medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Real time estimation of the risk of death from novel coronavirus (2019-nCoV) infection: Inference using exported cases

Sung-mok Jung[1], Andrei R. Akhmetzhanov[1], Katsuma Hayashi[1], Natalie M. Linton[1,*], Yichi Yang[1], Baoyin Yuan[1], Tetsuro Kobayashi[1,*], Ryo Kinoshita[1], Hiroshi Nishiura[1,2,†]

[1] *Graduate School of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan*

[2] *CREST, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan*

[*] These authors equally contributed to this work.

[†] Correspondence: nishiurah@med.hokudai.ac.jp; Tel.: +81-11-706-5066

Abstract

The exported cases of 2019 novel coronavirus (2019-nCoV) infection who were confirmed in other countries provide a chance to estimate the cumulative incidence and confirmed case fatality risk (cCFR) in China. Knowledge of the cCFR is critical to characterize the severity and understand pandemic potential of 2019-nCoV in the early stage of epidemic. Using the exponential growth rate of the incidence, the present study statistically estimated the cCFR and the basic reproduction number, i.e., the average number of secondary cases generated by a single primary case in a naïve population. As of 24 January 2020, with 23 exported cases, and estimating the growth rate from 8 December 2019 (scenario 1) and using the data since growth of exported cases (scenario 2), the cumulative incidence in China was estimated at 5433 cases (95% confidence interval (CI): 3883, 7160) and 17780 cases (95% CI: 9646, 28724), respectively. The latest estimates of the cCFR were 4.6% (95% CI: 3.1-6.6) for scenario 1 and 7.7% (95% CI: 4.9-11.3%) for scenario 2, respectively. The basic reproduction number was estimated to be 2.2 (95% CI: 2.1, 2.3) and 3.7 (95% CI: 3.1, 4.3) for scenarios 1 and 2, respectively. Based on the results, we note that current 2019-nCoV epidemic has a substation potential to cause a pandemic. The proposed approach can provide insights into early risk assessment using only publicly available data.

Keywords: mortality; censoring; travel; migration; importation; emerging infectious diseases

1 Introduction

Since 8 December 2019, clusters of pneumonia cases of unknown etiology have emerged in Wuhan City, Hubei Province, China [1]. Virological investigation suggests the causative agent of this pneumonia to be a novel coronavirus (2019-nCoV) [2]. As of 27 January 2020, a total of 4515 cases including 106 deaths have been confirmed [3]. There were also 41 cases of 2019-nCoV outside of mainland China, in other Asian countries, the United States, France, Australia, and Canada.

A local market selling seafood and wildlife in Wuhan was visited by many cases in the initial cluster, indicating that a common-source zoonotic exposure may have been the main mode of transmission [4]. However, even after shutting down the market the number of cases continued to grow across China and several instances of household transmission were reported [5]. It is now speculated that sustained human-to-human transmission aided in establishment of the epidemic [6] and that reported case counts greatly underestimated the actual number of infections in China [7].

Early assessment of the severity of infection and transmissibility can help quantify the pandemic potential of 2019-nCoV and anticipate the likely number of deaths by the end of the epidemic. One important epidemiological measure of severity is case fatality risk (CFR), which defines the risk of

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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

death among cases. However, the actual CFR tends to be underestimated due to real-time nature of the growth of fatal cases. For example, during the early stage of an epidemic failing to right-censor cases with respect to the time delay from illness onset to death may lead to underestimation of the CFR. This is because death due to infection may yet occur following case identification [8–10]. Moreover, when the CFR denominator only includes confirmed cases it is referred to as the cCFR [9] and may seriously overestimate the actual CFR among all infected individuals as a result of under-ascertainment of infections in the population. Nonetheless, the cCFR is still a valuable measure of the upper bound of the symptomatic CFR (sCFR) among all symptomatic cases, particularly in circumstances of high uncertainty such as the emergence of a new human pathogen (i.e. 2019-nCoV).

Using the growth rate of the estimated cumulative incidence from exported cases and accounting for the time from illness onset to death, the present study aims to estimate the cCFR and cumulative incidence of 2019-nCoV in real-time. The growth rate is also translated to an estimate of the transmissibility.

2 Methods

2.1 Epidemiological data

The information on exported 2019-nCoV cases who were confirmed in other countries and deaths due to 2019-nCoV infection in China were retrieved from the first announcement date of current outbreak (i.e.,∼31 December 2019) through 24 January 2020. All data were collected either from government websites or media quoting government announcements. Reporting dates and dates of illness onset were collected from 23 exportation events and dates of illness onset and death among 41 deceased cases in China were also retrieved from data sources.

2.2 Estimation of the delay distributions

The observed incidence $i(t)$ by date of illness onset t is modeled by an exponential growth model with the rate r: $i(t) = i_0 e^{rt}$, where i_0 is the expected number of infected cases at time $t = 0$. The cumulative incidence $I(t)$ is an integral of $i(t)$ over the time interval from zero to t that can be written as: $I(t) = i_0(e^{rt} - 1)/r$. The cumulative incidence is adjusted to the date of report by the factor u dependent on the parameters of the delay distribution. Assuming that the latter is described by a gamma distribution with the shape a and inverse scale b, u can be defined by a function of the growth rate r and those parameters a and b [9]:

$$u(r, a, b) = \left(1 + \frac{r}{b}\right)^{-a}. \tag{1}$$

If a_d and b_d are the shape and the inverse scale parameters for a distribution of time from illness onset to death, the adjusted cumulative incidence obeys $u(r, a_d, b_d)I(t)$. The cumulative number of deaths $D(t)$ reported by date t will be the result of Binomial sampling:

$$D(t) \sim \text{binom}\big(\text{size} = u(r, a_d, b_d)I(t),\ \text{prob} = \text{CFR}(t)\big)\,. \tag{2}$$

If a_e and b_e are the shape and inverse scale of the time from illness onset to report of the exported case, we adjust the cumulative incidence by factor $u(r, a_e, b_e)$, and define the number of exported cases $E(t)$

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to be sampled from another Binomial distribution:

$$E(t) \sim \text{binom}\big(\text{size} = u(r, a_e, b_e)I(t),\ \text{prob} = p\big)\,, \tag{3}$$

where p describes the probability of finding a traveler from Wuhan among all travelers from China subject to detection time window of the virus $T = 12.5$ days [7]. The total volume of inbound passengers from China $M = 5.56$ million of passengers per year, the fraction of Wuhan travelers $\phi = 2.1\%$, and the population of Wuhan $N = 11$ million, then the probability p is given by:

$$p = \frac{\phi M T}{365 N} \approx 0.0036\,. \tag{4}$$

2.3 Statistical inference

Firstly, we fit two delay distributions of the time from illness onset to death $\Delta T_{d,k}$ ($k \leq K_d$), and time from illness onset to report $\Delta T_{e,k}$ ($k \leq K_e$) to two gamma distributions. In each instance, we define the log-likelihoods as:

$$\log L_\circ\big(\theta_\circ = \{a_\circ, b_\circ\} \,|\, \Delta T_{\circ,k}\big) = \sum_k \log\big(\text{gamma}\big(\Delta T_{\circ,k} \,|\, \text{shape} = a_\circ, \text{scale} = b_\circ\big)\big), \tag{5}$$

with $\circ = \{$"*d*", "*e*"$\}$ and separately maximize them to find the mean values of the parameters $\tilde{\theta}_\circ = \{\tilde{a}_\circ, \tilde{b}_\circ\}$, used in the following step.

Second, we fit the observed counts of exported cases and deaths by considering two other likelihoods respectively to each process:

$$\log L_E\big(\{r, i_0\} \,|\, \{E(t), t_e \leq t \leq T\}\big) = \sum_{t=t_e}^{T} \log\Big(\text{binom}\big(E(t)|\text{size} = u(r, \tilde{a}_e, \tilde{b}_e)I(t),\ \text{prob} = p\big)\Big)\,, \tag{6}$$

$$\log L_D\big(\{r, i_0, \text{CFR(t)}\} \,|\, \{D(t), t_d \leq t \leq T\}\big) = \sum_{t=t_d}^{T} \log\Big(\text{binom}\big(D(t)|\text{size} = u(r, \tilde{a}_d, \tilde{b}_d)I(t), \text{prob} = \text{CFR(t)}\big)\Big), \tag{7}$$

where t_e and t_d are time moments of observing first exportation event and first death, respectively. The total log-likelihood as a sum of two log-likelihoods:

$$\log L_\Sigma\big(\{r, i_0, \text{CFR(t)}\} \,|\, D(t), E(t)\big) = \log L_E\big(\{r, i_0\} \,|\, E(t))\big) + \log L_D\big(\{r, i_0, \text{CFR(t)}\} \,|\, D(t)\big)\,, \tag{8}$$

is then maximized to determine the best-fit parameters $\{r, i_0, \text{CFR(t)}\}$.

We consider two main possible scenarios aimed to distinguish the fits relatively balanced for both low and high incidence counts and the fits aimed for better coincidence with the data at higher counts. Whereas, one of the parameters i_0 is fixed at one on the illness onset date of the first 2019-nCoV confirmed case (i.e., 8 December 2019) in the former "Scenario 1", all parameters $\{r, i_0, \text{CFR}(t)\}$ are variable for the fitting in the latter "Scenario 2".

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The basic reproduction number of 2019-nCoV, R_0, the average number of secondary cases generated by a single primary case in a fully susceptible population, was calculated as:

$$R_0 = 1 + rS, \tag{9}$$

where the r stands for the estimated growth rate from each estimation scenarios and S describe the mean serial interval of 2019-nCoV. However, as serial interval of 2019-nCoV is unknown information as of 24 January 2020, the serial interval of SARS was derived from empirical data with 182 confirmed cases during 2003 Singapore SARS outbreak (mean $= 8.4$ days, SD $= 3.8$ days) [11] was used for assumed serial interval of 2019-nCoV.

In all analyses, we employ Markov chain Monte-Carlo (MCMC) simulations (20 chains, 5000 samples each and same length for the tuning stage) with the No-U-Turn sampler (NUTS) and using the Python PyMC3 package. However, it should be noted that one of the input variables for Binomial distribution $u(r, a_e, b_e)I(t)$ or $u(r, a_d, b_d)I(t)$ is modeled as a continuous variable in our approach. We conveniently used the transformation of the discrete Binomial distribution to its continuous approximation by using the gamma distribution and matching first two moments [12]. In addition, the joint estimation of all parameters $\{\theta_\Sigma, \theta_e, \theta_d\}$ was avoided due to heterogeneity in the aggregated data. In this case, all three likelihoods would be incomparable in their relative weight. See a similar issue discussed in [9]. Instead we implemented a sequential fitting: first considering only the likelihoods L_e and L_d, second the likelihood L_Σ with the use of mean values of the estimated parameters obtained in the previous round. Finally, we verified the fit obtained by MCMC by obtaining pointwise estimates using the maximum likelihood estimation method with confidence intervals derived from the profile-likelihood. We found that both approaches were at complete agreement in their results.

3 Results

Figures 1A and 1B show the mean and standard deviation (SD) of the time from illness onset to reporting and death, respectively. Employing the gamma distribution, the mean time from illness onset to reporting was estimated to be 5.1 days (95% CI: 3.5, 7.5) and the mean time from illness onset to death was estimated as 15.2 days (95% CI: 13.1, 17.7).

Subsequently, the cumulative incidence was estimated from exported case data by fitting an exponentially growing incidence curve for both scenarios 1 and 2 (Figure 2). As of 24 January 2020, a total of 23 exported cases were observed, and the cumulative incidence in China was estimated as 5433 cases (95% CI: 3883, 7160) in scenario 1 and 17780 cases (95% CI: 9646, 28724) in scenario 2. Table 1 shows the real time update of the estimated cumulative incidence. The exponential growth rates (r), derived from the growth rate of cumulative incidence was estimated at 0.14 per day (95% CI: 0.13, 0.15) and 0.32 per day (95% CI: 0.25, 0.39) in scenarios 1 and 2, respectively.

Figures 2C and 2D show the estimated cCFR which accounted for the time delay from illness onset to death under scenarios 1 and 2, respectively. The cCFR as on 24 January 2020 when there were 41 confirmed deaths reported was estimated at 4.6% (95% CI: 3.1-6.6) for scenario 1 and 7.7% (95% CI: 4.9-11.3%) for scenario 2, respectively. We estimated the basic reproduction number for the 2019-nCoV infection, using the estimated exponential growth (r) and accounting for possible variations in the mean serial interval (Figure 3) [7]. Assuming that the mean generation time was 8.4 days, the basic reproduction number was estimated at 2.19 (95% CI: 2.11, 2.26) and 3.70 (95% CI: 3.07, 4.27) for scenarios

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

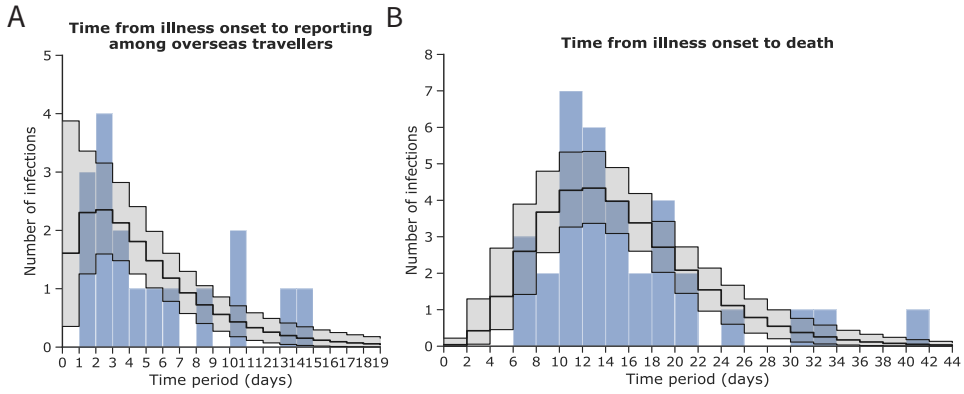


Figure 1. Estimates of the mean and standard deviation (SD) of the time from illness onset to reporting and death cases with novel coronavirus (2019-nCoV) infection in China, 2020. Inference of A and B was conducted among (A) exported cases and (B) deceased cases. A) Frequency distribution of the time from illness onset to reporting among exported cases employing a gamma distribution with a mean of 5.1 days (95% CI: 3.5, 7.5) and SD of 4.0 days (95% CI: 2.6, 6.6). B) Frequency distribution of the time from illness onset to death with a mean of 15.2 days (95% CI: 13.0, 17.7) and SD of 6.78 days (95% CI: 5.25, 8.94) employing a gamma distribution. Blue bars show empirically observed data, collected from governmental reports (as of 24 January 2020). Black lines and grey shades represent the median and its 95% credible intervals, respectively. Credible intervals were derived using the Markov Chain Monte Carlo simulations.

1 and 2, respectively. However, as the mean serial interval varies, the estimates can range from 2.1 to 2.3 and 3.5 to 3.9 for scenarios 1 and 2, respectively.

To address the uncertainty with respect to unobserved date of illness onset of index case in scenario 1, cCFR was estimated by varying the date of illness onset of the index case from 1–10 December 2019 (Supplementary figure 1 and Supplementary table 1). When we assumed the date of illness onset of the index case was 1 December 2019, the estimated incidence in China and the cCFR on 24 January 2020 were estimated at 4293 (95% CI: 3054, 5650) and 4.4% (95% CI: 2.9, 6.2). In addition, sensitivity analyses using a varying data cutoff dates (i.e., 17–25 January 2020) were conducted. Depending on the number of data points, the estimates of cumulative incidence and cCFR were shown to have similar values in scenario 1 (Supplementary figure 2) but were slightly decreased for scenario 2, as cutoff date is earlier (Supplementary figure 3).

4 Discussion

The present study estimated the risk of death among confirmed cases while addressing ascertainment bias by using exported case data and employing the likelihood that addresses right-censored data. We estimated the cCFR at 4.6% (95% CI: 3.1, 6.6) when calculating from an index case with illness onset on 8 December 2019 (Scenario 1), and 7.7% (95% CI: 4.9, 11.3%) from exported case data (Scenario 2). In addition, the estimated R_0 was in the range of 1.4–3.5 in Scenario 1 and 2.0–6.5 in Scenario 2. From either estimate, we can conclude that 2019-nCoV has substantial potential to cause a pandemic via

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

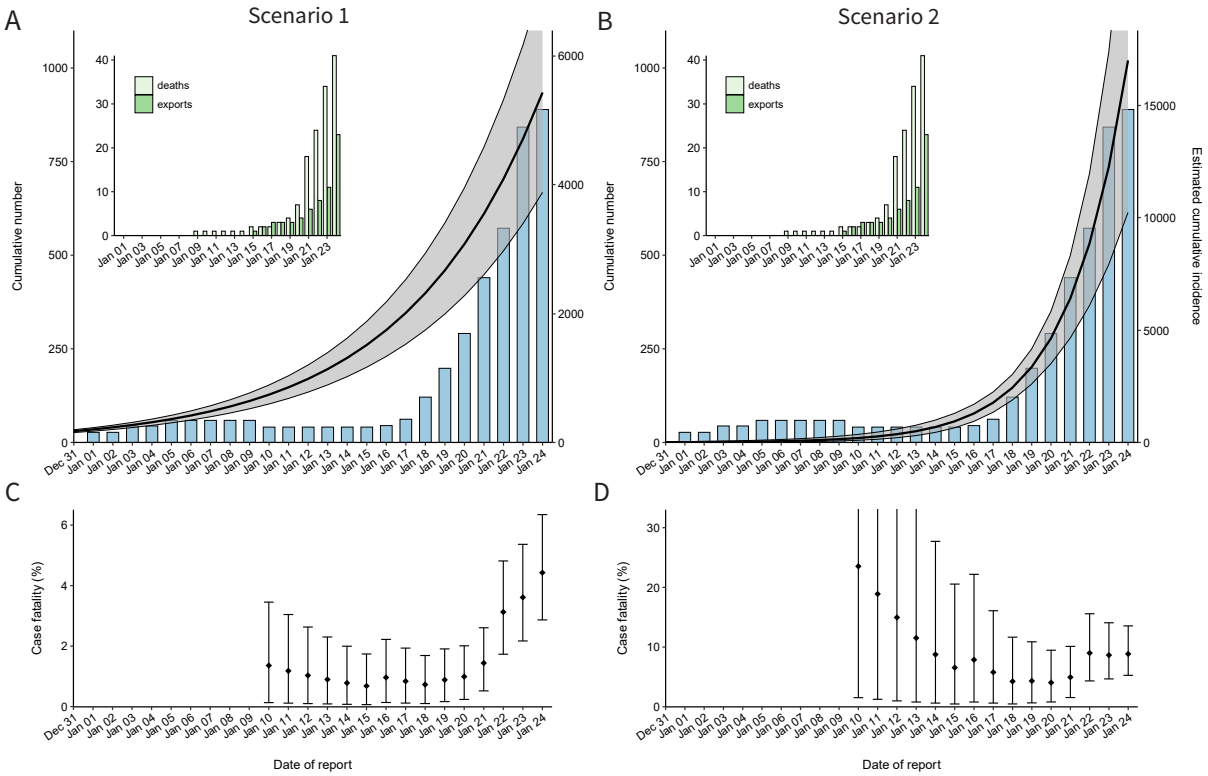


Figure 2. Cumulative incidence and the confirmed case fatality risk of the novel coronavirus (2019-nCoV) outbreak in China, 2020. A-B) Observed and estimated cumulative number of cases in China by the date of report. An exponential growth curve was extrapolated using the exported case data. Scenario 1 extrapolated the exponential growth from December to first case on 8 December 2019, while scenario 2 started the estimation of the exponential growth only from 15 Jan 2020. Black line and shaded area represent median and 95% credible interval of the cumulative incidence in China, respectively. Blue bars show the cumulative number of reported cases from the government of mainland China. There is a decrease in the cumulative number of reported cases in early January, because only 41 cases tested positive for the novel coronavirus among the reported 59 cases on 10 Jan 2020. Left-top panels on both A and B show the cumulative numbers of exported cases and the cumulative number of deaths, represented by dark and light green bars, respectively. C-D) Confirmed case fatality risk (cCFR) by the date of reporting. The points and error bars represent median and its 95% credible interval of cCFR. All 95% credible intervals were derived from Markov Chain Monte Carlo simulations.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

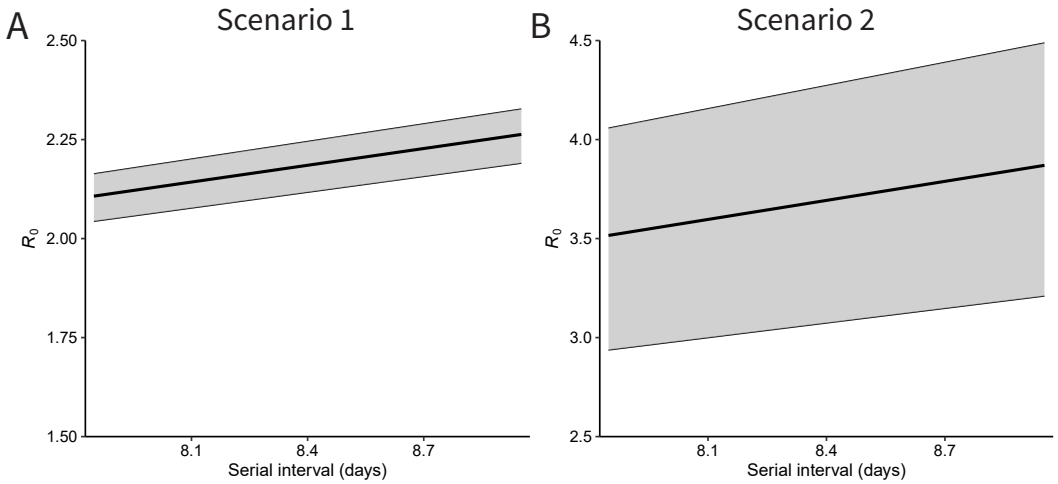


Figure 3. Basic reproduction number of novel coronavirus (2019-nCoV) infections in China, 2020. Black lines and grey shades represent the median and 95% credible intervals of the basic reproduction number. Panel A shows the result of Scenario 1 in which an exponential growth started from the assumed illness onset date of index case, while Panel B shows the result from exponential growth from the first exported case (Scenario 2). The 95% credible intervals were derived from Markov Chain Monte Carlo simulations.

human-to-human transmission.

Our estimates of cCFR at 4.6% and 7.7% indicate that the severity of 2019-nCoV is not as high as that estimated for severe diseases caused by coronaviruses, including SARS (CFR: 17%) in Hong Kong [8, 9, 13] and Middle East respiratory syndrome (CFR: 20%) in South Korea [14]. Nonetheless, considering the overall magnitude of the ongoing epidemic (which is likely to reach the order of 10–100 thousand confirmed cases), a 4–8% risk of death is not at all insignificant. In addition to quantifying the overall risk of death, real time research response must identify risk groups of death (e.g. the elderly and people with underlying comorbidities) in the future [15, 16].

Our estimated R_0 range of 1.4–6.5 for 2019-nCoV is consistent with other preliminary estimates posted on public domains [17–20], and are comparable to the R_0 of SARS, which was in the range of 2–5 during the 2003 outbreak in Singapore [11]. Between our two estimates, the latter scenario yielded a greater value than the former, and there was an increasingly improved ascertainment in early January 2020. The virus was identified and sequenced on 7 January 2020 and subsequently the primer was widely distributed, allowing for rapid laboratory identification of cases and contributing to a time-dependent increase in the number of confirmed cases out of China. Consequently, Scenario 2, which was fully dependent on the growth rate of exported cases, could have overestimated the intrinsic growth rate of cases.

From the technical side, it should be emphasized that our proposed approach can be especially useful during the early stage of an epidemic when local surveillance is affected by substantial ascertainment bias and export and death data are available and better ascertained. Nonetheless, caution must be used when implementing similar estimations for the 2019-nCoV epidemic, as all flights from Wuhan airport were grounded as of 23 January 2020 [21] and this intervention would abruptly change the human

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Importing locations	Date of report	Cumulative count	Estimated incidence in China (95% CI)	
			Scenario 1	Scenario 2
Thailand	15-Jan	1	1514 (1168, 1880)	939 (650, 1296)
Japan	16-Jan	2	1745 (1336, 2182)	1289 (939, 1698)
Thailand	17-Jan	3	2012 (1527, 2532)	1771 (1344, 2247)
South Korea	20-Jan	4	3080 (2279, 3954)	4632 (3507, 5836)
Taiwan, United States	21-Jan	6	3550 (2604, 4587)	6399 (4645, 8311)
Thailand	22-Jan	8	4091 (2975, 5321)	8852 (6091, 11969)
Singapore, Vietnam	23-Jan	11	4714 (3399, 6173)	12260 (7920, 17378)
France, Japan, Nepal, South Korea, Singapore, Thailand, United States	24-Jan	23	5433 (3,883, 7,160)	17780 (9646, 28724)

Table 1. Exportation events and estimated incidence in China, 2020. CI, confidence interval (the 95% CI was derived from the Monte Carlo Markov Chain simulations). Scenario 1 indicates the estimating the exponential growth rate with assumed the illness onset date of the first 2019-nCoV case (i.e., 8 December 2019), while scenario 2 presents the estimating the exponential growth rate from the date of first exportation event (i.e., 15 January 2020).

migration network. Despite abrupt decrease in the outbound flow of travelers from Wuhan, there will be a substantial risk that the next epidemic wave will originate from other cities. There are four limitations in the present study. First, our results present an estimate for the cCFR which only addresses fatality among confirmed cases. We expect that the risk of death among all infected individuals or all symptomatic cases would be smaller than 4–8%. CFR that includes infected individuals other than confirmed cases can only be estimated using additional pieces of data (e.g., seroepidemiological data or outpatient clinic visits). Second, our study relied on limited empirical data that were extracted from publicly available data sources. Thus, future studies with greater sample size and precision are needed. Nonetheless, we believe that this study will improve the situational assessment of the ongoing epidemic. Third, our assumed date of illness onset for the index case in Scenario 1 is based on initial reports of the earliest onset date for a case, and the continued exponential growth with the rate r is the authors' extrapolation. However, we did conduct a sensitivity analysis and ensured that the resulting statistical estimates would not greatly vary. Fourth, heterogeneous aspects of death (e.g. age and risk groups) need to be addressed in the future studies. In conclusion, the present study has successfully estimated cCFR to be on the order of 4-8% and R_0 to be 1.4–6.5, endorsing the notion that 2019-nCoV infection

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

in the ongoing epidemic possesses a substantial pandemic potential. The proposed approach can also help direct risk assessment in other settings using publicly available datasets.

Supplementary material: Supplementary figure 1: Sensitivity analysis with varying the illness onset date of first case from 1-10 December, 2019. Supplementary figure 2: Sensitivity analysis with varying the cutoff date from 15-24 January 2020 in estimation scenario 1, Supplementary figure 3: Sensitivity analysis with varying the cutoff date from 15-24 January 2020 in estimation scenario 2. Supplementary table 1: Sensitivity analysis with varying the illness onset date of first case from 1-10 December, 2019 in estimation scenario 1.

Author Contributions: S-m. J., A.R.A., and H.N. conceived the study and participated in the study design. All authors assisted in collecting the data. S-m. J. and A.R.A. analyzed the data and S-m. J., A.R.A., K.H. and H.N. drafted the manuscript. All authors edited the manuscript and approved the final version.

Funding: H.N. received funding from the Japan Agency for Medical Research and Development (AMED) [grant number: JP18fk0108050]; the Japan Society for the Promotion of Science (JSPS) KAKENHI [grant numbers, H.N.: 17H04701, 17H05808, 18H04895 and 19H01074; R.K.: 18J21587], the Inamori Foundation, and the Japan Science and Technology Agency (JST) CREST program [grant number: JPMJCR1413]. SMJ and NML receive graduate study scholarships from the Ministry of Education, Culture, Sports, Science and Technology, Japan. B.Y. wishes to thank China Scholarship Council.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. World Health Organization. Novel Coronavirus (2019-nCoV) situation report-1: (reference link) (accessed on 24 January 2020).

2. Cui, J.; Li, F.; Shi, Z.L. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 2019, 17, 181–192. (doi:10.1038/s41579-018-0118-9)

3. Center for Disease Control and Prevention. 2019 Novel Coronavirus, Wuhan, China: (reference link) (accessed on 24 January 2020).

4. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020; in press. (doi:10.1016/S0140-6736(20)30183-5)

5. Chan, J,F,W.; Yuan, S.; Kok, K.H.; To K.K.W.; Chu, H.; Yang, J.; Xing, J.; Xing, F.; Liu, J.; Yip, C.C.; Poon, R.W.S.; Tsoi, H.W.; et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020; in press. (doi:10.1016/S0140-6736(20)30154-9)

6. World Health Organization. Novel Coronavirus (2019-nCoV) situation report-3:(reference link) (accessed on 23 January 2020).

7. Nishiura, H.; Jung, S.-M.; Linton, N.M.; Kinoshita, R.; Yang, Y.; Hayashi, K.; Kobayashi, T.; Yuan, B.; Akhmetzhanov, A.R. The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. J. Clin. Med. 2020, 9, 330. (doi:10.3390/jcm9020330)

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

8. Ghani, A.C.; Donnelly, C.A.; Cox, D.R.; Griffin, J.T.; Fraser, C.; Lam, T.H.; Ho, L.M.; Chan, W.S.; Anderson, R.M.; Hedley, A.J.; Leung, G.M. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. Am J Epidemiol. 2005, 16, 479-86. (doi:10.1093/aje/kwi230)

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

10. Garske, T.; Legrand, J.; Donnelly, C.A.; Ward, H.; Cauchemez, S.; Fraser, C.; Ferguson, N.M.; Ghani, A.C. Assessing the severity of the novel influenza A/H1N1 pandemic. BMJ. 2009, 339, b2840. (doi:10.1136/bmj.b2840)

11. Lipsitch, M.; Cohen, T.; Cooper, B.; Robins, J.M.; Ma, S.; James, L.; Gopalakrishna, G.; Chew, S.K.; Tan, C.C.; Samore, M.H.; et al. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. Science. 2003, 300, 1966-1970. (doi:10.1126/science.1086616)

12. Li, M.; Dushoff, J.; Bolker, B.M. Fitting mechanistic epidemic models to data: A comparison of simple Markov chain Monte Carlo approaches. Stat Methods Med Res 2018, 27, 1956–1967. (doi:10.1177/0962280217747054)

13. Donnelly, C.A.; Ghani, A.C.; Leung, G.M.; Hedley, A.J.; Fraser, C.; Riley, S.; Abu-Raddad, L.J.; Ho, L.-M.; Thach, T.-Q.; Chau, P.; et al. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 2003, 361, 1761–1766 (doi:10.1016/S0140-6736(03)13410-1).

14. 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)

15. Majumder, M.S.; Kluberg, S.A.; Mekaru, S.R.; Brownstein, J.S. Mortality Risk Factors for Middle East Respiratory Syndrome Outbreak, South Korea, 2015. Emerg. Infect. Dis. 2015, 21, 2088–2090. (doi:10.3201/eid2111.151231)

16. 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)

17. Zhao, S.; Ran, J.; Musa, S.S.; Yang, G.; Lou, Y.; Gao, D.; Yang, L.; He, D. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak;, 2020; Available from: (reference link) (accessed on 29 January 2020).

18. Imperial College London-MRC Centre for Global Infectious Disease Analysis. News/Wuhan Coronavirus, 2020. (reference link) (accessed on 28 January 2020).

19. The University of Hong Kong-HKUMed WHO Collaborating Centre for Infectious Disease Epidemiology and Control. Nowcasting and forecasting the Wuhan 2019-nCoV outbreak, 2020. (reference link) (accessed on 28 January 2020)

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

20. Read, J.M.; Bridgen, J.R.; Cummings, D.A.; Ho, A.; Jewell, C.P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions; Infectious Diseases (except HIV/AIDS), 2020; Available from: (reference link) (accessed on 29 January 2020).

21. BBC NEWS. Coronavirus: Wuhan shuts public transport over outbreak, 2020. (reference link) (accessed 28 January 2020).

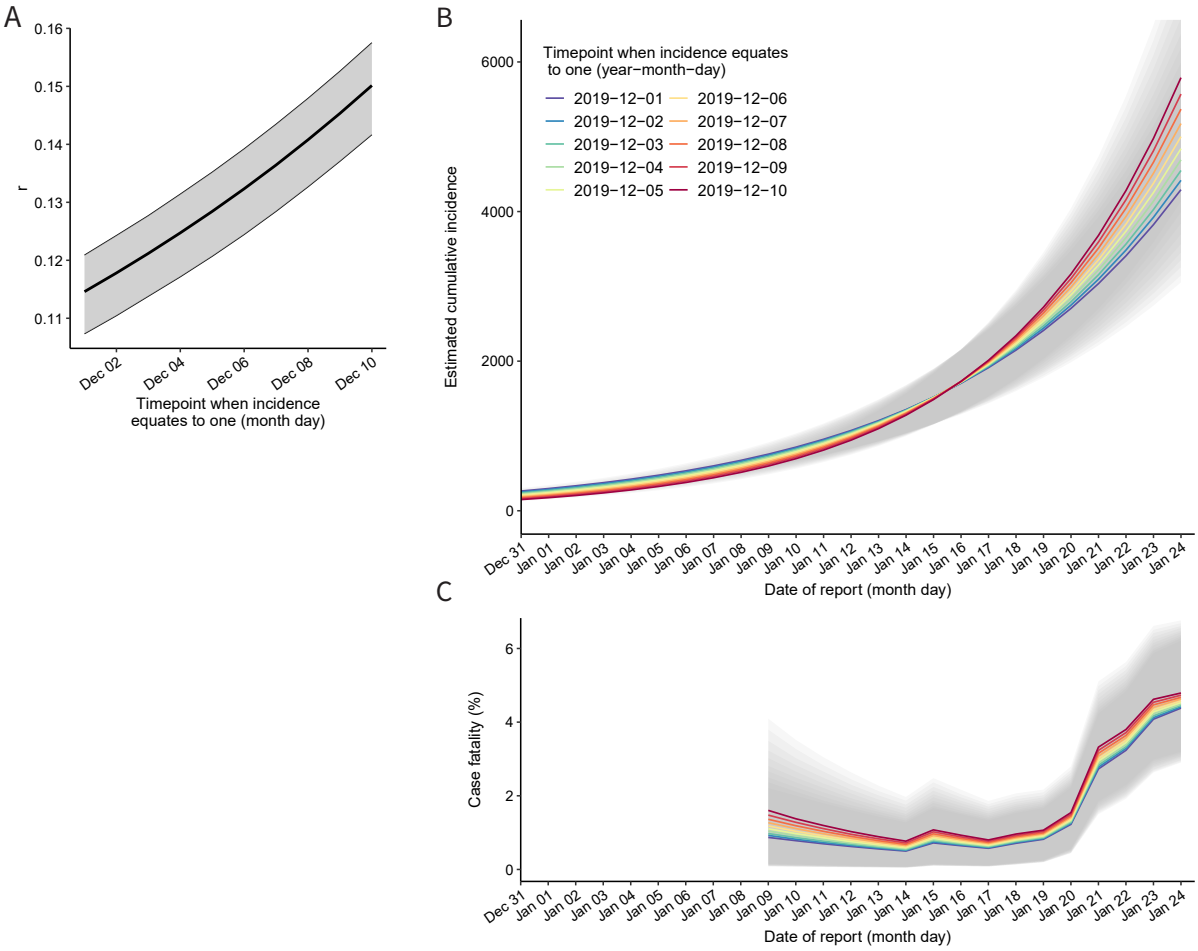
medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplementary material

Illness onset date of first case	Exponential growth rate (95% CI)	Estimated incidence in China on 24 January (95% CI)	cCFR on 24 January (95% CI)
1 December	0.11 (0.11, 0.12)	4293 (3054, 5650)	4.38% (2.90, 6.21)
2 December	0.11 (0.11, 0.12)	4417 (3147, 5826)	4.41% (2.92, 6.28)
3 December	0.12 (0.11, 0.13)	4549 (3256, 5985)	4.45% (2.95, 6.32)
4 December	0.12 (0.12, 0.13)	4691 (3349, 6184)	4.49% (2.98, 6.37)
5 December	0.13 (0.12, 0.14)	4841 (3452, 6372)	4.53% (3.02, 6.44)
6 December	0.13 (0.12, 0.14)	5002 (3567, 6582)	4.58% (3.05, 6.48)
7 December	0.14 (0.13, 0.14)	5176 (3699, 6817)	4.63% (3.08, 6.56)
8 December	0.14 (0.13, 0.15)	5433 (3833, 7160)	4.67% (3.12, 6.61)
9 December	0.15 (0.14, 0.15)	5571 (3989, 7323)	4.73% (3.16, 6.69)
10 December	0.15 (0.14, 0.16)	5790 (4137, 7600)	4.80% (3.20, 6.76)

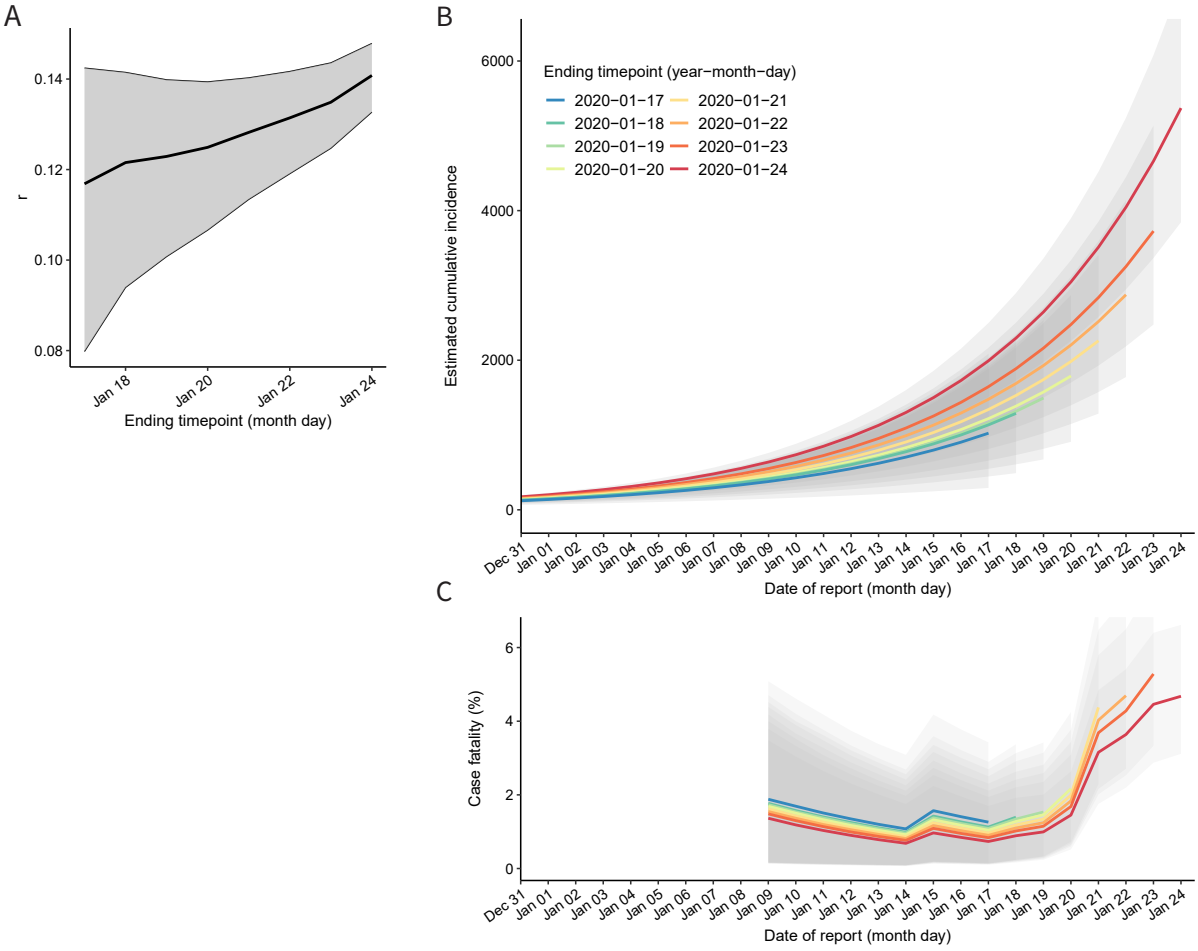
Supplementary table 1. Sensitivity analysis with varying the illness onset date of first case from 1-10 December, 2019 in estimation Scenario 1. cCFR, confirmed case fatality risk; CI, confidence interval (the 95% CI was derived from Markov Chain Monte Carlo simulations).

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



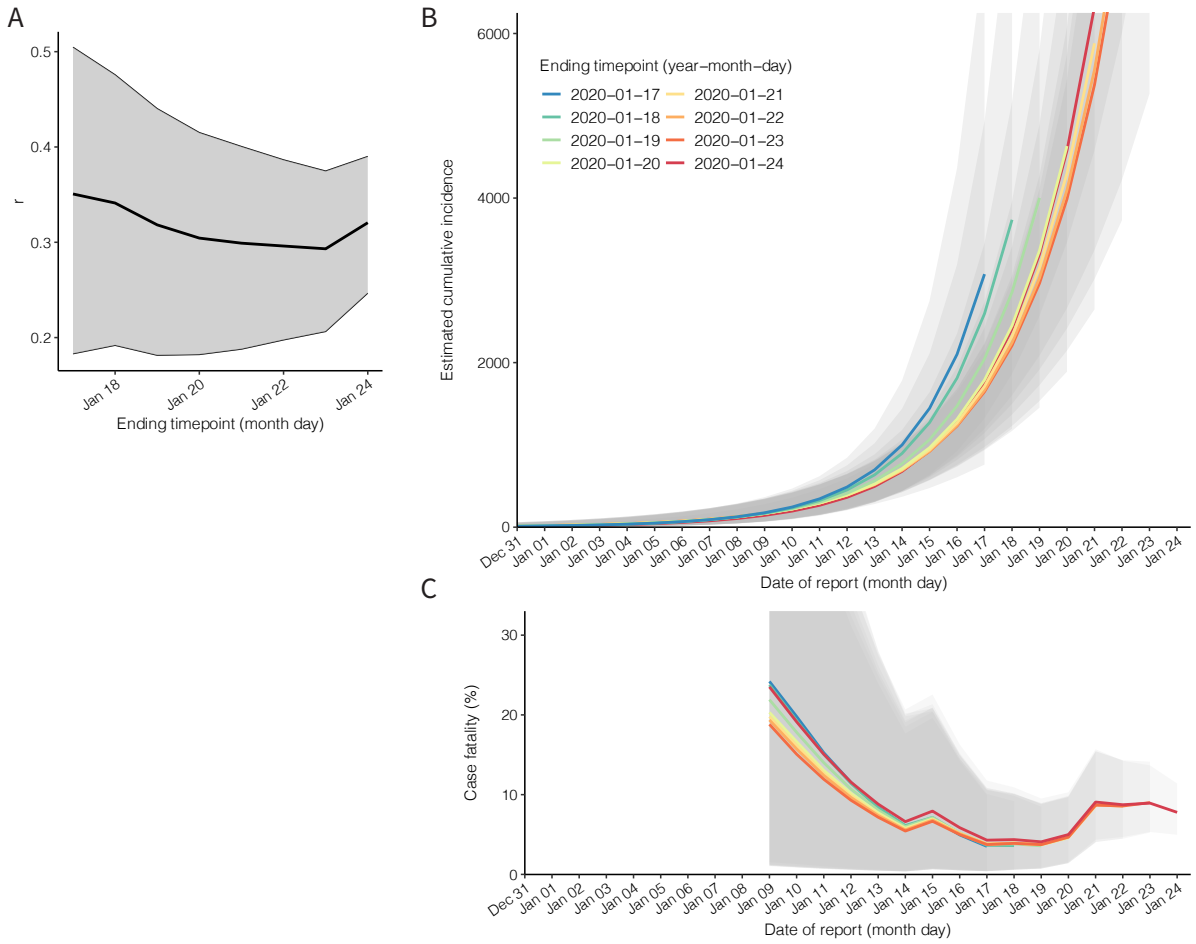
Supplementary figure 1. Sensitivity analysis with varying the illness onset date of first case from 1-10 December, 2019. The estimated values of A) exponential growth rate, B) cumulative incidence and C) confirmed case fatality risk were shown by varying the illness onset date of first case in Scenario 1. Each line and shade present the estimate and its 95% confidence interval and 95% confidence intervals were derived from Markov Chain Monte Carlo simulations.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



Supplementary figure 2. Sensitivity analysis in Scenario 1 with varying the cutoff date from 17-24 January, 2020. The estimated values of A) exponential growth rate, B) cumulative incidence and C) confirmed case fatality risk were estimated depends on the cutoff date (i.e., end point of each estimation) in Scenario 1. Each line and shade indicate the estimate and its 95% confidence interval.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.29.20019547; this version posted February 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



Supplementary figure 3. Sensitivity analysis in Scenario 2 with varying the cutoff date from 17-24 January, 2020. By varying the cutoff date, the estimated values of A) exponential growth rate, B) cumulative incidence and C) confirmed case fatality rate were presented by varying the illness onset date of first case in Scenario 2. Each line and shade present the estimate and its 95% confidence interval.