*International Journal of Infectious Diseases*
Published online: June 2021
doi:10.1016/j.ijid.2021.06.028

# A hospital-related outbreak of SARS-CoV-2 associated with variant Epsilon (B.1.429) in Taiwan: transmission potential and outbreak containment under intensified contact tracing, January–February 2021

Andrei R. Akhmetzhanov[1,*], Sung-mok Jung[2,3], Hao-Yuan Cheng[4], Robin N. Thompson[5,6]

[1] *College of Public Health, National Taiwan University, Taiwan*
[2] *School of Public Health, Kyoto University, Japan*
[3] *Graduate School of Medicine, Hokkaido University, Japan*
[4] *Epidemic Intelligence Center, Taiwan Centers for Disease Control, Taiwan*
[5] *Mathematics Institute, University of Warwick, U.K.*
[6] *Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, U.K.*

[*] Correspondence: akhmetzhanov@ntu.edu.tw

## Abstract

*Objectives*: A hospital-related cluster of 22 cases of coronavirus disease 2019 (COVID-19) occurred in Taiwan in January–February 2021. Rigorous control measures were introduced and could only be relaxed once the outbreak was declared over. Each day after the apparent outbreak end, we estimated the risk of future cases occurring was estimated in order to inform decision-making.

*Methods*: Probabilistic transmission networks were reconstructed, and transmission parameters (the reproduction number $R$ and overdispersion parameter $k$) were estimated. The reporting delay during the outbreak was estimated (Scenario 1). In addition, a counterfactual scenario with less effective interventions characterized by a longer reporting delay was considered (Scenario 2). Each day, the risk of future cases was estimated under both scenarios.

*Results*: The values of $R$ and $k$ were estimated to be 1.30 (95% credible interval (CI): 0.57–3.80) and 0.38 (95% CI: 0.12–1.20), respectively. The mean reporting delays considered were 2.5 days (Scenario 1) and 7.8 days (Scenario 2). Following the final case, the inferred probability of future cases occurring declined more quickly in Scenario 1 than Scenario 2.

*Conclusions*: Rigorous control measures allowed the outbreak to be declared over quickly following outbreak containment. This highlights the need for effective interventions, not only to reduce cases during outbreaks but also to allow outbreaks to be declared over with confidence.

## Introduction

As of 9 March 2021, there have been fewer than 1,000 confirmed cases of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection in Taiwan, of which only 77 were locally acquired [1].

Following stringent border control measures, proactive contact tracing and case isolation, Taiwan's largest individual outbreak to date was a hospital-related outbreak that involved 22 cases and occurred in January-February 2021. Despite successful containment of that outbreak, some aspects were concerning. First, the custom of wearing face masks, especially in hospital, was unable to fully prevent transmission in several instances [2]. Second, the source of infection for one of the infected inpatients was undetermined: that individual attended a hospital ward that was not included in a so-called "red zone", and he did not interact with other detected infected individuals [3]. Third, having implemented control measures at the time of the first suspected cases, additional cases continued to be seen for several weeks afterwards. This led to further investigations into possible causes of the outbreak and required an end-of-outbreak determination (i.e., assessment of the probability that the outbreak was over — or, conversely, the probability that additional reported cases would occur in future) after the last case was reported [4–6].

This article provides a descriptive analysis of the outbreak, and quantifies viral transmissibility during the outbreak. In addition, estimates of the probability that additional reported cases will occur in future are presented, as obtained in real-time after the final case had been observed. As the time since the last observed case increases, the certainty that the outbreak is over increases. Two distinct scenarios have been considered when estimating the probability of future cases. Scenario 1 describes containment of the outbreak under intensified contact tracing, as was the situation during this outbreak. Under Scenario 1, proactive testing and quarantine of all close contacts of confirmed cases (and suspected cases) after epidemiological investigations is considered, so that cases are found quickly, and transmission beyond individuals that attended hospital and their contacts is unlikely. Scenario 2 describes a situation with reduced contact tracing and testing, increasing the risk of transmission into the wider community, with some chains of transmission potentially remaining untraced. In this analysis, these scenarios are assessed by implementing two different reporting delays, which represent the time periods from symptom onset to case confirmation. The reporting delay under Scenario 1 is shorter than that under Scenario 2 due to efficient case identification, which is followed by isolation [7].

## Materials and Methods

### Outbreak investigation

A cluster of locally acquired SARS-CoV-2 infections occurred in Taiwan in January–February 2021. This cluster originated in a hospital and involved 22 reported cases Figure 1. The first two cases to be detected, a doctor (B1.1) and his household and work contact (B1.2), were suspected positive and tested on 11 January. They were then confirmed positive the following day. The authorities acted proactively by testing their close contacts on 11 January, ordering a two-week home isolation of all close contacts, restricting hospital admissions, and arranging for a second round of health inspections three days later. Regular press conferences raised public awareness and ensured that the local community remained vigilant throughout the outbreak.

The index case (A0) was a Taiwanese female in her 60s who travelled to the USA in October 2020 and returned to Taiwan on 27 December 2020. Having tested negative for SARS-CoV-2 infection in the three days before her flight, she developed initial symptoms on 29 December while in quarantine. She was later hospitalized and was placed on a ventilator. During her treatment, a doctor (B1.1) was exposed to the virus on 4 January 2021 and experienced initial mild symptoms on 8 January. The virus

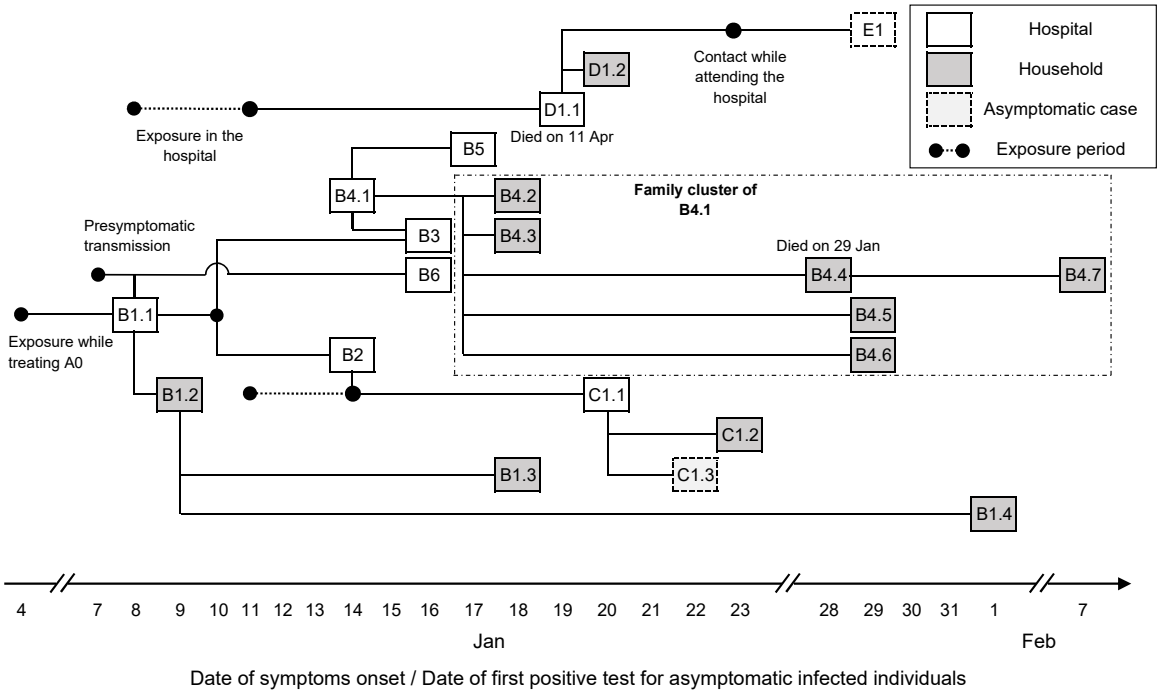



**Figure 1.** Timeline of exposure and possible connections between reported cases. Connections shown here were determined either by identifying the most probable infector via epidemiological investigation or by the earliest time of symptom onset among all close contacts if the most likely pair could not be determined (such as in family clusters of B1.2, C1.1, and B4.1).

further spread to his household contact (B1.2) and other medical personnel, most likely due to work-related interactions between B1.1, B2, and B3 on 10 January. The chains of transmission that followed then included three other work-related infections (B4–6), three infections of attending inpatients (C1.1, D1.1, E1) and transmission in their households. Household transmission accounted for 12 cases (57%), with the family cluster of B4.1 involving all seven family members including one death. In total, two deaths (B4.4, D1.1) occurred.

All cases were epidemiologically linked through contact tracing, except for an inpatient (D1.1) who had no record of contact with any known infected individual in the hospital. This suggests that his infection was likely due to either indirect transmission (e.g. via a contaminated surface from a known source) or from an undetected case. The same route of transmission could have occurred for infection of B2 by B1.1, since both individuals were wearing masks during their interaction (one of which was a highly effective surgical N95 mask).

One individual (C1.2) was pre-symptomatic when testing positive, with onset of symptoms two days later. Two infected individuals remained asymptomatic throughout infection. At least one pre-symptomatic transmission occurred: a foreign nurse (B6) was exposed to the virus on 7 January while interacting with B1.1, one day before B1.1 developed symptoms. Unlike the family cluster of B4.1, where the secondary attack rate was 100%, the employer of B6 and all his family members tested negative despite their close contact with infected case B6. Genetic sequencing of a subset of cases from the

outbreak identified variant Epsilon (also known as Cal.20C, lineage B.1.429), originated in Southern California in 2020 [8–10].

## Reconstruction of the transmission network and estimation of the transmission potential

The transmission potential of this novel variant was characterized by analysing the offspring distribution, which describes the number of secondary infections per primary case. A negative-binomial distribution was fitted with mean $R$ and overdispersion parameter $k$ [11]. The value of $R$ describes the average number of secondary infections per primary case, while $k$ measures variability in the number of secondary infections between primary cases and quantifies the potential for superspreading (which, for a fixed value of $R$, is more likely to occur for lower values of $k$).

First, the Wallinga-Teunis (WT) method [12] was applied to resolve the uncertainty in transmission patterns in family clusters by incorporating the serial interval distribution [13]. The serial interval distribution assumed here is similar to other reported estimates [14, 15]. The pairings of infectees to their infectors were known for 12 secondary cases as a result of epidemiological investigations: A0 $\rightarrow$ B1.1, B1.1 $\rightarrow$ $\{$B1.2, B2, B3, B6, D1.1$\}$, B1.2 $\rightarrow$ B1.3, B2 $\rightarrow$ C1.1, B4.1 $\rightarrow$ $\{$B4.2, B4.3$\}$ and D1.1 $\rightarrow$ $\{$D1.2, E1$\}$. The infection of case D1.1, an inpatient who attended the hospital in the first week of the outbreak, was assigned to case B1.1, given the timing (B1.1 was the only symptomatic at that time). All other potential infectors (B2, B3, B4.1, B5, and B6) developed initial symptoms more than two days after the visit of D1.1 to the hospital (Figure 1). The infectors of the other nine cases (excluding the index case A0) were uncertain, with the following possibilities: $\{$B1.2, B1.3$\}$ $\rightarrow$ B1.4, $\{$B1.1, B3, B2$\}$ $\rightarrow$ B4.1, $\{$B4.1, B4.5, B4.6$\}$ $\rightarrow$ B4.4, $\{$B4.1, B4.4, B4.6$\}$ $\rightarrow$ B4.5, $\{$B4.1, B4.4, B4.5$\}$ $\rightarrow$ B4.6, $\{$B4.4, B4.5, B4.6$\}$ $\rightarrow$ B4.7, $\{$C1.1, C1.3$\}$ $\rightarrow$ C1.2, and $\{$C1.1, C1.2$\}$ $\rightarrow$ C1.3. Almost all of these transmissions (except for infection of case B5) may have been due to household transmission, and so exact determination of who infected whom is impossible. The infector of case B5 could not be identified precisely as that transmission likely occurred in the workplace, where case B5 contacted multiple possible infectors. Under the WT method, for each of those nine infectees $i$, we selected an infector $j$ from their lists of potential infectors $J_i$ based on probabilistic sampling. The likelihood that case $j$ (with symptoms onset at time $t_j$) infected case $i$, relative to the likelihood that any other potential infector infected case $i$, was given by:

$$p_{ij} = \frac{g(t_i - t_j \mid \{\mu_{\mathrm{SI}}, k_{\mathrm{SI}}\})}{\sum_{v \in J_i} g(t_i - t_v \mid \{\mu_{\mathrm{SI}}, k_{\mathrm{SI}}\})}, \tag{1}$$

where $g(\circ \mid \{\mu_{\mathrm{SI}}, k_{\mathrm{SI}}\})$ represents the serial interval distribution modeled by a Weibull distribution with the mean $\mu_{\mathrm{SI}} = 4.8 \pm 0.6$ days (i.e. $\mu_{\mathrm{SI}} \sim \mathcal{N}(4.8, 0.6)$) and shape parameter $k_{\mathrm{SI}} = 2.3 \pm 0.4$ [13].

Second, the number of transmissions from each primary case in any probabilistic realization of the transmission network was determined. A negative binomial probability mass function was fitted to each resulting distribution, with mean $R$ and overdispersion parameter $k$.

## Generation-based reproduction number

The reconstructed transmission networks allowed us to make a probabilistic assignment of generation membership to cases and derive the generation-based reproduction number, $R_m$ [16, 17]. Given a

particular network, each node (i.e. each case) was assigned to a generation $m$, where the value of $m$ represents the number of links from that node to the index case A0. The node A0 was placed at the root of the network and assigned to generation zero. To derive the generation-based reproduction number, $R_m$, we divided the number of transmissions generated by cases in generation $m$ by the number of cases in that generation. Hence, the reproduction number for the final generation $M$ (so that other generations $m \leq M$) was exactly zero. The reproduction number for generation zero was equal to one. Because the transmission networks were generated probabilistically, each $R_m$ was also characterized by a posterior distribution.

## Estimation of the reporting delay

Fitting the reporting delay distribution with a mixture of three shifted distributions (gamma, Weibull, and lognormal), the mean reporting delay for this outbreak was estimated (under the intensive measures that were in place during this outbreak – Scenario 1). In addition, a counterfactual scenario (Scenario 2) in which public health measures are less rigorous was considered. Rather than attempting to model the wide range of possible effects of less rigorous contact tracing and case isolation, in Scenario 2 we simply set the reporting delay to be longer than in Scenario 1. In Scenario 2, we set the mean reporting delay by estimating its value using data for all local cases reported in Taiwan since the beginning of 2020.

Specifically, for each scenario, we extracted data describing dates of symptoms onset and confirmation for all symptomatic cases. The number of extracted cases was $N = 20$ for Scenario 1 and $N = 68$ for Scenario 2. The likelihood was given by a mixture of three component likelihoods with respective weights $w_l$ ($l = 1, 2, 3$) and potential right truncation at the time of the latest update $T$ was accounted for:

$$L(\theta \,|\, \{\Delta_i\}) = \sum_{l=\{1,2,3\}} w_l L^{(l)}(\theta \,|\, \{\Delta_i\}), \tag{2}$$

$$L^{(l)}(\theta \,|\, \{\Delta_i\}) = \prod_{i=1\ldots N} \frac{f(\Delta_i \,|\, \theta)}{F(T - o_i \,|\, \theta)}, \tag{3}$$

where $\Delta_i = c_i - o_i$ is the time difference between confirmation $c_i$ and symptom onset $o_i$ for case $i$. Because the extracted data contained only the dates of symptom onset $O_i$ and confirmation $C_i$, we assumed that the priors for the precise times of symptom onset $o_i$ and confirmation $c_i$ were uniformly distributed within those days: $o_i \sim \mathcal{U}(O_i, O_i + 1)$ and $c_i \sim \mathcal{U}(C_i, C_i + 1)$. Some observed values of $\Delta_i$ were negative, so that the reporting delay distributions were modeled by either shifted versions of the gamma, Weibull, or lognormal distributions ($l = 1, 2, 3$). The function $f_l(\Delta_i \,|\, \theta)$ denoted the probability density function (PDF):

$$f_l(\Delta_i \,|\, \theta = \{\tau, \mu, \sigma) = \mathrm{PDF}_l(\Delta_i + \tau \,|\, \mu, \sigma), \tag{4}$$

where $\tau$ is the shift of distribution $l$ ($\tau > 0$), $\mu$ and $\sigma$ are the mean and standard deviation of the distribution $l$. To improve the convergence of the mixture model, it was assumed that parameters $\{\tau, \mu, \sigma\}$ were common to the three distributions, as has been proposed elsewhere for Bayesian model averaging [18, 19]. The notation $F(T - o_i \,|\, \theta)$ was used to denote the cumulative distribution function (CDF) of the reporting delay. The truncation time $T$ was set corresponding to the beginning of 12 April 2021.

The relative weightings of the different component distributions were defined using the formula:

$$q_l = \frac{w_l L^{(l)}(\theta \mid \{\Delta_i\})}{L(\theta \mid \{\Delta_i\})}. \tag{5}$$

### End-of-outbreak probability

On a given day, to estimate whether or not the outbreak was already over, we used a previously described method devised by Linton *et al.* [20]. First, we considered the epidemic curve up to the time of report $t$ with dates of symptom onsets $o_i < t$ for all symptomatic cases $i = 1 \ldots N$. The probability that one or more new cases will be reported after day $t$ is given by the following expression:

$$\Pr(X(t) > 0) = 1 - \prod_{i=1}^{N} \sum_{y=1}^{\infty} p_y \left[H_i(t \mid \theta)\right]^y. \tag{6}$$

In this expression, $X(t)$ is the number of cases reported on day $t$ and $p_y$ is the probability of $y$ transmissions occurring from a primary case $i$, which follows a negative binomial distribution with mean $R$ and overdispersion parameter $k$ as described above. The function $H_i(t \mid \theta)$ represents the probability that an individual infected by case $i$ reports infection by time $t$. This function is therefore the CDF of a convolution of the serial interval and the reporting delay. For each potential infectee, the reporting delay was selected at random from the three distributions described above according to the probabilities $q_l$ (5).

### Technical details

R 4.1.0 [21] and CmdStan 2.27.0 [22] were used to conduct the main analysis; Python 3.6 was used for statistical inference of the generation-based reproduction number. Reproducible code for this study is available on GitHub at https://github.com/aakhmetz/Taiwan-COVID19-end-of-outbreak-JanFeb2021. All derived estimates of model parameters and the results of sensitivity analyses can also be found in Supplementary Materials (Supplementary Tables 1–2, Supplementary Figures 1–3).

## Results

The statistical inference of the offspring distribution identified the median estimate of $R$ to be 1.30 (95% credible interval (CI): 0.57–3.80) and the median estimate of $k$ to be 0.38 (95% CI: 0.12–1.20). The generation-based reproduction number (i.e., the expected number of transmissions arising from an infector in a specific generation of the transmission chain, where patient A0 represents generation 0) declined throughout the outbreak from generation 1 onwards. In generation 1, the generation-based reproduction number was estimated to be 6, falling below 1 by generation 3 (Supplementary Figures 4–5; Supplementary table 2). Inspection of probabilistic transmission networks (Supplementary figure 6) confirmed a high value of the case reproduction number, $R$, for B1.1, but also supported sequential transmission of the virus within households resulting in a greater estimated value of $k$ compared with most earlier studies [23–25]. A previous study by Ng *et al.* [26] involved an analysis of data from Taiwan from 2020, and found an estimated value of $k$ that was substantially larger, in part due to the small sample size in their analysis (the posterior mean of $k$ was 19.20).

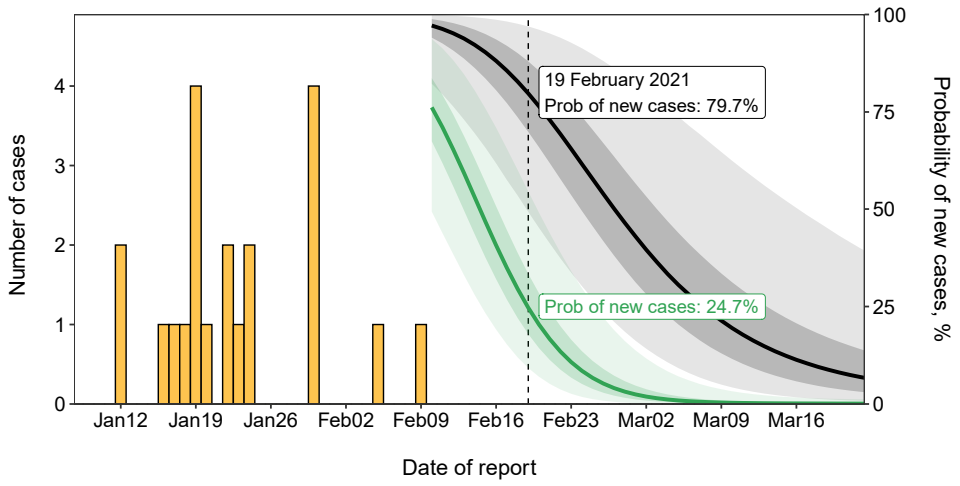



**Figure 2.** The estimated risk of cases being reported in future for Scenario 1 (under intensified contact tracing; green) and Scenario 2 (less rigorous public health measures; black). Bars in orange indicate the incidence of COVID-19 by confirmation date.

The mean reporting delay for the outbreak (under the intensive measures that were in place for this outbreak – Scenario 1) was estimated to be 2.5 days (95% CI: 1.8–3.5) with a standard deviation (SD) of 1.6 days (95% CI: 1.1–2.9). Under counterfactual Scenario 2, where the mean reporting delay was estimated for all local cases reported in Taiwan since the beginning of 2020, the mean reporting delay was 7.8 days (95% CI: 6.2–10.1) with a SD of 7.8 days (95% CI: 5.7–13.2). Due to a small number of negative delays (i.e. some individuals were detected prior to developing symptoms), the distributions were shifted approximately one day earlier as a result of the model fitting (1.0 day (95% CI: 0.2–2.9) for Scenario 1 and 0.8 days (95% CI: 0.1–2.0) for Scenario 2). The observed difference in mean reporting delays between Scenario 1 and Scenario 2 can be attributed to different ways in which cases were detected. Under Scenario 1, cases were detected quickly by rigorous contact tracing, whereas under Scenario 2 cases were detected by a combination of some contact tracing and symptom-based surveillance [23].

Incorporating the posterior distributions for $R$, $k$, the serial interval [13] and the reporting delay in the formula for the end-of-outbreak probability (equation (6)), following the final case reported in this outbreak, a sharper decline in the estimated probability that new cases will be reported in future was observed under Scenario 1 than Scenario 2 (green and black lines and regions in Figure 2). Ten days after the last reported case, on 19 February this probability (reported here as a percentage) dropped to 24.7% under Scenario 1 compared to 79.7% under Scenario 2. Depending on the policy-maker's "acceptable risk", different thresholds in this probability could be chosen before declaring an outbreak over [27]. For instance, if a threshold of 10% is chosen, the outbreak could have been declared over on 24 February under Scenario 1 compared to a later date of 18 March under Scenario 2. Sensitivity analyses are presented for different values of $R$ and $k$, as well as different reporting delay distributions for Scenario 2, in the Supplementary Material, indicating qualitatively similar results. In each case, more rigorous control measures (characterized by a shorter reporting delay) allow policy-makers to be confident that the outbreak is over sooner after the final reported case.

## Conclusions

In summary, our results suggest that the rigorous public health measures that were in place allowed the outbreak end to be declared around three weeks earlier than if these intensive measures were not introduced. More generally, stringent control measures allow policy-makers to be confident that outbreaks are over earlier compared to scenarios with less intense measures. However, in the outbreak considered here, even with strict control measures, public vigilance was required for several weeks after the final reported case until total confidence that the outbreak was over was achieved (Figure 2).

To conclude, proactive countermeasures and high public compliance contributed to efficient containment and a high confidence that the hospital-related outbreak in Taiwan was over by late February 2021. We note that later identification of the outbreak could have led to larger number of infections [28, 29], and therefore potentially a later end-of-outbreak declaration.

**Acknowledgements:** We thank Taiwan public health authorities and institutions for surveillance, laboratory testing, epidemiological investigations, and data collection. We are grateful to two anonymous reviewers for their helpful comments. A.R.A. also thanks Yin-Chin Fan and Yun-Chun Wu (National Taiwan University) for discussions about this topic.

**Conflict of interests:** We declare that we have no conflict of interest.

**Ethical standards:** The present study used publicly available data, and thus, did not require ethical approval.

**Financial source:** S-m.J. received funding from the Japan Society for the Promotion of Science (JSPS) via a KAKENHI grant (20J2135800).

## References

1. Taiwan Centers for Disease Control. 2021. Available from: https://www.cdc.gov.tw. [Accessed 12 April 2021].

2. Central Epidemic Command Center. COVID-19 press conference of 17 January 2021 (in Chinese); 2021b. Available from: https://youtu.be/LvGoNhe2lXw. [Accessed 12 April 2021].

3. Central Epidemic Command Center. COVID-19 press conference of 5 February 2021 (in Chinese); 2021a. Available from: https://youtu.be/SLMJIV6ncBI. [Accessed 12 April 2021].

4. Djaafara BA, Imai N, Hamblion E, Impouma B, Donnelly CA, Cori A. A quantitative framework for defining the end of an infectious disease outbreak: application to Ebola virus disease. Am J Epidemiol 2021;190(4):642-51 (doi:10.1093/aje/kwaa212)

5. Nishiura H, Miyamatsu Y, Mizumoto K. Objective determination of end of MERS outbreak, South Korea, 2015. Emerg Infect Dis 2016;22(1):146-8 (doi:10.3201/eid2201.151383)

6. Parag KV, Donnelly CA, Jha R, Thompson RN. An exact method for quantifying the reliability of end-of-epidemic declarations in real time. PLoS Comput Biol 2020;16(11):e1008478 (doi:10.1371/journal.pcbi.1008478)

7. Tian L, Li X, Qi F, Tang Q-Y, Tang V, Liu J, *et al.* Harnessing peak transmission around symptom onset for non-pharmaceutical intervention and containment of the COVID-19 pandemic. Nat Commun 2021;12(1):1147 (doi:10.1038/s41467-021-21385-z)

8. GISAID. Accession numbers of first 3 samples: EPI_ISL_956329, EPI_ISL_956330, EPI_ISL_1020315. 2021.

9. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 variant in Southern California. JAMA 2021;325(13):1324-6 (doi:10.1001/jama.2021.1612)

10. McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. Science 2021;eabi7994 (doi:10.1126/science.abi7994)

11. 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;25(4):2000058 (doi:10.2807/1560-7917.ES.2020.25.4.2000058)

12. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 2004;160:509-16 (doi:10.1093/aje/kwh255)

13. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020;93:284-6 (doi:10.1016/j.ijid.2020.02.060)

14. Biggerstaff M, Cowling BJ, Cucunubá ZM, Dinh L, Ferguson NM, Gao H, *et al.* Early insights from statistical and mathematical modeling of key epidemiologic parameters of COVID-19. Emerg Infect Dis 2020;26(11) (doi:10.3201/eid2611.201074)

15. Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID-19 symptom onset highlights the importance of continued contact tracing. eLife 2021;10:e65534 (doi:10.7554/eLife.65534)

16. Akhmetzhanov AR, Lee H, Jung S-m, Kinoshita R, Shimizu K, Yoshi K, *et al.* Real time forecasting of measles using generation-dependent mathematical model in Japan, 2018. PLoS Curr Outbreaks 2018;10:ecurrents.outbreaks.3cc277d133e2d6078912800748dbb492 (doi:10.1371/currents.outbreaks.3cc277d133e2d6078912800748dbb492)

17. Worden L, Ackley SF, Zipprich J, Harriman K, Enanoria WTA, Wannier R, *et al.* Measles transmission during a large outbreak in California. Epidemics 2020;30:100375 (doi:10.1016/j.epidem.2019.100375)

18. Keller M, Kamary K. Bayesian model averaging via mixture model estimation. arXiv 2018:1711.10016

19. Akhmetzhanov AR. Estimation of delay-adjusted all-cause excess mortality in the USA: March–December 2020. Epidemiol Infect 2021 (doi:10.1017/S0950268821001527)

20. Linton NM, Akhmetzhanov AR, Nishiura H. Localized end-of-outbreak determination for coronavirus disease 2019 (COVID-19): examples from clusters in Japan. Int J Infect Dis 2021;105:286-92 (doi:10.1016/j.ijid.2021.02.106)

21. R Development Core Team. R: a language and environment for statistical computing; 2021. Available from: https://www.r-project.org.

22. Stan Development Team. Stan modeling language users guide and reference manual, 2.27.0; 2021. Available from: https://mc-stan.org. [Accessed 3 June 2021].

23. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020;20(8):911-9 (doi:10.1016/S1473-3099(20)30287-5)

24. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China [version 3; peer review: 2 approved]. Wellcome Open Res 2020;5 (doi:0.12688/wellcomeopenres.15842.3)

25. Nakajo K, Nishiura H. Transmissibility of asymptomatic COVID-19: Data from Japanese clusters. Int J Infect Dis 2021;105:236-8 (doi:10.1016/j.ijid.2021.02.065)

26. Ng T, Cheng H, Chang H, Liu C, Yang C, Jian S, *et al.* Effects of case- and population-based COVID-19 interventions in Taiwan. medrxiv 2020 (doi:10.1101/2020.08.17.20176255)

27. Thompson RN, Morgan OW, Jalava K. Rigorous surveillance is necessary for high confidence in end-of-outbreak declarations for Ebola and other infectious diseases. Phil Trans Roy Soc B 2019;374(1776):20180431 (doi:10.1098/rstb.2018.0431)

28. Liu J, Huang J, Xiang D. Large SARS-CoV-2 outbreak caused by asymptomatic traveler, China. Emerg Infect Dis 2020;26:2260-3 (doi:10.3201/eid2609.201798)

29. Akhmetzhanov AR. Large SARS-CoV-2 Outbreak Caused by Asymptomatic Traveler, China. Emerg Infect Dis 2020;26(12):3106 (doi:10.3201/eid2612.203437)

# Supplementary Materials

## Results of statistical inference

The median values and 95% credible intervals for all model parameters are shown in Supplementary Tables 1–2. The results of the sensitivity analysis are shown in Supplementary Figures 1–3.

# Supplementary Tables

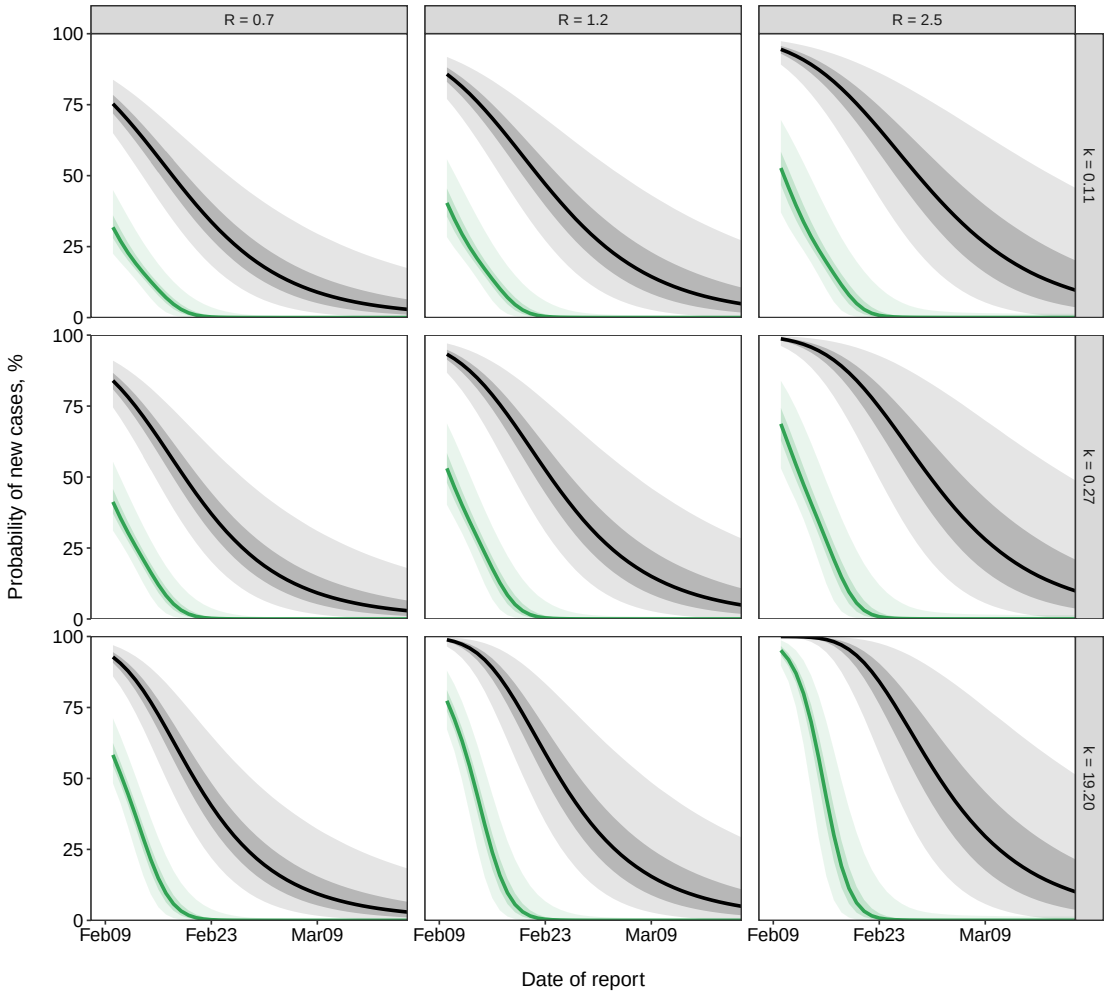
**Supplementary Table 1.** Median estimates and 95% credible intervals (CI) for inferred model parameters.

| Parameter | Median | 95% CI |
|---|---|---|
| Reproduction number $R$ | 1.30 | 0.57–3.80 |
| Overdispersion parameter $k$ | 0.38 | 0.12–1.20 |
| Mean reporting delay for Scenario 1, days (under intensified contact tracing) | 2.5 | 1.8–3.5 |
| SD of the reporting delay for Scenario 1, days | 1.6 | 1.1–2.9 |
| Negative shift in the reporting delay for Scenario 1, days | 1.0 | 0.2–2.9 |
| Mean reporting delay for Scenario 2, days | 7.8 | 6.2–10.1 |
| SD of the reporting delay for Scenario 2, days | 7.8 | 5.7–13.2 |
| Negative shift in the reporting delay for Scenario 2, days | 0.8 | 0.1–2.0 |

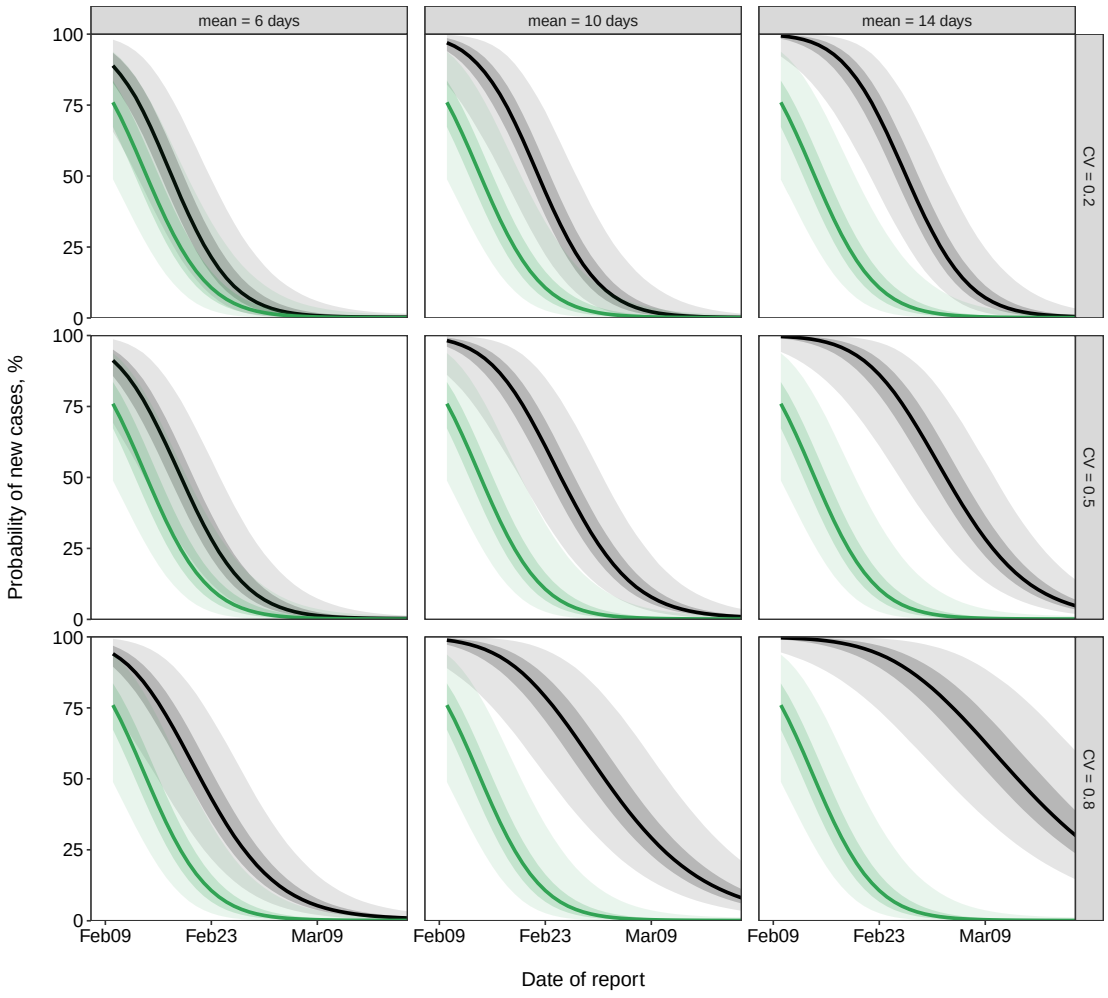
**Supplementary Table 2.** Median estimates and 95% credible intervals (CI) of the generation-based reproduction number.

| Generation number | Median | 95% CI |
|---|---|---|
| 0 | 1 | 1–1 |
| 1 | 6 | 6–7 |
| 2 | 1.33 | 1.0–1.5 |
| 3 | 0.50 | 0.33–0.75 |
| 4 | 0.33 | 0.2–0.67 |
| 5 | 0 | 0–0 |

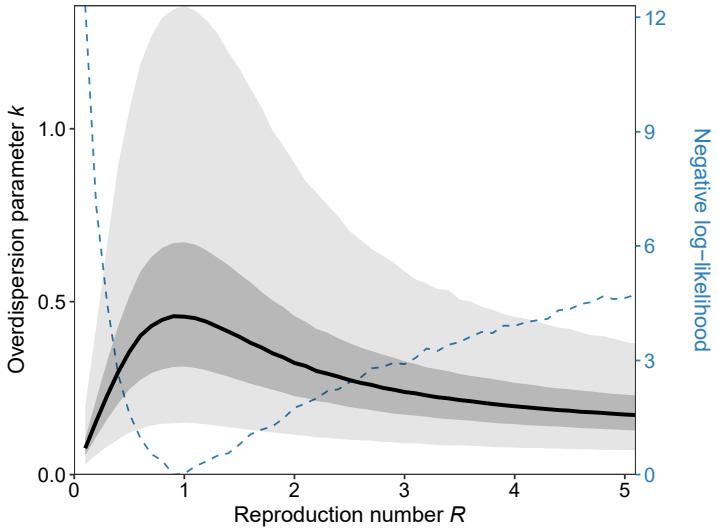
## Supplementary Figures



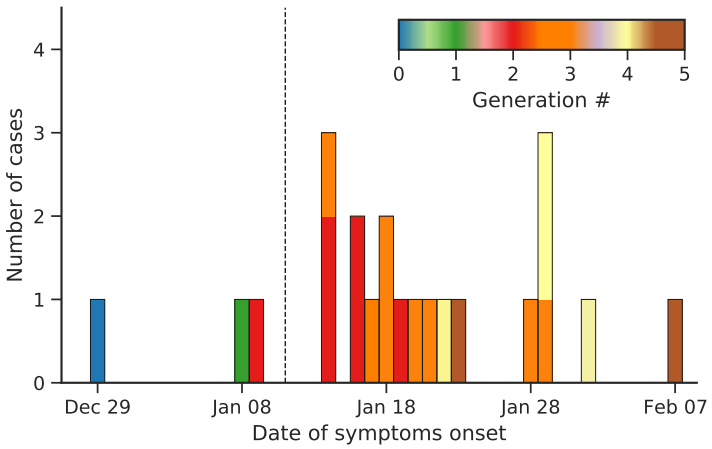

**Supplementary Figure 1.** Probability of new cases occurring in future for Scenario 1 (under intensified contact tracing; green) and Scenario 2 (less intense control measures; black) for values of $R = 0.7, 1.2, 2.5$ (columns) and $k = 0.11, 0.27, 19.20$ (rows) as found in previous studies [23, 24, 26]. The median values are indicated by the solid lines. The light shaded areas show 95% credible intervals, while the dark shaded areas show the interquartile ranges.



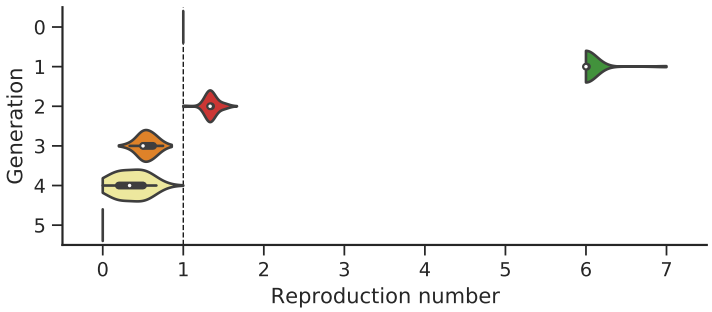

**Supplementary Figure 2.** Probability of new cases occurring in future for Scenario 1 (under intensified contact tracing; green) and Scenario 2 (less intense control measures; black) for different values of the reporting delay under Scenario 2 modeled by gamma distribution shifted to the left by one day with the mean $= 6, 10, 14$ days (columns) and coefficient of variation CV $= 0.2, 0.5, 0.7$ (rows). The median values are indicated by the solid lines. The light shaded areas show 95% credible intervals, while the dark shaded areas show the interquartile ranges.



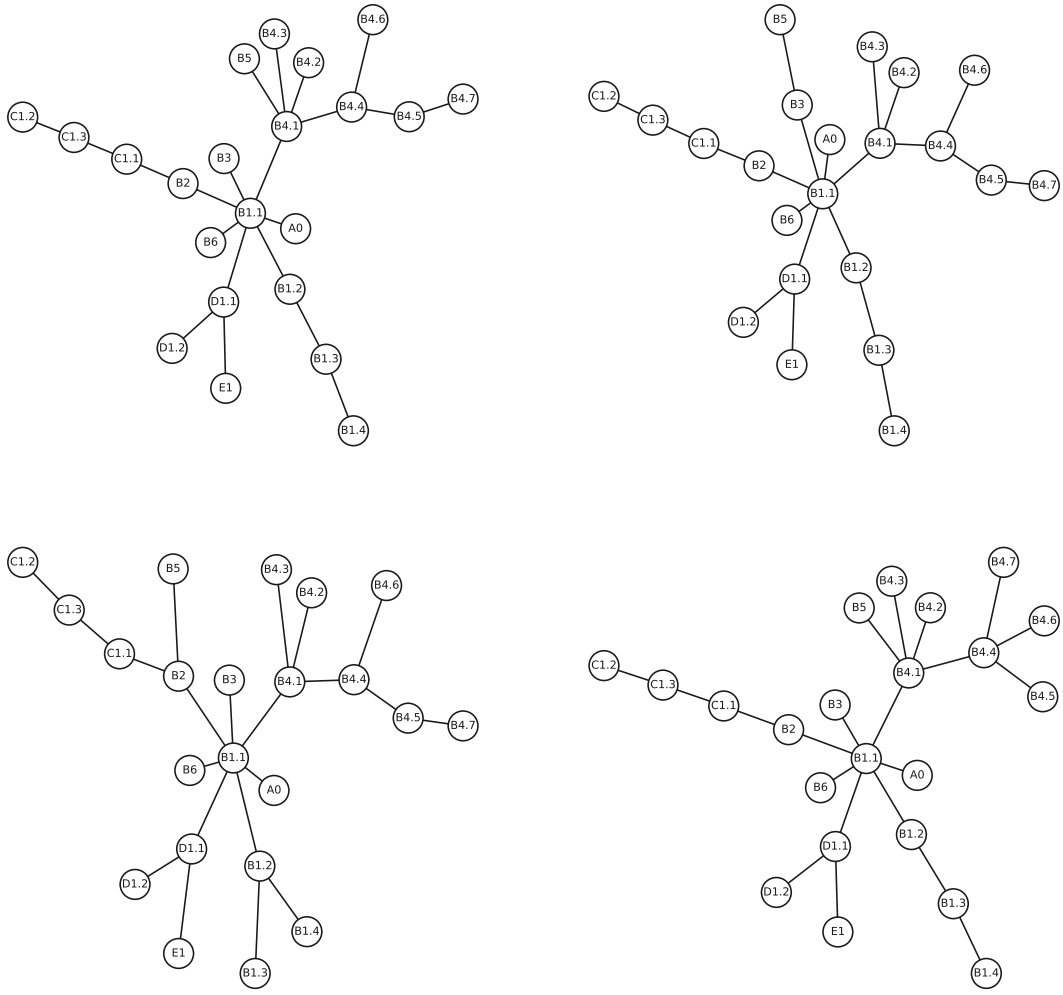

**Supplementary Figure 3.** Variation in the estimated overdispersion parameter, $k$, when the value of the reproduction number, $R$, is fixed ($x$-axis). The dashed blue line shows the change in the negative log-likelihood. The median estimate of $k$ is indicated by solid line. The light shaded area shows the 95% credible interval, while the dark shaded area shows the interquartile range.




**Supplementary Figure 4.** Epidemic curve indicating the dates on which cases developed symptoms, with cases classified by their estimated generation number (see color bar). The dashed line denotes the date that the outbreak was discovered and interventions were introduced.




**Supplementary Figure 5.** Generation-based reproduction number. The dashed vertical line indicates the threshold where the generation-based reproduction number is equal to 1.




**Supplementary Figure 6.** Possible realizations of transmission networks based on known epidemiological links using the Wallinga-Teunis method [12].