Japan Society for the Promotion of Science (JSPS)  
 369 KAKENHI (Grant Number 18K17368 and 20H03940) and from the Leading Initiative  
 370 for Excellent Young Researchers from the Ministry of Education, Culture, Sport,  
 371 Science & Technology of Japan. KK acknowledges support from the JSPS KAKENHI  
 372 Grant Number 18K19336 and 19H05330. GC acknowledges support from NSF grant  
 373 1414374 as part of the joint NSF–NIH–USDA Ecology and Evolution of Infectious  
 374 Diseases program.

## 375 **Authors' contributions**

376 KM and GC conceived the early study idea. KM and KK built the model. KM  
 377 implemented statistical analysis and wrote the first full draft. GC advised on and helped  
 378 shape the research. All authors contributed to the interpretation of the results and edited  
 379 and commented on several earlier versions of the manuscript. All authors read and  
 380 approved the final manuscript.

## 381 **Acknowledgements**

382 Not applicable.  
 383

## 384 **REFERENCES**

- 385 1. Jon Cohen. **Mining coronavirus genomes for clues to the outbreak's origins.**  
 386 **Science.** Jan 31, 2020.

- 387 [https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-](https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins)
- 388 [outbreak-s-origins](https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins). Accessed, Feb 3<sup>rd</sup>, 2020
- 389 2. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) situation
- 390 reports. Available from:
- 391 [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-rep](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports)
- 392 [orts](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports) Accessed, [cited 2020 May 10<sup>th</sup> ]
- 393 3. The State Council, The People's Republic of China. [cited 2020 May 10<sup>th</sup> ].
- 394 <http://www.gov.cn/>
- 395 4. Nishiura H, Jung SM, Linton NM, Kinoshita R, Yang Y, Hayashi K, et al. **The**
- 396 **Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020.** *J Clin*
- 397 *Med.* 2020; 9(2); 330
- 398 5. Wu JT, Leung K, Leung GM. **Nowcasting and forecasting the potential**
- 399 **domestic and international spread of the 2019-nCoV outbreak originating**
- 400 **in Wuhan, China: a modelling study.** *Lancet.* 2020. pii:
- 401 S0140-6736(20)30260-9. doi: 10.1016/S0140-6736(20)30260-9.
- 402 6. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung SM, et
- 403 al. **Epidemiological characteristics of novel coronavirus infection: A**

- 404            **statistical analysis of publicly available case data.** *medRxiv*
- 405            2020.01.26.20018754; doi: <https://doi.org/10.1101/2020.01.26.20018754>
- 406            7. Health Commission of Hubei Province, China. [cited 2020 Feb 7<sup>th</sup>].
- 407            <http://wjw.hubei.gov.cn/>
- 408            8. Health Commission of Wuhan City, Hubei Province, China [cited 2020 Feb 7<sup>th</sup>]
- 409            <http://wjw.hubei.gov.cn/>
- 410            9. Clinical guideline for COVID-19, version 5. The State Council, The People's
- 411            Republic of China. Available from
- 412            [http://www.gov.cn/zhengce/zhengceku/2020-02/05/5474791/files/de44557832a](http://www.gov.cn/zhengce/zhengceku/2020-02/05/5474791/files/de44557832ad4be1929091dcbcfca891.pdf)
- 413            [d4be1929091dcbcfca891.pdf](http://www.gov.cn/zhengce/zhengceku/2020-02/05/5474791/files/de44557832ad4be1929091dcbcfca891.pdf) [Accessed Feb 29<sup>th</sup>, 2020][in Chinese]
- 414            10. Ministry of Health, Labour and Welfare, Japan.
- 415            <https://www.mhlw.go.jp/index.html> [in Japanese]
- 416            11. 2020 Hubei lockdowns, Wikipedia.
- 417            [https://en.wikipedia.org/wiki/2020\\_Hubei\\_lockdowns](https://en.wikipedia.org/wiki/2020_Hubei_lockdowns)
- 418            12. Li R,a Weiskittel AR,a Kershaw Jr, JA. **Modeling annualized occurrence,**
- 419            **frequency, and composition of ingrowth using mixed-effects zero-inflated**
- 420            **models and permanent plots in the Acadian Forest Region of North**
- 421            **America.** *Can J For Res.* 2011; 41:2077–2089

- 422 13. Northeastern University. Laboratory for the Modeling of Biological and Socio -  
423 Technical Systems, 2020. Available online: [https://www.mobs -](https://www.mobs-lab.org/2019ncov.html)  
424 [lab.org/2019ncov.html](https://www.mobs-lab.org/2019ncov.html) [accessed on January 22<sup>nd</sup>, 2020).  
425 14. Li Q, Guan X, et. al. **Early Transmission Dynamics in Wuhan, China, of**  
426 **Novel Coronavirus–Infected Pneumonia.** *N Engl J Med.* 2020 Jan 29. DOI:  
427 10.1056/NEJMoa2001316 Available at:  
428 <https://www.nejm.org/doi/full/10.1056/NEJMoa2001316>  
429 15. Nishiura H, Linton NM, Akhmetzhanov AR. *Int J Infect Dis.* 2020  
430 Apr;93:284-286. doi: 10.1016/j.ijid.2020.02.060  
431 16. Ghani AC, Donnelly CA, Cox DR, Griffin JT, Fraser C, Lam TH, et al.  
432 **Methods for estimating the case fatality ratio for a novel, emerging**  
433 **infectious disease.** *Am J Epidemiol.* 2005; 162: 479-486  
434 17. Nishiura H, Klinkenberg D, Roberts M, Heesterbeek JA. **Early epidemiological**  
435 **assessment of the virulence of emerging infectious diseases: a case study of**  
436 **an influenza pandemic.** *PLoS One.* 2009;4(8):e6852. doi:  
437 10.1371/journal.pone.0006852.  
438 18. Tsuzuki S, Lee H, Miura F, Chan YH, Jung SM, Akhmetzhanov AR, Nishiura H.  
439 **Dynamics of the pneumonic plague epidemic in Madagascar, August to**

- 440           **October 2017. *Euro Surveill.* 2017;22(46). doi:**
- 441           10.2807/1560-7917.ES.2017.22.46.17-00710.
- 442       19. Mizumoto K, Chowell G. **Estimating the risk of 2019 Novel Coronavirus**
- 443           **death during the course of the outbreak in China, 2020. *Emerg Infect Dis.***
- 444           2020; 26.
- 445       20. Gamerman, D. & Lopes, H. F. **Markov Chain Monte Carlo: Stochastic**
- 446           **Simulation for Bayesian Inference.** 2nd edn (Chapman & Hall/CRC, 2006).
- 447       21. Gelman, A. & Rubin, D. B. **Inference from iterative simulation using**
- 448           **multiple sequences. *Stat Sci* 7:457-472, doi:10.1214/ss/1177011136 (1992).**
- 449       22. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. **Novel coronavirus**
- 450           **2019-nCoV: early estimation of epidemiological parameters and epidemic**
- 451           **predictions. *medRxiv.* doi: <https://doi.org/10.1101/2020.01.23.20018549>**
- 452       23. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, Ferguson NM.
- 453           **Report 3: Transmissibility of 2019-nCoV.**
- 454           [https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fello](https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf)
- 455           [wships/Imperial-2019-nCoV-transmissibility.pdf](https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf)

- 456 24. Mizumoto K, Chowell G. **Transmission potential of the novel coronavirus**  
457 **(COVID-19) onboard the Diamond Princess Cruises Ship, 2020.** *Infect Dis*  
458 *Model.* 2020. 264-270
- 459 25. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, Chowell G. **Real-time**  
460 **monitoring the transmission potential of COVID-19 in Singapore, February**  
461 **2020.** *medRxiv.* doi: <https://doi.org/10.1101/2020.02.21.20026435>
- 462 26. Shim E, Tariq A, Choi W, Lee Y, Chowell G. **Transmission potential of**  
463 **COVID-19 in South Korea.** *medRxiv.*  
464 doi:<https://doi.org/10.1101/2020.02.27.20028829>
- 465 27. Chowell G, Abdirizak F, Lee S, Lee J, Jung E, Nishiura H, Viboud C.  
466 **Transmission characteristics of MERS and SARS in the healthcare setting:**  
467 **a comparative study.** *BMC Med.* 2015;**13**:210. doi:  
468 10.1186/s12916-015-0450-0.
- 469 28. Abdirizak F, Lewis R, Chowell G. **Evaluating the potential impact of targeted**  
470 **vaccination strategies against severe acute respiratory syndrome**  
471 **coronavirus (SARS-CoV) and Middle East respiratory syndrome**  
472 **coronavirus (MERS-CoV) outbreaks in the healthcare setting.** *Theor Biol*  
473 *Med Model.* 2019;**16**(1):16. doi: 10.1186/s12976-019-0112-6.



- 474 29. Blake Essig, Brent Swails, Yoko Wakatsuki and Ben Westcott, CNN. **Top**  
475 **Japanese government adviser says Diamond Princess quarantine was**  
476 **flawed.** Updated 0708 GMT (1508 HKT) February 27, 2020.  
477 <https://edition.cnn.com/2020/02/27/asia/japan-diamond-princess-quarantine-cre>  
478 <w-intl-hnk/index.html>
- 479 30. Da-hae P, Dam-eun S, Jae-gu K. HANKYOREH. **The reasons why**  
480 **transmission is so prevalent among Shincheonji members.** Mar 2, 2020.  
481 [http://english.hani.co.kr/arti/english\\_edition/e\\_national/930749.html](http://english.hani.co.kr/arti/english_edition/e_national/930749.html) [Accessed  
482 Mar/10, 2020]
- 483 31. Mizumoto K, Kagaya K, Zarebski A, Chowell G. **Estimating the**  
484 **Asymptomatic Proportion of 2019 Novel Coronavirus onboard the Princess**  
485 **Cruises Ship, 2020.** *Euro Surveill.* 2020; 25.
- 486 32. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al.  
487 **Estimates of the severity of coronavirus disease 2019: a model-based**  
488 **analysis.** *Lancet Infect Dis.* 2020. pii: S1473-3099(20)30243-7.
- 489 33. Russell TW, Hellewell J, Jarvis CI, et al. **Estimating the Infection and Case**  
490 **Fatality Ratio for Coronavirus Disease (COVID-19) Using Age-Adjusted**  
491 **Data From the Outbreak on the Diamond Princess Cruise Ship, February**

- 492           **2020. *Euro Surveill.* 2020; 25(12): 2000256. doi:**  
493           **10.2807/1560-7917.ES.2020.25.12.2000256.**
- 494       34. Wu X, Fu B, Chen L, Feng Y. **Serological tests facilitate identification of**  
495           **asymptomatic SARS-CoV-2 infection in Wuhan, China.** *Journal of Medical*  
496           *Virology.* <https://doi.org/10.1002/jmv.25904>
- 497       35. Giugliano F. **Mass Coronavirus Antibody Tests Have Serious Limits.**  
498           *Bloomberg Opinion.* April 24, 2020. [Accessed on May 30<sup>th</sup>, 2020)]
- 499       36. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. **Clinical Characteristics of**  
500           **138 Hospitalized Patients With 2019 Novel Coronavirus-Infected**  
501           **Pneumonia in Wuhan, China.** *JAMA.* 2020. doi: 10.1001/jama.2020.1585.
- 502       37. Mizumoto K, Endo A, Chowell G, Miyamatsu Y, Saitoh M, Nishiura H.  
503           **Real-time characterization of risks of death associated with the Middle East**  
504           **respiratory syndrome (MERS) in the Republic of Korea, 2015.** *BMC Med.*  
505           2015;**13**:228. doi: 10.1186/s12916-015-0468-3.
- 506       38. Mizumoto K, Saitoh M, Chowell G, Miyamatsu Y, Nishiura H. **Estimating the**  
507           **risk of Middle East respiratory syndrome (MERS) death during the course**  
508           **of the outbreak in the Republic of Korea, 2015.** *Int J Infect Dis.* 2015;**39**:7-9.  
509           doi: 10.1016/j.ijid.2015.08.005.

- 510 39. J.Y. Wong, P. Wu, H. Nishiura, E. Goldstein, E.H. Lau, L. Yang, et al. **Infection**  
511 **fatality risk of the pandemic A(H1N1)2009 virus in Hong Kong.** *Am J*  
512 *Epidemiol.* 2013;**177** (8):pp. 834-840
- 513 40. Presanis AM, De Angelis D; New York City Swine Flu Investigation Team,  
514 Hagy A, Reed C, Riley S, Cooper BS, et al. **The severity of pandemic H1N1**  
515 **influenza in the United States, from April to July 2009: a Bayesian analysis.**  
516 *PLoS Med.* 2009;**6**(12):e1000207.
- 517 41. The Guardian. **Black people four times more likely to die from Covid-19,**  
518 **ONS finds.**  
519 <https://www.theguardian.com/world/2020/may/07/black-people-four-times-more>  
520 **-likely-to-die-from-covid-19-ONS-finds** [accessed May 29, 2020]
- 521 42. Mizumoto K, Dahal S, Chowell G. **Spatial variability in the risk of death**  
522 **from COVID-19 in 20 regions of Italy.** *medRxiv.* 2020.04.01.20049668.
- 523 43. Shim E, Mizumoto K, Choi W, et al. **Estimating the risk of COVID-19 death**  
524 **during the course of the outbreak in Korea, February- May, 2020.** *medRxiv.*  
525 2020.03.30.20048264.

- 526        44. Dahal S, Mizumoto K, Chowell G. **Investigating spatial variability in**
- 527                **COVID-19 pandemic severity across 19 geographic areas, Spain, 2020.**
- 528                *medRxiv*. 2020.04.14.20065524.
- 529        45. Undurraga EA, Chowell G, Mizumoto K. **Case fatality risk by age from**
- 530                **COVID-19 in a high testing setting in Latin America: Chile, March-May,**
- 531                **2020.** *medRxiv*. 2020.05.25.20112904.
- 532
- 533

## 534     **Figures**

### 535     **Figure 1. Observed and posterior estimates of the daily new cases and** 536     **cumulative cases of the COVID-19 cases in Wuhan, China, 2019–2020**

537     Observed and posterior estimates of laboratory–confirmed reported cases (A) and  
538     cumulative reported cases (B) are presented.  
539     Observed data are presented in the dot, while dashed line indicates 50 percentile, and  
540     areas surrounded by light grey and deep grey indicates 95% and 50% credible intervals  
541     (CrI) for posterior estimates, respectively. Epidemic day 1 corresponds to the day that  
542     starts at January 1<sup>st</sup>, 2020.

### 544     **Figure 2. Observed and posterior estimates of the cumulative deaths of the** 545     **COVID-19 in Wuhan, China, 2019–2020**

546     Observed and posterior estimates of the cumulative deaths of the COVID-19 in Wuhan  
547     is presented. Observed data are presented in the dot, while dashed line indicates 50  
548     percentile, and areas surrounded by light grey and deep grey indicates 95% and 50%  
549     credible intervals (CrI) for posterior estimates, respectively. Epidemic day 1  
550     corresponds to the day that starts at January 1<sup>st</sup>, 2020.

### 552     **Figure 3. Temporal variation of the infection fatality risks caused by COVID-19 in** 553     **Wuhan, China, 2019–2020**

554

555 (A) Posterior estimates of crude infection fatality ratio in Wuhan City. (B) Posterior  
556 estimates of time–delay adjusted infection fatality ratio in Wuhan City.  
557 Black dots shows observed data, and light and dark indicates 95% and 50% credible  
558 intervals for posterior estimates, respectively. Epidemic day 1 corresponds to the day  
559 that starts at January 1<sup>st</sup>, 2020.  
560

## 561 Tables

562 Table 1 – Death risk by COVID-19 in Wuhan City, China, 2020 (As of  
563 February 12, 2020)

Death Risk	Latest estimate	Range of median estimates
Crude CFR (Observed)	4.06%	2.0 – 9.0%
Crude CFR (Estimated)	4.3% (95%CrI <sup>‡</sup> : 3.9 – 5.0%)	3.4 – 7.1%
Time delay adjusted CFR	12.2% (95%CrI: 11.4 – 13.1%)	4.0 – 34.5%
Crud IFR	0.04% (95%CrI: 0.03 – 0.06%)	0.02 – 0.07%
Time delay adjusted IFR	0.12% (95%CrI: 0.08 – 0.17%)	0.03 – 0.33%

564 CrI: Credibility intervals, CFR: Case fatality ratio, IFR: Infection fatality  
565 ratio

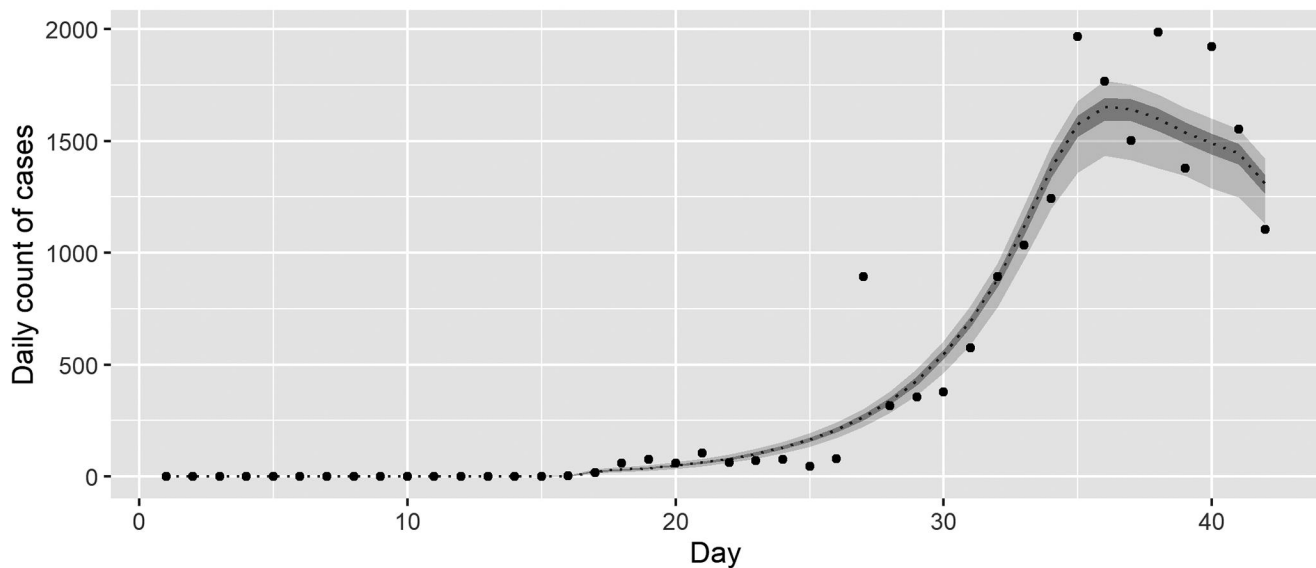
566 <sup>‡</sup>Upper and lower 95% credibility interval

567

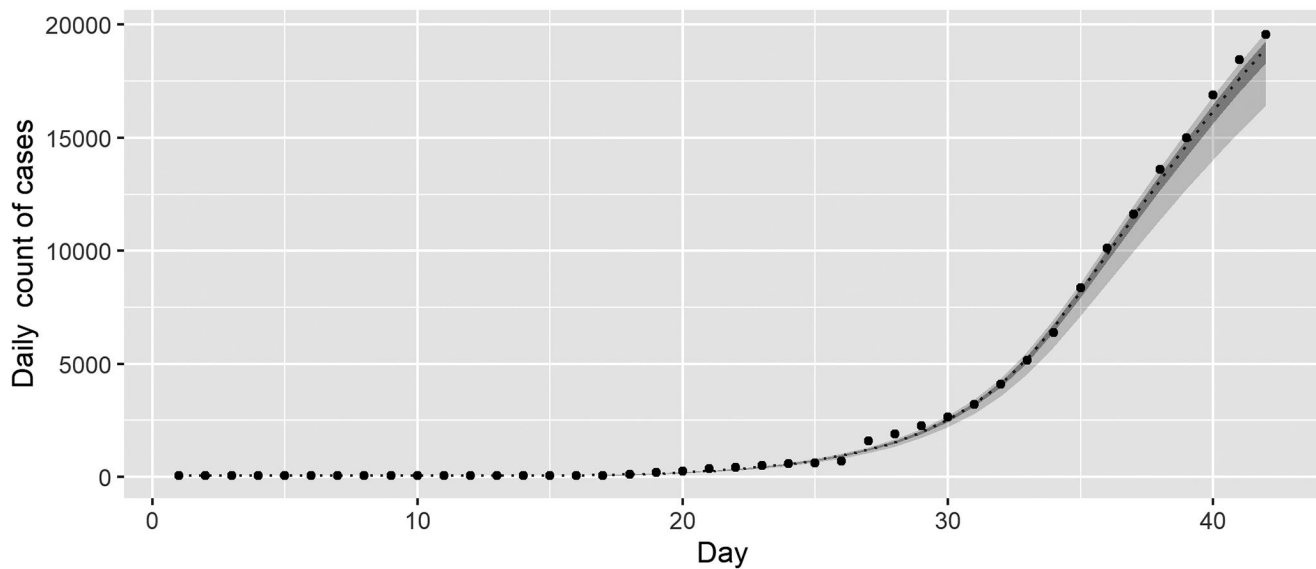
568

569

**A** Observed and estimated number of reported cases  
Wuhan

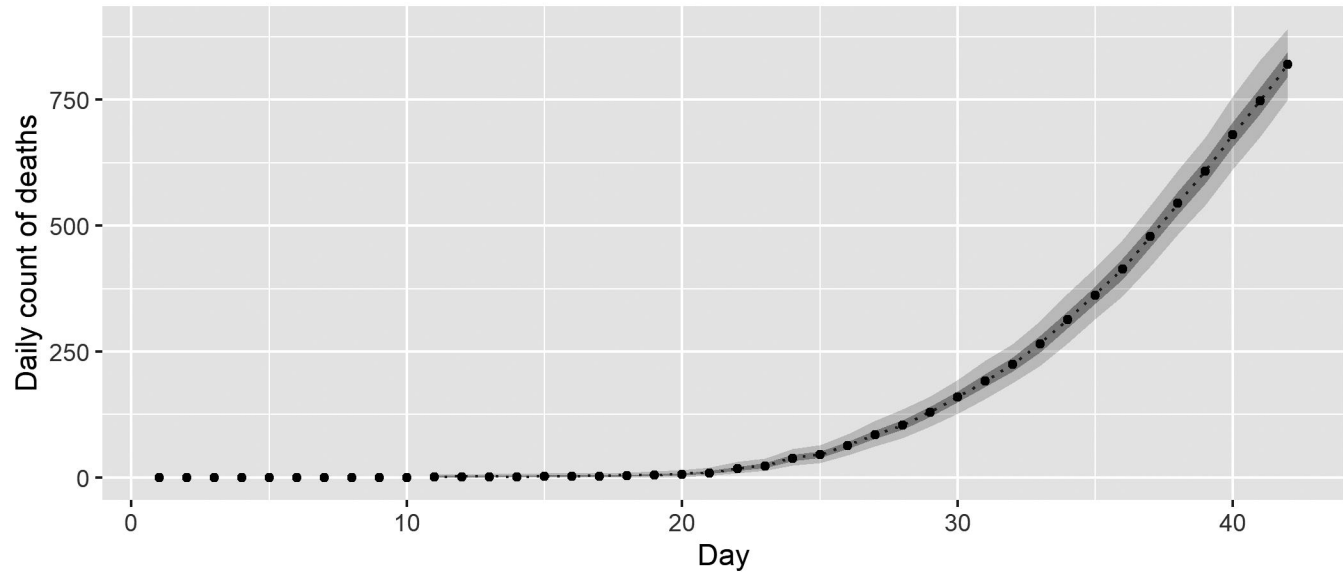


**B** Observed and estimated number of cumulative reported cases  
Wuhan



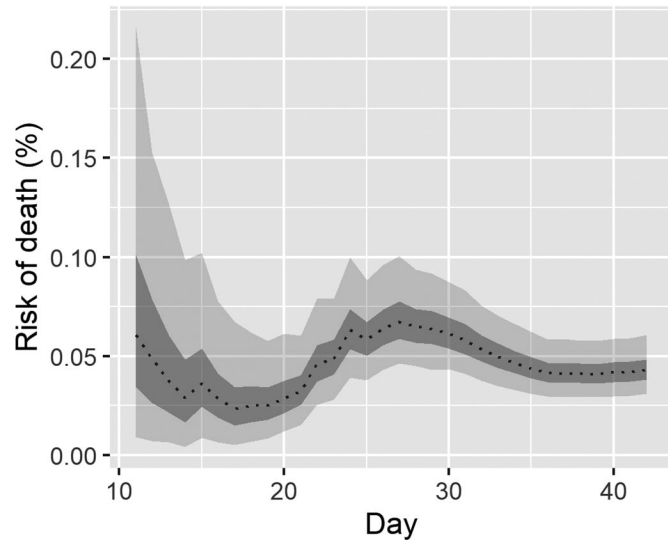


Observed and estimated number of reported cumulative death  
Wuhan



**A**

Infection fatality ratio  
Crude  
Wuhan

**B**

Infection fatality ratio  
Time-delay adjusted  
Wuhan

