

# Early dynamics of transmission and control of COVID-19: a mathematical modelling study



Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo,  
on behalf of the Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group\*



## Summary

**Background** An outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to 95 333 confirmed cases as of March 5, 2020. Understanding the early transmission dynamics of the infection and evaluating the effectiveness of control measures is crucial for assessing the potential for sustained transmission to occur in new areas. Combining a mathematical model of severe SARS-CoV-2 transmission with four datasets from within and outside Wuhan, we estimated how transmission in Wuhan varied between December, 2019, and February, 2020. We used these estimates to assess the potential for sustained human-to-human transmission to occur in locations outside Wuhan if cases were introduced.

**Methods** We combined a stochastic transmission model with data on cases of coronavirus disease 2019 (COVID-19) in Wuhan and international cases that originated in Wuhan to estimate how transmission had varied over time during January, 2020, and February, 2020. Based on these estimates, we then calculated the probability that newly introduced cases might generate outbreaks in other areas. To estimate the early dynamics of transmission in Wuhan, we fitted a stochastic transmission dynamic model to multiple publicly available datasets on cases in Wuhan and internationally exported cases from Wuhan. The four datasets we fitted to were: daily number of new internationally exported cases (or lack thereof), by date of onset, as of Jan 26, 2020; daily number of new cases in Wuhan with no market exposure, by date of onset, between Dec 1, 2019, and Jan 1, 2020; daily number of new cases in China, by date of onset, between Dec 29, 2019, and Jan 23, 2020; and proportion of infected passengers on evacuation flights between Jan 29, 2020, and Feb 4, 2020. We used an additional two datasets for comparison with model outputs: daily number of new exported cases from Wuhan (or lack thereof) in countries with high connectivity to Wuhan (ie, top 20 most at-risk countries), by date of confirmation, as of Feb 10, 2020; and data on new confirmed cases reported in Wuhan between Jan 16, 2020, and Feb 11, 2020.

**Findings** We estimated that the median daily reproduction number ( $R_t$ ) in Wuhan declined from 2.35 (95% CI 1.15–4.77) 1 week before travel restrictions were introduced on Jan 23, 2020, to 1.05 (0.41–2.39) 1 week after. Based on our estimates of  $R_t$ , assuming SARS-like variation, we calculated that in locations with similar transmission potential to Wuhan in early January, once there are at least four independently introduced cases, there is a more than 50% chance the infection will establish within that population.

**Interpretation** Our results show that COVID-19 transmission probably declined in Wuhan during late January, 2020, coinciding with the introduction of travel control measures. As more cases arrive in international locations with similar transmission potential to Wuhan before these control measures, it is likely many chains of transmission will fail to establish initially, but might lead to new outbreaks eventually.

**Funding** Wellcome Trust, Health Data Research UK, Bill & Melinda Gates Foundation, and National Institute for Health Research.

**Copyright** © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

## Introduction

As of Feb 13, 2020, an outbreak of coronavirus disease 2019 (COVID-19) has resulted in 46 997 confirmed cases.<sup>1</sup> The outbreak was first identified in Wuhan, China, in December, 2019, with most early cases being reported in the city. Most internationally exported cases reported to date have history of travel to Wuhan.<sup>2</sup> In the early stages of a new infectious disease outbreak, it is crucial to understand the transmission dynamics of the infection. Estimation of changes in transmission over time can

provide insights into the epidemiological situation<sup>3</sup> and identify whether outbreak control measures are having a measurable effect.<sup>4,5</sup> Such analysis can inform predictions about potential future growth,<sup>6</sup> help estimate risk to other countries,<sup>7</sup> and guide the design of alternative interventions.<sup>8</sup>

However, there are several challenges to such analyses, particularly in real time. There can be a delay to symptom appearance resulting from the incubation period and delay to confirmation of cases resulting from detection

*Lancet Infect Dis* 2020

Published Online  
March 11, 2020  
[https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)

See Online/Comment  
[https://doi.org/10.1016/S1473-3099\(20\)30161-4](https://doi.org/10.1016/S1473-3099(20)30161-4)

\*Members are listed in the appendix

Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK (A J Kucharski PhD, T W Russell PhD, C Diamond MSc, Y Liu PhD, J Edmunds PhD, S Funk PhD, R M Eggo PhD)

Correspondence to:  
Dr Adam J Kucharski, Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK  
[adam.kucharski@lshtm.ac.uk](mailto:adam.kucharski@lshtm.ac.uk)

See Online for appendix

### Research in context

#### Evidence before this study

We searched PubMed, BioRxiv, and MedRxiv for articles published in English from inception to Feb 10, 2020, with the keywords “2019-nCoV”, “novel coronavirus”, “COVID-19”, “SARS-CoV-2” AND “reproduction number”, “R0”, “transmission”. We found several estimates of the basic reproduction number ( $R_0$ ) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), including average exponential growth rate estimates based on inferred or observed cases at a specific timepoint and early growth of the outbreak in China. However, we identified no estimates of how  $R_0$  had changed in Wuhan since control measures were introduced in late January or estimates that jointly fitted data within Wuhan to international exported cases and evacuation flights.

#### Added value of this study

Our study combines available evidence from multiple data sources, reducing the dependency of our estimates on a single

timepoint or dataset. We estimate how transmission has varied over time, identify a decline in the reproduction number in late January to almost 1, coinciding with the introduction of large scale control measures, and show the potential implications of estimated transmission for outbreak risk in new locations.

#### Implications of all the available evidence

Coronavirus disease 2019 is currently showing sustained transmission in China, creating a substantial risk of outbreaks in other countries. However, if SARS-CoV-2 has Middle East respiratory syndrome coronavirus-like or SARS-CoV-like variability in transmission at the individual level, multiple introductions might be required before an outbreak takes hold.

and testing capacity.<sup>9</sup> Modelling approaches can account for such delays and uncertainty by explicitly incorporating delays resulting from the natural history of infection and reporting processes.<sup>10</sup> Additionally, individual data sources might be biased, incomplete, or only capture certain aspects of the outbreak dynamics. Evidence synthesis approaches, which fit to multiple data sources rather than a single dataset (or datapoint) can enable more robust estimation of the underlying dynamics of transmission from noisy data.<sup>11,12</sup> Combining a mathematical model of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission with four datasets from within and outside Wuhan, we estimated how transmission in Wuhan varied between December, 2019, and February, 2020. We used these estimates to assess the potential for sustained human-to-human transmission to occur in locations outside Wuhan if cases were introduced.

## Methods

### Data sources

To estimate the early dynamics of transmission in Wuhan, we fitted a stochastic transmission dynamic model<sup>13</sup> to multiple publicly available datasets on cases in Wuhan and internationally exported cases from Wuhan. The four datasets we fitted to were: daily number of new internationally exported cases (or lack thereof), by date of onset, as of Jan 26, 2020; daily number of new cases in Wuhan with no market exposure, by date of onset, between Dec 1, 2019, and Jan 1, 2020; daily number of new cases in China, by date of onset, between Dec 29, 2019, and Jan 23, 2020; and proportion of infected passengers on evacuation flights between Jan 29, 2020, and Feb 4, 2020 (appendix p 3). We used an additional two datasets for comparison with model outputs: daily number of new exported cases from Wuhan (or lack

thereof) in countries with high connectivity to Wuhan (ie, top 20 most at-risk countries), by date of confirmation, as of Feb 10, 2020; and data on new confirmed cases reported in Wuhan between Jan 16, 2020, and Feb 11, 2020 (appendix p 3).

### Procedures

In the model, we divided individuals into four infection classes, as follows: susceptible, exposed (but not yet infectious), infectious, and removed (ie, isolated, recovered, or otherwise no longer infectious; figure 1). The model accounted for delays in symptom onset and reporting by including compartments to reflect transitions between reporting states and disease states. The model also incorporated uncertainty in case observation, by explicitly modelling a Poisson observed process of newly symptomatic cases, reported onsets of new cases, reported confirmation of cases, and a binomial observation process for infection prevalence on evacuation flights (appendix pp 1–3). The incubation period was assumed to be Erlang distributed with mean 5.2 days<sup>14</sup> (SD 3.7) and delay from onset to isolation was assumed to be Erlang distributed with mean 2.9 days (2.1).<sup>2,15</sup> The delay from onset to reporting was assumed to be exponentially distributed with mean 6.1 days (2.5).<sup>2</sup> Once exposed to infection, a proportion of individuals travelled internationally and we assumed that the probability of cases being exported from Wuhan to a specific other country depended on the number of cases in Wuhan, the number of outbound travellers (assumed to be 3300 per day before travel restrictions were introduced on Jan 23, 2020, and zero after), the relative connectivity of different countries,<sup>16</sup> and the relative probability of reporting a case outside Wuhan, to account for differences in clinical case definition, detection, and reporting within

Wuhan and internationally. We considered the 20 countries outside China most at risk of exported cases in the analysis.

We modelled transmission as a geometric random walk process, and we used sequential Monte Carlo simulation to infer the transmission rate over time, as well as the resulting number of cases and the time-varying basic reproduction number ( $R_t$ ), defined here as the mean number of secondary cases generated by a typical infectious individual on each day in a full susceptible population. The model had three unknown parameters, which we estimated: magnitude of temporal variability in transmission, proportion of cases that would eventually be detectable, and relative probability of reporting a confirmed case within Wuhan compared with an internationally exported case that originated in Wuhan. We assumed the outbreak started with a single infectious case on Nov 22, 2019, and the entire population was initially susceptible. Once we had estimated  $R_t$ , we used a branching process with a negative binomial offspring distribution to calculate the probability an introduced case would cause a large outbreak. We also did a sensitivity analysis on the following three key assumptions: we assumed the initial number of cases was ten rather than one; we assumed connectivity between countries followed WorldPop rather than MOBS Lab estimates; and we assumed that cases were infectious during the second half of their incubation period rather than only being infectious while symptomatic. All data and code required to reproduce the analysis is available online.

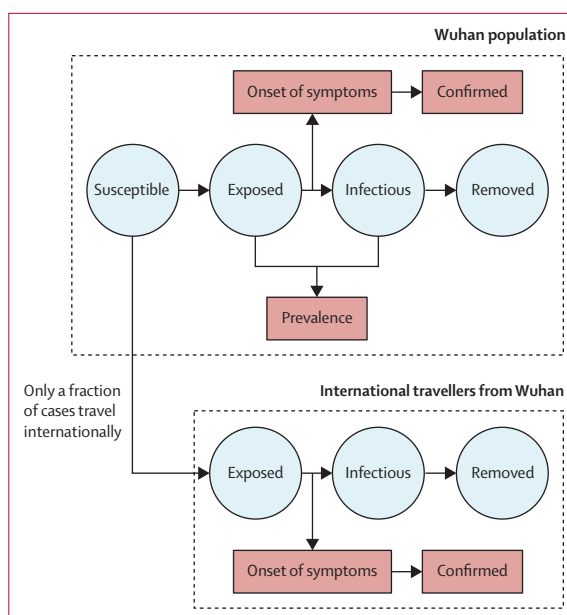
### Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Results

We estimated that  $R_t$  varied during January, 2020, with median values ranging from 1.6 to 2.6 between Jan 1, 2020, and the introduction of travel restrictions on Jan 23, 2020 (figure 2). We estimated a decline in  $R_t$  in late January, from 2.35 (95% CI 1.15–4.77) on January 16, 1 week before the restrictions, to 1.05 (0.41–2.39) on January 31.

The model reproduced the observed temporal trend of cases within Wuhan and cases exported internationally. The model captured the exponential growth in case onsets in early January, the rising number of exported case onsets between Jan 15, and Jan 23, 2020, and the prevalence of infection measured on ten evacuation flights from Wuhan to seven countries. We estimated that 94.8% (95% CI 93.1–96.1%) of the Wuhan population were still susceptible on Jan 31, 2020 (figure 2). Our results suggested there were around



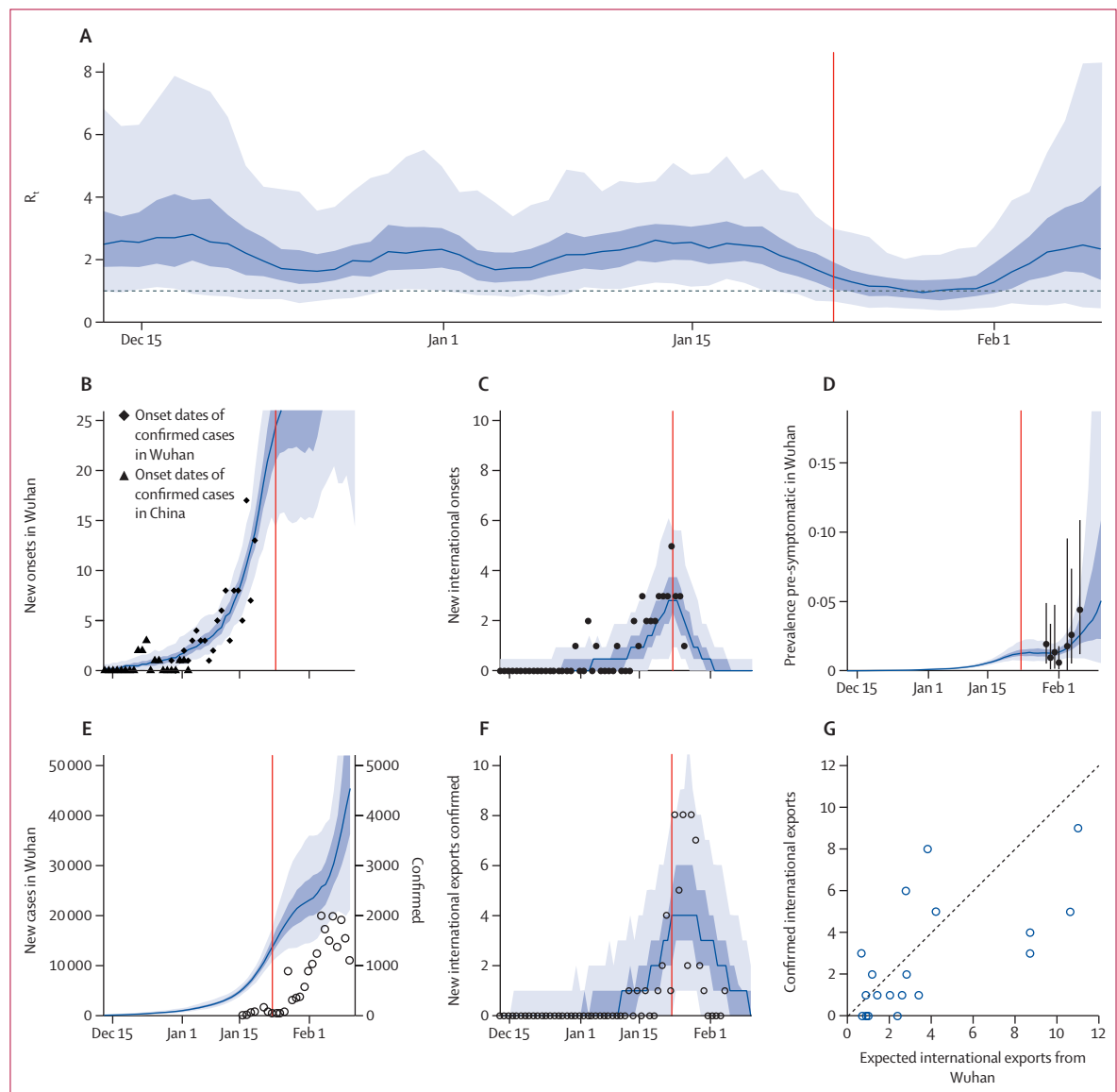
**Figure 1: Model structure**

The population is divided into the following four classes: susceptible, exposed (and not yet symptomatic), infectious (and symptomatic), and removed (ie, isolated, recovered, or otherwise non-infectious). A fraction of exposed individuals subsequently travel and are eventually detected in their destination country.

ten times more symptomatic cases in Wuhan in late January than were reported as confirmed cases (figure 2), but the model did not predict the slowdown in cases that was observed in early February. The model could also reproduce the pattern of confirmed exported cases from Wuhan, which was not explicitly used in the model fitting (figure 2). We found that confirmed and estimated exported cases among the 20 countries most connected to China generally corresponded with each other, with the USA and Australia as notable outliers, having had more confirmed cases reported with a travel history to Wuhan than would be expected in the model (figure 2). There was evidence that the majority of cases were symptomatic. We estimated that 100% (95% CI 51–100) of cases would eventually have detectable symptoms, implying that most infections that were exported internationally from Wuhan in late January were in theory eventually detectable. As a sensitivity analysis, we repeated the analysis with a large number of initial cases, different mobility data, and the assumption that pre-symptomatic cases could transmit. In these analyses, we observed the same result of a decline in  $R_t$  from more than 2 to almost 1 in the last 2 weeks of January, 2020 (appendix pp 10–13).

To examine the potential for new outbreaks to establish in locations outside Wuhan, we used our estimates of the  $R_t$  to simulate new outbreaks with potential individual-level variation in transmission (ie, so called superspreading events).<sup>17–19</sup> Such variation increases the fragility of transmission chains, making it

For data and code required to reproduce the analysis see <https://github.com/adamkucharski/2020-ncov/>



**Figure 2: Dynamics of transmission in Wuhan, fitted up to Feb 11, 2020**

The red line marks travel restrictions starting on Jan 23, 2020. For parts (A) to (F) blue lines represent median, light blue shading represents 50% confidence intervals of the model estimate, and dark blue shading represents 95% confidence intervals of the model estimate. In all panels, datasets that were fitted to are shown as solid points; non-fitted data are shown as empty circles. (A) Estimated  $R_t$  over time. The dashed line represents an  $R_t$  of 1. (B) Onset dates of confirmed cases in Wuhan and China. (C) Reported cases by date of onset (black points) and estimated internationally exported cases from Wuhan by date of onset (blue line). (D) Estimated prevalence of infections that did not have detectable symptoms (blue line), and proportion of passengers on evacuation flights that tested positive for severe acute respiratory syndrome coronavirus 2 (black points; error bars show 95% binomial CIs). (E) New confirmed cases by date in Wuhan (circles, right hand axis) and estimated new symptomatic cases (blue line, left hand axis). (F) International exportation events by date of confirmation of case, and expected number of exports in the fitted model. (G) Estimated number of internationally exported cases from Wuhan confirmed up to Feb 10, 2020 and observed number in 20 countries with the highest connectivity to China.  $R_t$ =daily reproduction number.

less likely that an outbreak will take hold following a single introduction. If transmission is more homogeneous, with all infectious individuals generating a similar number of secondary cases, it is more likely than an outbreak will establish.<sup>18</sup> Based on the median  $R_t$  estimated during January before travel restrictions were introduced, we estimated that a single introduction of SARS-CoV-2 with SARS-like or Middle East

respiratory syndrome (MERS)-like individual-level variation in transmission would have a 17% to 25% probability of causing a large outbreak (figure 3). Assuming SARS-like variation and Wuhan-like transmission, we estimated that once four or more infections have been introduced into a new location, there is an over 50% chance that an outbreak will occur (figure 3).

## Discussion

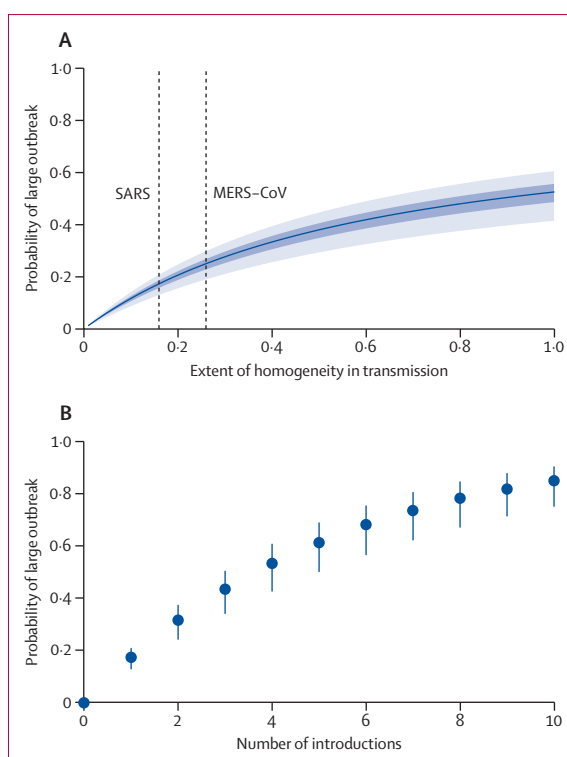
Combining a mathematical model with multiple datasets, we found that the median daily  $R_t$  of SARS-CoV-2 in Wuhan probably varied between 1.6 and 2.6 in January, 2020, before travel restrictions were introduced. We also estimated that transmission declined by around half in the 2 weeks spanning the introduction of restrictions.

The estimated fluctuations in  $R_t$  were driven by the rise and fall in the number of cases, both in Wuhan and internationally, as well as prevalence on evacuation flights. Such fluctuations could be the result of changes in behaviour in the population at risk, or specific superspreading events that inflated the average estimate of transmission.<sup>17–19</sup> We found some evidence of a reduction in  $R_t$  in the days before the introduction of travel restrictions in Wuhan, which might have reflected outbreak control efforts or growing awareness of SARS-CoV-2 during this period. The uncertainty in our estimates for  $R_t$  following the decline in early February, 2020, results from a paucity of data sources to inform changes in transmission during this period.

Comparing model predictions with observed confirmed cases reported in Wuhan, we found that the model predicted at least ten times higher cases than were reported in early February, 2020. The model also did not predict the more recent slowdown in cases, suggesting that transmission might have declined more than our model—which did not fit to this case data—estimated during early February, 2020. Our estimates for international cases in specific countries were broadly consistent with the number of subsequently confirmed exported cases outside Wuhan. However, there were notably more cases exported to France, USA, and Australia compared with what our model predicted. This could be the result of increased surveillance and detection as awareness of SARS-CoV-2 increased in late January, which would suggest earlier exported cases might have been missed, or could be the result of increased travel out of Wuhan immediately before introduction of travel restrictions on Jan 23, 2020.

Based on our estimated reproduction number and published estimates of individual-level variation in transmission for SARS-CoV and MERS-CoV, we found that a single case introduced to a new location would not necessarily lead to an outbreak. Even if the reproduction number is as high as in Wuhan in early January, it could take several introductions for an outbreak to establish, because high individual-level variation in transmission makes new chains of transmission more fragile, and hence it becomes less likely that a single infection will generate an outbreak. This factor highlights the importance of rapid case identification and subsequent isolation and other control measures to reduce the chance of onward chains of transmission.<sup>20</sup>

Our analysis highlights the value of combining multiple data sources in analysis of COVID-19. For



**Figure 3: Risk that introduced infections will establish in a new population** (A) Probability that a single case will lead to a large outbreak for different assumptions about the extent of homogeneity in individual-level transmission (ie, the dispersion parameter  $k$  in a negative binomial offspring process). Results are shown for the median reproduction number estimated for severe acute respiratory syndrome coronavirus 2 in Wuhan between Jan 1, 2020, and Jan 23, 2020. (B) Probability that a given number of introductions will result in a large outbreak, assuming SARS-like superspreading events can occur. Points show the median estimated reproduction number between Jan 1, 2020, and Jan 23, 2020; bars show 95% quantile of the range of median values of  $R_t$  during this period. SARS=severe acute respiratory syndrome. MERS=Middle East respiratory syndrome.  $R_t$ =daily reproduction number.

example, the rapid growth of confirmed cases globally during late January, 2020, with case totals in some instances apparently doubling every day or so, would have had the effect of inflating  $R_t$  estimates to implausibly large values if only these recent datapoints were used in our analysis. Our results also have implications for estimation of transmission dynamics using the number of exported cases from a specific area.<sup>21</sup> Once extensive travel restrictions are introduced, as they were in Wuhan, the signal from such data gets substantially weaker. If restrictions and subsequent delays in detection of cases are not accounted for, this could lead to artificially low estimates of  $R_t$  or inferred case totals from the apparently declining numbers of exported cases. Our model estimates benefited from the availability of testing data from evacuation flights, which allowed us to estimate current prevalence. Having such information for other settings, either through widespread testing or serological surveillance, will be valuable to reduce reliance on case reports alone.



There are several other limitations to our analysis. We used plausible biological parameters for SARS-CoV-2 based on current evidence, but these values might be refined as more comprehensive data become available. However, by fitting to multiple datasets to infer model parameters, and conducting sensitivity analyses on key areas of uncertainty, we have attempted to make the best possible use of the available evidence about SARS-CoV-2 transmission dynamics. Furthermore, we used publicly available connectivity and risk estimates based on international travel data to predict the number of cases exported into each country. These estimates have shown good correspondence with the distribution of exported cases to date,<sup>22</sup> and are similar to another risk assessment for COVID-19 with different data.<sup>23</sup> We also assumed that the latent period is equal to the incubation period (ie, individuals become infectious and symptomatic at the same time) and all infected individuals will eventually become symptomatic. However, there is evidence that transmission of SARS-CoV-2 can occur with few reported symptoms.<sup>24</sup> Therefore, we did a sensitivity analysis in which transmission could occur in the second half of the incubation period, but this did not change our overall conclusions of a decline in  $R_t$  from around 2.4 to almost 1 during the last 2 weeks of January. We also explored having a larger initial spillover event and using different sources for flight connectivity data, both of which produced the same conclusion about the decline in transmission. In our analysis of new outbreaks, we also used estimates of individual-level variation in transmission for SARS and MERS-CoV to illustrate potential dynamics. However, it remains unclear what the precise extent of such variation is for SARS-CoV-2.<sup>17</sup> If transmission were more homogenous than SARS-CoV or MERS-CoV, it would increase the risk of outbreaks following introduced cases. As more data become available, it will be possible to refine these estimates; therefore we have made an online tool so that users can explore these risk estimates if new data become available (appendix p 4).

Our results show that there was probably substantial variation in SARS-CoV-2 transmission over time, and suggest a decline in transmission in Wuhan in late January, 2020, around the time that control measures were introduced. If COVID-19 transmission is established outside Wuhan, understanding the effectiveness of control measures in different settings will be crucial for understanding the dynamics of the outbreak, and the likelihood that transmission can eventually be contained or effectively mitigated.

#### Contributors

Data analysis was led by AJK, who programmed the model with help from TWR. AJK, SF, and RME planned the inference framework. CD provided the data from online sources. The CMMID 2019-nCoV working group members contributed to processing, cleaning, and interpretation of data, interpreted the study findings, contributed to the manuscript, and approved the work for publication. All authors interpreted the

findings, contributed to writing the manuscript, and approved the final version for publication.

#### Declaration of interests

We declare no competing interests.

#### Acknowledgments

We would like to thank Motoi Suzuki for his help in identifying the flight evacuation data sources.

#### References

- 1 WHO. Coronavirus disease 2019 (COVID-19). Situation report 24. February 13, 2020. Geneva: World Health Organization, 2020.
- 2 nCoV-2019 Data Working Group. Epidemiological data from the nCoV-2019 outbreak: early descriptions from publicly available data. 2020. <http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-available-data/337> (accessed Feb 13, 2020).
- 3 Camacho A, Kucharski A, Aki-Sawyer Y, et al. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLoS Curr* 2015; 7.
- 4 Funk S, Ciglenecki I, Tiffany A, et al. The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. *Philos Trans R Soc Lond B Biol Sci* 2017; 372: 20160302.
- 5 Riley S, Fraser C, Donnelly CA, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* 2003; 300: 1961–66.
- 6 Viboud C, Sun K, Gaffey R, et al. The RAPIDD Ebola forecasting challenge: synthesis and lessons learnt. *Epidemics* 2018; 22: 13–21.
- 7 Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the international spread of pandemic influenza. *PLoS Med* 2006; 3: e212.
- 8 Kucharski AJ, Camacho A, Checchi F, et al. Evaluation of the benefits and risks of introducing Ebola community care centers, Sierra Leone. *Emerg Infect Dis* 2015; 21: 393–99.
- 9 Aylward B, Barboza P, Bawo L, et al. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014; 371: 1481–95.
- 10 Nishiura H, Klinkenberg D, Roberts M, Heesterbeek JAP. Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. *PLoS One* 2009; 4: e6852.
- 11 Birrell PJ, De Angelis D, Presanis AM. Evidence synthesis for stochastic epidemic models. *Stat Sci* 2018; 33: 34–43.
- 12 Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Med* 2013; 10: e1001527.
- 13 Dureau J, Kalogeropoulos K, Baguelin M. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics* 2013; 14: 541–55.
- 14 Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; published online Jan 29. DOI:10.1056/NEJMoa2001316.
- 15 Liu T, Hu J, Kang M, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv* 2020; published online Feb 13. DOI:10.1101/2020.01.25.919787.
- 16 MOBS Lab. Situation report mainland China. <https://datastudio.google.com/u/0/reporting/3fdd36c3-0272-4510-a140-39e288a9f15c> (accessed Jan 30, 2020).
- 17 Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Euro Surveill* 2020; published online Jan 30. DOI:10.2807/1560-7917.ES.2020.25.4.2000058.
- 18 Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005; 438: 355–59.
- 19 Kucharski AJ, Althaus CL. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Euro Surveill* 2015; 20: 14–18.
- 20 Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis C. Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts. *medRxiv* 2020; published online Feb 11. DOI:10.1101/2020.02.08.20021162.

- 
- 21 Imai N, Dorigatti I, Cori A, et al. Report 2: estimating the potential total number of novel coronavirus cases in Wuhan City, China. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/2019-nCoV-outbreak-report-22-01-2020.pdf> (accessed Feb 20, 2020).
- 22 Pullano G, Pinotti F, Valdano E, Boëlle P-Y, Poletto C, Colizza V. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Euro Surveill* 2020; **25**: 2000057.
- 23 Lai S, Bogoch II, Watts A, Khan K, Li Z, Tatem A. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China. <https://www.worldpop.org/events/china> (accessed Jan 30, 2020).
- 24 Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 2020; published online Jan 30. DOI:10.1056/NEJMc2001468.

# THE LANCET

## Infectious Diseases

### **Supplementary appendix**

This appendix formed part of the original submission and has been peer reviewed.  
We post it as supplied by the authors.

Supplement to: Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* 2020; published online March 11. [http://dx.doi.org/10.1016/S1473-3099\(20\)30144-4](http://dx.doi.org/10.1016/S1473-3099(20)30144-4).



## **Appendix for Early dynamics of transmission and control of COVID-19: a mathematical modelling study**

*Adam J Kucharski<sup>1</sup>, Timothy W Russell<sup>1</sup>, Charlie Diamond<sup>1</sup>, Yang Liu<sup>1</sup>, CMMID COVID-19 working group\*, John Edmunds<sup>1</sup>, Sebastian Funk<sup>1</sup>, Rosalind M Eggo<sup>1</sup>*

*<sup>1</sup>Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London.*

*\*The following authors were part of the CMMID nCoV working group: Fiona Sun, Mark Jit, James D Munday, Nicholas Davies, Amy Gimma, Kevin van Zandvoort, Hamish Gibbs, Joel Hellewell, Christopher I Jarvis, Sam Clifford, Billy J Quilty, Nikos I Bosse, Sam Abbott, Petra Klepac and Stefan Flasche.*

### **Supplementary methods**

#### **Transmission model structure**

We formulated an extended version of an SEIR model to reflect the disease-specific dynamics. The model extensions include: a class for asymptomatic cases, two cumulative classes - one for symptomatic cases and the other for confirmed cases – and two identical classes for exposed and infected individuals resulting in Erlang distributed incubation and infectious periods respectively. We simulated the dynamics using the Euler-Maruyama numerical scheme with the noise term corresponding to a stochastic transmission rate, which we modelled as standard geometric Brownian motion. The resulting system of stochastic difference equations are described schematically in Figure 1 in the main text and are given explicitly below:

*Model compartments for individuals in Wuhan:*

$$S(t+1) = S(t) - \beta(t) S(t) [I_{1w}(t) + I_{2w}(t)] / N$$

$$E_{1w}(t+1) = E_{1w}(t) + (1 - f) \beta(t) S(t) [I_{1w}(t) + I_{2w}(t)] / N - 2\sigma E_{1w}(t)$$

$$E_{2w}(t+1) = E_{2w}(t) + 2\sigma E_{1w}(t) - 2\sigma E_{2w}(t)$$

$$I_{1w}(t+1) = I_{1w}(t) + 2\sigma E_{2w}(t) - 2\gamma I_{1w}(t)$$

$$I_{2w}(t+1) = I_{2w}(t) + 2\gamma I_{1w}(t) - 2\gamma I_{2w}(t)$$

$$Q_w(t+1) = Q_w(t) + 2\sigma E_{2w}(t) e^{-\gamma\kappa} - \kappa Q_w(t)$$

$$D_w(t+1) = D_w(t) + 2\sigma E_{2w}(t) e^{-\gamma\kappa}$$

$$C_w(t+1) = C_w(t) + \kappa Q_w(t)$$

Here  $S(t)$  is the number of individuals in Wuhan susceptible at time  $t$ ;  $E_{1w}(t)$  and  $E_{2w}(t)$  are individuals in Wuhan the first and second period of their Erlang distributed incubation period respectively;  $I_{1w}(t)$  and  $I_{2w}(t)$  are individuals in Wuhan in the first and second period of their Erlang distributed infectious period respectively;  $Q_w(t)$  is the number of symptomatic cases in Wuhan yet to be reported at time  $t$ ;  $D_w(t)$  is the

cumulative number of cases with symptoms in Wuhan at time  $t$ ;  $C_w(t)$  is the cumulative number of confirmed cases in Wuhan at time  $t$ . We include two compartments for I and E because the combination of these compartments with exponentially distributed waiting times creates an Erlang distribution with shape parameter 2, which is a more realistic representation of the peaked distribution of incubation and infectious period for 2019-nCoV (Figure S1).

Here  $\beta(t)$  is the transmission rate at time  $t$ ;  $\sigma$  is the rate of becoming symptomatic (i.e. 1/incubation period);  $\gamma$  = rate of isolation (i.e. 1/delay from onset-to-hospitalisation);  $\kappa$  is rate of reporting (i.e. 1/delay from onset-to-confirmation);  $f$  is the fraction of cases that travel;  $N$  is the population size in Wuhan. We therefore implicitly assume that all individuals become symptomatic, and this happens at the same time as they become infectious. The  $e^{-\gamma\kappa}$  term in the equation for  $Q_w(t)$  accounts for the probability that some cases may recover before being confirmed. A full parameter description is provided in Table S1.

*Model compartments for traveller cases from Wuhan:*

$$E_{1T}(t+1) = E_{1T}(t) + f\beta(t) S(t) [I_{1w}(t) + I_{2w}(t)] - 2\sigma E_{1T}(t)$$

$$E_{2T}(t+1) = E_{2T}(t) + 2\sigma E_{1T}(t) - 2\sigma E_{2T}(t)$$

$$I_{1T}(t+1) = I_{1T}(t) + 2\sigma E_{2T}(t) - 2\gamma I_{1T}(t)$$

$$I_{2T}(t+1) = I_{2T}(t) + 2\gamma I_{1T}(t) - 2\gamma I_{2T}(t)$$

$$Q_T(t+1) = Q_T(t) + 2\sigma E_{2T}(t) e^{-\gamma\kappa} - \kappa Q_T(t)$$

$$D_T(t+1) = D_T(t) + 2\sigma E_{2T}(t) e^{-\gamma\kappa}$$

$$C_T(t+1) = C_T(t) + \kappa Q_T(t)$$

Here  $E_{1T}(t)$  and  $E_{2T}(t)$  are individuals who have travelled from Wuhan and who are in the first and second period of their Erlang distributed incubation period respectively;  $I_{1T}(t)$  and  $I_{2T}(t)$  are individuals who have travelled from Wuhan and who are in the first and second period of their Erlang distributed infectious period respectively;  $Q_T(t)$  is the number of symptomatic cases among travellers from Wuhan yet to be reported at time  $t$ ;  $D_T(t)$  is the cumulative number of cases among travellers from Wuhan with symptoms at time  $t$ ;  $C_T(t)$  is the cumulative number of confirmed cases among travellers from Wuhan at time  $t$ .

Transmission is modelled as geometric Brownian motion:

$$d \log(\beta(t)) = a dB_t$$

where  $a$  is the volatility of transmission over time and  $B_t$  is Brownian motion. As the Euler-Maruyama algorithm works in discrete time, at each step, we implement this geometric Brownian motion by sampling a random normally distributed number  $X$  with mean 0 and standard deviation  $a$ , then define  $\beta(t+1) = \beta(t)e^X$ .

### Internationally exported cases

Cases that travelled internationally and became symptomatic,  $D_T(t)$ , and were later confirmed,  $C_T(t)$ , were distributed among other countries based on proportional risk inferred from connectivity to those countries (1,2). For example, if a country had a relative risk of export  $W(country)$  from Wuhan, then we would expect there to be  $D_T(t) W(country)$  new symptomatic cases in this country at time  $t$ . We assumed no travel out of Wuhan occurred after 23rd January 2020, when extensive restrictions were put in place in Wuhan.

### Model fitting

We estimated the time-varying transmission rate,  $\beta(t)$ , using sequential Monte Carlo (SMC) by jointly fitting to three datasets, with one used for validation:

1. Daily number of exported new cases from Wuhan (or lack thereof) in countries with high connectivity to Wuhan (i.e. top 20 most at risk), by date of onset. We only consider onsets up to 26<sup>th</sup> January 2020, as many of the cases detected after this point were not travellers from Wuhan (3).
2. Daily number of new cases in Wuhan (subsequently confirmed) with no market exposure, by date of onset, between 1st December 2019 and 1st January 2020 (4). This is combined with daily number of new cases in China, by date of onset (also subsequently confirmed), between 29th December 2019 and 23rd January 2020, which we assume that these are all in Wuhan (5). The most recent five data points from the China data were omitted during fitting as they were likely to be strongly influenced by delays in reporting.
3. Prevalence of infection on evacuation flights from Wuhan (see Table S3 for more details).
4. Validation dataset (not used for fitting): Data on confirmed cases in Wuhan from 16<sup>th</sup> January to 12<sup>th</sup> February (6)
5. Validation dataset (not used for fitting): Daily number of new exported infection from Wuhan (or lack thereof) in countries with high connectivity to Wuhan (i.e. top 20 most at risk), by date of confirmation, up to 10<sup>th</sup> February (7). We only considered individual export events (i.e. a family of travellers was counted as a single export).

To calculate the likelihood, we first used the model outputs to calculate expected values for each of the datasets we were fitting to:

1. International onset data was not disaggregated by country and so we used the total daily exported cases from Wuhan as our expected value, scaled by the proportion that would have detectable symptoms and proportion with known onsets:  $D_T(t) \omega \rho_T$ .
2. The number of new onsets per unit time reported in Wuhan was given the number of new onsets in the model, scaled by the relative reporting probability within Wuhan, the proportion that were detectable, and the proportion of confirmed cases with known onset:  $D_W(t) \omega \delta \rho_W$ .

3. The number of new confirmed cases in Wuhan was given by the number of new confirmed cases in the model, scaled by the proportion that were detectable and relative reporting probability within Wuhan  $C_w(t) \omega \delta$ .
4. The proportion of positive evacuated cases was assumed to be equivalent to the proportion of infectious individuals in Wuhan that would not be detected by thorough symptom screening at departure and were infected sufficiently long ago to have a positive test on arrival (as it takes time to test positive (8)). We therefore assumed infected individuals in Wuhan in the second half of their incubation period would be detectable at arrival, either through testing or subsequent appearance of symptoms during travel. This fraction is equivalent to:  $[E_{2w}(t) + (1 - \omega) (I_{1w}(t) + I_{2w}(t))] / N$ .
5. As destination country was known for confirmed exported cases, we generated 20 timeseries for cases exported (or not) to most at-risk countries each day (see Table S2). Hence we would expect  $C_T(t) W(\text{country})$  cases to be confirmed in a specific country at time  $t$ ; the total number of confirmed cases to date is therefore the sum of these values up to the date being considered.

Here  $\omega$  is the proportion of cases that would be eventually have detectable symptoms,  $\rho_w$  and  $\rho_T$  are the proportion of confirmed cases in Wuhan and exported from Wuhan that have known onsets and is  $\delta$  the relative proportion of cases that are reported in Wuhan compared to internationally. We used a Poisson observation model fitted jointly to datasets 1–3, using the expected values from model outputs 1–3 above.

The time-varying  $R_t$  was defined by  $\beta(t)/\gamma$ . Estimates for  $\beta(t)$  were generated by running 200 repetitions of SMC with 2000 particles. The transmission volatility ( $\alpha$ ) and relative reporting of cases inside Wuhan ( $\delta$ ) were selected based on a grid search. Profile likelihoods for each parameter were constructed based on the joint likelihood distribution, with splines fitted to the estimates to calculate the MLE and 95% CI, defined as the point at which the spline crossed a line 1.92 units below the MLE (Figure S2–S4). Results shown in the main text use the MLE. We assumed the outbreak started on 22nd November with 1 infectious individual (9) and the population was initially fully susceptible. We assumed that the population in Wuhan was 11m. We also assumed all internationally exported infections would be eventually be detected in destination country if they travelled by plane.

### Branching process simulation model

We used a branching process with a negative binomial offspring distribution (10, 11) to calculate the probability  $P$  that an outbreak starting with a single imported case would fail to go extinct (i.e. would cause a large outbreak). We also calculated the probability that an outbreak would occur after  $n$  introductions:  $1 - (1 - P)^n$ . We used the following representative parameter values in our analysis: 0.16 (95% CI: 0.11–0.64) for SARS-CoV and 0.25 (95% CI: 0.09–0.91) for MERS-CoV (12), and 0.54 (95% CI: 0.014 –  $\infty$ ) for SARS-CoV-2 (13). We have built a Shiny online tool so these results can be explored further:

[https://cmmid-lshtm.shinyapps.io/new\\_outbreak\\_probability/](https://cmmid-lshtm.shinyapps.io/new_outbreak_probability/)

### Data and code availability

All data and code required to reproduce the analysis is available at:

[https://github.com/adamkucharski/2020-ncov/stoch\\_model\\_V2\\_paper](https://github.com/adamkucharski/2020-ncov/stoch_model_V2_paper)

### Sensitivity analysis

To check the robustness of our results, we repeated our analysis with a larger assumed initial number of cases (i.e. 10 rather than 1), but this did not change our overall conclusions (Figure S5). As another sensitivity analysis, we used different flight data, taken from WorldPop rather than MOBS lab. However, again this did not change our overall conclusions (Figure S6). We also considered the effect of allowing individuals in the second half of their incubation period (i.e.  $E_2$ ) to be infectious; this did not change our overall conclusions either (Figure S7).

Parameter	Value	Distribution	Source
Incubation period	5.2 days	Erlang (rate=2)	(15)
Infectious period	2.9 days	Erlang (rate=2)	(5)
Delay onset-to-confirmation	6.1 days	Exponential	(7)
Daily outbound passengers	3300		(14)
Population of Wuhan	11m		(14)
Fraction of cases that travel	3300/11m		
Initial cases	1		(9)
Introduction date	22nd November 2019		(9)
Proportion of Wuhan cases with onsets known ( $\rho_w$ )	0.16		(7)
Proportion of internationally exported cases with onsets known ( $\rho_T$ )	0.47		(3)
Relative reporting inside Wuhan compared to internationally ( $\delta$ )	0.0066 (95% CI: 0.0029-0.009)		Fitted
Transmission volatility ( $a$ )	0.395 (95% CI: 0.138-0.693)		Fitted
Proportion of infections that are eventually detectable ( $\omega$ )	1 (95% CI: 0.518-1)		Fitted

Table S1: Parameters used in the model, and assumed values.

<b>Country</b>	<b>Relative risk</b>
Thailand	0.1382
Japan	0.1334
Taiwan	0.1095
Korea	0.1094
USA	0.0533
Singapore	0.0483
Malaysia	0.0430
Vietnam	0.0358
Australia	0.0353
Philippines	0.0330
Indonesia	0.0303
Cambodia	0.0258
United Kingdom	0.0182
Canada	0.0151
Russian Federation	0.0127
Germany	0.0119
United Arab Emirates	0.0118
India	0.0113
Italy	0.0092
France	0.0088

*Table S2: List at risk countries used in main analysis, and relatively probability of export to this country.*



<b>Destination (country)</b>	<b>Date</b>	<b>No. of passengers</b>	<b>Confirmed cases</b>	<b>Source</b>
Singapore	30/01/2020	92	2	<a href="https://www.straitstimes.com/singapore/all-92-evacuated-from-wuhan-on-scoot-flight-being-tested">https://www.straitstimes.com/singapore/all-92-evacuated-from-wuhan-on-scoot-flight-being-tested</a>
Belgium	02/02/2020	9	1	<a href="https://www.channelnewsasia.com/news/world/wuhan-coronavirus-virus-belgium-tests-positive-china-12394176">https://www.channelnewsasia.com/news/world/wuhan-coronavirus-virus-belgium-tests-positive-china-12394176</a>
Korea	31/01/2020	368	1	<a href="https://www.cdc.go.kr/board/board.es?mid=a30402000000&amp;bid=0030&amp;act=view&amp;list_no=365941&amp;tag=&amp;nPage=2">https://www.cdc.go.kr/board/board.es?mid=a30402000000&amp;bid=0030&amp;act=view&amp;list_no=365941&amp;tag=&amp;nPage=2</a>
Malaysia	04/02/2020	107	2	<a href="https://www.channelnewsasia.com/news/asia/wuhan-coronavirus-malaysia-two-positive-evacuated-12396282">https://www.channelnewsasia.com/news/asia/wuhan-coronavirus-malaysia-two-positive-evacuated-12396282</a>
Japan	29/01/2020	206	4	<a href="https://www.mhlw.go.jp/stf/newpage_09199.html">https://www.mhlw.go.jp/stf/newpage_09199.html</a>
Japan	30/01/2020	210	2	<a href="https://www.mhlw.go.jp/stf/newpage_09205.html">https://www.mhlw.go.jp/stf/newpage_09205.html</a>
Japan	31/01/2020	149	4	<a href="https://www.mhlw.go.jp/stf/newpage_09278.html">https://www.mhlw.go.jp/stf/newpage_09278.html</a>
Germany	01/02/2020	128	2	<a href="https://www.dw.com/en/coronavirus-german-evacuation-flight-lands-in-frankfurt/a-52223609">https://www.dw.com/en/coronavirus-german-evacuation-flight-lands-in-frankfurt/a-52223609</a>
Japan	30/01/2020	206	3	<a href="https://www.reuters.com/article/china-health-japan/update-5-three-japanese-evacuees-from-wuhan-test-positive-for-virus-2-had-no-symptoms-idUSL4N29Z0CS">https://www.reuters.com/article/china-health-japan/update-5-three-japanese-evacuees-from-wuhan-test-positive-for-virus-2-had-no-symptoms-idUSL4N29Z0CS</a>
Italy	02/02/2020	56	1	<a href="https://www.thelocal.it/20200203/coronavirus-italians-evacuated-from-china-quarantined-rome">https://www.thelocal.it/20200203/coronavirus-italians-evacuated-from-china-quarantined-rome</a>

*Table S3: Data on prevalence among evacuated individuals from Wuhan.*

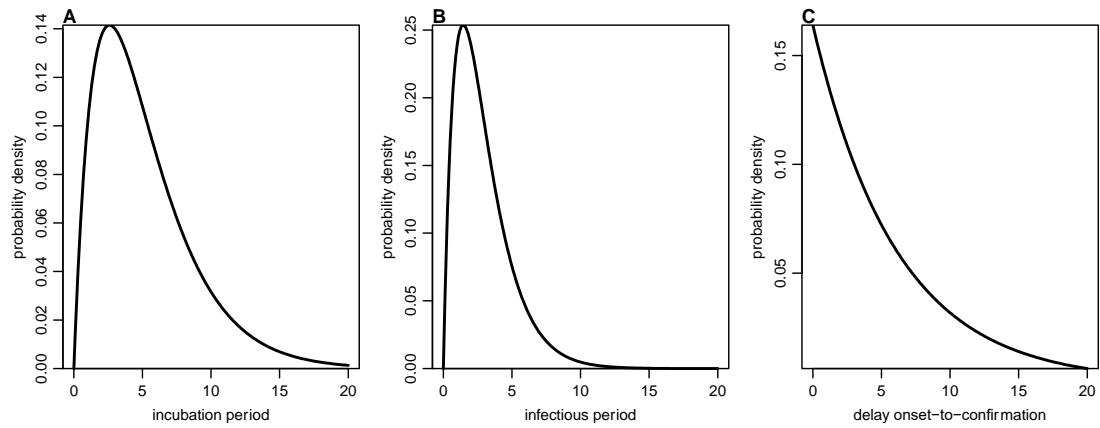


Figure S1: Assumed distribution of parameters. A) Erlang distributed incubation period. B) Erlang distributed infectious period. C) Exponentially distributed time from onset to report.

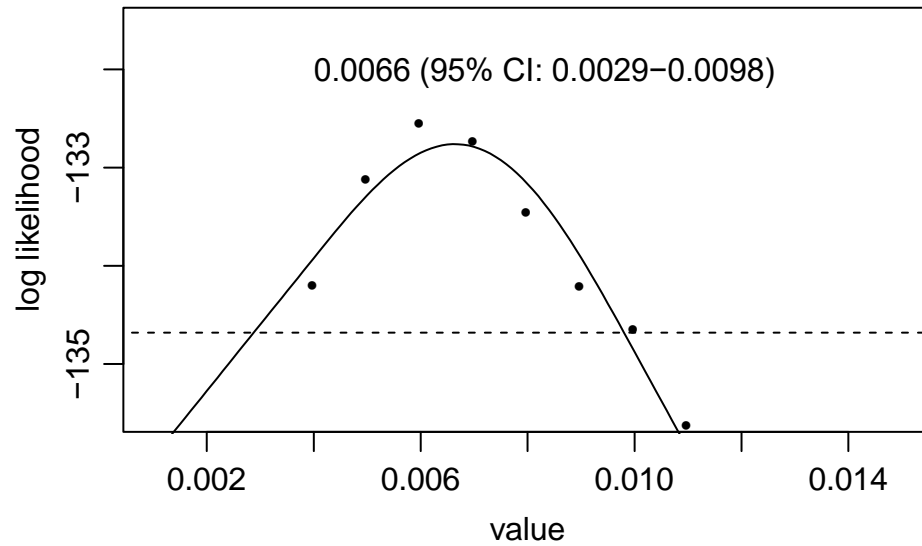


Figure S2: Profile likelihood for the relative reporting of confirmed cases within Wuhan compared to internationally exported cases ( $\delta$ ).

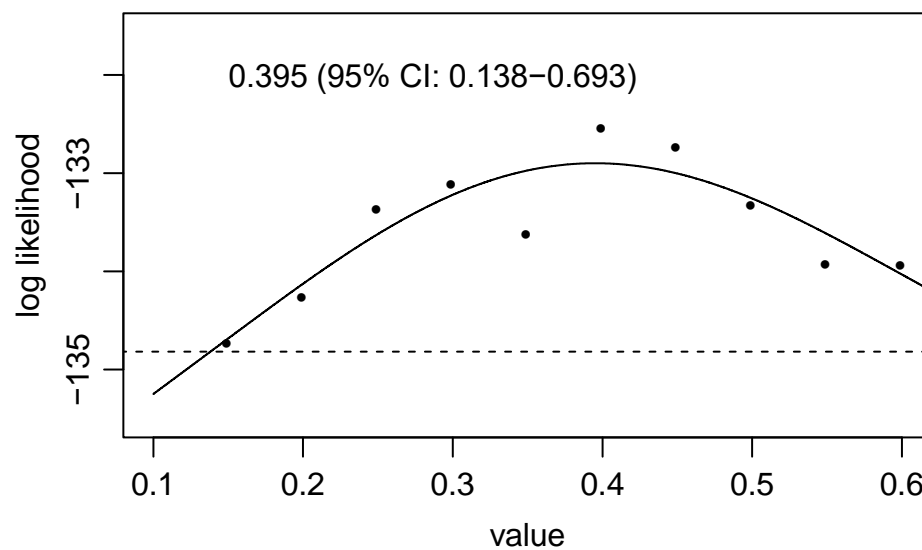


Figure S3: Profile likelihood for the transmission volatility parameter ( $a$ ).

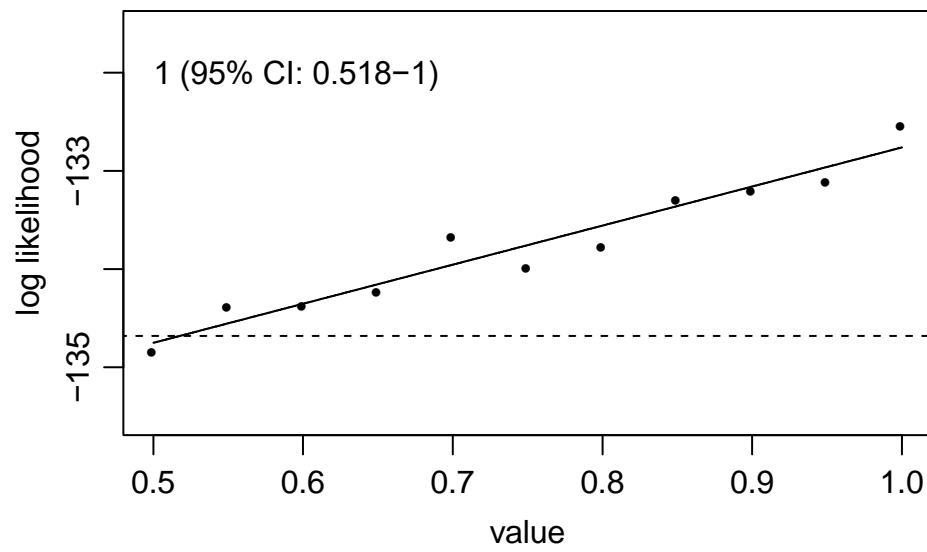


Figure S4: Profile likelihood for the proportion of cases that eventually have detectable symptoms ( $\omega$ ).

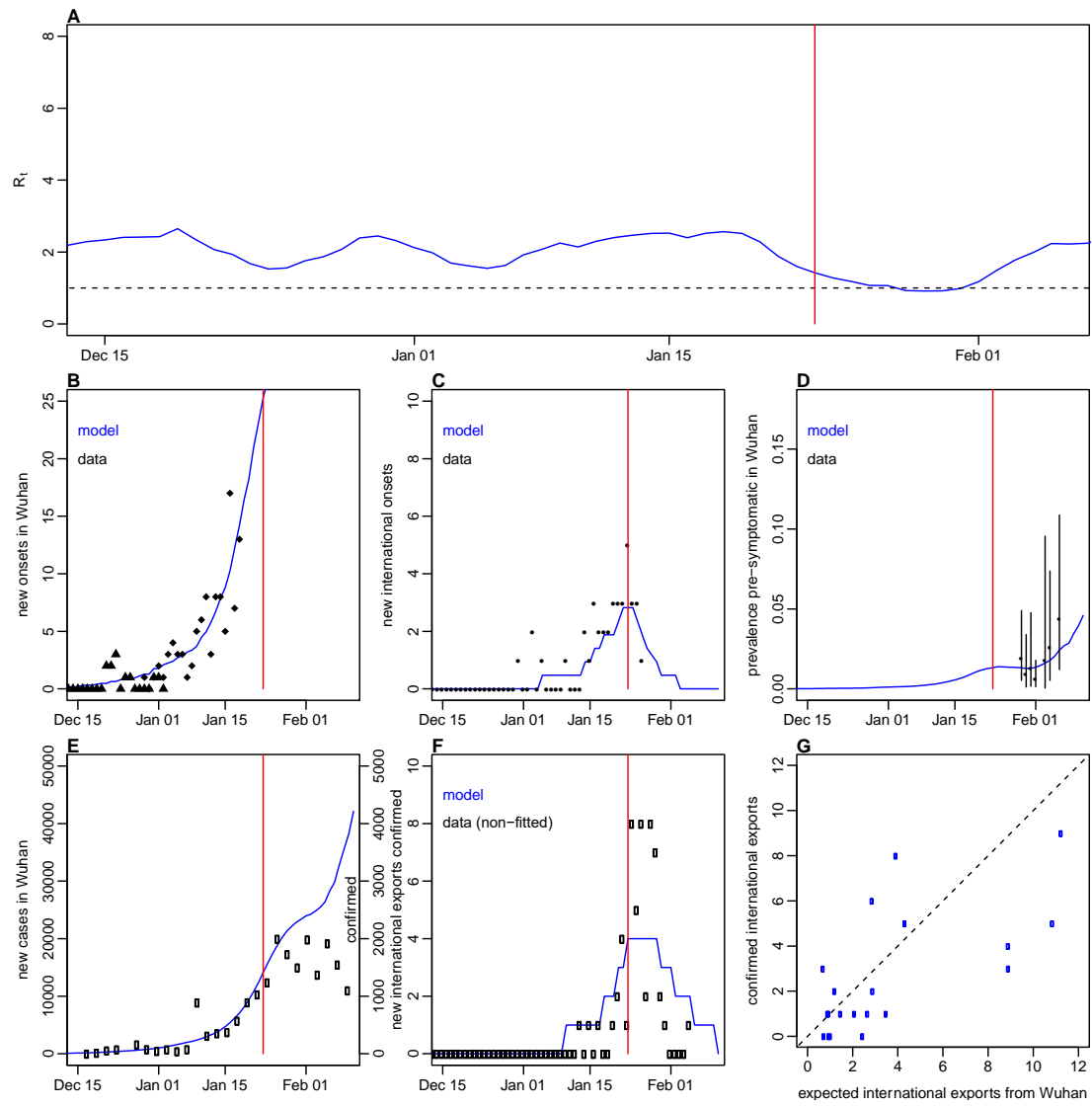


Figure S5: Model outputs when we assumed 10 initial cases rather than 1 on 22nd November 2019.

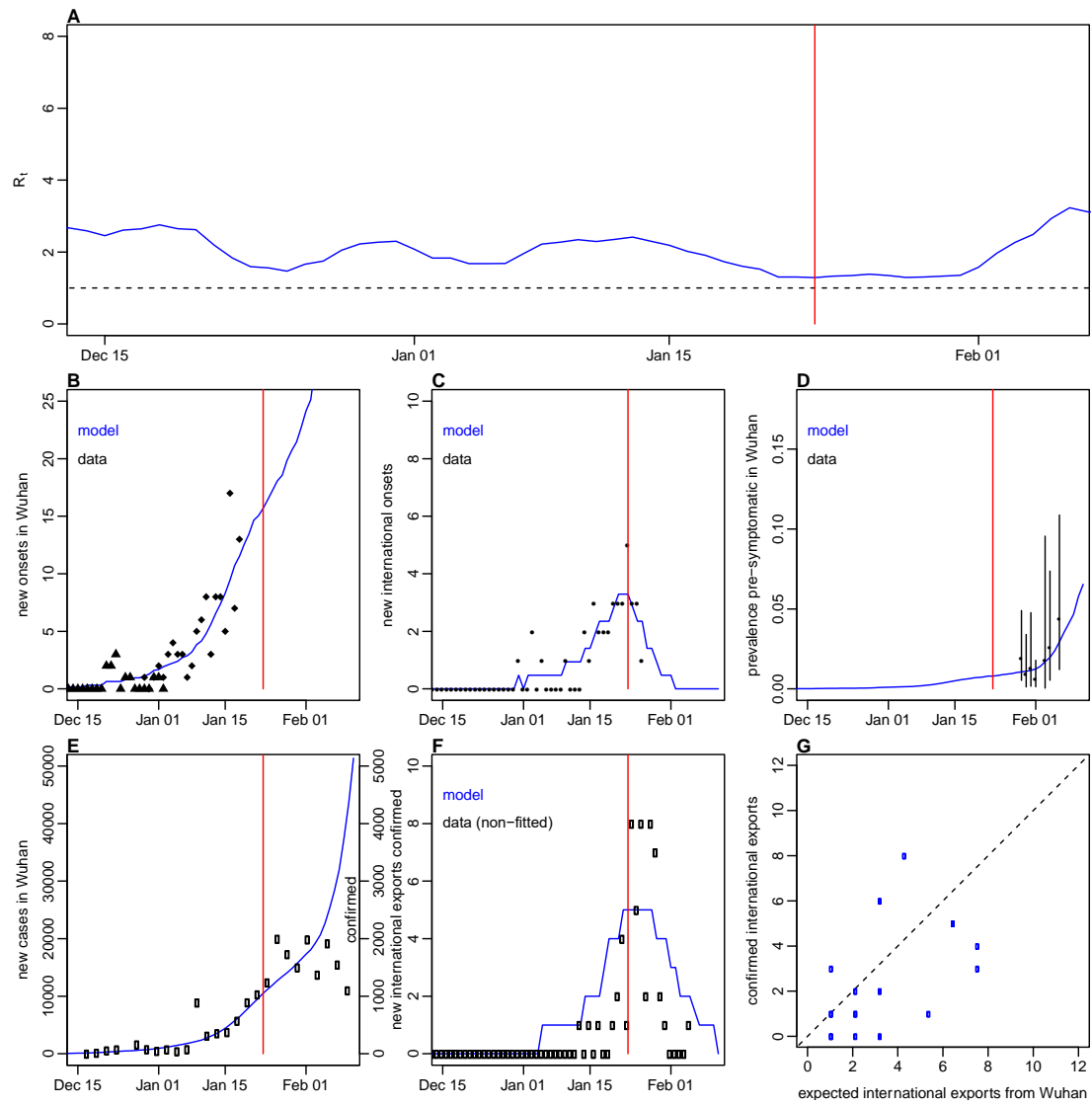
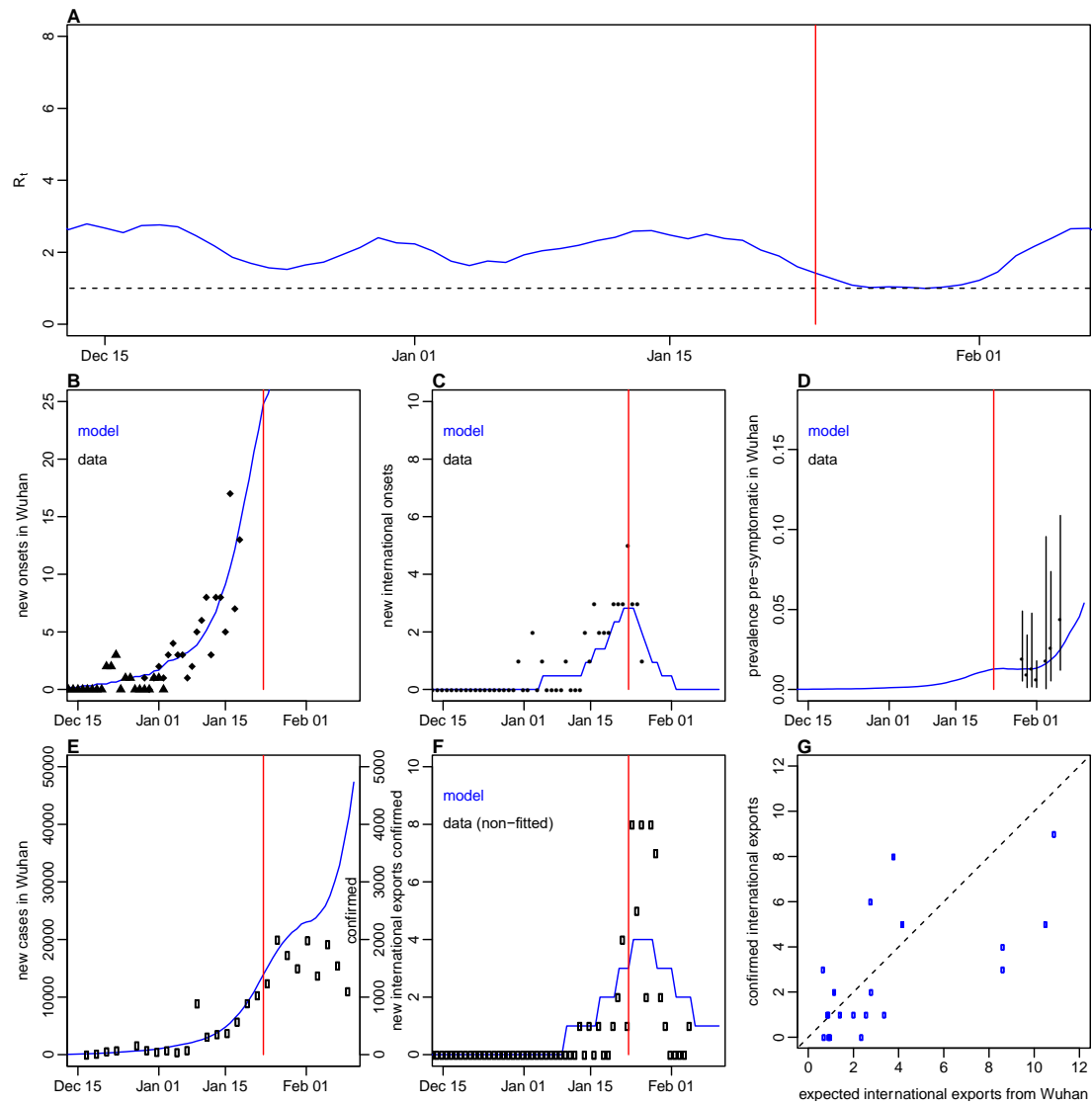


Figure S6: Model outputs when international traveller data from WorldPop is used instead of MOBS Lab estimates.



*Figure S7: Model outputs when we assume infected individuals can transmit in the second half of their incubation period.*

## References

1. MOBS Lab. Situation Report Mainland China [Internet]. [cited 2020 Jan 31]. Available from: <https://www.mobs-lab.org/2019ncov.html>
2. Lai S, Bogoch II, Watts A, Khan K, Li Z, Tatem A. Preliminary risk analysis of 2019 novel coronavirus spread within and beyond China:20.
3. World Health Organisation. Coronavirus disease 2019 (COVID-19). Situation Report 24. WHO; 2020.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Jan;S0140673620301835.



5. Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV) [Internet]. *Systems Biology*; 2020 Jan [cited 2020 Jan 31]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.01.25.919787>
6. National data statistics for COVID-19. Available from: <https://ncov.dxy.cn/ncovh5/view/pneumonia>
7. nCoV-2019 Data Working Group. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020.
8. World Health Organization. Use of laboratory methods for SARS diagnosis. Available from: <https://www.who.int/csr/sars/labmethods/en/>
9. Rambaut A. Phylodynamic Analysis | 44 genomes | 29 Jan 2020. *Virological*; 2020.
10. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005 Nov;438(7066):355–9.
11. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday JD, Kucharski AJ, Edmunds WJ, CMMID nCoV working group, Funk S, Eggo RM. *Medrxiv*. 2020 February; doi: <https://doi.org/10.1101/2020.02.08.20021162>
12. Kucharski AJ, Althaus CL. The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission. *Eurosurveillance* [Internet]. 2015 Jun 25 [cited 2020 Jan 31];20(25). Available from: <http://www.eurosurveillance.org/content/10.2807/1560-7917.ES2015.20.25.21167>
13. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* [Internet]. 2020 Jan 30 [cited 2020 Jan 31];25(4). Available from: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.4.2000058>
14. Imai N, Dorigatti I, Cori A, Donnelly C, Riley S, Ferguson NM. Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. 2020;6.
15. 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 Jan 29;NEJMoa2001316.