Early dynamics of transmission and control of 2019-nCoV: a mathematical modelling study

Adam J Kucharski, Timothy W Russell, Charlie Diamond, CMMID nCoV working group, Sebastian Funk, Rosalind M Eggo

Abstract

An outbreak of the novel coronavirus 2019-nCoV has led to 7818 cases as of 30th January 2020. Understanding the transmission dynamics of the infection is crucial for evaluating the likely effectiveness of control measures and potential for sustained transmission to occur in new areas. We combined a stochastic transmission model with data on cases 2019-nCoV in Wuhan and exported cases originating in Wuhan to estimate how transmission had varied over time and the likely prevalence of symptomatic cases in the city as of 23rd January 2020. Based on these estimates, we then calculated the probability that newly introduced cases would generate outbreaks in other areas. We estimated that the median reproduction number, *R*, fluctuated between 1.6–2.9 from mid-December to mid-January 2020. We found that the US, Australia and France had more confirmed cases with travel history to Wuhan than the model predicted, and estimated that there were 29,500 (14,300-85,700) prevalent symptomatic cases in Wuhan on 23rd January 2020, when travel restrictions were introduced. Based on our estimates of *R*, we calculated that in locations with similar transmission potential as Wuhan, once there are more than three introduced cases, there is a more than 50% chance the infection will establish within that population. Our results show that 2019-nCoV has substantial potential for ongoing human-to-human transmission, and exported cases from Wuhan may have increased prior to travel restrictions being introduced on 23rd January 2020. As more cases arrive in international locations, it is likely many chains of transmission will fail to establish initially, but may still cause new outbreaks eventually.

Introduction

As of 30th January 2020, an outbreak of the novel coronavirus 2019-nCoV has resulted in 7,818 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 nCoV transmission with four datasets from within and outside Wuhan, we estimated how transmission in Wuhan varied during December and January 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.

Methods

To estimate the early dynamics of transmission in Wuhan, we fitted a stochastic transmission dynamic model (18) to multiple publicly available datasets on cases in Wuhan and internationally exported cases from Wuhan. The three datasets we fitted to were: daily incidence of internationally exported cases (or lack thereof) in countries with high connectivity to Wuhan (i.e. top 20 most at risk), by date of onset, as of 21st January 2020 (1); daily incidence of initial cases in Wuhan with no market exposure, by date of onset, between 1st December 2019 and 1st January 2020 (19); daily incidence of initial cases in China, by date of onset, between 29th December 2019 and 23rd January 2020 (1,2). We also used an additional dataset as an out-of-sample validation of the model outputs: daily incidence of 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 30th January 2020 (2).

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 Supplementary Materials 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 (20) and zero after), the relative connectivity of different countries (21), and the relative probability of reporting a case outside Wuhan. 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 reproduction number, R, defined as the average number of secondary cases generated by a typical infectious individual on each day (18). The model had two unknown parameters, which we estimated: magnitude of temporal variability in transmission 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 2nd December 2019 and the entire population was initially susceptible. Once we had estimated R, we used a branching process simulation with a negative binomial offspring distribution to calculate the probability an introduced case would cause a large outbreak. More details of methodology and sensitivity analysis are provided in the Supplementary Materials.

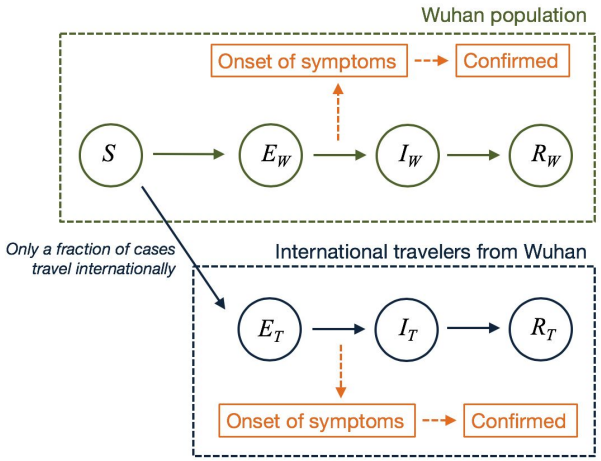

Figure 1: Model structure. The population is divided into four classes: susceptible, exposed (and not yet symptomatic), infectious (and symptomatic), removed (i.e. isolated, recovered, or

otherwise non-infectious). A fraction of exposed individuals subsequently travel and are eventually detected in their destination country.

Results

The model reproduced the observed temporal trend of cases observed within Wuhan and cases exported internationally (Figure 2A–B). It could also reproduce the pattern of confirmed exported cases from Wuhan, which was not explicitly used in the model fitting (Figure 2C). 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 2D). Accounting for under-reporting and delays to confirmation, our estimates suggest that the prevalence of symptomatic 2019-nCoV cases in Wuhan was 29,500 (14,300-85,700) on 23rd January 2020 (Figure 2E). We estimated that the daily reproduction number, R, varied during December and January, with median values ranging from 1.6–2.9 between 15th December 2019 and the introduction of travel restrictions on 23rd January 2020 (Figure 2F).

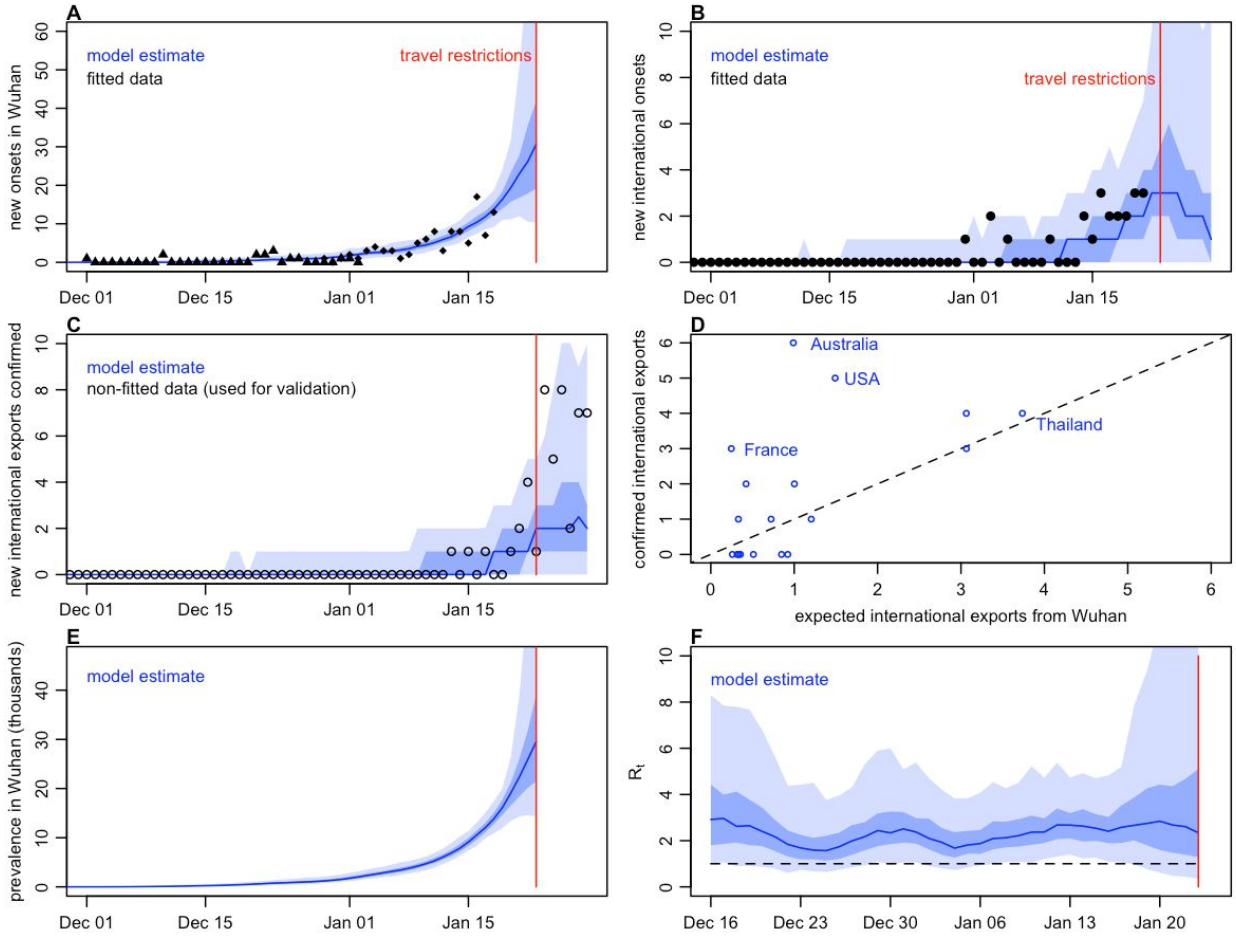

Figure 2: Dynamics of transmission in Wuhan, fitted up to 28 January 2020. Red line marks travel restrictions starting on 23 January 2020. A) Onset dates of confirmed cases in Wuhan (triangles) and China (diamonds). Blue lines and shaded regions: median, 50% and 95% credible intervals of model estimate. B) Reported cases by date of onset (black) and estimated internationally exported cases from Wuhan by date of onset (blue line). C) Cumulative confirmed cases by date in Wuhan (circles) and estimated cumulative cases (blue line). D) International exportation events by date of confirmation of case, and expected number of exports in the fitted model. E) Estimated prevalence of symptomatic cases over time in Wuhan. F) Estimated daily reproduction number (R) over time. Datasets that were fitted to shown as solid points; non-fitted data shown as circles.

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) (14,21,22). 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 (21). Based on the median reproduction number observed 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 more than three infections have been introduced into a new location, there is an over 50% chance that an outbreak will occur (Figure 3B).

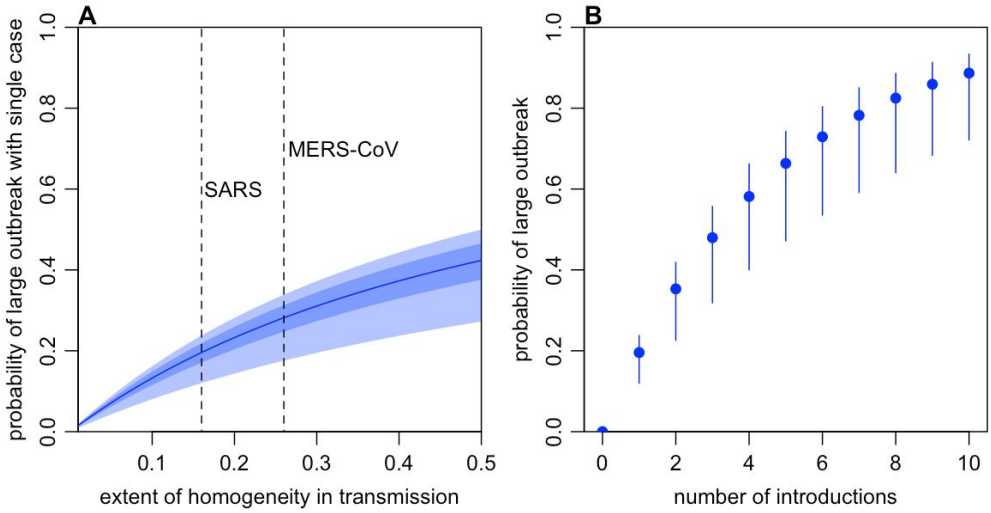

Figure 3: Risk that introduced infections will establish in a new population. A) Probability a single case will lead to a large outbreak for different assumptions about the extent of homogeneity in individual-level transmission (i.e. the dispersion parameter k in a negative binomial offspring process). 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.

Discussion

Combining a mathematical model with multiple datasets, we found that the median daily reproduction number, *R*, of 2019-nCoV in Wuhan likely varied between 1.6–2.9 during December and January 2020. The estimated fluctuations in *R* were driven by the rise and fall in number of cases both in Wuhan and internationally (Figures 2A–B). 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 (14,21,22). We did not find evidence of a significant reduction in *R* in the days prior to the introduction of travel restrictions in Wuhan,

which may have been expected if outbreak control efforts and growing awareness of 2019-nCoV during this period were having a substantial effect on transmission. Our estimates for international cases 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 2019-nCoV 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, it may take several introductions for an outbreak to establish. This highlights the importance of rapid case identification, and subsequent isolation and other control measures to reduce the chance of onward chains of transmission.

Our analysis highlights the value of combining multiple data sources in analysis of 2019-nCoV. 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 the effect of inflating R 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 (13). 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 or inferred case totals from the apparently declining numbers of exported cases.

There are several limitations to our analysis. We used plausible biological parameters for 2019-nCoV based on current evidence, but these values may be refined as more comprehensive data become available. However, because we fitted to multiple datasets to infer model parameters, and performed sensitivity analyses on key areas of uncertainty, we would not expect our findings to change substantially if new information becomes available. 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 2019-nCoV with different data (17,24).

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 some limited evidence that transmission of 2019-nCoV can

occur without reported symptoms (25). If this only happens for a small proportion of cases, we would not expect our results to change noticeably. If a large fraction are infectious prior to showing symptoms, but the difference between latent and incubation period is small (e.g. one or two days), then we would expect the overall patterns in our predicted incidence to be broadly consistent, but the estimates of transmissibility would shift slightly to reflect the different delay in time-to-infectiousness. If it turns out that a proportion of cases are asymptomatic throughout their infection and contribute to transmission, the estimated prevalence of infectious individuals shown in Figure 1E should be interpreted as showing only the prevalence of the proportion of infectious individuals who will eventually become symptomatic cases. 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 2019-nCoV (14); if transmission were more homogenous than SARS of MERS-CoV, it would increase the risk of outbreaks following introduced cases.

Our results demonstrate that there is likely to be substantial variation in 2019-nCoV transmission over time. Understanding this range of transmission dynamics in different settings will be crucial for evaluating the effectiveness of control measures, and determining the likelihood of secondary transmission becoming established in new locations.

Funding

This work was supported by the Wellcome Trust (grants 206250/Z/17/Z and 210758/Z/18/Z) and HDR UK (grant: MR/S003975/1).

References

1. World Health Organisation. Novel Coronavirus(2019-nCoV). Situation Report 10. WHO; 2020.
2. nCoV-2019 Data Working Group. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020.
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. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S. Report 3: Transmissibility of 2019-nCoV. :5.
14. 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
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. 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
18. 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.
19. 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.
20. 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.
21. 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.
22. 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
23. MOBS Lab. Situation Report Mainland China [Internet]. [cited 2020 Jan 31]. Available from: https://www.mobs-lab.org/2019ncov.html
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.
26. Group nCoV-2019 DW. Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020.
27. Rambaut A. Phylodynamic Analysis | 44 genomes | 29 Jan 2020. Virological; 2020.

Supplementary materials

Supplementary methods

Transmission model structure

We used a stochastic SEIR model implemented using the Euler-Maruyama algorithm with a 6 hour timestep, with the transmission rate following geometric brownian motion. The model structure is described below and shown in Figure 1 in the main text.

Model compartments for individuals in Wuhan:

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

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

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

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

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

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

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

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

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 incubation 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. Here $\beta(t)$ is the transmission rate at time t; σ is the rate of becoming symptomatic (i.e. 1/incubation period); γ = rate of isolation (i.e. 1/delay from onset-to-hospitalisation); κ 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.

Model compartments for traveller cases from Wuhan:

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

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

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

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

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

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

$$C_{T}(t+1)= C_{T}(t) + \kappa\, Q_{T}(t)$$

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 incubation 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:

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

where a is the volatility of transmission over time and B_t is Brownian motion.

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 (23,24). For example, if a country had a relative risk of export W from Wuhan, then we would expect $W\, D_T(t)\, f\, N$ 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.

Branching process simulation model

We used a branching process with a negative binomial offspring distribution (21) to calculate the probability P_E 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: $1–(1–P_E)^n$

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 incidence of exported 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 21st January, as many of the cases detected after this point were not travellers from Wuhan.
2. Daily incidence of initial cases in Wuhan with no market exposure, by date of onset. Source: Huang et al
3. Daily incidence of early cases in China, by date of onset. We assume that these are all in Wuhan. The most recent two data points were omitted during fitting as they are likely to be strongly influenced by delays in reporting.
4. Validation dataset (not used for fitting): Daily incidence of 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. We only considered individual export events (i.e. a family of travellers was counted as a single export).

We created a single timeseries for case onset data by in Wuhan by combining datasets 3 and 4 above. To calculate the likelihood, we used a Poisson observation model fitted jointly to expected values based on three model outputs: $D_W(t)$, $D_T(t)$, $C_T(t)$. To calculate the daily expectation for each Poisson observation process, we converted these outputs into new case onset and new reported cases inside Wuhan and travelling internationally. We assumed a different relative reporting probability for Wuhan cases compared to international cases, as assumed only a proportion of confirmed Wuhan cases had known onset dates (fixed at 0.15 based on available line list data (26)). As destination country was known for confirmed exported cases, we used 20 timeseries for cases exported (or not) to most at-risk countries each day and calculated the probability of obtaining each of these datasets given the model outputs. International onset data was not disaggregated by country and so we used the total daily exported cases in our Poisson probability calculation.

Estimates for time-varying *R* were generated by running 200 repetitions of SMC with 1000 particles. The transmission volatility and relative reporting of cases outside Wuhan were selected

based on a grid search to find the marginal MLE (Figure S1–S2). We assumed the outbreak started on 22nd November with 1 infectious individual (27) and the population was initially fully susceptible. We also assumed all infectious people eventually became symptomatic and would be eventually be detected in destination country if they travelled by plane. We assumed that the population in Wuhan was 11m. All data and code required to reproduce the analysis is available here: https://github.com/adamkucharski/2020-ncov/stoch_model_V1_paper

Parameter	**Value**	**Distribution**
Incubation period	4.8 days	Erlang (rate=2)
Infectious period	2.9 days	Erlang (rate=2)
Delay onset-to-confirmation	6.1 days	Exponential
Daily outbound passengers	3300	
Population of Wuhan	11m	
Initial cases	1	
Introduction date	22nd November 2019	
Proportion of cases with onsets known	0.16	
Relative reporting outside of Wuhan (W)	67 (fitted)	
Transmission volatility (a)	0.4 (fitted)	

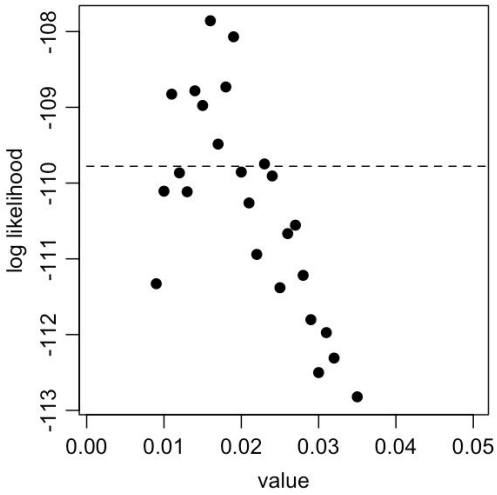

Figure S1: Sliced likelihood for the relative reporting of confirmed cases within Wuhan compared to internationally exported cases (equal to parameter 1/W in the calculation of exported cases from Wuhan confirmed).

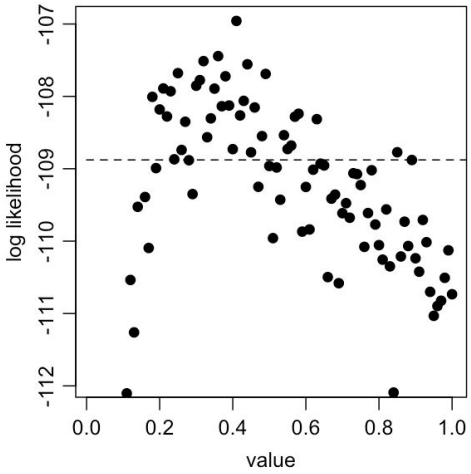

Figure S2: Sliced likelihood for the transmission volatility parameter, a.

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 S3). As another sensitivity analysis, we also used flight data from WorldPop rather than MOBS lab, but again this did not change our overall conclusions (Figure S4).

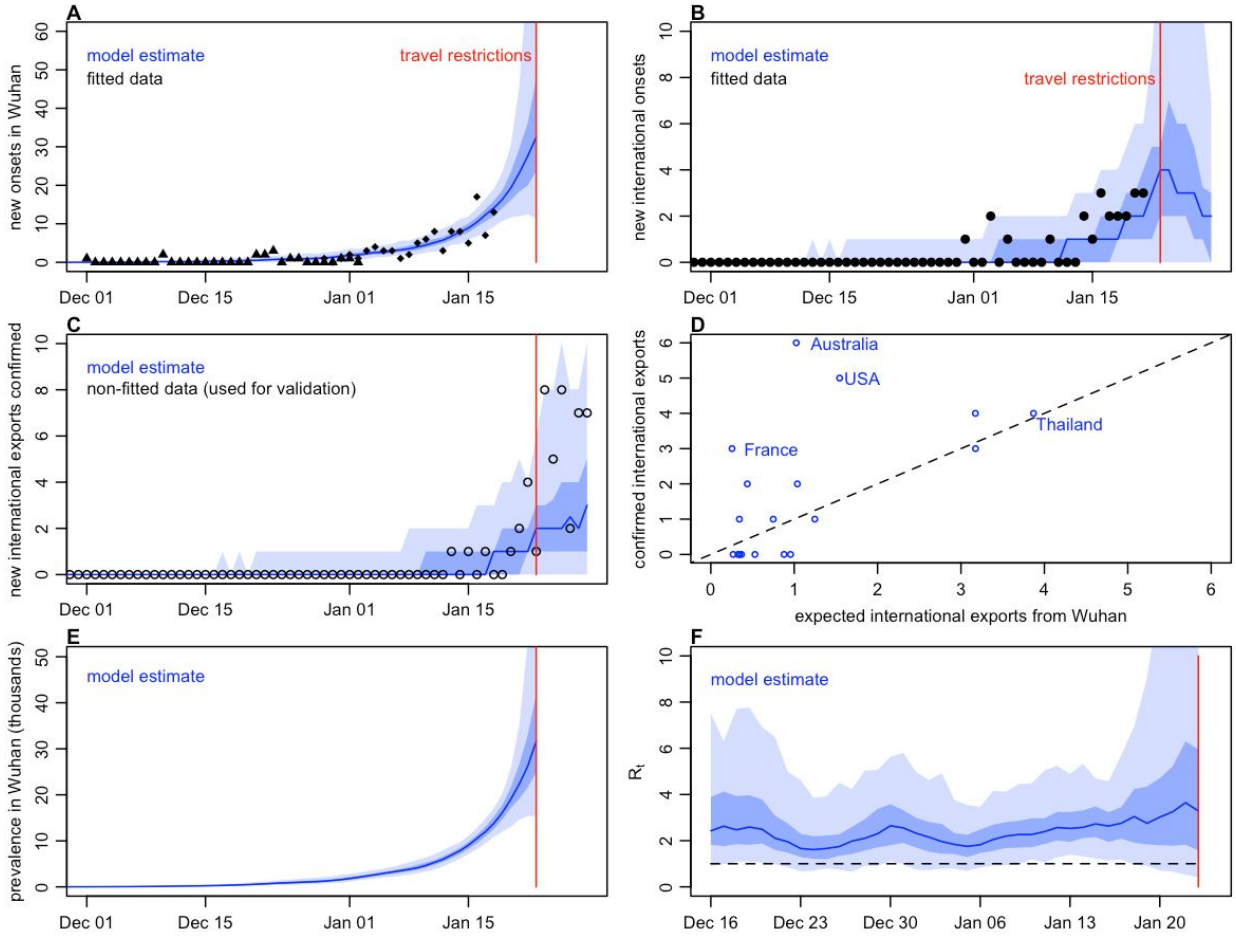

Figure S3: Model outputs when we assumed 10 initial cases rather than 1 on 22nd November 2019.

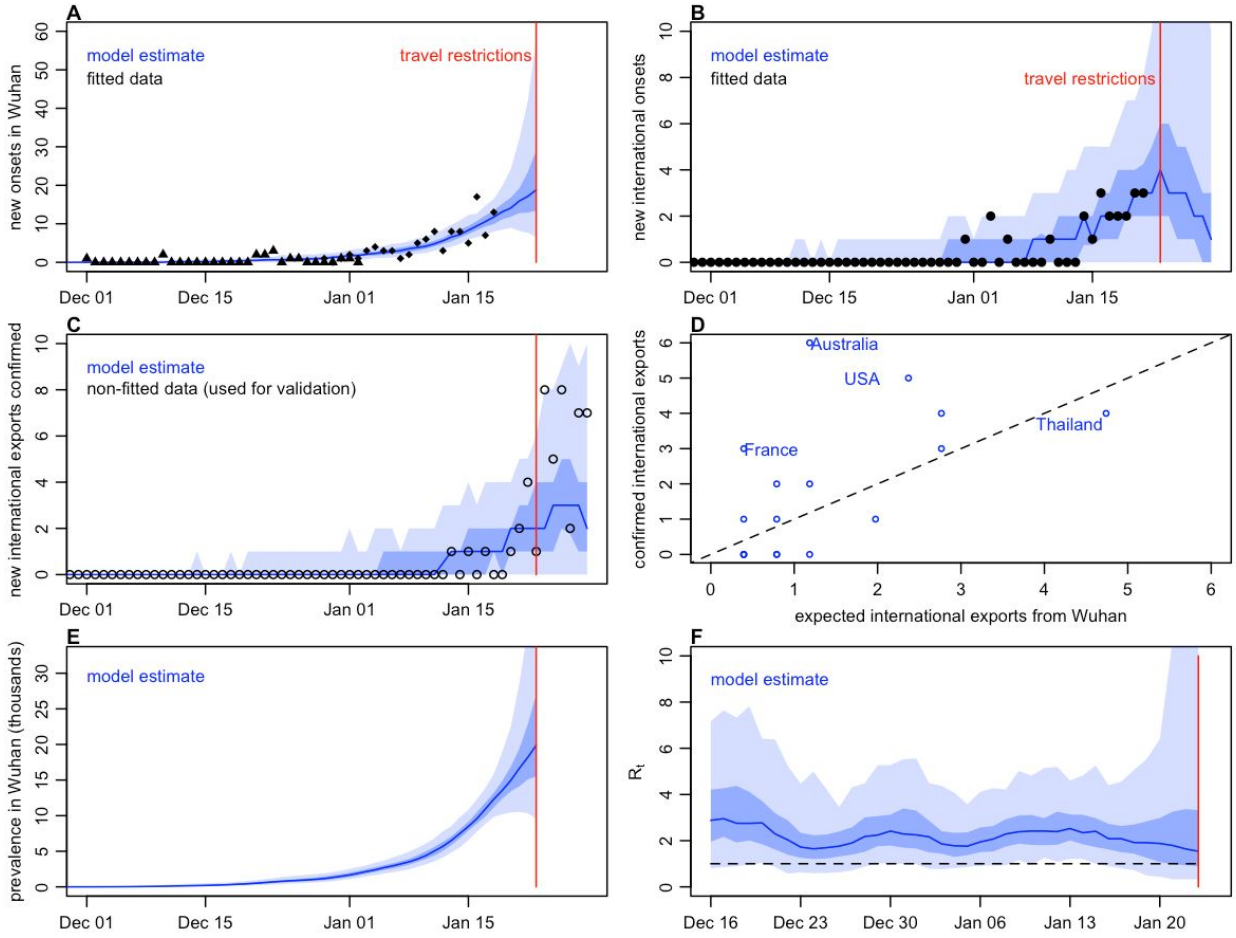

Figure S4: Model outputs when international traveller data from WorldPop is used instead of MOBS Lab estimates.