Evaluating the transmissibility of the coronavirus in the 2019-20 Wuhan Outbreak: Exploring initial point-source exposure sizes and durations using scenario analysis

Authors

S. Abbott (1), J. Hellewell (1), J. Munday (1), CMMID nCoV working group (1), S. Funk (1)

Correspondence to: sam.abbott@lshtm.ac.uk

Affiliations

1. Center for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, United Kingdom

Introduction

The ongoing pneumonia outbreak appears to have originated from an initial point-source exposure event at Huanan seafood wholesale market in Wuhan, China which was closed on the 31st of December 2019 [1,2]. As of the 26th of January 2020 there have been over 2000 confirmed cases with the majority in China [3]. Globally, countries are on high alert with wide implementation of airport checks and contact tracing. In China, officials have restricted travel across a wide area. There is still uncertainty around the precise scale and duration of the initial exposure event [4]. This has implications for the likely transmissibility of coronavirus and as such it is important that these potential scenarios are further explored.

We used a stochastic branching process model - parameterised with available data where possible and otherwise informed by the 2002-2003 SARS outbreak - to simulate the Wuhan outbreak. We considered a range of parameters where data were not available, quantifying how likely these scenarios were to occur using reported cases. We focused on the size and duration of the initial exposure event in particular, and the impact this has on the estimated level of human-to-human transmission. We aimed to provide decision makers, and researchers, with probability estimates for each scenario considered, along with estimates of the reproduction number (R0) across all scenarios.

Methods

Branching process model

We modelled the outbreak using a stochastic branching comparable to those used elsewhere [4]. We assumed that cases from the initial transmission event were uniformly distributed over the duration of the event. Each case then resulted in a subsequent generation of cases with the number that each case generated being drawn from a negative binomial distribution - to account for overdisperion - with a dispersion parameter k of 0.16 (assuming SARS-like dispersion) [5]. The mean number of cases generated by each case (R0) was sampled from a uniform distribution once per model simulation with a lower and upper bound determined by the scenario being evaluated. Generations were then sampled iteratively until the maximum simulation time was reached. The serial interval between each generation was assumed to be normal with a mean varied during the scenario analysis and a standard deviation of 3.8 (assumed SARS-like) [5]. After the simulation of the branching process reporting delays were added as reported in a line-list of cases compiled from media and other reports [6]. We fitted a geometric, Poisson, and a negative binomial distribution to these observed delays and selected the best fit using the Chi-squared statistic. If no good fit was determined - using a

p-value threshold of 0.05 - then the reporting delay was instead sampled from the empirical delays in the line-list.

Scenario analysis

We simulated the branching process model 10,000 times for all combinations of the following parameters: number of confirmed cases resulting from the initial exposure (20, 40, 60, 80, 200, 400), initial expsoure event duration (7 days, 14 days, 21 days, and 28 days), the mean of the serial interval (4 days, 8.4 days [5], 12)), and R0 (lower and upper bounds of a uniform distribution: 0-1, 1-2, 2-3, 3-4). Parameter values used in the scenario analysis were either assumptions based on the current knowledge of the Wuhan outbreak or based on those used previously for SARS [5]. We ran the model from the beginning of the outbreak for each scenario until the 25th of January 2020. The start date was determined by combining the duration of the transmission event with the date the fish market in Wuhan - the source of the outbreak - closed (31st of December 2019). We evaluated each scenario sample based on the cases observed on the 25th of January (1975). Samples were rejected if their simulated cumulative case estimates were outside a 5% interval on either side of this. Outbreak simulation was stopped if a sample exceeded the upper bound on the number of observed cases.

Analysis

We compared the percentage of samples that were accepted stratified by the transmission event size, transmission event duration, mean serial interval, and R0 using a heat map. We then compared the distribution of R0 for accepted samples by transmission event size, transmission event duration and mean serial interval. We reported 90% credible intervals (CrI) for R0, stratified by the transmission event size, transmission event duration and the assumed mean serial interval.

Implementation

All analysis was carried out using R version 3.6.2 [7]. The branching process model was implemented using the `bpmodels` package [8]. The analysis is available as an open-source R package [9]. A dockerfile has been made available with the code to ensure reproducibility [10].

Results

Percentage of outbreak simulations accepted

Overall, the highest acceptance rate was for scenarios with a large event size (200), short duration (7 days), an R0 between 3 and 4,and a 12 day serial interval. (Figure 1). Scenarios with a SARS-like mean serial interval (8.4 days), an R0 bounded between 2 and 3, a short duration, and a relatively large event size (100) also had a high acceptance rate compared to other scenarios. Across all scenarios a higher acceptance rate was correlated with a larger event size, a shorter event duration, and a larger mean serial interval. This is may be related to the influence these parameters have on the degree of volatility in outbreak simulations. Based on this trends in Figure 1 should be interpreted with care using prior knowledge. For example if the event size, serial interval and event duration is assumed the percentage of acceptance may be used to infer the most likely R0 scenario.

There were very few scenarios in which an upper bound on the reproduction of 1 resulted in scenarios that were accepted after conditioning on observed data, regardless of the mean serial interval, event size, or event duration. There were also very few scenarios that were accepted after conditioning on data for scenarios with an upper bound of 2 on the R0 with this scenario being most likely if the transmission event was large (200+), a longer duration, and a long mean serial interval (12 days). For a short serial interval (4 days) the percentage of accepted samples was low for all scenarios with the highest accepted proportion for scenarios with an upper bound on the R0 of 2. Across all R0 bounds this scenario was most likely if the transmission event was of a medium size (40 - 80 cases) and a short duration (7-14 days).

Estimated reproduction numbers

Uncertainty in the R0 estimate increased as the event size decreased, and decreased as the mean serial interval increased (Figure 2). Large event sizes resulted in the lowest R0 estimates across all scenarios evaluated.

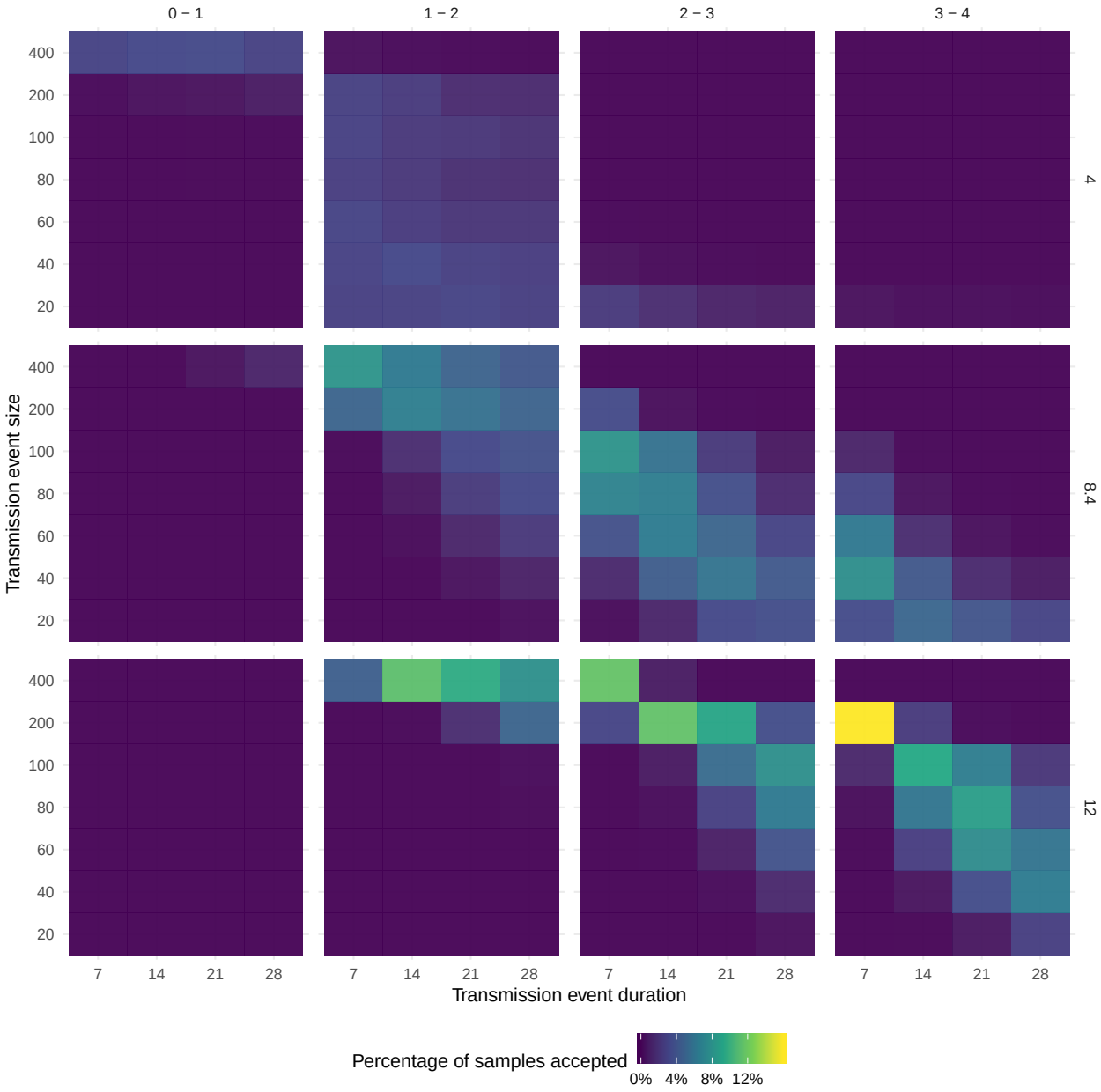

Figure 1: Heatmaps of the percentage of samples accepted by scenario. The figure is stratified by the upper bound on the R0 (columns) and the mean of the serial interval (rows).

The estimated R0 decreased as the event size and duration increased for both mean serial interval scenarios (Table 1 and Table 2). Assuming a longer serial interval (12 days) resulted in an approximate increase of 1 to the R0 estimates across all scenarios when compared to the SARS-like (8.4 days) serial interval. For the SARS-like interval the most likely scenario - with an event size of 100 and a duration of a week) (Figure 1) - resulted in an estimated R0 of 2.2 - 3.2 (90% CrI, Table 1). The most likely scenario with a mean serial interval of 12 days - with an outbreak size of 200 and a duration of 7 days - resulted in an estimated R0 of 2.8 - 3.9 (90% CrI, Table 2). Across all mean serial interval scenarios R0 estimates were comparable when event size was decreased and event duration was increased.

Table 1: Median, minimum, and maximum reproduction numbers of the Wuhan outbreak conditioned on case data from the 25th of January - for scenarios with a mean serial interval of 8.4 (SARS-like). Stratified by initial transmission event size and duration.

Transmission event size vs. Transmission event duration	7	14	21	28
20	3 - 4	2.7 - 3.9	2.3 - 3.9	2.1 - 3.8
40	2.8 - 3.9	2.3 - 3.8	2 - 3.4	1.8 - 3.2
60	2.5 - 3.8	2.1 - 3.4	1.8 - 3	1.7 - 2.7
80	2.3 - 3.5	1.9 - 3	1.7 - 2.6	1.5 - 2.3
100	2.2 - 3.2	1.8 - 2.7	1.6 - 2.4	1.5 - 2.2
200	1.7 - 2.3	1.5 - 2	1.3 - 1.8	1.2 - 1.6
400	1.2 - 1.6	1.1 - 1.4	1 - 1.3	0.9 - 1.2

Table 2: Median, minimum, and maximum reproduction numbers of the Wuhan outbreak conditioned on case data from the 25th of January - for scenarios with a mean serial interval of 12. Stratified by initial transmission event size and duration.

Transmission event size vs. Transmission event duration	7	14	21	28
20	-	3.6 - 3.6	3.2 - 4	2.8 - 4
40	-	3.5 - 4	3 - 4	2.7 - 3.9
60	3.5 - 4	3.2 - 4	2.8 - 3.9	2.4 - 3.9
80	3.4 - 4	3.1 - 4	2.7 - 3.9	2.2 - 3.7
100	3.5 - 4	3 - 3.9	2.5 - 3.8	2.1 - 3.4
200	2.8 - 3.9	2.2 - 3.2	1.9 - 2.7	1.7 - 2.4
400	1.8 - 2.5	1.5 - 2	1.3 - 1.7	1.2 - 1.6

Discussion

In this study, we explored a range of scenarios for the initial event size and duration of the exposure event which initiated the 2019-20 Wuhan coronavirus outbreak. We conditioned on observed cases to establish the probability of each scenario, given our model, and then estimated the R0 of coronavirus. We found that there was a very low probability that the reproduction numbers was less than 1 for any scenario considered. Across all mean serial interval scenarios larger exposure events over a shorter time horizon were most plausible. The most probable SARS-like serial interval scenarios resulted in an estimated R0 of 2.2 - 3.2 (90% CrI), whilst the most probable longer serial interval scenarios resulted in an estimated R0 of 2.8 - 3.9 (90% CrI). Reducing the event size led to estimates of the R0 increasing but also reduced the proportion of samples accepted. Similarly, increasing the event duration reduced the estimated R0 whilst decreasing the proportion of accepted samples. Decreasing the event size whilst increasing the duration resulted in R0 estimates that were comparable to those from the most plausible scenarios.

Our study used a stochastic model to capture the transmission dynamics of the outbreak with parameters

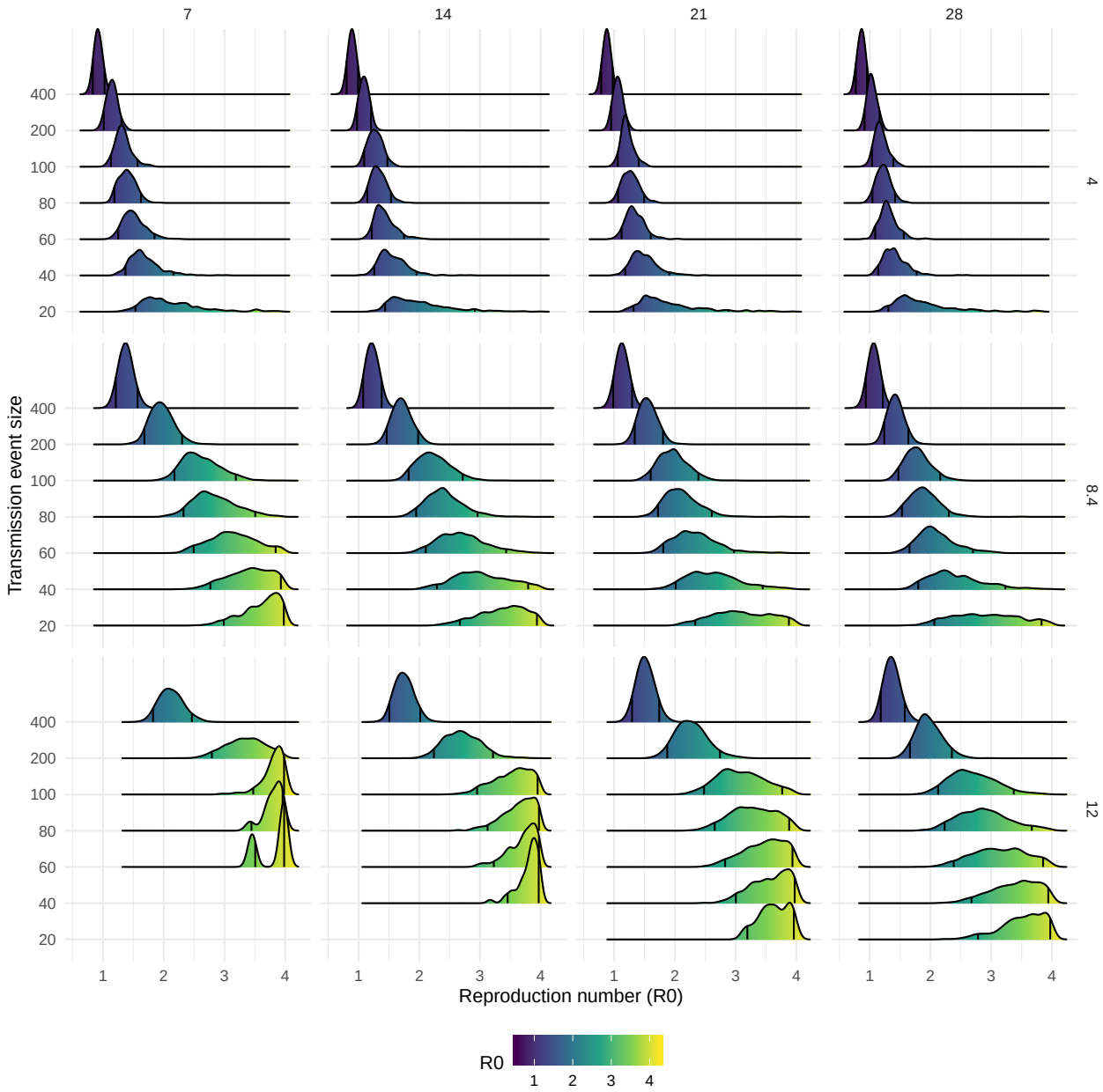

Figure 2: Density plot of reproduction number (R0) estimates from each accepted sample stratified by transmission event size, event duration (columns), and the mean serial interval used (rows). The black lines on each density plot represent the 90% credible interval

informed from data were possible or assumed to be similar to those estimated for SARS [5]. As the outbreak progresses direct simulation may become too computationally demanding to be practical so other approaches may be required. More data is likely to become available, in particular disease specific estimates for the serial interval, and case onset data, during the course of the outbreak. This will make it possible to estimate the R0 with greater precision with less risk of bias due to unknown parameters. The number of scenarios that need to be evaluated may also be reduced as additional information about cases connected to the initial exposure event becomes available. Though our estimates had wide credible intervals it is possible that we could not fully account for the numerous sources of bias and uncertainty present in the available data. This means that our model estimates may be both spuriously precise and potentially biased. There is some evidence of this in our results as the scenarios with the highest acceptance rate were on the edge of our scenario grid both for event size, event duration, and mean serial interval. This may be the result of these scenarios reducing volatility and therefore having narrower distributions of estimated cases. Indeed, we found that R0 estimates were comparable as event size decreased and event duration increased. Expert knowledge relating to the size and duration of the intial event may help clarify this issue. Alternatively, other estimates of R0 may be used to indicate which event size and event duration scenarios are most plausible. A previous study also looked at varying the event size and the impact that this had on R0 estimates using a branching process [4]. Our work builds on this by also looking at event duration, including reporting delays, and using a different approach to condition on observed cases. For comparable scenarios, our results were similar to those previously published but we found that R0 estimates were highly sensitive to variation in the assumed serial interval, event size, and event duration. We made use of a highly reproducible framework (an R package) and have released all of our code as open-source [9]. This means that this analysis may be repeated - both by the authors and others - as more data becomes available. In addition, subject area experts may be able to adapt our analysis using this open-source code to reduce the potential for bias using their expert knowledge or privately held data.

As the outbreak progress more data will become available on the number of cases, and the duration of the serial interval. These data are likely to improve our estimates of the R0 and also alter the likelihood of scenarios. Additional data may also lead to a reevaluation of the suitability of the negative binomial distribution for generating new cases - or at least provide an outbreak specific estimate of the dispersion. More data may also allow other more complex analysis to be done without risking incorporating significant biases. Our analysis did not include interventions, such as case isolation, doing so may improve our estimates. The R package we have developed alongside our analysis may be generalisable to other point source outbreaks when time series data on cases is unavailable or difficult to verify. Additional work is needed to ensure the robustness of this tool but this may allow this analysis to be repeated during future outbreaks with little additional overhead.

This analysis use a stochastic branching process to explore scenarios around the duration and size of the initial exposure event at the Huanan seafood wholesale market in Wuhan. Despite the scarcity of data currently available our estimates may be used to rule out some scenarios and to assess the likelihood of others. Our results indicate that it is very unlikely that the infectious agent responsible for the Wuhan outbreak has a R0 of less than 1, unless the size of the transmission event was much greater than currently reported. We also found that a large initial exposure event was likely with a relatively short duration. This corresponds with the evidence of rapid detection by Public Health Officials in Wuhan. These scenarios resulted in R0 estimates that are comparable to those estimated during the 2002-2003 SARS outbreak. Unfortunately, we could not identify whether scenarios with a SARS-like or longer serial interval were more likely. As more information becomes available it may be possible to further refine our results and establish the R0 of the outbreak more firmly. Providing clear quantitative information for decision makers on the transmissibility of coronavirus is of clear public health importance. Our work to make this process reproducible may reduce the time these estimates take to be made available in future outbreaks and increase knowledge sharing across response teams.

Contributors

All authors conceived and designed the work. SA undertook the analysis with advice from all other authors. SF developed the branching process model. SA wrote the first draft of the paper and all authors contributed to subsequent drafts. JH reviewed the analysis code. All authors approve the work for publication and agree

to be accountable for the work.

Funding

Need funding statement

Competing interests

There are no competing interests.

Accessibility of data and programming code

The code for this analysis, interim results, and final results can be found at: *need zenodo DOI link*

References

1 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

2 Thompson RN. 2019-20 Wuhan coronavirus outbreak: Intense surveillance is vital for preventing sustained transmission in new locations. *bioRxiv* 2020;1–14.

3 **China coronavirus 'spreads before symptoms show'**. https://www.bbc.co.uk/news/world-asia-china-51254523

4 Imai N, Cori A, Dorigatti I *et al.* **Report 3: Transmissibility of 2019-nCoV**. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf

5 Lipsitch M. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science* 2003;**300**:1966–70.

6 Xu B, Gutierrez B, Hill S *et al.* Epidemiological Data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data. 2020.

7 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria:: R Foundation for Statistical Computing 2019. https://www.R-project.org/

8 Funk S. *Bpmodels: Analysing chain statistics using branching process models*. 2020. https://github.com/sbfnk/bpmodels

9 Sam Abbott SF James Munday. *Bpmodels: Analysing chain statistics using branching process models*. 2020. https://github.com/epiforecasts/WuhanSeedingVsTransmission

10 Boettiger C. An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review* 2015;**49**:71–9.