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

Adam J Kucharski[1], Timothy W Russell[1], Charlie Diamond[1], Yang Liu[1], CMMID nCoV working group, John Edmunds[1], Sebastian Funk[1], Rosalind M Eggo[1]*

[1]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.*

Abstract

Background: An outbreak of the novel coronavirus SARS-CoV-2 has led to 46,997 confirmed cases as of 13th February 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.

Methods: We combined a stochastic transmission model with data on cases of novel coronavirus disease (COVID-19) in Wuhan and international cases that originated in Wuhan to estimate how transmission had varied over time during January and February 2020. Based on these estimates, we then calculated the probability that newly introduced cases might generate outbreaks in other areas.

Findings: We estimated that the median daily reproduction number, R_t , declined from 2.35 (95% CI: 1.15-4.77) one week before travel restrictions were introduced on 23rd January to 1.05 (95% CI: 0.413-2.39) one week after. Based on our estimates of R_t, we calculated that in locations with similar transmission potential as 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 likely declined in Wuhan during late January 2020, coinciding with the introduction of control measures. As more cases arrive in international locations with similar transmission potential to Wuhan pre-control, it is likely many chains of transmission will fail to establish initially, but may still cause new outbreaks eventually.

Funding: Wellcome Trust (206250/Z/17/Z, 210758/Z/18/Z), HDR UK (MR/S003975/1), Gates Foundation (INV-003174), NIHR (16/137/109)

—

Introduction

As of 13th February 2020, an outbreak of COVID-19 has resulted in 46,997 confirmed cases (1). The outbreak was first identified in Wuhan, China, in December 2019, with the majority of early cases being reported in the city. The majority of internationally exported cases reported to date have a travel history to Wuhan (2). In the early stages of a new infectious disease outbreak, it is crucial to understand the transmission dynamics of the infection. In particular, estimation of changes in transmission over time can provide insights into the current epidemiological situation (3) and identify whether outbreak control measures are having a measurable effect (4,5). Such analysis can inform predictions about potential future growth (6), help estimate risk to other countries (7), and guide the design of alternative interventions (8).

There are several challenges to such analysis, however, 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 and testing capacity (9). Modelling approaches can account for such delays and uncertainty, by explicitly incorporating delays resulting from the natural history of infection and reporting processes (10). In addition, individual data sources may 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 data point) can enable more robust estimation of the underlying dynamics of transmission from noisy data (11,12). Combining a mathematical model of SARS-CoV-2 transmission with four datasets from within and outside Wuhan, we estimated how transmission in Wuhan varied between December and February 2020. We then used these estimates to assess the potential for sustained human-to-human transmission to occur in locations outside Wuhan if cases are introduced.

Research in Context

Evidence before this study: We searched PubMed, BioRxiv and MedRxiv for articles published up to 10th February 2020 using 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, including average exponential growth rate estimates based on inferred or observed cases at a recent time point (13,14) and early growth of the outbreak in China (15,16). 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 to data within Wuhan with 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 time point or dataset. We estimate how transmission has varied over time, identify a decline in the reproduction number in late January to near 1, coinciding with the introduction of large scale control measures, and show the potential implications of estimated transmission for outbreak risk new locations.

Implications of all the available evidence: COVID-19 is currently exhibiting sustained transmission in China. This creates a substantial risk of outbreaks in other countries, although if SARS-CoV-2 has MERS-CoV or SARS-CoV-like variability in

transmission at the individual-level, multiple introductions may be required before an outbreak takes hold.

Methods

To estimate the early dynamics of transmission in Wuhan, we fitted a stochastic transmission dynamic model (17) 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 26th January 2020; daily number of new cases in Wuhan with no market exposure, by date of onset, between 1st December 2019 and 1st January 2020; daily number of new cases in China, by date of onset, between 29th December 2019 and 23rd January 2020; proportion of infected passengers on evacuation flights between 29th January and 4th February 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 (i.e. top 20 most at risk), by date of confirmation, as of 10th February 2020; data on new confirmed cases reported in Wuhan between 16th January and 11th February 2020 (full details in the Appendix).

In the model, individuals were divided into four infection classes (Figure 1): susceptible, exposed (but not yet infectious), infectious, and removed (i.e. isolated, recovered or otherwise no longer infectious). The model accounted for delays in symptom onset and reporting, as well as uncertainty in case observation (see Appendix for full model details). The incubation period was assumed to be Erlang distributed with mean 5.2 days (16) and delay from onset-to-isolation Erlang distributed with mean 2.9 days (2,15). The delay from onset-to-reporting was assumed to be exponentially distributed with mean 6.1 days (2). 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 23rd January 2020 and zero after), the relative connectivity of different countries (18), 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.

Transmission was modelled as a geometric random walk process, and we used sequential Monte Carlo 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 average 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 to an internationally exported case originated in Wuhan. We assumed the outbreak started with a single infectious case on 22nd November 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 conducted sensitivity analysis on

assumptions about the initial number of cases, connectivity between countries and proportion of cases that were infectious before showing symptoms. More details of methodology, sensitivity analysis, data and code availability are provided in the Appendix.

Role of the funding source
The sponsor 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 the daily reproduction number, R_t, varied during January 2020, with median values ranging from 1.6–2.6 between 1st January 2020 and the introduction of travel restrictions on 23rd January (Figure 2A). We estimated a decline in R_t in late January, from 2.35 (95% CI: 1.15-4.77) on 16th of January, one week before the restrictions, to 1.05 (95% CI: 0.413-2.39) on 31st January.

The model reproduced the observed temporal trend of cases within Wuhan and cases exported internationally, capturing all of the dynamics reflected by these different data streams (Figure 2B–D). Our results suggested there were around tenfold more symptomatic cases in Wuhan in late January than were reported as confirmed cases (Figure 2E), but the model 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 2F). We found that confirmed and estimated exported cases among the twenty countries most connected to China were generally in good correspondence, 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 2G). We estimated that 100% (95% CI: 51–100%) of cases would be eventually had detectable symptoms, implying that most infections that were exported internationally from Wuhan in late January were eventually detected.

To examine the potential for new outbreaks to establish in locations outside of Wuhan, we used our estimates of the reproduction number to simulate new outbreaks with potential individual-level variation in transmission (i.e. 'superspreading events') (13,19,20). Such variation increases the fragility of transmission chains, making it less likely that an outbreak will take off 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 (19). Based on the median reproduction number estimated during January before travel restrictions were introduced, we estimated that a single introduction of 2019-nCoV with SARS-like or MERS-like individual-level variation in transmission would have a 20–28% probability of causing a large outbreak (Figure 3A). 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 3B).

Discussion

Combining a mathematical model with multiple datasets, we found that the median daily reproduction number, R_t , of SARS-CoV-2 in Wuhan likely varied between 1.6–2.6 in January 2020 prior to travel restrictions being introduced. We also estimated that transmission declined by around half in the two weeks spanning the introduction of restrictions.

The estimated fluctuations in R_t were driven by the rise and fall in number of cases both in Wuhan and internationally, as well as prevalence on evacuation flights (Figures 2B–D). 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 (13,19,20). We found some evidence of reduction in R_t in the days prior to the introduction of travel restrictions in Wuhan, which may have been 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 (Figure 2A) results from limited data sources to inform changes in transmission during this period.

Comparing model predictions to observed confirmed cases reported in Wuhan during late January and early February, we found that the model predicts ten-fold higher cases than have been reported; the model also does not predict the recent slowdown in cases, suggesting a potential change in reporting rather than a genuine slowdown in transmission in early February. Our estimates for international cases in specific countries were broadly consistent with the number of subsequently confirmed exported cases outside of Wuhan. However, there were notably more cases exported to France, US, and Australia compared to what our model predicted. This may be the result of increased surveillance and detected as awareness of SARS-CoV-2 grew in late January, which would suggest earlier exported cases may have missed; it may also be the result of increased travel out of Wuhan immediately prior to travel restrictions being introduced on 23rd January.

Based our on estimated reproduction number, and published estimates of individual-level variation in transmission for SARS 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 it has been in Wuhan in early January, it may take several introductions for an outbreak to establish; this is because high individual-level variation in transmission makes new chains of transmission more fragile, and hence less likely that a single infection will generate out outbreak. This highlights the importance of rapid case identification, and subsequent isolation and other control measures to reduce the chance of onward chains of transmission (21).

Our analysis highlights the value of combining multiple data sources in analysis of COVID-19. For 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 data points were used in analysis. Our results also have implications for the estimation of transmission dynamics using the number of exported cases from a specific area (22). Once extensive 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 is not accounted for, it could lead to artificially low

estimates of R_t or inferred case totals from the apparently declining numbers of exported cases. Our model estimates benefitted 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 may be refined as more comprehensive data become available. However, by fitting to multiple datasets to infer model parameters, and performing 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. Further, we used publicly available connectivity and risk estimates based on international travel data to predict the number of exported cases into each country. These estimates have shown good correspondence with the distribution of exported cases to date (23), and are similar to another risk assessment for COVID-19 with different data (24). We also assumed that the latent period is equal to the incubation period (i.e. 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 limited reported symptoms (25). We therefore conducted a sensitivity analysis in which transmission could occur in the second half of the incubation period, but this did not change our overall conclusions (Appendix, page 7). We also explored having a larger initial spillover event and also using different sources for flight connectivity data, neither of which changed the conclusions of the analysis. 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 (13); if transmission were more homogenous than SARS of MERS-CoV, it would increase the risk of outbreaks following introduced cases. As more data becomes available, it will be possible to refine these estimates, and therefore we made an online tool so users can explore these risk estimates if new data become available (Appendix, page 4)

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

Author Contributions

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 following authors were part of the CMMID 2019-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. Each contributed in processing, cleaning and interpretation of data, interpreted 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 submission.

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

Conflict of Interest Statement
The authors have no conflicts of interest to declare.

References

1. World Health Organisation. Coronavirus disease 2019 (COVID-19). Situation Report 24. WHO; 2020.
2. nCoV-2019 Data Working Group. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020. XX ADD LINK XX
3. Camacho A, Kucharski A, Aki-Sawyerr Y, White MA, Flasche S, Baguelin M, et al. Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study. PLoS Curr Outbreaks. 2014;7.
4. Funk S, Ciglenecki I, Tiffany A, Gignoux E, Camacho A, Eggo RM, et al. The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. Philos Trans R Soc B Biol Sci. 2017 May 26;372(1721):20160302.
5. Riley S. Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. Science. 2003 Jun 20;300(5627):1961–6.
6. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, et al. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics. 2018 Mar;22:13–21.
7. Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the International Spread of Pandemic Influenza. Sepulveda-Amor J, editor. PLoS Med. 2006 May 2;3(6):e212.
8. Kucharski AJ, Camacho A, Checchi F, Waldman R, Grais RF, Cabrol J-C, et al. Evaluation of the Benefits and Risks of Introducing Ebola Community Care Centers, Sierra Leone. Emerg Infect Dis. 2015 Mar;21(3):393–9.
9. WHO Ebola Response Team. Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. N Engl J Med. 2014 Oct 16;371(16):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. Peccoud J, editor. PLoS ONE. 2009 Aug 31;4(8):e6852.
11. Birrell PJ, De Angelis D, Presanis AM. Evidence Synthesis for Stochastic Epidemic Models. Stat Sci. 2018 Feb;33(1):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. Leung GM, editor. PLoS Med. 2013 Oct 8;10(10):e1001527.

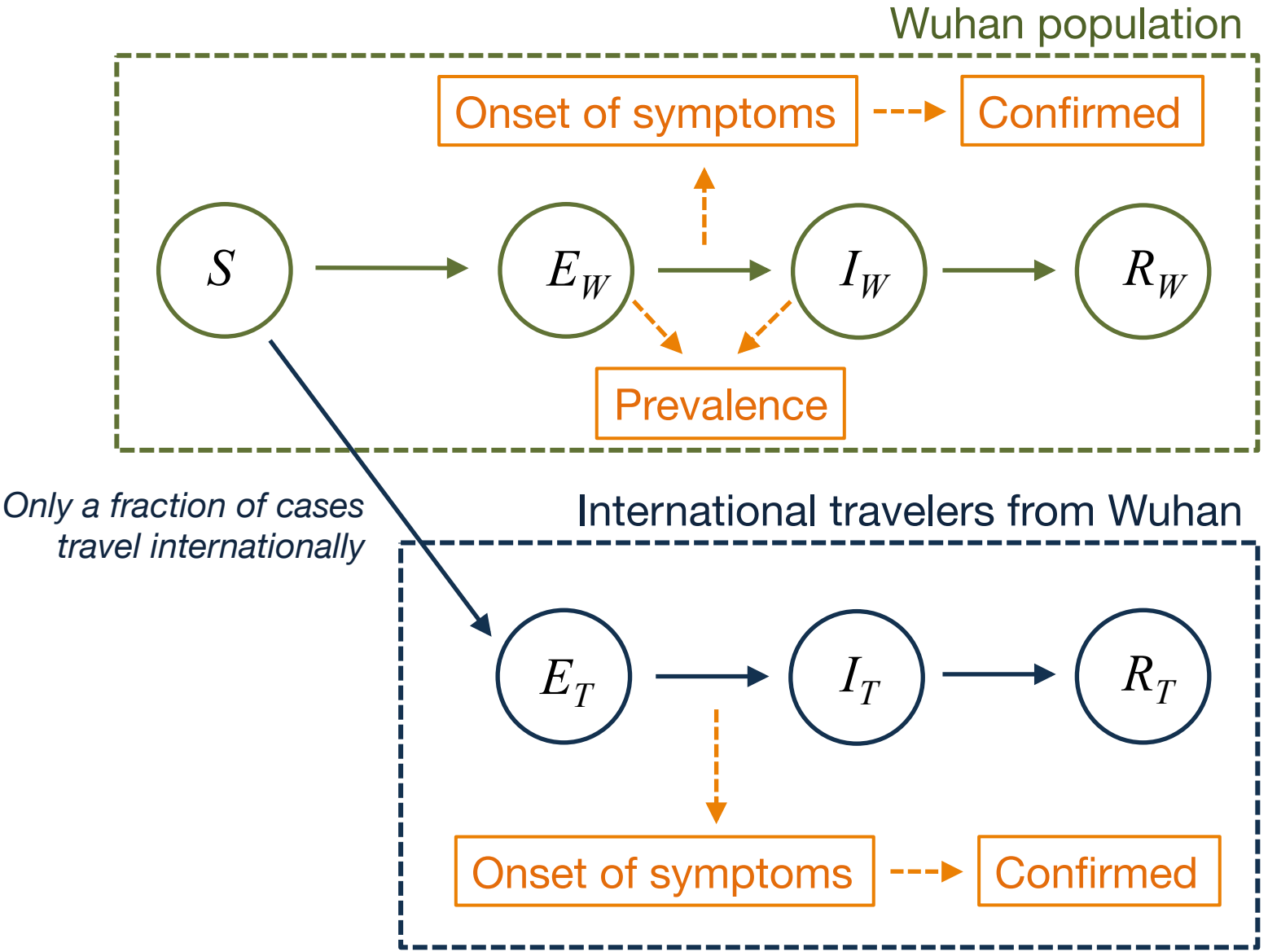
13. Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance [Internet]. 2020 Jan 30 [cited 2020 Jan 31];25(4). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.4.2000058
14. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S. Report 3: Transmissibility of 2019-nCoV. :5.
15. 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
16. 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.
17. Dureau J, Kalogeropoulos K, Baguelin M. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. Biostatistics. 2013 Jul 1;14(3):541–55.
18. MOBS Lab. Situation Report Mainland China [Internet]. [cited 2020 Jan 31]. Available from: https://www.mobs-lab.org/2019ncov.html
19. 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.
20. 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
21. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis C. Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts. medRxiv, 2020. 02.08.20021162
22. 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.
23. 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. Eurosurveillance [Internet]. 2020 Jan 30 [cited 2020 Jan 31];25(4). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.4.2000057
24. 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.
25. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med. 2020 Jan 30;NEJMc2001468.

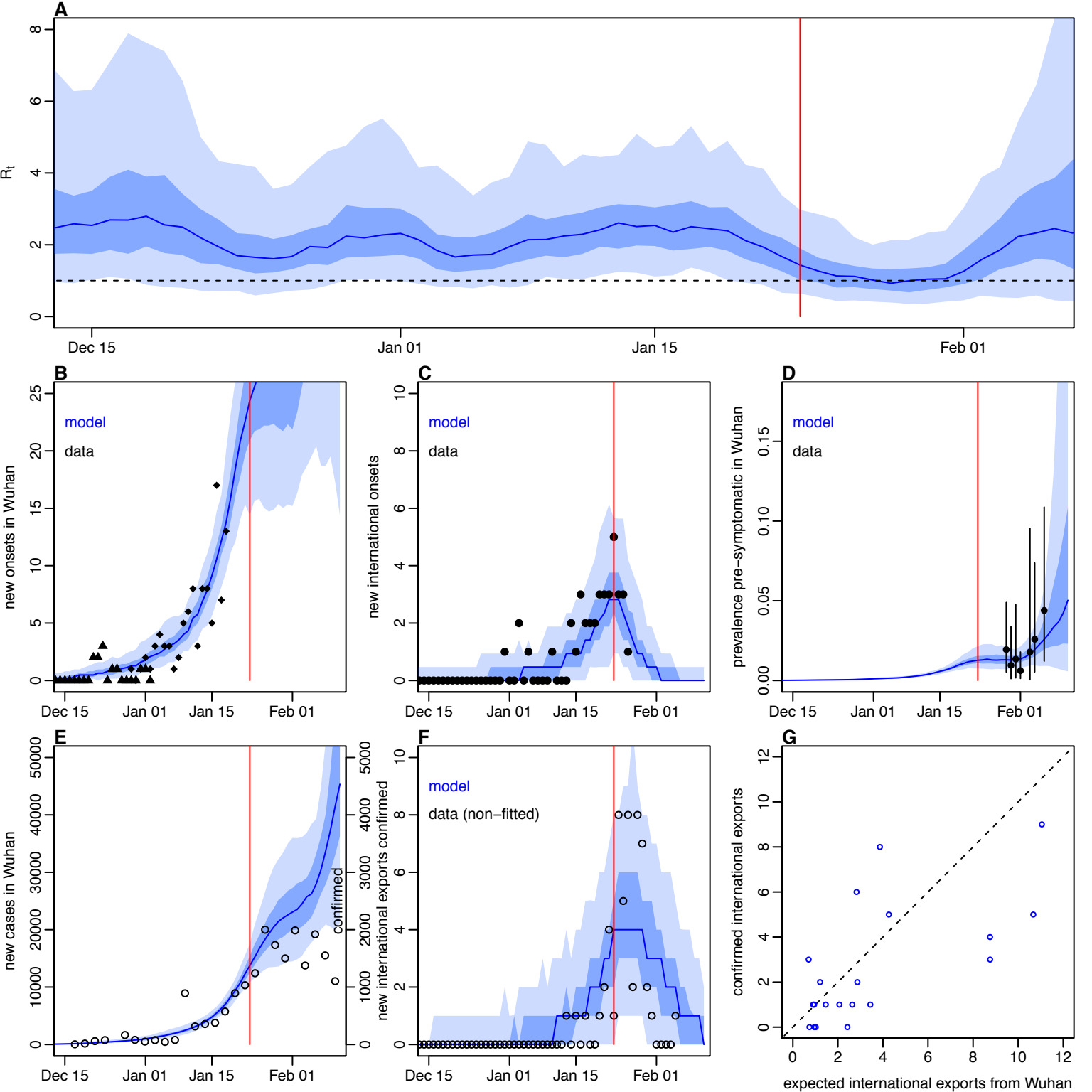
Figure 1: Model structure. The population is divided into four classes: susceptible, exposed (and not yet symptomatic), infectious (and symptomatic), removed (i.e. isolated, recovered, or otherwise non-infectious). A fraction of exposed individuals subsequently travel and are eventually detected in their destination country.

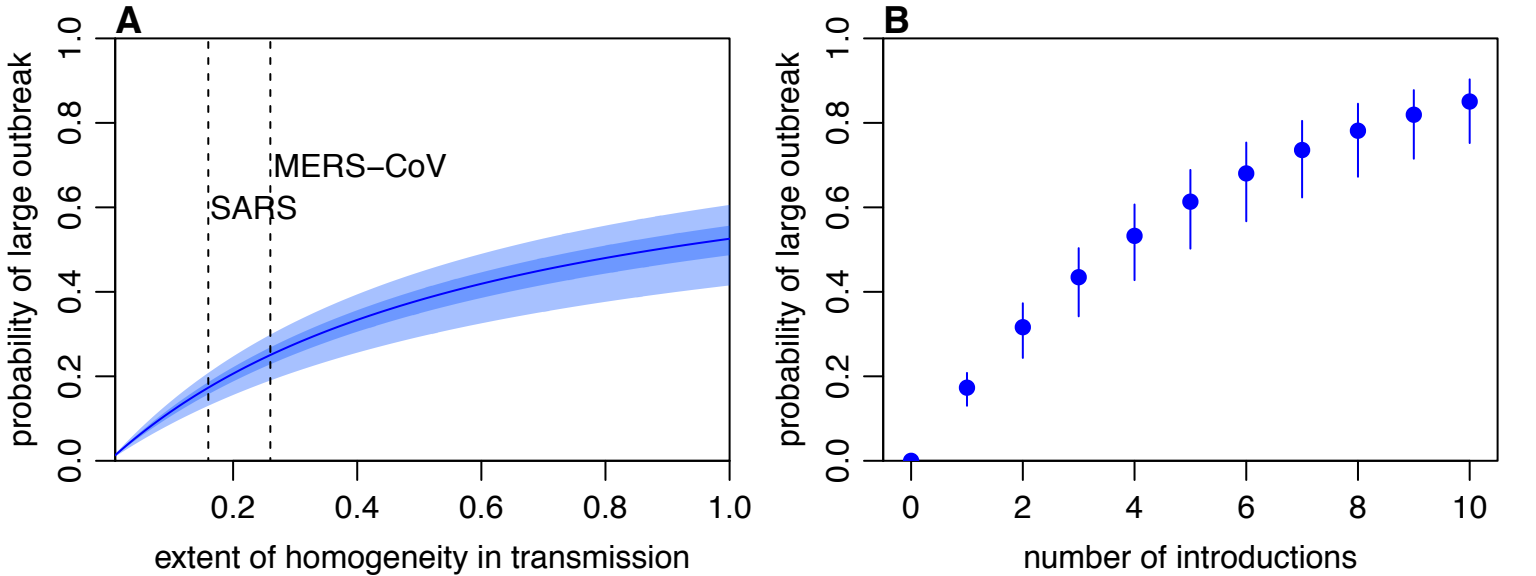
Figure 2: Dynamics of transmission in Wuhan, fitted up to 13th February 2020. Red line marks travel restrictions starting on 23rd January 2020. A) Estimated daily

reproduction number (R_t) over time. B) Onset dates of confirmed cases in Wuhan (triangles) and China (diamonds). Blue lines and shaded regions: median, 50% and 95% credible intervals of model estimate. C) Reported cases by date of onset (black) and estimated internationally exported cases from Wuhan by date of onset (blue line). D) Estimated prevalnece of infections that do not have detectable symptoms (blue line), and proportion of passengers on evacuation flights that tested positive for SARS-CoV-2 (black points, with 95% binomial CIs shown). 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 from Wuhan confirmed up to 10th Feburary 2020 and observed number in 20 countries with highest connectivity to China. In all panels, datasets that were fitted to shown as solid points; non-fitted data shown as circles.

Figure 3: Risk that introduced infections will establish in a new population. A) Probability a single case will lead to a large outbreak for different assumptions about the extent of homogeneity in individual-level transmission (i.e. the dispersion parameter k in a negative binomial offspring process). Results are shown for the median reproduction number estimated for nCoV-2019 in Wuhan between 1st January and 23rd January 2020 and. B) Probability a given number of introductions will result in a large outbreak, assuming SARS-like superspreading events can occur.


Wuhan population
Onset of symptoms
Confirmed
S
E_W
I_W
R_W
Prevalence
Only a fraction of cases travel internationally
International travelers from Wuhan
E_T
I_T
R_T
Onset of symptoms
Confirmed


A
R_t
0
2
4
6
8
Dec 15
Jan 01
Jan 15
Feb 01
B
model
data
new onsets in Wuhan
0
5
10
15
20
25
Dec 15
Jan 01
Jan 15
Feb 01
C
model
data
new international onsets
0
2
4
6
8
10
Dec 15
Jan 01
Jan 15
Feb 01
D
model
data
prevalence pre−symptomatic in Wuhan
0.00
0.05
0.10
0.15
Dec 15
Jan 01
Jan 15
Feb 01
E
new cases in Wuhan
0
10000
20000
30000
40000
50000
confirmed
0
1000
2000
3000
4000
5000
Dec 15
Jan 01
Jan 15
Feb 01
F
model
data (non−fitted)
new international exports confirmed
0
2
4
6
8
10
Dec 15
Jan 01
Jan 15
Feb 01
G
confirmed international exports
0
2
4
6
8
10
12
expected international exports from Wuhan


A
probability of large outbreak
0.0
0.2
0.4
0.6
0.8
1.0
SARS
MERS−CoV
0.2
0.4
0.6
0.8
1.0
extent of homogeneity in transmission
B
probability of large outbreak
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
number of introductions