

RESEARCH ARTICLE

Open Access



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

Kenji Mizumoto<sup>1,2,3\*</sup> , Katsushi Kagaya<sup>4</sup> and Gerardo Chowell<sup>3</sup>

## Abstract

**Background:** Since the first cluster of cases was identified in Wuhan City, China, in December 2019, coronavirus disease 2019 (COVID-19) rapidly spreads globally. Scientists have made strides in estimating key transmission and epidemiological parameters. In particular, accumulating evidence points to a substantial fraction of asymptomatic or subclinical infections, which influences our understanding of the transmission potential and severity 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 23 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 1,906,634 (95% CrI 1,373,500–2,651,124) 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 substantially smaller than the crude CFR estimated at 4.06%.

(Continued on next page)

\* Correspondence: [mizumoto.kenji.5a@kyoto-u.ac.jp](mailto:mizumoto.kenji.5a@kyoto-u.ac.jp)

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

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

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**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 modeling approach that is suitable to infer epidemiological parameters with quantified uncertainty from partial observations collected by surveillance systems. Our estimate of time–delay adjusted IFR falls in the range of the median IFR estimates based on multiple serological studies conducted in several areas of the world.

**Keywords:** Epidemic, Transmissibility, Statistical 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 12, 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 82,918 and 4633, respectively, as of May 10, 2020 [3].

Because the morbidity and mortality burden associated with the novel coronavirus has disproportionately 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 23, 2020. Indeed, out of the 82,918 COVID-19 cases reported in China, 50,339 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 the date of symptom onset or reporting date from several sources [3, 7, 8]. Our analysis relies on epidemiological data reported prior to February 11, 2020, because of the change in case definition that was announced on February 12, 2020 [9]. As of February 11, 2020, a total of 19,559 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 [10]. As of February 11, 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).

### 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  $G(\cdot)$  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  $t$  and 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 modeled by introducing a parameter  $\delta_1$ , which is given by

$$\delta_1 = \begin{cases} 1 & \text{otherwise} \\ \beta_1 & \text{if } t = \text{period}_1 \\ \beta_2 & \text{if } t = \text{period}_2 \end{cases}$$

where  $\text{period}_1$  and  $\text{period}_2$  represent the corresponding period from January 23 to February 2, 2020, and from February 3 to February 11, 2020, respectively. January 23, 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 epicenter of the coronavirus (COVID-19) to mitigate transmission [11]. Furthermore, we evenly divide the interval into two periods to incorporate the time-dependent effects on  $R$  using the parameters  $\beta_1$  and  $\beta_2$  which scale the effects of the intervention, taking values smaller than 1.

To account for the probability of occurrence,  $\theta$  [12], we assume that the number of observed cases on day  $t$ ,  $h(t)$ , occurred according to a Bernoulli sampling process, with the expected values  $E(c_i; H_{t-1})$ , where  $E(c_i; H_{t-1})$  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:

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

Further, we model the time-dependent variation in the reporting probability. This time dependence was modeled by introducing a parameter  $\delta_2$ , which is given by

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

where  $\text{period}_3$  and  $\text{period}_4$  represent the corresponding periods from the start of our study period to January 16 and from January 17 to January 22, respectively, while  $\alpha_1$  and  $\alpha_2$  scale the extent of the reporting probability (where  $\alpha_1$  and  $\alpha_2$  are 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

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

We assume the incidence,  $h(t)$ , is the result of the binomial sampling process with the expectation  $E[h]$ . The likelihood function for the time series of observed cases that we employ to estimate the effective reproduction number and other relevant parameters is given by:

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

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

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

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

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

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

### Infection fatality ratio

Crude CFR and crude IFR are defined as the number of cumulative deaths 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].

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

where  $c_t$  represents the number of new cases with reported day  $t$  and  $D_{it}$  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  $L = L_1 L_2 L_3$ , 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 was evaluated using the potential scale reduction statistic [20, 21]. Estimates and 95% credibility intervals for these estimates are based on the posterior probability distribution of each parameter and based on the samples drawn from the posterior distributions. All statistical analyses were conducted in R version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria) using the “rstan” package.

**Results**

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

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

We conducted sensitivity analyses to examine how varying the mean serial interval between 5.0 and 7.0 days affects our  $R$  estimates.  $R$  estimates are sensitive to 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 18,967 (95% CrI 16,428–19,680), while the actual number of reported laboratory-confirmed cases during our study period is 19,559 as of February 11, 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 1,906,634 (95% CrI 1,373,500–2,651,124).

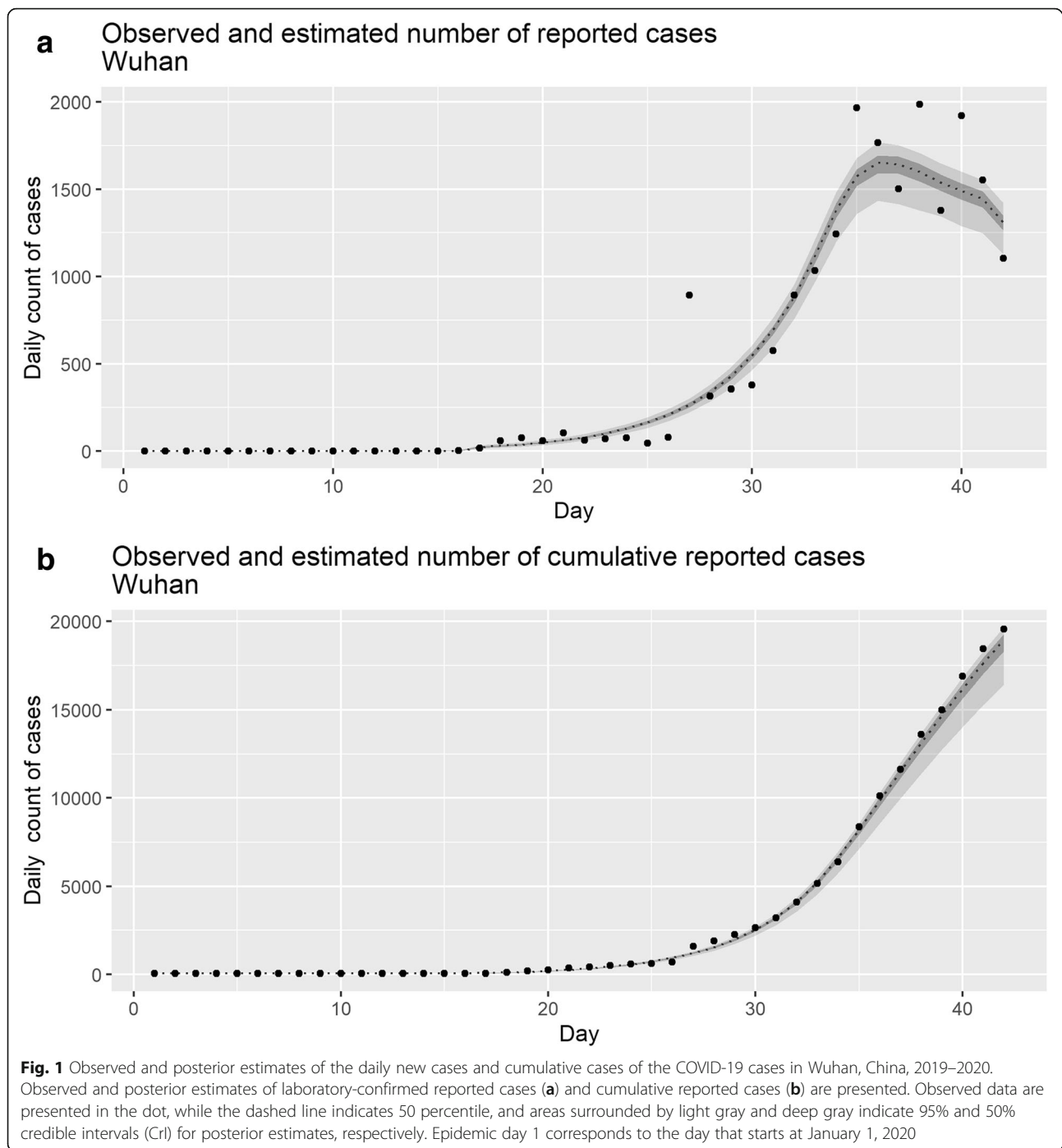
The observed and posterior estimates of the cumulative number of deaths from COVID-19 in Wuhan are displayed in Fig. 2, and model-based posterior estimates of the cumulative number of deaths are 821 (95% CrI 751–892), while the actual number of reported deaths is 820. The estimated temporal variation in the death risk caused by COVID-19 in Wuhan, China, 2019–2020 is shown in Fig. 3 and Figure S2. The observed and posterior estimates of the crude CFR in Wuhan City is presented in Fig. 2a, while the observed and posterior estimates of time–delay-adjusted CFR is shown in Fig. 2b. Furthermore, Fig. 3a and b illustrates time–delay not-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 Fig. 3c 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**

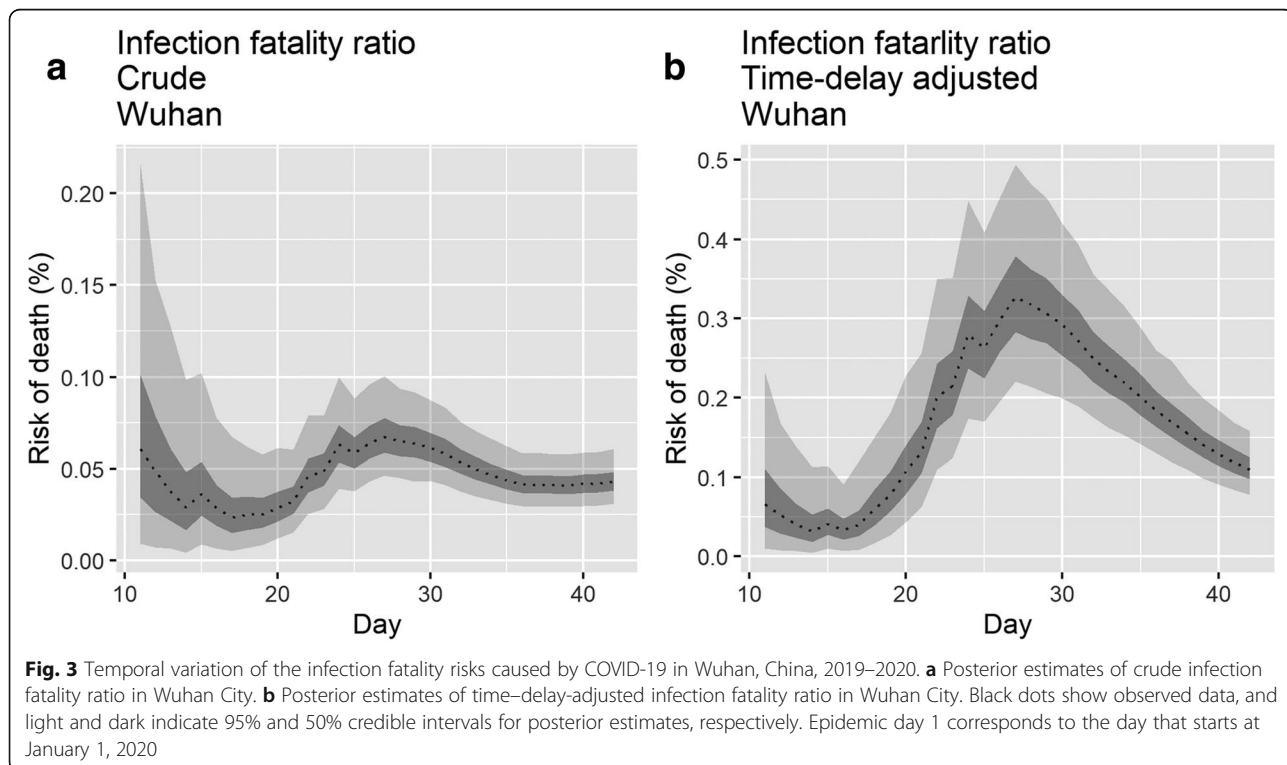
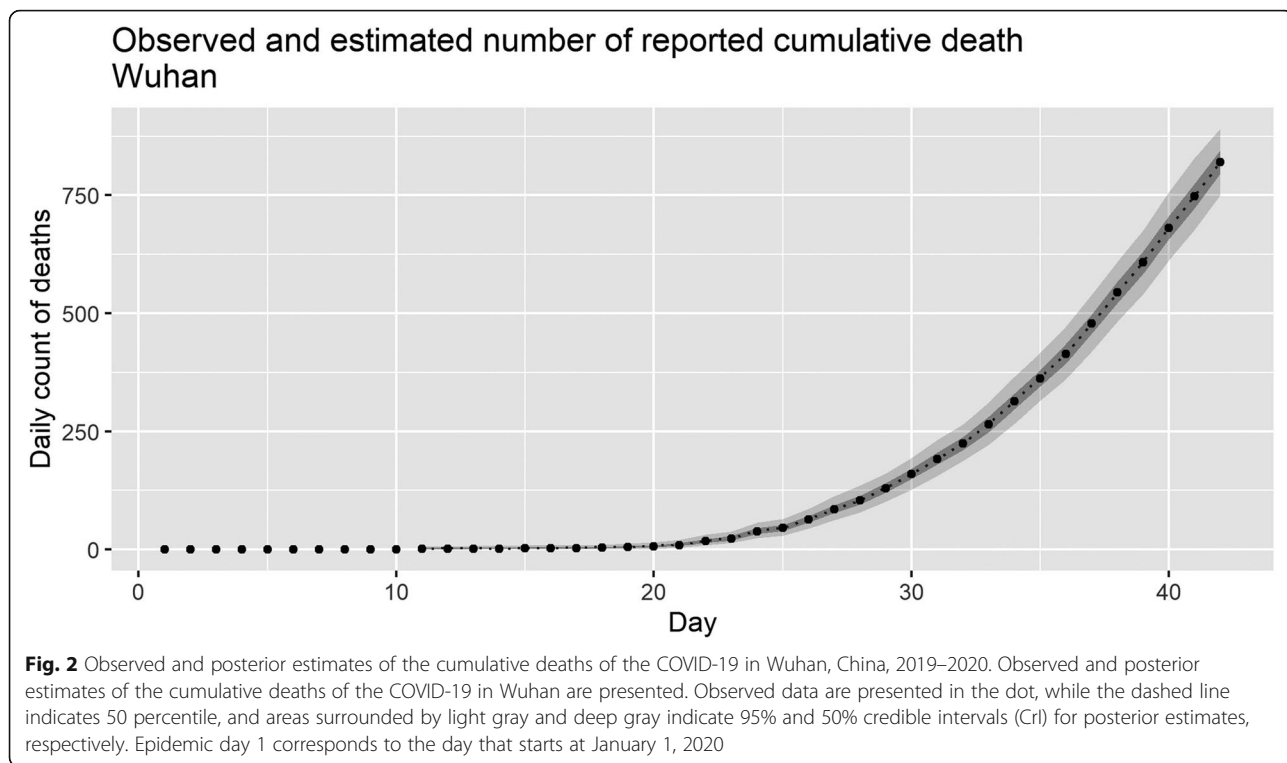
In this study, we derived estimates of the transmissibility and virulence of COVID-19 in Wuhan City, China, by reconstructing the underlying transmission dynamics using multiple data sources. Applying dynamic modeling, the reproduction number, death risks, and probabilities of occurrence and reporting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

Our posterior estimates of basic reproduction number ( $R$ ) in Wuhan City, China, in 2019–2020 are 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 23 to February 1, 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 2 to February 11, 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 1,906,634 (95% CrI 1,373,500–2,651,124) 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  $\sim 11$  [24]. Further, these estimates are higher than recent mean  $R$  estimates derived from the growth rates of the COVID-19 outbreaks in Singapore ( $R \sim 1.1$ ) [25] and Korea ( $R \sim 1.5$ ) [26].



**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 <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%
Crude 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

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

The sustained high  $R$  values in Wuhan City even after the lockdown and mobility restrictions suggest 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 3700 passengers as of February 28, 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 are estimated to be 0.04% (95% CrI 0.03–0.06%) and 0.12% (95% CrI 0.08–0.17%), respectively, which are substantially smaller than the crude CFR estimated at 4.06%. Our estimates are also lower compared to other recent estimates of the infection fatality ratio at 0.66% (95% CrI 0.39–1.33) and 0.6% (95% CI 0.2–1.3) based on data from China and the Diamond Princess ship, respectively [32, 33]. Several data and methodological differences can explain these differences, which we list in Table S2. For instance, Verity et al. conduct an age adjustment based on the data of age-stratified COVID-19 deaths from mainland China, assume a constant attack rate by age, and adjust for demographic structure. Our IFR estimates will be compared with estimates emerging from ongoing several mass serological studies. Our estimate of time–delay adjusted IFR of 0.12% (95% CrI 0.08–0.17%) falls in the range of mean IFR estimates based on 23 serological studies (0.02 to 0.78% (median 0.25%)) [34] and is consistent with an IFR estimate at 0.16% based on the 452 residents evacuated from Hubei [35], whereas it is smaller than an

estimate of 0.35% based on a serological study conducted in Wuhan [36], which is probably slightly overestimated since its target population includes hospitalized patients. Yet, it is important that the external validity of these serological studies, e.g., whether the results can offer a representative picture of the population in the region where they are performed, is closely examined when the results are employed to derive estimates of the infection fatality ratio as discussed elsewhere [34, 36, 37]. 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 [38]. 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 [39, 40].

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 [17, 41, 42].

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 are 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 16, 2020) and about the same during the 2nd period (January 17–22, 2020), relative to that estimated after January 23, 2020.

Our results are not free from the limitations. First, our methodology aims to capture the underlying transmission dynamics using multiple data sources. 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 is 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 underascertainment 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 [12, 43–48]. As CFR is influenced 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 modeling 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.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12916-020-01691-x>.

**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 : Figure 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 gray and deep gray 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 : Figure 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.

## Abbreviations

CFR: Case fatality ratio; IFR: Infection fatality ratio; MERS: Middle East respiratory syndrome; SARS: Severe acute respiratory syndrome

## Acknowledgements

Not applicable.

## Authors' contributions

KM and GC conceived the early study idea. KM and KK built the model. KM implemented the 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.

## Funding

KM acknowledges support from the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant Number 18K17368 and 20H03940), from the Leading Initiative for Excellent Young Researchers from the Ministry of Education, Culture, Sport, Science & Technology of Japan, and from the Japan Science and Technology Agency (JST) as part of J-RAPID (grant JPMJSC). 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.

## 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.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

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

Received: 11 February 2020 Accepted: 1 July 2020

Published online: 15 July 2020

## References

1. Jon Cohen. Mining coronavirus genomes for clues to the outbreak's origins. *Science*. 2020. <https://www.sciencemag.org/news/2020/01/mining-coronavirus-genomes-clues-outbreak-s-origins>. Accessed 3 Feb 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 10<sup>th</sup>].
3. The State Council, The People's Republic of China. <http://www.gov.cn/>. [cited 2020 May 10<sup>th</sup>].
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. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
6. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung SM, Yuan B, Kinoshita R, Nishiura H. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med*. 2020;9(2):538.



7. Health Commission of Hubei Province, China. <http://wjw.hubei.gov.cn/>. [cited 2020 Feb 7<sup>th</sup>].
8. Health Commission of Wuhan City, Hubei Province, China <http://wjw.hubei.gov.cn/>. [cited 2020 Feb 7<sup>th</sup>].
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 29 Feb 2020. [in Chinese].
10. Ministry of Health, Labour and Welfare, Japan. <https://www.mhlw.go.jp/index.html> [in Japanese]. Accessed 7 Feb 2020.
11. 2020 Hubei lockdowns, Wikipedia. [https://en.wikipedia.org/wiki/2020\\_Hubei\\_lockdowns](https://en.wikipedia.org/wiki/2020_Hubei_lockdowns). Accessed 7 Feb 2020.
12. Li R, Weiskittel AR, 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;41:2077–89.
13. Northeastern University. Laboratory for the modeling of biological and socio-technical systems, 2020. Available online: <https://www.mobs-lab.org/2019ncov.html>. Accessed 22 Jan , 2020.
14. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020;382:1199–207. <https://doi.org/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;93:284–6. <https://doi.org/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;162:479–86.
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. <https://doi.org/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). <https://doi.org/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(6):1251–6.
20. Gamerman, D. & Lopes, H. F. Markov Chain Monte Carlo: stochastic simulation for Bayesian inference. 2nd edn. London: CRC Press (Chapman & Hall); 2006.
21. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457–72. <https://doi.org/10.1214/ss/1177011136>.
22. Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv*. <https://doi.org/10.1101/2020.01.23.20018549>.
23. Imai N, Cori A, Dorigatti I, Baguein 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>. Accessed 12 Mar 2020.
24. Mizumoto K, Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the Diamond Princess Cruises Ship, 2020. *Infect Dis Model*. 2020;5:264–70.
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, March 2020. *BMC Med*. 2020;18(1):166. <https://doi.org/10.1186/s12916-020-01615-9>.
26. Shim E, Tariq A, Choi W, Lee Y, Chowell G. Transmission potential and severity of COVID-19 in South Korea. *Int J Infect Dis*. 2020;93:339–44.
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. <https://doi.org/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. <https://doi.org/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>. Accessed 12 Mar 2020.
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](http://english.hani.co.kr/arti/english_edition/e_national/930749.html). Accessed 10 Mar 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(10):2000180.
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;20(6):669–77.
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. <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256>.
34. Ioannidis J. The infection fatality rate of COVID-19 inferred from seroprevalence data. *medRxiv* 2020.05.13.20101253.
35. KKW T, Cheng VCC, Cai JP, Chan KH, Chen LL, Wong LH, et al. *Lancet Microbe*. [https://doi.org/10.1016/S2666-5247\(20\)30053-7](https://doi.org/10.1016/S2666-5247(20)30053-7).
36. Wu X, Fu B, Chen L, Feng Y. Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China. *J Med Virol*. <https://doi.org/10.1002/jmv.25904>.
37. Giugliano F. Mass coronavirus antibody tests have serious limits. *Bloomberg Opin*. 2020. Accessed 30 May 2020.
38. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020;323(11):1061–9. <https://doi.org/10.1001/jama.2020.1585>.
39. 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. <https://doi.org/10.1186/s12916-015-0468-3>.
40. 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. <https://doi.org/10.1016/j.ijid.2015.08.005>.
41. Wong JY, Wu P, Nishiura H, Goldstein E, Lau EH, Yang L, et al. Infection fatality risk of the pandemic A(H1N1)2009 virus in Hong Kong. *Am J Epidemiol*. 2013;177(8):834–40.
42. Presanis AM, De Angelis D, New York City Swine Flu Investigation Team, Hagy A, Reed C, Riley S, Cooper BS, et al. The severity of pandemic H1N1 influenza in the United States, from April to July 2009: a Bayesian analysis. *PLoS Med*. 2009;6(12):e1000207.
43. The Guardian. Black people four times more likely to die from Covid-19, ONS finds <https://www.theguardian.com/world/2020/may/07/black-people-four-times-more-likely-to-die-from-covid-19-ons-finds>. Accessed 29 May 2020.
44. Mizumoto K, Dahal S, Chowell G. Spatial variability in the risk of death from COVID-19 in 20 regions of Italy. *Int J Tuberc Lung*. 2020; in press.
45. Shim E, Mizumoto K, Choi W, et al. Estimating the risk of COVID-19 death during the course of the outbreak in Korea, February–May, 2020. *J Clin Med*. 2020;9(6):E1641.
46. Dahal S, Mizumoto K, Chowell G. Investigating spatial variability in COVID-19 pandemic severity across 19 geographic areas, Spain, 2020. *medRxiv*. 2020. 04.14.20065524.
47. Undurraga EA, Chowell G, Mizumoto K. Case fatality risk by age from COVID-19 in a high testing setting in Latin America: Chile, March–May, 2020. *medRxiv*. 2020.05.25.20112904.
48. Munayco C, Chowell G, Tariq A, Undurraga EA, Mizumoto K. Risk of death by age and gender from COVID-19 in Peru, March–May, 2020. *medRxiv*. 2020.06.14.20123315.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.