

Article type

Original Research Article

Title

A novel analysis of the epidemic outbreak of coronavirus disease 2019

Authors

Qingyuan Zhao*, PhD, University Lecturer, Statistical Laboratory, University of Cambridge

Yang Chen, PhD, Assistant Professor, Department of Statistics, University of Michigan

Michael Z Levy, PhD, Associate Professor, Department of Biostatistics and Epidemiology, University of Pennsylvania

Dylan S Small, PhD, Professor, Department of Statistics, University of Pennsylvania

*Correspondence to: Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0W2, United Kingdom; gyzhao@statslab.cam.ac.uk

Conflicts of interest

None declared.

Sources of funding

None.

Data and reproducibility

Data used in this study are from public source. We have kept the data and statistical programs for our analysis public available in the Online Supplement and also at

<https://github.com/qingyuanzhao/2019-nCov-Data>.

A novel analysis of the epidemic outbreak of coronavirus disease 2019

Abstract

Background: On January 23, 2020, outbound travel was banned from Wuhan, China, where the coronavirus disease 2019 (COVID-19) outbreak originated. Initial estimates of the basic epidemiological parameters of COVID-19 could be severely biased by delayed diagnosis and the complex effects of the travel ban on available data.

Methods: We obtained information on the 50 cases of COVID-19 who left Wuhan before the travel ban and were subsequently diagnosed in Hong Kong, Japan, Korea, Macau, Singapore, and Taiwan. Most cases have detailed travel history. We used this data to informatively simulate infection times using a previously reported incubation interval. We then fitted an exponential growth model to the distribution of the infection time adjusting for the January 23 travel ban. We assumed that the travel rate to the international destinations is constant before January 23. We used a Bayesian analysis with diffuse priors to quantify the sampling uncertainty.

Results: We found that the epidemic was doubling in size every 2.8 days (95% credible interval [CrI], 2.0 days—4.0 days). Using a mean serial interval of 7.5 days, we estimated the basic reproduction number of COVID-19 at 5.7 (95% CrI, 3.4—9.3). Our estimates did not change substantially in sensitivity analyses.

Conclusions: Our results indicate that COVID-19 could have been spreading much faster than initial estimates.

Key messages

1. We use internationally confirmed cases to study the COVID-19 outbreak. This avoids selection bias due to delayed diagnosis in Wuhan.
2. We directly model the infection time instead of the noisier symptom onset. Travel history of the international cases allows us to narrow down the range of possible infection time.
3. Our model includes a key term to adjust for the effect of the January 23 travel ban on international travel.
4. Our estimated epidemic growth of the early COVID-19 outbreak are substantially higher than initial estimates.

Introduction

On December 31, 2019, the Health Commission in Wuhan, China, announced 27 cases of unknown viral pneumonia and alerted the World Health Organization. The causative pathogen was subsequently identified as a novel coronavirus and the disease was later designated as the coronavirus disease 2019 (COVID-19) (WHO, 2020b). As of February 17, 2020, COVID-19 has already infected more than 40,000 patients in Wuhan and 70,000 patients worldwide, taking the lives of more than 1,700 (WHO, 2020a).

The international research community responded quickly and obtained initial estimates of some key epidemiological parameters (Table 1). Initially, the early epidemic was estimated to be doubling in size every 6 to 7 days and the basic reproduction number was estimated to be between 2 to 3 (Imai et al., 2020; Li et al., 2020; Read, Bridgen, Cummings, Ho, & Jewell, 2020; Wu, Leung, & Leung, 2020). Subsequent reports (including a preprint of this manuscript) have found the epidemic growth could have been much faster than the initial estimates (Sanche et al., 2020; Yang et al., 2020; Q. Zhao, Chen, & Small, 2020).

Two unique features of this outbreak pose serious challenges to its epidemiological analyses. The first key challenge is the significant diagnostic delay during the early outbreak. In fact, because many patients in Wuhan could not be diagnosed in time due to limited number of real-time RT-PCR assays, the Chinese Health Commission decided to introduce a new clinical diagnostic criterion on February 12, 2020. The new criteria led to a spike of 14,108 new cases in mainland China on February 12, compared to only 2,467 new cases in the previous day (Figure 1). Therefore, any epidemiological analysis using the early cases in Wuhan almost certainly missed a large number of unconfirmed patients, resulting in underestimation of the epidemic size and growth.

The second key challenge for analyzing this outbreak is the abrupt quarantine of Wuhan on January 23, 2020. With very short notice, all outbound public transportation from Wuhan, including domestic and international flights, was suspended. Therefore, any epidemiological analysis using cases outside Wuhan needs to take this travel ban into account.

To the best of our knowledge, none of the existing analyses of the COVID-19 outbreak has convincingly addressed these challenges. In this article, we present an analysis to tackle these problems in two novel ways:

1. We used international cases of COVID-19 that were exported from Wuhan. In particular, we focused on the Wuhan-exported cases in six countries and regions (Hong Kong, Japan, Korea, Macau, Singapore, and Taiwan) that have good public health surveillance systems and have reported detailed case information (Figure 1). Because we collected the data three weeks after the travel ban, it is safe to assume that our dataset contains all the Wuhan-exported cases in these destinations.
2. Unlike the majority of other analyses which model the noisier symptom onset time, we directly model when the patients were infected. Doing so allows us to incorporate the

patients' travel history and, more importantly, take into account the January 23 quarantine in our model.

Details of our analysis can be found in the Methods section below.

Methods

Our analysis of the COVID-19 outbreak consists of six main steps:

1. We obtained a dataset of internationally confirmed case exported from Wuhan.
2. We developed a novel way of inferring the infection time of these cases that take into account their travel history.
3. We developed an exponential growth model for the infection counts with adjustment to the January 23 travel ban.
4. We obtained a point estimator of the growth exponent by fitting a log-linear regression with a model-derived offset.
5. We cross checked a qualitative prediction of our model with the inferred infection time distribution.
6. We obtained uncertainty quantification for our estimated exponential growth using a Bayesian model.

Dataset preparation and visualization

We obtained information for 233 cases of COVID-19 in Hong Kong, Japan, Korea, Macau, Singapore, and Taiwan that have been confirmed as of February 16, 2020. These internationally confirmed cases can be separated into three categories: those who were exported from Wuhan before the January 23 travel ban (Wuhan-exported), those who were evacuate from Wuhan by charter flights after the January 23 travel ban (evacuated), and those who were infected outside Wuhan (domestic). To be specific, we consider a case as Wuhan-exported if

1. The case had been to Wuhan during the COVID-19 outbreak (from Dec 1, 2019 to Jan 23, 2020).
2. The case had left Wuhan before the January 23 travel ban took effect.
3. The case had no known exposure to other confirmed COVID-19 cases outside Wuhan. If two or more confirmed cases have traveled from Wuhan together, we only considered the first confirmed case as Wuhan-exported as the others were likely infected during travel.

Notice that we collected the data more than 3 weeks after the January 23 travel ban. Previous reports have estimated that the 95% quantile of the incubation period is about 11–14 days (Backer, Klinkenberg, & Wallinga, 2020; Li et al., 2020), so it is safe to assume that our dataset contains all Wuhan-exported cases to these destinations.

In total, we obtained 50 Wuhan-exported cases, 20 evacuated cases, and 163 domestic cases (Figure 2). Our statistical analysis below focused on the $n = 50$ Wuhan-exported cases, whose symptom onset date and detailed travel history were available.

Inferring the infection time

A key methodological novelty of our analysis is that we directly model the infection time. Because infection time is usually unobserved, most epidemic analyses instead model the noisier symptom onset. However, the nature of the COVID-19 outbreak and in particular the January 23 travel ban provide us a rare opportunity to narrow down the infection time. Because all the cases considered are Wuhan-exported, we know that they had been infected before they arrived at the destination country and certainly no later than January 23. For a few cases, it is also known that the patient stayed in Wuhan for a limited period of time. These constraints allow us to narrow down and informatively simulate the infection time.

Next we introduce some notation to formalize the idea. In this analysis we modeled the transmission of COVID-19 in Wuhan from December 1, 2019 (7 days before the symptom onset of the first known patient of COVID-19) to January 23, 2020 (outbound travel ban from Wuhan), a total of $T = 54$ days. For case i , we observed the interval $[B_i, E_i]$ for which patient i stayed in Wuhan ($E_i \leq T$ due to the travel ban) and the symptom onset time S_i . Let I_i be the unobserved infection time. Because the patients must have been infected during their stay in Wuhan, we know $B_i \leq I_i \leq E_i$. We thus assumed that the incubation period $S_i - I_i$ follows a truncated Gamma distribution:

$$S_i - I_i \mid S_i, B_i, E_i \sim \text{Gamma}(\alpha, \beta; S_i - B_i, S_i - E_i),$$

where $\text{Gamma}(\alpha, \beta; l, u)$ is the conditional distribution of a $\text{Gamma}(\alpha, \beta)$ -distributed random variable given that it is between l and u . This reduces to the usual assumption of a Gamma-distributed incubation period if there were no constraint on the infection time ($[B_i, E_i]$ were unknown).

Based on a previous estimate for the incubation period of COVID-19, in the main analysis we used a Gamma distribution of mean = 5.2 days and standard deviation = 3.7 days (Li et al., 2020). We also used another slightly longer and more concentrated estimate (mean = 6.5 days, standard deviation = 2.6 days) as a sensitivity analysis (Backer et al., 2020). We applied a uniform smoothing of the observed discretized dates to obtain the continuous time S_i, B_i, E_i in the model (Obadia, Haneef, & Boëlle, 2012); see the Online Supplement for more details.

Exponential growth model with adjustment to the January 23 travel ban

After obtaining the inferred distribution of infection time I_i as described above, we used it to estimate the growth exponent r in an exponential-growth model that is commonly used for early epidemic outbreak (Wallinga & Lipsitch, 2007). Let OI_t be the number of Wuhan-exported cases whose infection date was t for $1 \leq t \leq T$, so

$$OI_t = \sum_{i=1}^n 1_{\{t-1 < I_i \leq t\}},$$

where 1 is the indicator function and n is the total number of Wuhan-exported cases. Let the number of new infections in Wuhan on day t be WI_t , so OI_t is a small fraction of WI_t who travelled to Hong Kong, Japan, Korea, Macau, Singapore, and Taiwan from day t to day T (before outbound travel was banned on January 23). Let OR_t be the (very small) proportion of people in Wuhan traveling to those countries and destinations on day t . We assumed

$$OI_t \sim \text{Poisson}(\lambda_t), 1 \leq t \leq T,$$

where the mean λ_t is given by

$$\lambda_t = WI_t \times P(\text{travel from day } t \text{ to day } T) = WI_t \times \left(1 - \prod_{s=t}^T (1 - OR_s)\right) \approx WI_t \times \sum_{s=t}^T OR_s.$$

As international travel from Wuhan was relatively stable during the study period, we assumed

$$OR_t \equiv OR, 1 \leq t \leq T.$$

Finally, we assumed the number of new infections in Wuhan WI_t was growing exponentially in this period

$$WI_t = WI_0 \times e^{rt}, 1 \leq t \leq T.$$

Under these two assumptions, we have

$$\lambda_t \approx (WI_0 \times OR) (T - t + 1) e^{rt}.$$

There are three parameters in this model: the growth exponent r , the travel rate OR , and the baseline epidemic size WI_0 . We are mainly interested in estimating the growth exponent r .

Model fitting --- Point estimator

We estimated λ_t by simulating the infection time and averaging the simulated OI_t . Notice that the estimated λ_t only depend on the observed data and assumed incubation period and does not depend on the adjusted exponential growth model. By taking logarithm of λ_t in the adjusted exponential growth model, we obtain

$$\log(\lambda_t) - \log(T - t + 1) \approx rt + \log(WI_0 \times OR).$$

This motivates the following point estimator of r by fitting a linear regression of $\log(\lambda_t)$ on time t with an offset term $\log(T - t + 1)$ to adjust for the effect of the travel ban.

In our analysis, we obtained our point estimator of r using estimated λ_t values between January 1, 2020 and January 15, 2020. We then examined whether the estimated r provides a good fit before January 1 and after January 15.

Model checking --- Implied stationary point of λ_t

A key feature of our adjusted exponential growth model is that it predicts λ_t to have a stationary point t^* that satisfies

$$0 = \frac{d}{dt} \lambda_t \big|_{t=t^*} \propto e^{rt^*} (r(T - t^* + 1) - 1).$$

In other words, our model predicts that λ_t decreases after time $t^* = T + 1 - 1/r$. We used this qualitative behavior to check the fit of our statistical model, by comparing the empirical stationary point of the estimated λ_t (which only relies on the assumed incubation period but not the exponential growth model) with the model predicted stationary point.

Model fitting --- Bayesian inference

Our point estimator of r does not take into account the uncertainty in estimating λ_t . To further quantify the sampling uncertainty, we performed a Bayesian analysis for our statistical model.

We used a diffuse prior for our model parameters; more details can be found in the Online Supplement. We expected that the posterior of the number of new infections WI_t would be very sensitive to the choice of the prior mean of OR , but the posterior of the growth exponent r would be insensitive to the choice of prior.

Notice that our statistical model depends on the latent infection time I_i , whose distribution need to be integrated out in a Bayesian analysis:

$$\pi(r, OR, WI_0 | \text{Data}) = \int \pi(r, OR, WI_0 | OI) \prod_{i=1}^n P(I_i | S_i, B_i, E_i) dI.$$

To obtain posterior samples of r , we first simulated 100 realizations of the infection time using the truncated Gamma distribution described above. We then pooled the posterior samples of r over the different realizations of the infection time.

In our main analysis, we computed the posterior distribution of r using the simulated infection counts OI_t from January 1 to January 20. We did not use infections before January 1 because the simulated infection time could be unreliable. We did not use infections after January 20 because of a public announcement of human-to-human transmission of COVID-19 on that day. We expected that the public announcement lowered the epidemic growth after January 20. As a sensitivity analysis, we also fitted our model using infection counts from January 1 to January 23. The Bayesian posterior samples were drawn using the *rstan* software package (Carpenter et al., 2017).

Computing the basic reproduction number

The basic reproduction number R_0 (expected number of cases generated by one case) is a crucial parameter for an epidemic outbreak. In the early outbreak, it can be estimated from the growth exponent r and the serial interval (time between successive cases in a chain of transmission) by $R = 1/M(-r)$, where $M(\cdot)$ is the moment generating function of the distribution of the serial interval (Wallinga & Lipsitch, 2007). For our analysis, we used a Gamma-distributed serial interval of mean = 7.5 days and standard deviation = 3.4 days reported by a previous article (Li et al., 2020).

Results

Inferred infection time, point estimator of r , and the implied stationary point of λ_t

Using the truncated Gamma distribution described in Methods, we obtained the distribution of the individual infection time I_i (grey bars in Figure 2). Among the 50 internationally confirmed cases being considered, 35 arrived at their destinations before symptom onset with an average of 3.6 days. This allowed us to substantially narrow down the distribution for those I_i .

We obtained an estimate of the mean infection count λ_t (Figure 3, Panel A) by summing over the simulated distribution of I_i in Figure 2. Error bars in Panel A show that the uncertainty in estimated λ_t due to Monte-Carlo simulation is small.

In the linear regression of $\log(\lambda_t)$ on time t with the offset $\log(T - t + 1)$, the growth exponent r was estimated to be 0.26 (Figure 2, Panel B). This estimate provides a good fit to the estimated $\log(\lambda_t)$ outside the training period. We further used the stationary point of λ_t to check the fit of our model as explained in Methods. The calculations in “Model checking” suggest that our model with this $r = 0.26$ would predict that λ_t starts decreasing from time $t = 51.1$ (close to January 20). This is largely consistent with the stationary point of the estimated λ_t in Figure 2, Panel A.

Bayesian inference for epidemic growth

In our main Bayesian model (Table 2), the posterior mean of the growth exponent r was 0.25 (95% CrI, 0.17 – 0.34). This corresponds to a doubling time of 2.8 days (95% CrI, 2.0 days – 4.0 days) and basic reproduction number R_0 of 5.7 (95% CrI, 3.4 – 9.3).

We further performed a sensitivity analysis of our results (Table 2 and Online Supplement) using a more concentrated incubation period and a different period of infection counts as described in Methods. We found that our estimated growth exponent r are insensitive to these choices. The smallest estimated r in our sensitivity analysis is 0.21 (95% CrI, 0.15 – 0.27), which corresponds to a doubling time of 3.4 days (95% CrI 2.6 days – 4.5 days) and basic reproduction number R_0 of 4.4 (95% CrI, 3.0 – 6.2). We also performed a sensitivity analysis to the prior for the rate of travel OR . Not surprisingly, we found that the total number of infections are very sensitive to the choice but the estimated growth exponent r is not sensitive to the choice (Online Supplement).

Discussion

Our analysis reported a much faster growth for the early outbreak of COVID-19 in Wuhan than initial estimates (Table 1). A direct comparison can be drawn with a previous analysis that

also used internationally confirmed cases to infer the epidemic size and growth in Wuhan (Wu et al., 2020). However, their estimated doubling time was 6.4 days (95% CrI, 5.8 days – 7.1 days). Another report (Li et al., 2020) using the cases in Wuhan estimated the doubling time at 7.4 days (4.2 days – 14 days). Other analyses using more recent data have reported larger estimates of the epidemic growth and the basic reproduction number (Sanche et al., 2020; Yang et al., 2020; S. Zhao et al., 2020).

Given the motivations of our study in the Introduction, the large difference between the previous epidemiological estimates and ours should hardly be surprising. One obvious reason is the different study sample being used. The previous reports have used internationally confirmed cases as of January 28 and the first 425 confirmed cases in Wuhan. Many COVID-19 cases had yet been confirmed when those reports were written. In comparison, we collected all confirmed cases in 6 selected countries and regions as of February 16, which most likely included all the Wuhan-exported cases that we chose to model.

Another key novelty of our analysis is that it uses the inferred infection time instead of the noisier symptom onset as in the previous analyses. By using the cases' travel history and accounting for the January 23 travel ban in Wuhan, we were able to narrow down the infection time.

Compared to (Wu et al., 2020), a third distinction is that our model takes into account the restricted outbound travel from Wuhan since January 23, 2020 --- a milestone event in this epidemic. The analysis by (Wu et al., 2020) used a more complicated susceptible-exposed-infectious-recovered (SEIR) model but the authors only accounted for the quarantine of Wuhan in their forecast, not in their estimation. If we had made the same simplifications we would have had a very poor fit to our data. For example, had we ignored the January 23 travel ban in our model, we would have predicted that the mean infection count λ_t is always increasing over time, contradicting the stationary point observed in Figure 2, Panel A.

Our analysis should be viewed in terms of its limitations. The international cases are only “shadows” of the epidemic in Wuhan and we relied on the assumption that they form a representative sample. We used a simple exponential growth model for the new infections, appropriate for early in an epidemic (Wallinga & Lipsitch, 2007), but did not account for the complexities that accrue as an epidemic proceeds. We assumed a constant rate of travel in the study period; our assumption may be poor if the epidemic itself changed travel behavior of those in Wuhan. We also assumed that the internationally confirmed cases were no more or less likely to be infected than citizens of Wuhan at the time that they left the city. Finally, we recommend cautious interpretation of our estimated basic reproduction number as it relies on a preliminary estimate of the serial interval.

Despite these potential limitations, our simple model --- exponential growth with adjustment for the travel ban on January 23 --- provides a very good fit to the internationally confirmed cases. Our results suggest that the early outbreak of COVID-19 could have been spreading much faster in Wuhan than initial estimates.

Acknowledgement

The authors have no conflict of interest to declare. We thank Sergio Bacallado, Cindy Chen, Yunjin Choi, Hera He, Nianqiao Ju, Marc Lipsitch, James Robins, Andrew Rosenfeld, and Yachong Yang for providing helpful comments on the data collection and statistical analysis.

References

- Backer, J. A., Klinkenberg, D., & Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*, 25(5). [https://doi.org/https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062](https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062)
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software; Vol 1, Issue 1 (2017)*. Retrieved from <https://www.jstatsoft.org/v076/i01>
- Imai, N., Cori, A., Dorigatti, I., Baguelin, M., Donnelly, C. A., Riley, S., & Ferguson, N. M. (2020). Report 3: Transmissibility of 2019-nCoV. *Online Preprint*.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMoa2001316>
- Obadia, T., Haneef, R., & Boëlle, P.-Y. (2012). The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Medical Informatics and Decision Making*, 12(1), 147. <https://doi.org/10.1186/1472-6947-12-147>
- Read, J. M., Bridgen, J. R. E., Cummings, D. A. T., Ho, A., & Jewell, C. P. (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *MedRxiv*, 2020.01.23.20018549. <https://doi.org/10.1101/2020.01.23.20018549>
- Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. *MedRxiv*, 2020.02.07.20021154. <https://doi.org/10.1101/2020.02.07.20021154>
- Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings. Biological Sciences*, 274(1609), 599–604. <https://doi.org/10.1098/rspb.2006.3754>
- WHO. (2020a). *Coronavirus disease 2019 (COVID-19) Situation Report – 28*. Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200217-sitrep-28-covid-19.pdf?sfvrsn=a19cf2ad_2
- WHO. (2020b). *WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China*. Retrieved from <https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china>
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet (London, England)*, 0(0). <https://doi.org/10.1016/S0140->

6736(20)30260-9

Yang, Y., Lu, Q., Liu, M., Wang, Y., Zhang, A., Jalali, N., ... Fang, L. (2020). Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. *MedRxiv*, 2020.02.10.20021675. <https://doi.org/10.1101/2020.02.10.20021675>

Zhao, Q., Chen, Y., & Small, D. S. (2020). Analysis of the epidemic growth of the early 2019-nCoV outbreak using internationally confirmed cases. *MedRxiv*, 2020.02.06.20020941. <https://doi.org/10.1101/2020.02.06.20020941>

Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., ... Wang, M. H. (2020). 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. *International Journal of Infectious Diseases*. <https://doi.org/10.1016/J.IJID.2020.01.050>

Figure 1 Number of daily confirmed cases of COVID-19 in Mainland China and 6 countries and regions (Hong Kong, Japan, Korea, Macau, Singapore, Taiwan). Two key dates in the outbreak: January 23, when outbound travel from Wuhan was banned and February 12, when a new diagnostic criterion is used for cases in the Hubei province.

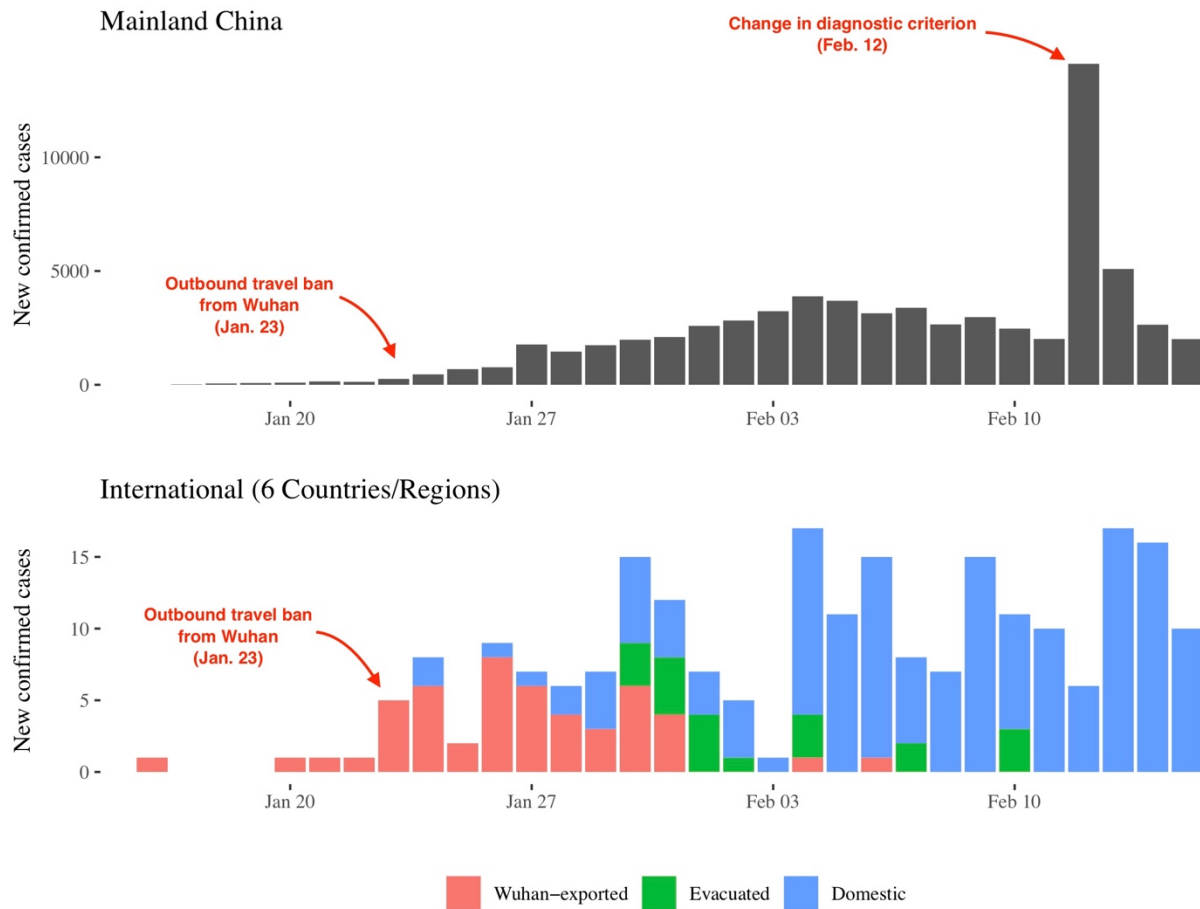


Figure 2 Key dates and the distribution of inferred infection date of 50 confirmed COVID-19 cases in 6 Asian countries and regions (HK: Hong Kong, JP: Japan, KR: Korea, MO: Macau, SG: Singapore, TW: Taiwan). Arrival: when the patient arrived at the destination; Symptom: when the patient first showed symptoms; Confirmed: when the case was confirmed as COVID-19 positive; Infected: distribution of the inferred infection time, simulated using Symptom onset minus previously reported incubation period with the constraints that the infection happened before the patient left Wuhan.

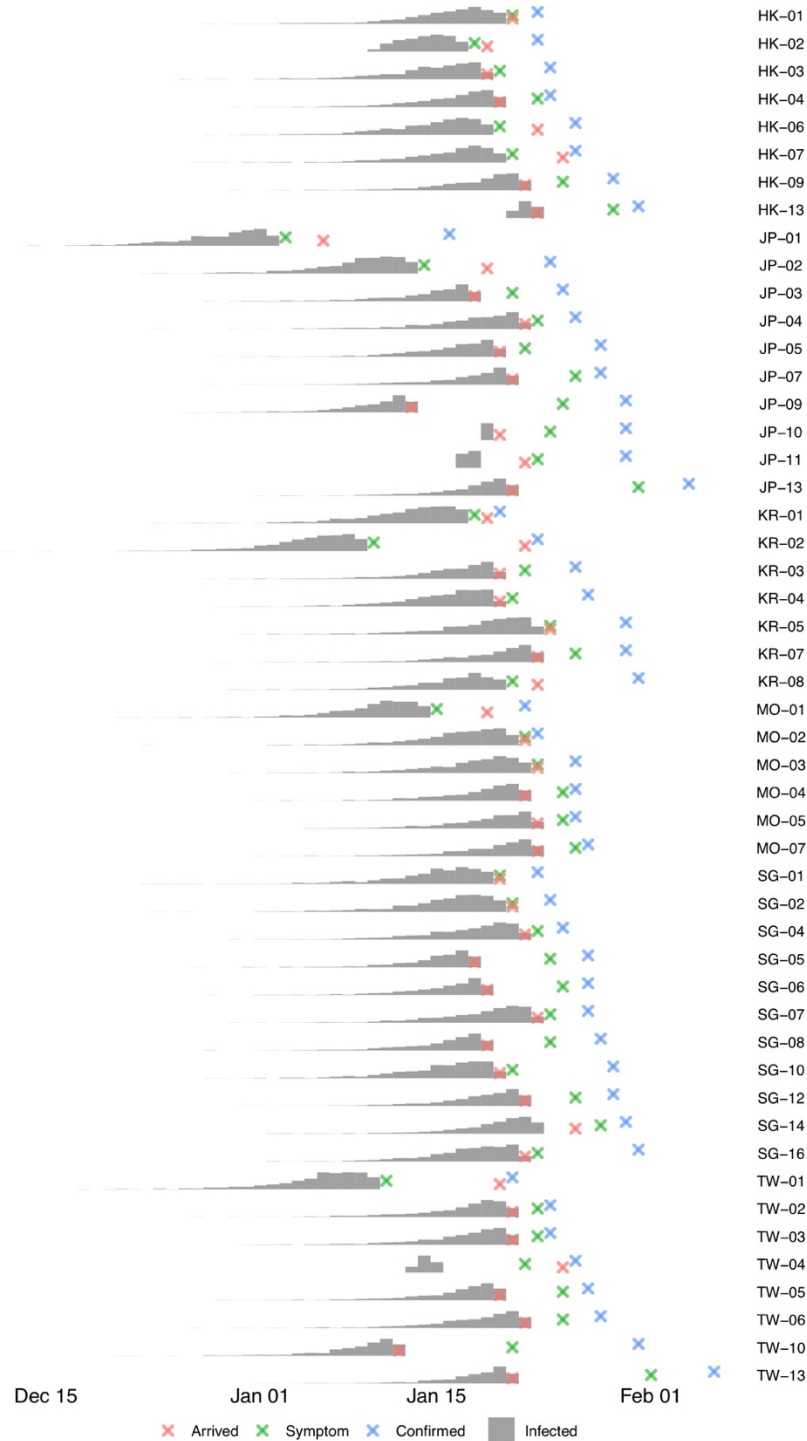


Figure 3 Model fitting using estimated λ_t (mean counts of new infections). Panel A: Grey bars indicate the estimated λ_t computed by aggregating the inferred infection dates in Figure 2. Red and blue bars are one standard error-bars of λ_t due to Monte-Carlo simulation. Grey dots are 20 realizations of the simulated infection counts OI_t . Panel B: Fitted linear regression for logarithm of the mean counts versus time with offset adjusting for the January 23 travel ban. Red line is a linear regression fitted using the data in panel A from January 1 to January 15.

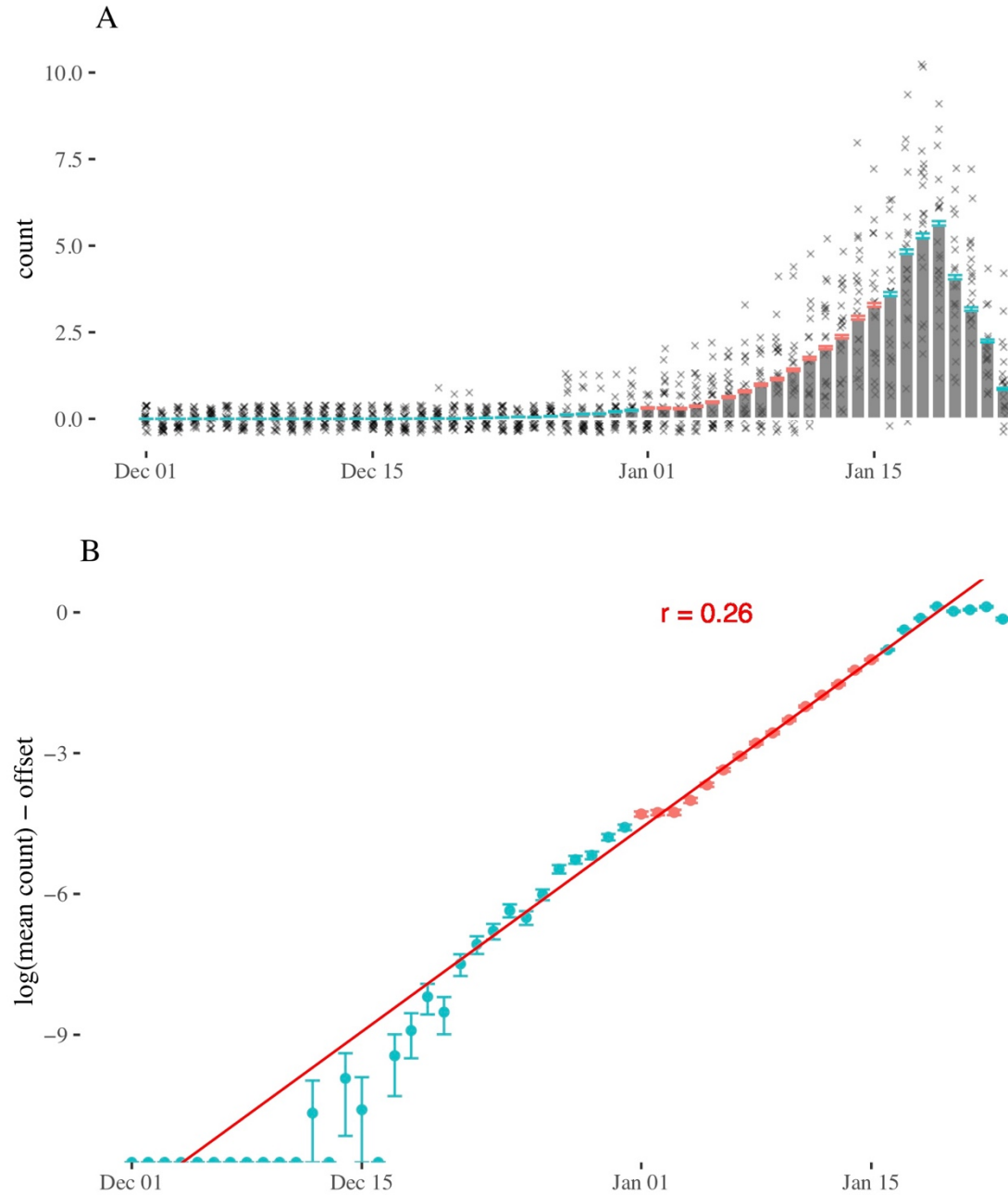


Table 1 Early estimates of the epidemiological parameters of COVID-19 in some selected publications and preprints.

Date published or posted as preprint	When was the dataset collected?	Sample	Does the analysis account for the quarantine?	Estimated parameters (95% CI or CrI)	Reference
January 25	January 22	First 7 international cases exported from Wuhan	No	R_0 : 2.6 (1.5 – 3.5)	(Imai et al., 2020)
January 28	January 22	Case counts	No	R_0 : 3.11 (2.39 – 4.13)	(Read et al., 2020)
January 29	January 22	First 425 confirmed cases in Wuhan	No	Doubling time: 7.4 (4.2 – 14) days; R_0 : 2.2 (1.4 – 3.9)	(Li et al., 2020)
January 30	January 24	Case counts	No	R_0 : 2.24 (1.96 – 2.55) to 3.58 (2.89 – 4.39)	(S. Zhao et al., 2020)
January 31, 2020	January 25, 2020	78 international cases exported from Wuhan	No	Doubling time: 6.4 (5.8 – 7.1) days; R_0 : 2.68 (2.47 – 2.86)	(Wu et al., 2020)
February 9, 2020	February 5, 2020	46 international cases	Yes	Doubling time: 2.9 (2.0 – 4.1) days; R_0 : 5.7 (3.4 – 9.2)	(Q. Zhao et al., 2020) (preprint of this article)
February 11, 2020	January 26, 2020	First 4021 confirmed cases in China	No	R_0 : 3.77 (3.51 – 4.05)	(Yang et al., 2020)
February 11, 2020	End of January, 2020	140 cases (first or first few confirmed cases in other provinces of China)	Yes	<i>Method “first arrival”</i> Doubling time: 2.4 (1.9 – 3.3) days; R_0 : 6.3 (3.3 – 11.3) <i>Method “case count”</i> Doubling time: 2.3 (2.0 – 2.6) days R_0 : 6.6 (4.0 – 10.5)	(Sanche et al., 2020)

Table 2 Estimated parameters for epidemic growth rate using different periods of infection and different assumptions on the travel rate. Estimated parameters are in the form of “posterior mean (95% credible interval)”.

		Incubation period	
		Mean = 5.2 days Std. deviation = 3.7 days (Li et al., 2020)	Mean = 6.5 days Std. deviation = 2.6 days (Wu et al., 2020)
Using infections on January 1---20	Growth exponent	0.25 (0.17 – 0.34)	0.25 (0.18 – 0.32)
	Doubling time (days)	2.8 (2.0 – 4.0)	2.9 (2.2 – 4.0)
	Basic reproduction number	5.7 (3.4 – 9.3)	5.5 (3.5 – 8.3)
Using infections on January 1---23	Growth exponent	0.22 (0.16 – 0.29)	0.21 (0.15 – 0.27)
	Doubling time (days)	3.2 (2.4 – 4.3)	3.4 (2.6 – 4.5)
	Basic reproduction number	4.8 (3.2 – 7.0)	4.4 (3.0 – 6.2)