

Estimating Transmission Parameters for COVID-19 Clusters by Using Symptom Onset Data, Singapore, January–April 2020

Sheryl Hui-Xian Ng, Palvinder Kaur, Cécile Kremer, Woan Shin Tan, Aidan Lyanzhiang Tan, Niel Hens, Matthias Paul Toh, Kiok Liang Teow, Palvannan Kannapiran

We estimated the generation interval distribution for coronavirus disease on the basis of serial intervals of observed infector–infectee pairs from established clusters in Singapore. The short mean generation interval and consequent high prevalence of presymptomatic transmission requires public health control measures to be responsive to these characteristics of the epidemic.

A systematic review estimated that the basic reproduction number (R_0) for coronavirus disease (COVID-19) is 2–3 (1). However, alone is insufficient to characterize an epidemic. The distribution of the serial interval (i.e., the length of time between symptom onset of 2 cases) has been estimated for COVID-19; mean intervals range from 3.1 to 7.5 days (2,3). Estimation of the generation interval (T_g) (i.e., the length of time between the points of infection for 2 linked cases) is less common. Although studies have reported means of 3.3 and 5.0 days (4; Li et al., unpub. data, <https://doi.org/10.1101/2020.02.26.20028431>), Ganyani et al. (5) estimated the mean (\pm SD) of T_g to be 3.9 (\pm 2.7) days on the basis of which they estimated that 66% (95% credible interval [CrI] 45%–84%) of transmission occurred before symptoms. Another study of 77 pairs estimated the same proportion to be 44% (95% CI 25%–69%) (6). Because conventional outbreak control measures are centered around isolation, contact tracing, and treatment of symptomatic case-patients, a high prevalence of presymptomatic transmission (p) would warrant shifting

measures to address potential transmission among persons with no apparent symptoms (7). Hence, to inform control measures for the outbreak in Singapore, we generated estimates of T_g , R_0 , and p by using published symptom onset data for COVID-19 cases in Singapore.

The Study

We implemented a cross-sectional study design to estimate T_g , R_0 , and p for the COVID-19 outbreak in Singapore during January 23–April 6, 2020. Given that containment measures were initiated over the duration of the study, we considered R_0 to be the effective reproduction number of the outbreak. All confirmed COVID-19 cases classified by the Ministry of Health of Singapore (MOH) as linked to a local cluster were included in this analysis. Information on case number, cluster, patient age and sex, imported status, date of symptom onset (DOO), and known contacts who have also been confirmed as case-patients were extracted from daily press releases published by MOH. DOOs for cases that were not available from press releases were extracted from a similar anonymized dataset of COVID-19 admissions to the National Centre for Infectious Diseases, Singapore. Cases with DOOs not available from that dataset were subsequently excluded from analysis. Our study was approved by the ethics review board of National Healthcare Group, Singapore.

We identified index cases and potential infectors of each case-patient on the basis of available information of the case-patients' known contacts, published case links, and a heuristic to sensibly include potential infectors who could have transmitted the infection to the case-patients (Appendix, <https://wwwnc.cdc.gov/EID/article/27/2/20-3018-App1.pdf>). We subsequently used the infector–infectee pairs constructed to estimate the serial and generation interval distribution.

Author affiliations: National Healthcare Group, Singapore (S.H.-X. Ng, P. Kaur, W.S. Tan, A.L. Tan, K.L. Teow, P. Kannapiran); Hasselt University, Hasselt, Belgium (C. Kremer, N. Hens); University of Antwerp, Antwerp, Belgium (N. Hens); National Centre for Infectious Diseases, Singapore (M.P. Toh); National University of Singapore, Singapore (M.P. Toh)

DOI: <https://doi.org/10.3201/eid2702.203018>

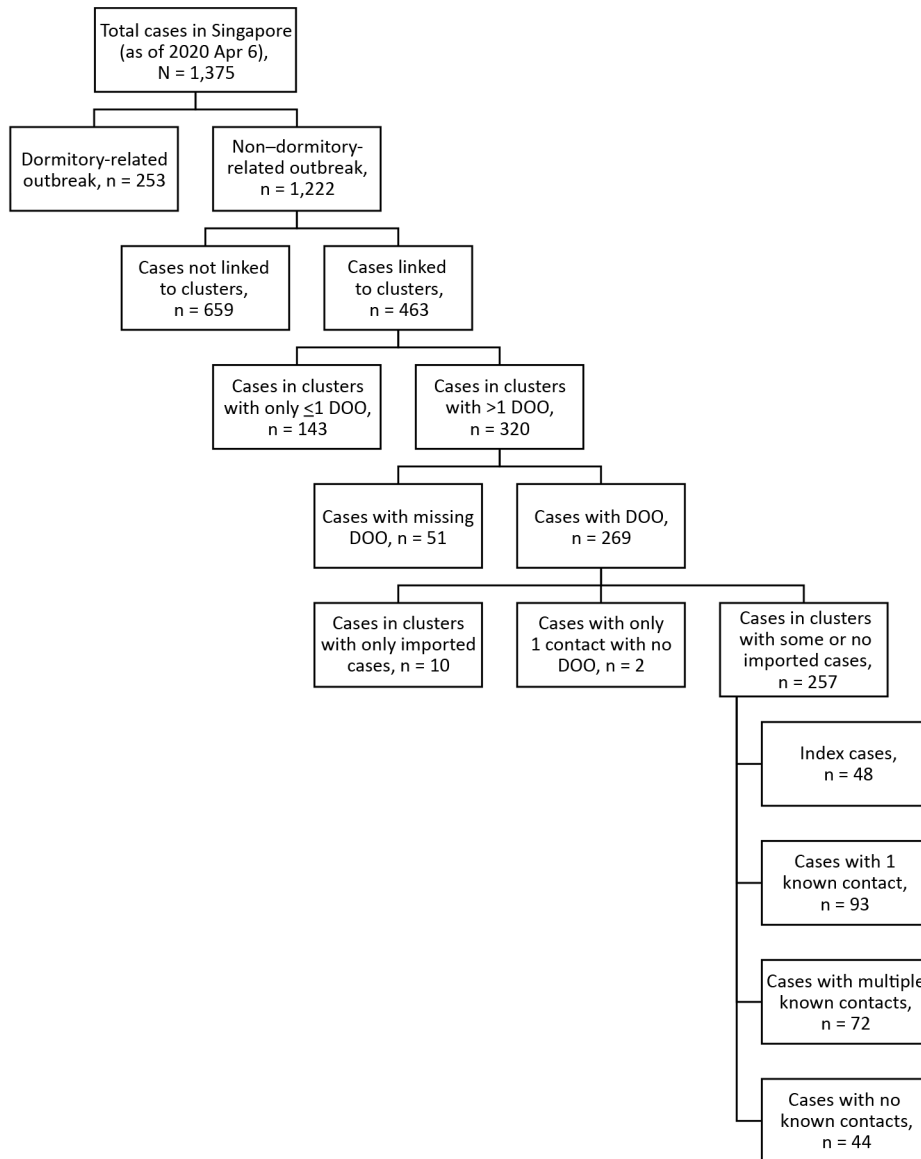


Figure 1. Inclusion and exclusion of coronavirus disease case-patients for analysis, Singapore, January–April 2020. DOO, date of symptom onset.

Assuming the same incubation period with mean (\pm SD) of 5.2 (\pm 2.8) days, we replicated the Bayesian Markov chain Monte Carlo procedure detailed in Ganyani et al. (5) to estimate the mean (SD) of the T_g (Appendix). With the estimated parameters, we constructed the distribution of R_0 and subsequently p by simulating infections and computing the proportion of presymptomatic transmissions. We conducted subgroup analyses for case-patients with multiple, family, or no contacts, and for clusters with no missing DOO. We conducted sensitivity analyses estimating the distribution of R_0 by using resampled values from a 95% CI of the epidemic growth rate and group-specific rates. For each distribution, we reported the median and 95% CrI. All analyses were conducted by using RStudio 1.2.5033 (<https://rstudio.com>).

A total of 1,375 confirmed cases had been reported as of April 6, 2020, and we applied our exclusion criteria to obtain a final sample size of 257 cases (Figure 1). We have summarized sample characteristics (Table 1) and the spread of cases over time (Figure 2). Because 48 index case-patients had no known infector, a maximum of 209 infector-infectee pairs were constructed for analysis.

Analyzing the 209 pairs, we estimated the mean T_g to be 3.44 (95% CrI 2.79–4.11) days, with an SD of 2.39 (95% CrI 1.27–3.45) days (Table 2). This estimate corresponded to an R_0 of 1.09 (95% CrI 1.08–1.11) and p of 0.72 (95% CrI 0.64–0.80). We estimated the serial interval distribution (Appendix Table 1) and convergence plots for all analyses (Appendix Figure 3).

Table 1. Characteristics of coronavirus disease case-patients in study estimating transmission parameters for coronavirus disease clusters by using symptom onset data, Singapore, January–April 2020

Characteristic	No. (%) [*]
Age, y, median (25th–75th percentile)	47 (30–59)
Sex	
M	121 (47.1)
F	136 (52.9)
Imported	24 (9.3)
Cluster size, N = 51	
2 cases	28 (54.9)
3 cases	12 (23.5)
>4 cases and above	11 (21.6)

^{*}Values are no. (%) unless otherwise indicated.

Examining the 93 pairs with only 1 known contact, the estimates for mean T_g , SD T_g and R_0 increased, whereas p decreased (Table 2). The 116 pairs that required identification of potential infectors had a shorter mean T_g and a higher p in comparison (Table 2). Subgroup analyses are summarized in Appendix Table 2. However, the chains for pairs with family or no known contact exhibited poor convergence, and estimates were not reported. Sensitivity analyses using resampled growth rates and group-specific rates did not yield estimates differing from those of the main analyses (Appendix Table 3).

Conclusions

The mean generation interval of the COVID-19 outbreak in Singapore was estimated at 3.44 days, suggesting that an infected person would be expected to pass on an infection to another person in 3 days,

within the range of 3.3–5.0 days reported by other studies (4,5; Li et al., unpub. data). Pairs with only 1 known contact yielded a larger estimate of 3.93 days, whereas pairs for whom infectors were identified had a shorter mean generation interval of 3.03 days. These results suggest that we might best report the upper bound of estimates, accounting for the presence of unclear transmission links within the clusters.

The R_0 estimated was slightly >1 , higher than other estimates reported as of March 31, 2020 (8). We observed a high p , potentially a result of prompt isolation of symptomatic case-patients (M. Casey et al., unpub. data, <https://doi.org/10.1101/2020.05.08.20094870>). This higher proportion might also be attributable to our allowance of infector DOOs to be up to 3 days after their infectees' DOOs, establishing the plausibility of presymptomatic transmission. We acknowledge that this cutoff would have an influence on our eventual estimates. Although negative serial intervals >3 days have occurred in other studies (5; Z. Du et al., unpub. data, <https://doi.org/10.1101/2020.02.19.20025452>), we chose a conservative cutoff of 3 days consistent with He et al. (6), where 9% of transmissions would occur before 3 days before DOO.

Nonetheless, the high prevalence of presymptomatic transmission in the community requires public health strategies to be responsive to this characteristic to remain effective. Universal wearing of masks in the community might reduce the likelihood of transmission through saliva and respiratory droplets (9). In place of testing when symptoms are observed,

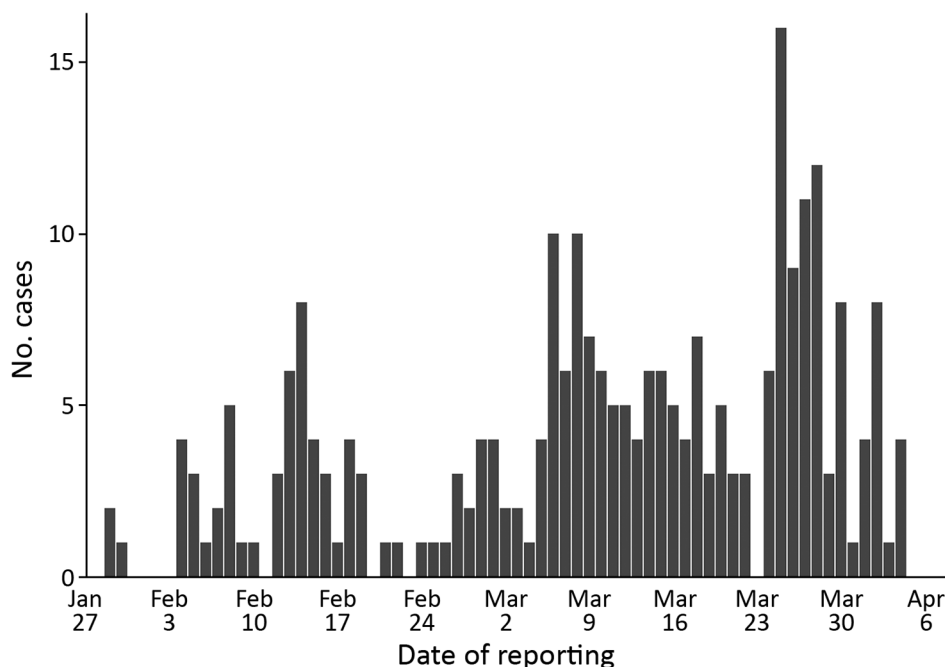


Figure 2. Epidemic curve of coronavirus disease clusters, Singapore, January–April 2020.

Table 2. Estimates of transmission parameters of coronavirus disease clusters, Singapore, January–April 2020 *

Infected type	Median (95% credible interval)			
	Mean T_g	SD T_g	R_0	p
All case-patients, N = 209	3.44 (2.79–4.11)	2.39 (1.27–3.45)	1.09 (1.08–1.11)	0.72 (0.64–0.80)
Case-patients with only 1 known contact, n = 93	3.93 (3.00–4.93)	2.63 (1.10–4.31)	1.11 (1.08–1.14)	0.65 (0.54–0.76)
Case-patients with only multiple or no known contacts, n = 116	3.03 (2.13–3.97)	2.45 (0.86–4.21)	1.08 (1.06–1.11)	0.76 (0.65–0.86)

* p , presymptomatic proportion; R_0 , basic reproduction number; T_g , generation time.

universal testing of persons living in or working with confined populations should be prioritized to mitigate the risk for transmission of the infection into these populations (10). Contact tracing should be modified to include the period before symptom onset (6,7) and should adopt a digital approach to be more comprehensive and less labor intensive (4).

Our study generated estimates that accounted for the uncertainty arising from multiple potential infectors and a small sample size, which contributes to the scarce information about disease characteristics. Because we dropped cases without a reported DOO, and DOO data and contact information were self-reported, our estimates might be subject to selection, self-report, and recall biases. Our estimation approach assumed equal probability of infecting among potential infectors, although a higher likelihood of transmission among household contacts has been suggested (11). We also did not account for the potential formation of cyclical infector networks, although their effects on the estimates have been demonstrated to be limited (12). Nevertheless, our estimates contribute to knowledge about the transmission dynamics of COVID-19 and have implications for control measures.

Acknowledgments

We thank the Singapore Ministry of Health for their tireless efforts in outbreak control and publication of the data for this manuscript.

N.H. acknowledges funding provided by the EpiPose project from the European Union's SC1-PHE-CORONA-VIRUS-2020 programme (project no. 101003688) and by the European's Horizon 2020 research and innovation programme (grant agreement no. 682540–TransMID).

About the Author

Ms. Ng is a research analyst with the National Healthcare Group in Singapore. Her research interests include the application of biostatistics in epidemiology and health services research.

References

1. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A systematic review of COVID-19 epidemiology based on current evidence. *J Clin Med.* 2020;9:967. <https://doi.org/10.3390/jcm9040967>
2. Zhao S, Cao P, Chong MKC, Gao D, Lou Y, Ran J, et al. The time-varying serial interval of the coronavirus disease (COVID-19) and its gender-specific difference: a data-driven analysis using public surveillance data in Hong Kong and Shenzhen, China from January 10 to February 15, 2020. *Infect Control Hosp Epidemiol.* 2020;41:750–1.
3. 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;382:1199–207. <https://doi.org/10.1056/NEJMoa2001316>
4. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science.* 2020 Mar 31 [Epub ahead of print]. <https://doi.org/10.1126/science.abb6936>
5. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill.* 2020;25:2000257. <https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257>
6. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Author correction: temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* 2020;26:1491–3. <https://doi.org/10.1038/s41591-020-1016-z>
7. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic transmission of SARS-CoV-2 – Singapore, January 23–March 16, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69:411–5. <https://doi.org/10.15585/mmwr.mm6914e1>
8. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, et al. Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. *BMC Med.* 2020;18:166. <https://doi.org/10.1186/s12916-020-01615-9>
9. Cheng VCC, Wong SC, Chuang VWM, So SYC, Chen JHK, Sridhar S, et al. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *J Infect.* 2020;81:107–14. <https://doi.org/10.1016/j.jinf.2020.04.024>
10. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' Heel of current strategies to control Covid-19. *N Engl J Med.* 2020;382:2158–60. <https://doi.org/10.1056/NEJMe2009758>
11. 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;20:911–9. [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5)
12. Kremer C, Ganyani T, Chen D, Torneri A, Faes C, Wallinga J, et al. Authors' response: estimating the generation interval for COVID-19 based on symptom onset data. *Euro Surveill.* 2020;25: 2001269. <https://doi.org/10.2807/1560-7917.ES.2020.25.29.2001269>

Address for correspondence: Sheryl Hui-Xian Ng, National Healthcare Group, Health Services and Outcomes Research, 3 Fusionopolis Link, 03-08, Singapore 138543; email: sheryl_hx_ng@nhg.com.sg

Estimating Transmission Parameters for COVID-19 Clusters by Using Symptom Onset Data, Singapore, January–April 2020

Appendix

Identifying index cases within each cluster

Within each cluster, an index case was either

(i) a primary case determined by epidemiologic investigations by the Ministry of Health, Singapore,

(ii) an imported case,

(iii) a case with the earlier date of symptom onset (DOO) in a cluster with only one other case,

(iv) a case with the earliest DOO in the cluster, and no subsequent cases with a DOO within 3 days after it.

The index cases would not have any possible infectors by definition. Clusters with no cases satisfying the criteria (i) to (iv) would not have a defined index case.

Heuristic to identify potential infectors of cases

We identified potential infectors of each case based on available information of the cases' known contacts, published case links (*1,2*), and a heuristic to sensibly include a pool of candidates who could have transmitted the infection to the cases.

We defined a potential infector as an infector with a DOO within a period spanning 14 days before and 3 days after the DOO of the case. These thresholds were chosen to ensure that an infector-infectee pair would generate a plausible serial interval. We described two scenarios in Appendix Figure 1 to illustrate the implications of the thresholds. The first scenario would occur when the DOO of the infectee is after that of the infector (Appendix

Figure 1). As the maximum incubation period of COVID-19 has been suggested to be 14 days (3), we have considered the maximum plausible serial interval to be 14 days. The second scenario would occur when the DOO of the infectee is before that of the infector (Appendix Figure 1). As analyses by He *et al* found only 9% of transmission to occur 3 days before an infector's DOO and recommended for contact tracing to include this window (4), the minimum serial interval we would consider would be -3 days.

We then constructed infector-infectee pairs of cases by assigning an infector to each case. Within clusters that have only two cases, the index case was assigned as the infector for the other case in the cluster. Within larger clusters, cases that have only one known contact were assigned this contact as their infector (Appendix Figure 2).

For cases with multiple contacts, a potential infector was randomly selected from the pool of contacts using an independence sampler in the Markov Chain Monte Carlo (MCMC) algorithm. For cases with no known contacts, a potential infector was similarly selected from other cases within the same cluster. If a case had no potential infectors with a DOO within the period for plausible serial intervals, the case's known contacts were nonetheless assigned as potential infectors.

Estimating the distributions of transmission parameters

Assuming the same incubation period with mean of 5.2 days and standard deviation of 2.8 days (5), we replicated the Bayesian MCMC procedure as detailed in Ganyani *et al* (6), to estimate the mean and standard deviation of the generation interval (T_g) distribution. Briefly, the infectors assigned and parameters of the T_g distribution were updated in a two-step process. The unknown infector-infectee links were updated using an independence sampler. Uniform priors were assigned to the parameters of the generation interval distribution and updated using a random-walk Metropolis-Hastings algorithm with a uniform proposal distribution (7). The posterior distribution was modeled using 3,000,000 iterations of which the first 500,000 were discarded as burn-in, thinning applied every 200 iterations, resulting in 12,500 iterations of accepted parameter estimates.

With the 12,500 instances of estimates and the spread of cases, we computed the corresponding effective R_0 estimates using the relationship $R_0 = e^{r\mu - (\frac{1}{2}r^2\sigma^2)}$, where r is the growth rate of the epidemic and μ, σ are the mean and standard deviation of the generation interval distribution (8). Using the same parameter estimates and the incubation period

parameters assumed above, we simulated 1000 transmissions per iteration and computed the pre-symptomatic proportion as the proportion of transmissions that occurred before the end of the infector's incubation period. For each distribution, the median and 95% credible interval were reported.

References

1. Abu Baker J, Yong M. Chinese New Year gathering identified as missing link between COVID-19 church clusters. CNA [cited 2020 Jun 12].
<https://www.channelnewsasia.com/news/singapore/covid19-coronavirus-missing-link-found-church-clusters-12469236>
2. Kurohi R, Chee K, Goh T, Khalik S. 9 in 10 coronavirus patients in Singapore are linked to worker dormitories. The Straits Times [cited 2020 Jun 12].
<https://www.straitstimes.com/singapore/health/novel-coronavirus-cases-in-singapore>
3. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med.* 2020;172:577–82. [PubMed https://doi.org/10.7326/M20-0504](https://doi.org/10.7326/M20-0504)
4. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Author correction: temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* 2020;26:1491–3. [PubMed https://doi.org/10.1038/s41591-020-1016-z](https://doi.org/10.1038/s41591-020-1016-z)
5. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei Province, China: a descriptive and modelling study. *Lancet Infect Dis.* 2020;20:793–802. [PubMed https://doi.org/10.1016/S1473-3099\(20\)30230-9](https://doi.org/10.1016/S1473-3099(20)30230-9)
6. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill.* 2020;25:2000257. [PubMed https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257](https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257)
7. Robert C, Casella G. Monte Carlo statistical methods. New York: Springer Science and Business Media; 2013.
8. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci.* 2007;274:599–604. [PubMed https://doi.org/10.1098/rspb.2006.3754](https://doi.org/10.1098/rspb.2006.3754)

Appendix Table 1. Estimated serial interval (SI) distributions*

Type of infector–infected pairs	Median (95% credible interval)	
	Simulated SI	SD SI
All cases (N = 209)	3.28 (−5.41, 13.11)	4.62 (4.16, 5.25)
Cases with only 1 known contact (N = 93)	3.74 (−5.10, 13.91)	4.75 (4.11, 5.86)
Cases with multiple or no contacts (N = 116)	2.83 (−5.83, 12.83)	4.66 (4.05, 5.78)

*SI, Serial interval; DOO, Date of symptom onset.

Appendix Table 2. Subgroup analyses*

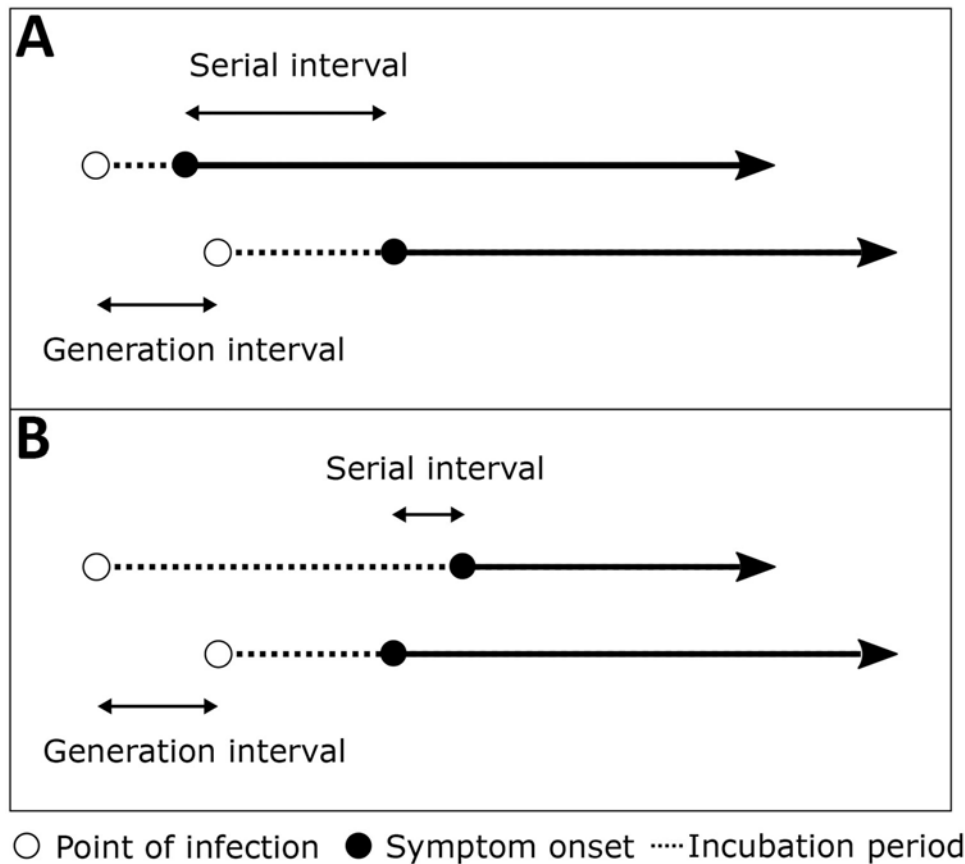
Infected type	Median (95% credible interval)			
	Mean T_g	SD T_g	R_0	p
Cases with multiple known contacts (N = 72)	3.72 (2.57, 5.04)	3.22 (1.49, 5.43)	1.10 (1.07, 1.14)	0.69 (0.56, 0.81)
Cases in clusters with no missing DOOs (N = 103)	4.05 (3.13, 5.06)	2.90 (1.36, 4.42)	0.98 (0.97, 0.98)	0.64 (0.53, 0.74)

* T_g , generation time; SD, standard deviation; R_0 , basic production number; p: pre-symptomatic proportion; DOO: Date of symptom onset.

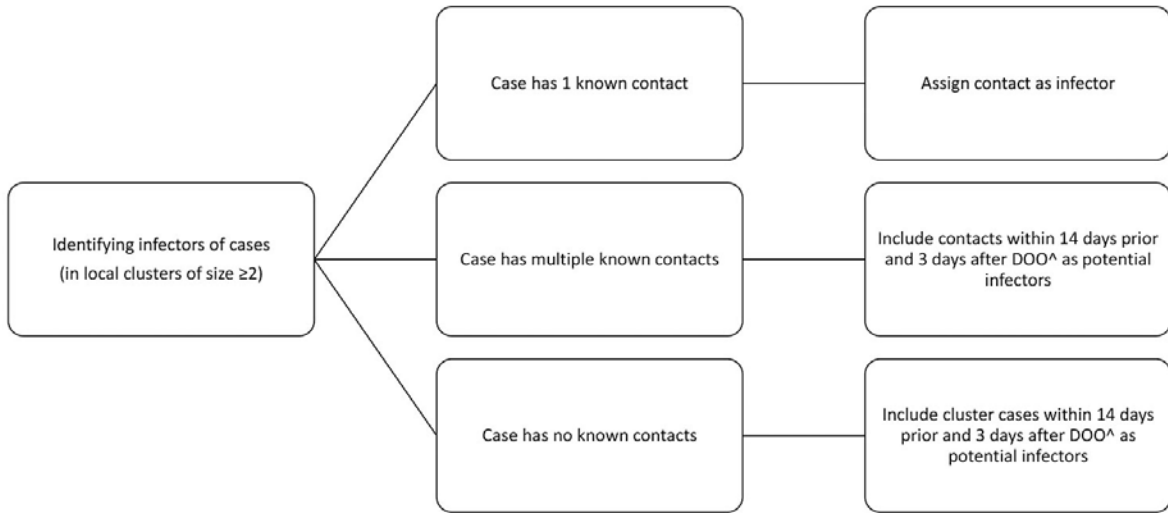
Appendix Table 3. Sensitivity analyses of reproduction number R_0 distribution

Infected type	No. time points	Max daily incident cases	Growth rate of log (cases)	Median (95% credible interval)	
				Original R_0	Revised R_0
Main analyses					
All cases (N = 209)	37	13	0.027	1.09 (1.08, 1.11)	1.90 (1.03 to 1.17)*
Cases with only 1 known contact (N = 93)	32	7	0.017	1.11 (1.08, 1.14)	1.07 (1.05, 1.09)
Cases with only multiple or no known contact (N = 116)	35	13	0.021	1.08 (1.06, 1.11)	1.06 (1.04, 1.08)
Subgroup analyses					
Cases with multiple known contacts (N = 72)	8	5	0.104	1.10 (1.07, 1.14)	1.39 (1.22, 1.57)
Cases in clusters with no missing DOOs (N = 103)	34	7	−0.003	0.98 (0.97, 0.98)	0.99 (0.99, 0.99)

*Estimates computed using resampled growth rates.

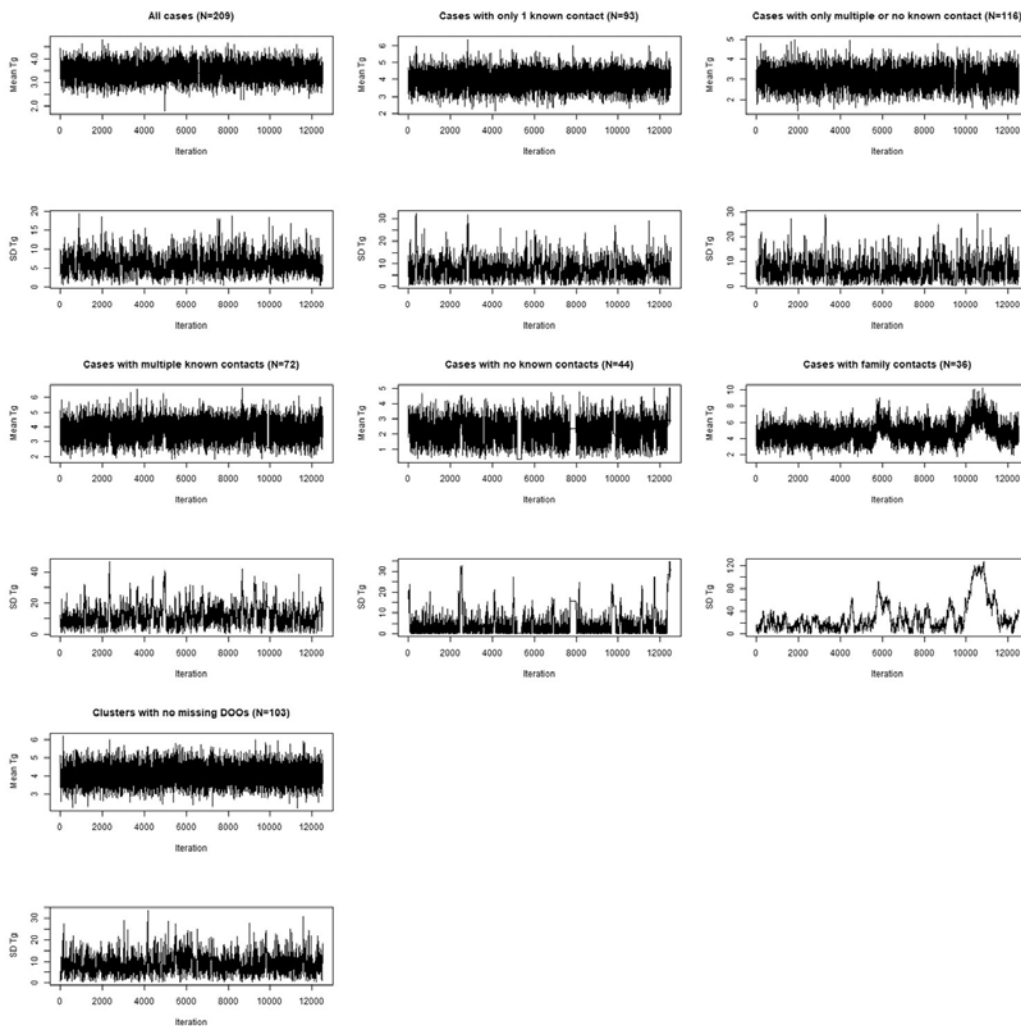


Appendix Figure 1. Time intervals within a transmission chain.



^DOO: Date of symptom onset

Appendix Figure 2. Heuristic for identifying potential infectors of a case.



Appendix Figure 3. Convergence plots.