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

Authors: Kenji Mizumoto[1,2,3] §, Katsushi Kagaya[4], Gerardo Chowell [3]

Affiliations:

[1] Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University Yoshida–Nakaadachi–cho, Sakyo–ku, Kyoto, Japan
[2] Hakubi Center for Advanced Research, Kyoto University, Yoshidahonmachi, Sakyo–ku,Kyoto, Japan;
[3] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, Georgia, USA
[4] Center for Education and Research in Information Science and Technology (CERIST), Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan

§Corresponding author
Email addresses:
KM: mizumoto.kenji.5a@kyoto-u.ac.jp, KK: kagaya.katsushi.8e@kyoto-u.ac.jp,
GC: gchowell@gsu.edu

Article type:

Original Research

Word count:

Abstract: 347 (Max 350)
Main: 3103

Abstract

Background:

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

Methods:

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

Results:

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

Conclusions:

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

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

Background

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

Because the morbidity and mortality burden associated with the novel coronavirus has disproportionally affected the city of Wuhan, the center of the epidemic in China, the central government of the People's Republic of China imposed a lockdown and social distancing measures in this city and surrounding areas starting on January 23rd 2020. Indeed, out of the 82918 COVID-19 cases reported in China, 50339 cases (60.7%) are from Wuhan City. In terms of the death count, a total of 3869 deaths (83.5%) have been recorded in Wuhan city out of the 4633 deaths reported throughout China. To guide the effectiveness of interventions, it is crucial to gauge the uncertainty relating to key epidemiological parameters characterizing the transmissibility and the severity of the disease. Despite the scarcity of publicly available data, scientists around the world have made strides in estimating the magnitude of the epidemic, the basic 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 11th, 2020 because of the change in case definition that was announced on February 12th, 2020 [9]. As of February 11th, 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 11th, a total of four flights with the Japanese evacuees left Wuhan City. We collected information on the timing of the evacuee fights 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).

Statistical analysis

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

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

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

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

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

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

Subsequently, we also employed the time–dependent variation in *R* to estimate the impact of enhanced interventions on the reproduction number. This time dependence was modelled by introducing a parameter δ_1, which is given by

$$\delta_1 = \begin{cases} 1 & otherwise \\ \beta_1 & \text{if } t = period_1 \\ \beta_2 & \text{if } t = period_1 \end{cases}$$

where $period_1$ and period2 represent the corresponding period from January 23rd to February 2nd 2020 and from February 3rd to February 11th, 2020, respectively. January 23rd 2020 is the date when the central government of the People's Republic of China

imposed a lockdown in Wuhan and other cities in Hubei in an effort to quarantine the epicentre of the coronavirus (COVID-19) to mitigate transmission. Furthermore, we evenly divide the interval into two periods to incorporate the time-dependent effects on R using the parameters β_1 and β_2 which scale the effects of the intervention, taking values smaller than 1[10].

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

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

Further, we model the time–dependent variation in the reporting probability. This time dependence was modelled by introducing a parameter δ_2, which is given by

$$\delta_2 = \begin{cases} \alpha_1, & \text{if } t = period_3, \\ \alpha_2, & \text{if } t = period_4, \\ 1, & \text{otherwise,} \end{cases}$$

where period_3 and period_4 represent the corresponding periods from the start of our study period to Jan 16 and from Jan 17 to Jan 22, respectively, while α_1 and α_2 scale the extent of the reporting probability (where α_1 and α_2 is expected to be smaller than 1). We evenly divide the time interval before the lockdown was put in place into two periods in order to incorporate the time dependency of the reporting probability. The number of expected newly observed cases should be updated as

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

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:

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

155 Where M_{ti} and m_{ti} is the number of government charted flight passengers and
156 non–infected passengers at the date of t_i, respecitively, 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].

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

where c_t represents the number of new cases with reported day t, and D_{ti} 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 π. Subsequently, crude IFR and time–delay adjusted IFR are calculated using the estimated π 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 23rd
201 to February 1st and 0.24 (95%CrI: 0.23–0.26), declining *R* to 0.84 (95%CrI: 0.81–0.88)
202 from February 2nd to February 11th, 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, α, is estimated to be 0.07 (95% CrI: 0.03–0.18) before January
206 16th and to be 0.99 (95% CrI: 0.96–1.00) from January 17th to January 22nd.
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

(95%CrI: 3.96–4.38).

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

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

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

Discussion

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

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

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

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

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

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

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

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

Conclusion

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

List of abbreviations

CFR: Case fatality ratio, IFR: Infection Fatality ratio, SARS: Severe Acute Respiratory Syndrome, MERS: Middle East Respiratory Syndrome

Additional files

Additional file 1: Table S1. Information related to Japanese evacuees from Wuhan City on board government–chartered flights. **Table S2.** Main differences between our study and previous study.

Additional file 2: Fig. S1. Observed daily new cases and posterior estimates of the daily new infections of the COVID-19 in Wuhan, China, 2019–2020.

Observed daily new cases and posterior estimates of infections of the COVID-19 are presented. Observed data are presented in the dot, while dashed line indicates 50 percentile, and areas surrounded by light grey and deep grey indicates 95% and 50% credible intervals (CrI) for posterior estimates, respectively. Epidemic day 1 corresponds to the day that starts at January 1st, 2020.

Additional file 3: Fig. S2. Temporal variation of the case fatality risks caused by COVID-19 in Wuhan, China, 2019–2020.

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

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Funding

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

Authors' contributions

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

Acknowledgements

Not applicable.

REFERENCES

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

https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins.Accessed, Feb 3rd, 2020

2. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) situation reports. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports Accessed, [cited 2020 May 10th]

3. The State Council, The People's Republic of China. [cited 2020 May 10th]. http://www.gov.cn/

4. Nishiura H, Jung SM, Linton NM, Kinoshita R, Yang Y, Hayashi K, et al. **The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020**. *J Clin Med*. 2020; 9(2); 330

5. Wu JT, Leung K, Leung GM. **Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.** *Lancet.* 2020. pii: S0140-6736(20)30260-9. doi: 10.1016/S0140-6736(20)30260-9.

6. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung SM, et al. **Epidemiological characteristics of novel coronavirus infection: A**

statistical analysis of publicly available case data. *medRxiv* 2020.01.26.20018754; doi: https://doi.org/10.1101/2020.01.26.20018754

7. Health Commission of Hubei Province, China. [cited 2020 Feb 7th]. http://wjw.hubei.gov.cn/

8. Health Commission of Wuhan City, Hubei Province, China [cited 2020 Feb 7 th] http://wjw.hubei.gov.cn/

9. Clinical guideline for COVID-19, version 5. The State Council, The People's Republic of China. Available from http://www.gov.cn/zhengce/zhengceku/2020-02/05/5474791/files/de44557832ad4be1929091dcbcfca891.pdf [Accessed Feb 29th, 2020][in Chinese]

10. Ministry of Health, Labour and Welfare, Japan. https://www.mhlw.go.jp/index.html [in Japanese]

11. 2020 Hubei lockdowns, Wikipedia. https://en.wikipedia.org/wiki/2020_Hubei_lockdowns

12. Li R,a Weiskittel AR,a Kershaw Jr, JA. **Modeling annualized occurrence, frequency, and composition of ingrowth using mixed-effects zero-inflated models and permanent plots in the Acadian Forest Region of North America.** *Can J For Res*. 2011; 41:2077–2089

13. Northeastern University. Laboratory for the Modeling of Biological and Socio‐Technical Systems, 2020. Available online: https://www.mobs‐lab.org/2019ncov.html [accessed on January 22nd, 2020).

14. Li Q, Guan X, et. al. **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia**. *N Engl J Med*. 2020 Jan 29. DOI: 10.1056/NEJMoa2001316 Available at: https://www.nejm.org/doi/full/10.1056/NEJMoa2001316

15. Nishiura H, Linton NM, Akhmetzhanov AR. Int J Infect Dis. 2020 Apr;93:284-286. doi: 10.1016/j.ijid.2020.02.060

16. Ghani AC, Donnelly CA, Cox DR, Griffin JT, Fraser C, Lam TH, et al. **Methods for estimating the case fatality ratio for a novel, emerging infectious disease.** *Am J Epidemiol*. 2005; 162: 479-486

17. Nishiura H, Klinkenberg D, Roberts M, Heesterbeek JA. **Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic**. *PLoS One*. 2009;4(8):e6852. doi: 10.1371/journal.pone.0006852.

18.Tsuzuki S, Lee H, Miura F, Chan YH, Jung SM, Akhmetzhanov AR, Nishiura H. **Dynamics of the pneumonic plague epidemic in Madagascar, August to**

October 2017. *Euro Surveill*. 2017;22(46). doi: 10.2807/1560-7917.ES.2017.22.46.17-00710.

19. Mizumoto K, Chowell G. **Estimating the risk of 2019 Novel Coronavirus death during the course of the outbreak in China, 2020**. *Emerg Infect Dis*. 2020; 26.

20. Gamerman, D. & Lopes, H. F. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Interference**. 2nd edn (Chapman & Hall/CRC, 2006).

21. Gelman, A. & Rubin, D. B. **Inference from iterative simulation using multiple sequences**. *Stat Sci* **7:**457-472, doi:10.1214/ss/1177011136 (1992).

22. Read JM, Bridgen JR, Cummings DA, Ho A,Jewell CP. **Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions**. *medRxiv*. doi: https://doi.org/10.1101/2020.01.23.20018549

23. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, Ferguson NM. **Report 3: Transmissibility of 2019-nCoV**. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf

24. Mizumoto K, Chowell G. **Transmission potential of the novel coronavirus (COVID-19) onboard the Diamond Princess Cruises Ship, 2020.** *Infect Dis Model*. 2020. 264-270

25. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, Chowell G. **Real-time monitoring the transmission potential of COVID-19 in Singapore, February 2020**. *medRxiv*. doi: https://doi.org/10.1101/2020.02.21.20026435

26. Shim E, Tariq A, Choi W, Lee Y, Chowell G. **Transmission potential of COVID-19 in South Korea.** *medRxiv*. doi:https://doi.org/10.1101/2020.02.27.20028829

27. Chowell G, Abdirizak F, Lee S, Lee J, Jung E, Nishiura H, Viboud C. **Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study.** *BMC Med.* 2015;**13**:210. doi: 10.1186/s12916-015-0450-0.

28. Abdirizak F, Lewis R, Chowell G. **Evaluating the potential impact of targeted vaccination strategies against severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) outbreaks in the healthcare setting.** *Theor Biol Med Model*. 2019;**16(1)**:16. doi: 10.1186/s12976-019-0112-6.

29. Blake Essig, Brent Swails, Yoko Wakatsuki and Ben Westcott, CNN. **Top Japanese government adviser says Diamond Princess quarantine was flawed**. Updated 0708 GMT (1508 HKT) February 27, 2020. https://edition.cnn.com/2020/02/27/asia/japan-diamond-princess-quarantine-crew-intl-hnk/index.html

30. Da-hae P, Dam-eun S, Jae-gu K. HANKYOREH. **The reasons why transmission is so prevalent among Shincheonji members**. Mar 2, 2020. http://english.hani.co.kr/arti/english_edition/e_national/930749.html [Accessed Mar/10, 2020]

31. Mizumoto K, Kagaya K, Zarebski A, Chowell G. **Estimating the Asymptomatic Proportion of 2019 Novel Coronavirus onboard the Princess Cruises Ship, 2020**. *Euro Surveill*. 2020; 25.

32. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. **Estimates of the severity of coronavirus disease 2019: a model-based analysis.** *Lancet Infect Dis*. 2020. pii: S1473-3099(20)30243-7.

33. Russell TW, Hellewell J, Jarvis CI, et al. **Estimating the Infection and Case Fatality Ratio for Coronavirus Disease (COVID-19) Using Age-Adjusted Data From the Outbreak on the Diamond Princess Cruise Ship, February**

2020. *Euro Surveill*. 2020; 25(12): 2000256. doi: 10.2807/1560-7917.ES.2020.25.12.2000256.

34. Wu X, Fu B, Chen L, Feng Y. **Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China**. Journal of Medical Virology. https://doi.org/10.1002/jmv.25904

35. Giugliano F. **Mass Coronavirus Antibody Tests Have Serious Limits.** *Bloomberg Opinion.* April 24, 2020. [Accessed on May 30th, 2020)]

36. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. **Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China**. *JAMA*. 2020. doi: 10.1001/jama.2020.1585.

37. Mizumoto K, Endo A, Chowell G, Miyamatsu Y, Saitoh M, Nishiura H. **Real-time characterization of risks of death associated with the Middle East respiratory syndrome (MERS) in the Republic of Korea, 2015.** *BMC Med.* 2015;**13**:228. doi: 10.1186/s12916-015-0468-3.

38. Mizumoto K, Saitoh M, Chowell G, Miyamatsu Y, Nishiura H. **Estimating the risk of Middle East respiratory syndrome (MERS) death during the course of the outbreak in the Republic of Korea, 2015**. *Int J Infect Dis.* 2015;**39**:7-9. 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

Figures

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

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

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

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

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

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

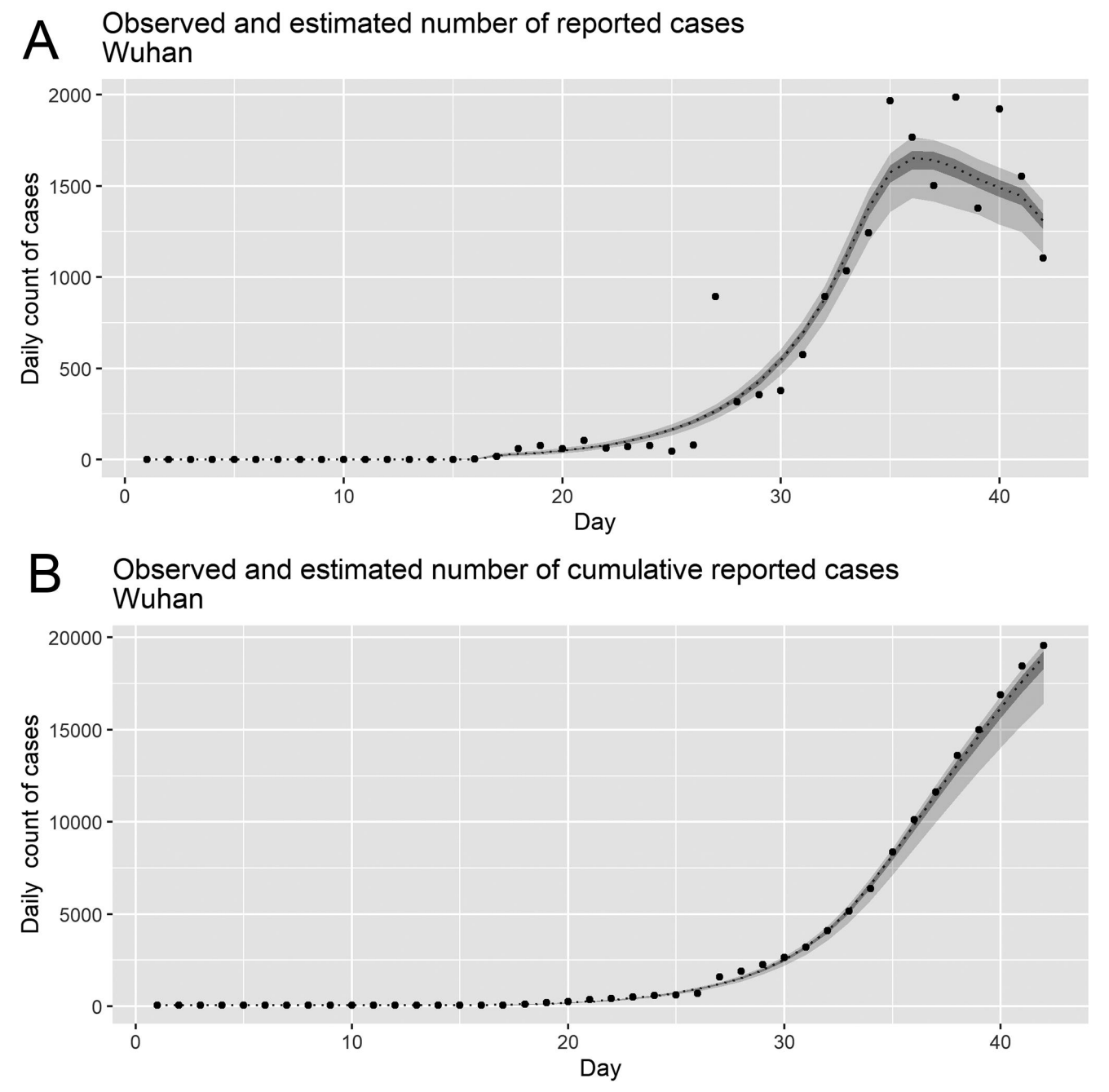
Tables

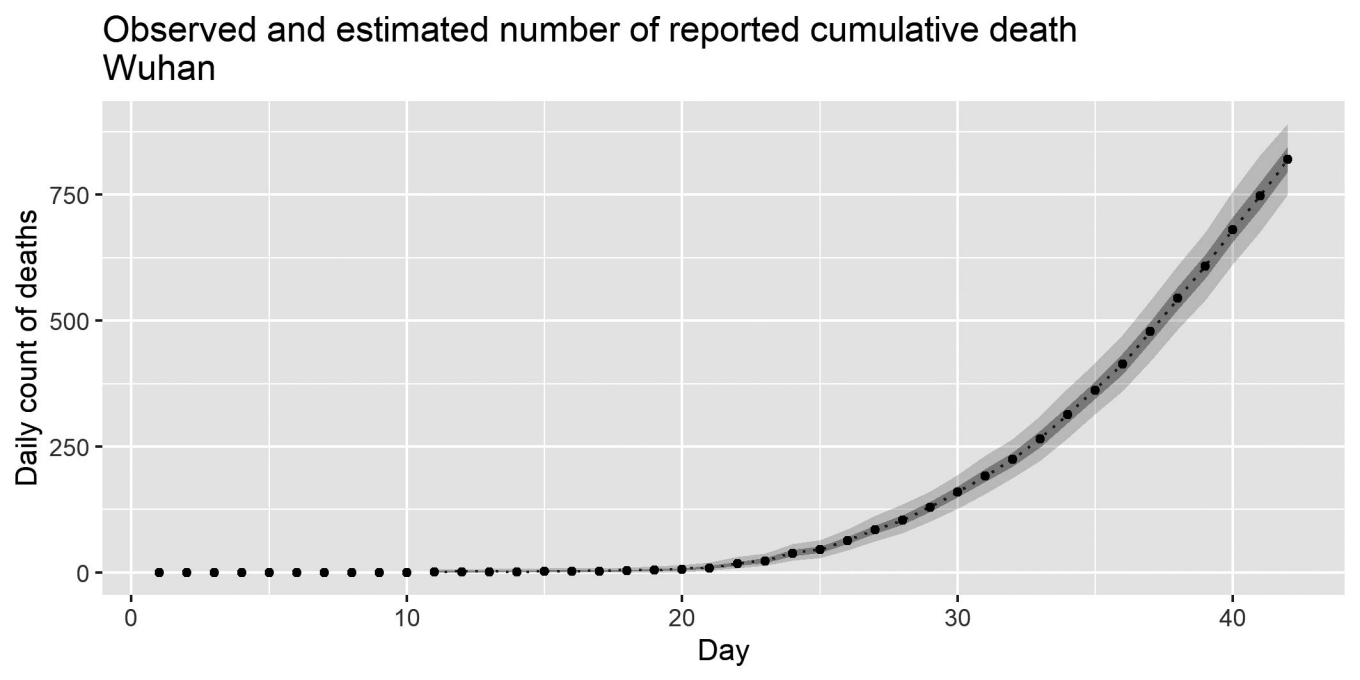
Table 1 – Death risk by COVID-19 in Wuhan City, China, 2020 (As of February 12, 2020)

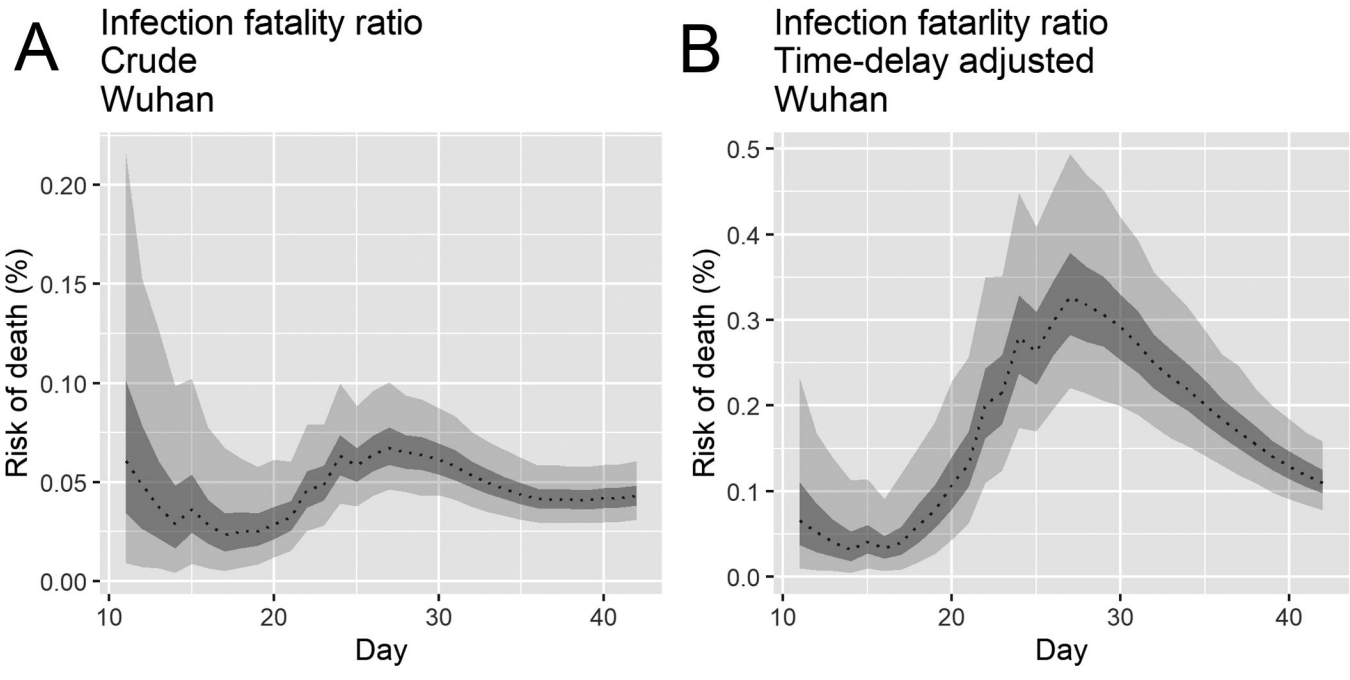
Death Risk	Latest estimate	Range of median estimates
Crude CFR (Observed)	4.06%	2.0 – 9.0%
Crude CFR (Estimated)	4.3% (95%CrI‡: 3.9 – 5.0%)	3.4 –7.1%
Time delay adjusted CFR	12.2% (95%CrI: 11.4 – 13.1%)	4.0 – 34.5%
Crud IFR	0.04% (95%CrI: 0.03 – 0.06%)	0.02 – 0.07%
Time delay adjusted IFR	0.12% (95%CrI: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

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

‡Upper and lower 95% credibility interval


A
Observed and estimated number of reported cases
Wuhan
Daily count of cases
2000
1500
1000
500
0
0
10
20
30
40
Day
B
Observed and estimated number of cumulative reported cases
Wuhan
Daily count of cases
20000
15000
10000
5000
0
0
10
20
30
40
Day


Observed and estimated number of reported cumulative death
Wuhan
Daily count of deaths
750
500
250
0
0
10
20
30
40
Day


A
Infection fatality ratio
Crude
Wuhan
Risk of death (%)
0.00
0.05
0.10
0.15
0.20
10
20
30
40
Day
B
Infection fatarlity ratio
Time-delay adjusted
Wuhan
Risk of death (%)
0.0
0.1
0.2
0.3
0.4
0.5
10
20
30
40
Day