

## Supplementary information

### 1. Previous estimates of the generation interval of COVID-19

Some early estimates of the generation interval of COVID-19 were published during the first half of 2020.<sup>1-3</sup> However, they were not estimated directly from data on exposure dates but derived from estimates of the serial interval and incubation period, as indicated in Supplementary Table 1 below. We found only one other study (Li et al.) that estimated the generation interval directly from reported dates of exposure, although this study assumed infection occurred at the middle of a given exposure period,<sup>4</sup> which may have been intended to address the censored observational data. None of the studies specifically aimed to estimate the generation interval for pairs with asymptomatic infectors.

The studies approached the problem of pair ascertainment (determining directionality of transmission) and estimation of the timing of exposure and contact in different ways.

- Ferretti et al.<sup>1</sup> selected transmission pairs based on high confidence of direct transmission inferred from publicly available sources reported.
- Ganyani et al.<sup>2</sup> used datasets of cases reported in Singapore and Tianjin. For cases linked to clusters, they imputed links between cases to determine transmission pairs and assign directionality.
- Tindale et al.<sup>3</sup> used the same datasets as Ganyani et al. Linkages, when not explicitly available from the data, were established using the methods described in te Beest et al.<sup>5</sup>
- Bushman et al.<sup>6</sup> combined transmission pair data from four published studies.<sup>7-10</sup>
- Li et al.<sup>4</sup> limited their analysis to transmission pairs where the infector had travel history to Hubei Province, China (where the original epicenter of COVID-19, Wuhan, is located). They did not apply doubly-interval censoring, as has been done elsewhere,<sup>11-13</sup> but assumed infection occurred at the exact middle of a given exposure period.
- Hart et al. (worldwide data)<sup>14</sup> combined data from five published studies,<sup>1,7,15-17</sup> and considered a “mechanistic” model where infectors who developed symptoms progressed through different stages of infection.
- Hart et al. (UK data)<sup>18</sup> estimated the generation interval from cases reported in households in the United Kingdom. Their methods considered asymptomatic infectors, and they considered a model where infectiousness and symptom onset were independent, as well as their previously published mechanistic model (see above).
- Zhao et al.<sup>19</sup> used data collected from a previous study<sup>19</sup> that included pairs from China, Japan, and Singapore. The authors used the serial interval and incubation period of the infectee to estimate the generation interval.
- Lau et al.<sup>20</sup> reanalyze data from a previously published study.<sup>21</sup> Their methods do not use the coarse exposure periods to estimate generation interval directly, but assume that symptom onset of the infector is independent of infectiousness conditioning on infection time of the infector.

Among these studies, only Bushman et al. and Hart et al. considered dependence between the generation time and incubation period of the infector. The mechanistic model of Hart et al. considers that symptoms and transmission are not independent, but do not directly quantify possible interdependence. Rather, the model conditions infectiousness on the duration of the incubation period. Bushman et al., although they considered “incubation-dependent” models, found that the best fits for their data were “incubation-independent” models, suggesting low- or no correlation between the generation interval and incubation period. Their method for defining “incubation-dependent” models was to vary the rate parameter of the gamma distribution by dividing it by the length of the incubation period, and it is possible this method—as it did not include correlation as a parameter—could not capture existing correlation. As well, it is possible that the method they used to calculate the generation interval from the serial interval—which assumes independence from the incubation period<sup>1,2,6,22</sup>—does not accurately reflect .

1 **Supplementary Table 1.** Estimates of the generation interval of COVID-19 from previously published studies

Estimation method	Mean (95% CI)	SD (95% CI)	Distribution	Geographic scope	Time period	Pairs	Reference
SI and IP*	7.50 (6.81, 8.20)	3.95 (3.32, 4.74)	Gamma	China, pre-NPI	Jan 2020	873 <sup>†</sup>	Bushman <sup>6</sup>
SI and IP*	3.90 (3.59, 4.24)	3.15 (2.78, 3.53)	Gamma	China, post-NPI	Jan–Feb 2020	873 <sup>†</sup>	Bushman
SI and IP	4.83 (4.31, 5.40)	1.73 (0.98, 2.55)	Gamma	Worldwide	Jan–Mar 2020	-	Challen <sup>23</sup>
SI and IP*	5.04 (4.19, 6.31)	1.93 (1.52, 2.47)	Weibull	Worldwide	Dec 2019–Feb 2020	40	Ferretti <sup>1</sup>
SI and IP‡	5.20 (3.78, 6.78)	1.72 (0.91, 3.93)	Gamma	Singapore	Jan–Feb 2020	54	Ganyani <sup>2</sup>
SI and IP‡	3.95 (3.01, 4.91)	1.51 (0.74, 2.97)	Gamma	Tianjin, China	Jan–Feb 2020	45	Ganyani
Mechanistic*	5.57 (5.08, 6.09)	2.32 (1.83, 2.91)	Gamma	Worldwide	Dec 2019–Mar 2020	191	Hart <sup>14</sup>
Independent§	4.2 (3.3, 5.3)	4.9 (3.0, 8.3)	Lognormal	United Kingdom	Mar–Nov 2020	¶	Hart <sup>18</sup>
Mechanistic§	6.0 (5.2, 7.0)	4.9 (4.0, 6.3)	Gamma	United Kingdom	Mar–Nov 2020	¶	Hart
Exposure and onset times	5.7 (4.8, 6.5)	1.7 (0.7, 2.5)	Gamma	China	Jan 2020	81 <sup>**</sup>	Lau <sup>20</sup>
Exposure times	4.81 (4.13, 5.58)	2.52 (1.93, 3.32)	Gamma	China	Jan–Feb 2020	67	Li <sup>4</sup>
IP intermediates	3.71 (2.36, 4.91)	-	Gamma	Singapore	Jan–Feb 2020	56 <sup>**</sup>	Tindale <sup>3</sup>
IP intermediates	2.82 (1.82, 3.52)	-	Gamma	Tianjin, China	Jan–Feb 2020	72 <sup>**</sup>	Tindale
SI and infectee IP	6.7 (5.4, 7.6)	1.8 (0.3, 3.8)	Gamma	China, Japan, Singapore	Dec 2019–Apr 2020	254	Zhao <sup>19</sup>

2 CI: confidence/credible interval; COVID-19: coronavirus disease 2019; IP: incubation period; NPI: non-pharmaceutical intervention; SD: standard deviation; SI: serial interval. \*IP  
3 distribution based on Lauer et al.<sup>24</sup> †Total pairs for both pre- and post-NPI. ‡IP distribution based on Zhang et al.<sup>10</sup> §IP distribution based on McAloon et al.<sup>25</sup> ¶172 households with  
4 603 cases. \*\*Manually counted from figure.

5

6

7 Supplementary Table 2 lists previously published studies where correlation between transmission  
8 intervals (generation or serial interval) and the incubation period were assessed. Dependence between the  
9 two parameters has rarely been estimated.

10 **Supplementary Table 2.** Estimates of correlation between transmission intervals and incubation periods  
11 from other studies

Disease	Transmission interval	Kendall's tau (CrI or p-value)	Geographic scope	Timeframe	Method	Reference
Measles	GI	0.27 (-0.34, 0.76)*	Rhode Island	1917–1923	Copula, gamma	Klinkenberg <sup>26</sup>
Measles	GI	0.63 (0.14, 0.88)*	Rhode Island	1929–1934	Copula, gamma	Klinkenberg
Measles	GI	0.72 (0.52, 0.84)*	Rhode Island	1917–1923	Copula, lognormal	Klinkenberg
Measles	GI	0.84 (0.67, 0.95)*	Rhode Island	1929–1934	Copula, lognormal	Klinkenberg
COVID-19	SI†	0.13 (p=0.2)	Singapore	2020	Case pairs	Tindale <sup>3</sup>
COVID-19	SI†	0.19* (-)	Tianjin, China	2020	Case pairs	Tindale

12 CrI: credible interval; COVID-19: coronavirus disease 2019; GI: generation interval; SI: serial interval. \*Original results  
13 were presented as correlation coefficient  $\rho$ , which is defined in terms of Kendall's tau as  $\tau_K = \frac{2}{\pi} \arcsin(\rho)$ . †Estimates  
14 of the serial interval included only positive values.

15

16 Klinkenberg and Nishiura<sup>26</sup> estimated the correlation between the generation interval and incubation period  
17 of measles in Rhode Island prior to development of the measles vaccine. They found very different results  
18 depending on whether they were using bigamma or bilognormal marginals. The better fit to the bilognormal  
19 marginals may reflect the better fit to the lognormal distribution commonly seen in incubation period data  
20 for respiratory diseases,<sup>27</sup> including COVID-19.<sup>12,13,24</sup> Tindale et al. assessed covariance between the serial  
21 interval and incubation period of COVID-19 using empirical data from their selected transmission pairs.  
22 However, the pairs ignored negative serial intervals.

23

## 24 2. Case definition and determination of transmission pairs

25 Nearly all COVID-19 cases in Japan had a positive viral test for SARS-CoV-2, with a few exceptions made  
26 for cases with positive antibody tests based on clinical judgement. These viral tests were either nucleic acid  
27 amplification tests (NAATs) or antigen tests. Not all public health jurisdictions shared the type of test in their  
28 case reports, but typically the NAATs were either reverse transcription polymerase chain reaction (RT-PCR)  
29 or loop-mediated isothermal amplification (LAMP) tests.

30 Presymptomatic and asymptomatic transmission are possible for SARS-CoV-2 infections.<sup>3,7</sup> Thus,  
31 directionality of transmission for epidemiologically linked cases that is determined solely based on dates of  
32 onset among linked cases is likely to include some misclassification of infector and infectee status, and may  
33 ignore possible intermediate infectors. We collected information on COVID-19 cases reported in Japan and  
34 looked for epidemiological information (i.e. contact and travel history) that would provide insight into  
35 directionality of transmission between linked cases. Criteria for ascertainment of directionality of  
36 transmission generally fell into one of four categories:

- 37 1. Import: If a potential infector had onset during or following international travel to a country with  
38 COVID-19 cases they were considered an imported case.
- 39 2. Cluster: A cluster was defined as five or more cases. If a potential infector was linked directly or  
40 via a chain of infections to a cluster (common exposure) the link was classified as cluster-based.
- 41 3. Domestic travel: If onset of an infector occurred during or within 10 days after travel to another  
42 prefecture and there were no more obvious possible exposures these cases were labelled as  
43 having domestic travel as their possible exposure, and the dates of travel form the left- and  
44 right-hand bounds of their exposure.
- 45 4. Contact pattern: the contact pattern between cases, typically supplemented by some reported  
46 dates of contact or exposure, provided insight into directionality of infection for linked infector-  
47 infectee pairs.

48 We supply a variable “link\_basis” in the dataset of transmission pairs as a means of assessing whether  
49 there was any difference between estimates given the differing rationales for ascertaining directionality of  
50 transmission for pairs for each of the above categories.

51

### 52 3. Cleaning dates of exposure and contact

53 Reported dates on exposure, contact, and symptom onset were cleaned to obtain EL and ER (left- and right-  
54 hand exposure times for the infector), CL and CR (left- and right-hand times of contact between the infector  
55 and infectee), as well as S1 and S2 (symptom onset times of the infector and infectee). All times are reported  
56 as number of days relative to a “time zero” of 1 January 2020. An abridged data dictionary for calculated  
57 values is provided in

58 Supplementary Table 3.

59 For most case reports that included temporal information on potential exposures or contact between  
60 cases, exact dates or date ranges were explicitly stated. However, some of the data may include inferences  
61 made from statements related to parts of the month, with data entry performed as follows:

- 62 - Beginning of the month = until the seventh day of the month. More common may be just the  
63 first day of the month, or until the third day.
- 64 - End of the month = from the 25<sup>th</sup> day of the month. An informal survey found that  
65 approximately one-third of respondents felt that the phrase “end of the month” could include  
66 dates as early as the 25<sup>th</sup> (see <https://mainichi-kotoba.jp/enq-255>).
- 67 - Beginning, middle, and end of the month = days 1–10, 11–20, 21–last day of a given month,  
68 respectively.

69 In cases where left- or right-hand bounds for exposure and contact were missing from the data, we used  
70 other epidemiological information to substitute for these bounds when plausible to do so. The assumptions  
71 were as follows:

72 1) Data cleaning applied to ER/CR:

- 73 a) For some cases, recorded right-hand exposure (e.g., travel, contact with a case) for infector and/or  
74 infectee may exceed symptom onset (S1 or S2). As we are only interested in exposure and contact  
75 dates as they related to the possible infection time for infector and infectee, we set ER and CR equal  
76 to S1 and S2.
- 77 b) Similarly, for some infectors, their recorded right-hand exposure may exceed their time of contact  
78 with the infectee. In these scenarios we presume the infector is infected by the time of final contact  
79 the infectee, so we censor ER to CR.
- 80 c) Missing ER were sequentially assigned to the minimum of either:
  - 81 i) S1,
  - 82 ii) CR,
  - 83 iii) date of infector laboratory confirmation,
  - 84 iv) date of infector isolation,
  - 85 v) or 14 days\* after EL.
- 86 d) Missing CR were sequentially assigned to the minimum of either:
  - 87 i) S2,
  - 88 ii) date of infectee laboratory confirmation,
  - 89 iii) or 14 days\* after CL.

90 2) Data cleaning applied to EL/CL:

- 91 a) Missing EL were assigned to 14 days\* before ER.
- 92 b) In some cases, infectees may have been reported to have contact with the infector prior to the  
93 infector’s left-hand exposure (EL). For example, the infector and infectee may have met several

94 times. It is possible that the left-hand exposure for the infectee (CL) was assigned to a date that was  
95 earlier than EL. However, as pairs included in the dataset were selected for having a relatively high  
96 likelihood of the directionality of infection being true, in cases where  $CL < EL$ , we set  $CL = EL$ .

97 c) If the infector exposure type was labelled "International travel" and the infectee was missing CL,  
98 then CL was set to ER.

99 d) If contact type between infector and infectee is "household," link basis was "cluster" or "domestic  
100 contact", infectee was missing CL, and EL was not missing, CL was set to EL.

101 e) Missing CL were assigned to EL, as an infector cannot be infectious (and therefore CL cannot be a  
102 valid date of contact for transmission to occur from infector to infectee) unless the infector's own  
103 transmission has occurred.

104 \*14 days was selected as this  $\sim$ >95% of the incubation period.<sup>12,13,24</sup> Although the above assumptions made it  
105 possible to obtain EL and ER even in a left- or right-hand bound of exposure was not reported, we only  
106 included cases where a left- or right-hand bound of exposure was explicitly reported.

107

108

Variable	Description	Class	Values	Details
EL	Left-hand bound of infector exposure	Numeric	Days	
ER	Right-hand bound of infector exposure	Numeric	Days	
CL	Left-hand bound of infectee contact with infector	Numeric	Days	
CR	Right-hand bound of infectee contact with infector	Numeric	Days	
S1	Symptom onset day of infector in days	Numeric	Days	
S2	Symptom onset day of infectee in days	Numeric	Days	
link_basis	What the basis was for determining directionality of transmission between linked pairs	Factor	Cluster	Infector was linked to a cluster or part of a transmission chain linked back to a cluster
			Contact pattern	Timings of contact and onset between cases provide plausible evidence for directionality of transmission
			Domestic travel	Infector travelled domestically to a location with ongoing transmission and timing of travel is plausibly related to onset of disease
			International travel	Infector travelled internationally part of a transmission chain linked back to international travel, and said travel is believed to have been the source of infection
exposure_type	Transmission setting for infector exposure.	Factor	Household	A household member or family member (when household status was not specified)
			Social-contact based interaction	Venue for interaction is based on social interaction. Restaurants, nightlife, karaoke, sports events, live music, gyms, friends, relatives, acquaintances, etc., or type of contact is not specified.
contact_type	Transmission setting for contact between infector and infectee.		Core community interaction	Venue for exposure are schools, general workplaces, essential workplaces (care facilities, medical facilities, government services, etc.), or exposure is related to travel to another area and source of infection is unknown (community infection).

111 4. Bivariate joint distribution

112 We employed a Bayesian approach combining copulas (multivariate cumulative distribution functions)  
 113 with doubly-interval censoring to obtain estimates of the generation interval and incubation period, as well  
 114 as a measure of correlation those two parameters. For  $i \in \{1, \dots, N\}$  transmission pairs, we obtain the  
 115 following doubly-interval censored likelihoods for the generation interval and incubation period for data  $D$ :

$$L1(\theta_j; D) = \prod_i \int_{E_{L,i}}^{E_{R,i}} \int_{C_{L,i}}^{C_{R,i}} j(e)f(c - e)dcde, \quad (1)$$

$$L2(\theta_j; D) = \prod_i \int_{E_{L,i}}^{E_{R,i}} \int_{S_{L,i}}^{S_{R,i}} j(e)g(s - e)dsde.$$

116 Here,  $e$  is the time of infection of the infector and  $j(\cdot)$  is the probability distribution function (PDF) of  
 117 the time of infection of the infector following a uniform distribution across exposure time  $E_L$  to  $E_R$ .  
 118 Similarly,  $c$  is the time of transmission of the pathogen from infector and infectee occurring between  $C_L$   
 119 and  $C_R$ , with  $f(\cdot)$  representing the PDF of the generation interval. Finally,  $s$  is the time of symptom onset of  
 120 the infector occurring between  $S_L$  and  $S_R$ , with  $g(\cdot)$  representing the PDF of the incubation period.

121 Combining doubly-interval censoring with a copula function allowed us to obtain the bivariate joint  
 122 distribution of the generation interval and incubation period. Copulas provide a correlation structure to  
 123 sets of marginal distributions, allowing for the marginal distributions and dependence structure to be  
 124 modeled separately. In accordance with Sklar's theorem,<sup>28</sup> the joint cumulative distribution function (CDF)  
 125 can be decomposed into the copula and univariate marginal CDFs. As such, if  $H(x, y)$  is a joint bivariate CDF  
 126 with marginal CDFs  $F(x)$  and  $G(y)$ , there exists a copula  $C: [0,1]^2 \rightarrow [0,1]$  such that:

$$H(x, y) = C(F(x), G(y)) \quad (2)$$

127 for all  $(x, y) \in [-\infty, \infty]^2$ . From relations  $F(x) = u$  and  $G(y) = v$  we obtain  $x = F^{-1}(u)$  and  $y = G^{-1}(v)$ .  
 128 Substituting  $F^{-1}(u)$  and  $G^{-1}(v)$  into (2) we obtain the copula:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)), \quad (3)$$

129 for all  $(u, v) \in [0,1]^2$ . The marginal distribution functions are given by  $F(x) = \int_{-\infty}^x f(u)du$  and  $G(y) =$   
 130  $\int_{-\infty}^y g(v)dv$ . From the marginal distributions we have  $U = F(X)$  and  $V = G(Y)$  which are uniformly  
 131 distributed in  $[0,1]$ . The joint distribution of  $(U, V)$  is the copula  $C(u, v) = P(U \leq u, V \leq v)$ . The density of  
 132 the bivariate copula  $C$  is:

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \quad (4)$$

133 Setting  $x$  and  $F(x)$  to be the data and CDF of the generation interval, and  $y$  and  $G(y)$  to be the data  
 134 and CDF of the generation interval, the joint distribution described in equation (2) is therefore the joint  
 135 distribution of the incubation period and generation interval. Given a copula parameter  $\theta$ , the overall log-

136 likelihood is expressed as the sum of the log-likelihood of each marginal distribution plus the log-likelihood  
 137 of the copula:

$$\log L(\mu_X, \sigma_X, \mu_Y, \sigma_Y, \theta|x, y) = \log L(\mu_X, \sigma_X|x) + \log L(\mu_Y, \sigma_Y|y) + \log L(\theta|u, v). \quad (5)$$

138

### 139 *Copula selection*

140 The Gaussian copula was utilized by Klinkenberg and Nishiura to assess correlation between the  
 141 generation interval and incubation period for measles data,<sup>26</sup> and therefore was of interest for inclusion in  
 142 this study. However, the Gaussian copula does not account for tail dependence. To consider lower tail  
 143 dependence we included the Clayton copula, while to consider upper tail dependence we included the  
 144 Gumbel copula. The independence copula, which is the copula that results from a dependency structure in  
 145 which each individual variable is independent of each other, was also considered. Supplementary Table 4  
 146 introduces various properties of these four copulas.

147

### 148 *Gaussian copula*

149 The Gaussian copula is an elliptical copula. It is defined by

$$C(u, v) = \Phi_\theta(\Phi^{-1}(u), \Phi^{-1}(v)), \quad (6)$$

150 where  $\Phi$  is the standard normal distribution function and  $\theta \in (-1, 1)$  is the correlation between the  
 151 components. The density of the bivariate Gaussian copula is given by

$$c(u, v; \theta) = \frac{1}{\sqrt{1 - \theta^2}} \exp\left\{-\frac{\theta^2(s^2 + t^2) - 2\theta st}{2(1 - \theta^2)}\right\}, \quad (7)$$

152 where  $s = \Phi^{-1}(u)$  and  $t = \Phi^{-1}(v)$ . As such, the log-likelihood function for the Gaussian copula is

$$\log(L(\rho|u, v)) = -\frac{1}{2} \log(1 - \theta^2) - \frac{\theta^2(s^2 + t^2) - 2\theta st}{2(1 - \theta^2)}. \quad (8)$$

153 For Gaussian copula, Kendall's tau is defined as  $\frac{2}{\pi} \sin^{-1}(\theta)$ . Independence is reached when  $\theta = 0$ .

154

### 155 *Independence, Gumbel, and Clayton copulas*

156 The Gumbel and Clayton copulas are single-parameter Archimedean copulas.<sup>29</sup> As a family, Archimedean  
 157 copula are defined as

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \quad (9)$$

158 where  $\varphi$  is the generator function of the copula  $C$ .  $\varphi: [0, 1] \rightarrow [0, \infty]$  is a continuous strictly decreasing

159 convex function such that  $\varphi(1) = 0$  and  $\varphi^{[-1]}$  is the pseudo-inverse of  $\varphi$ , defined as  $\varphi^{[-1]}: [0, \infty] \rightarrow [0, 1]$

160 with

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \quad (10)$$

161 For the bivariate Gumbel copula, the generator is  $\varphi(t) = (-\log(t))^\theta$  and inverse generator  $\varphi(t)^{-\theta} =$   
 162  $\exp(-t^{1/\theta})$  for copula parameter  $\theta \in [1, \infty)$ , where  $t$  varies from 0 to 1 regardless of whether it is equal  
 163 to  $u$  or  $v$ . The bivariate Gumbel copula is given as

$$C(u, v) = \exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right\}, \quad \theta \in [1, \infty), \quad (11)$$

164 where  $\theta \rightarrow \infty$  indicates full dependence, while  $\theta = 1$  corresponds to independence. The Gumbel copula  
 165 has upper-tail dependence and is useful for datasets where the dependence between high values of the  
 166 univariate distributions is stronger than the dependence between their low values. The Gumbel copula  
 167 does not allow negative dependence. Its density is given by

$$c(u, v; \theta) = \frac{C(u, v)}{(uv)} (\log(u) \log(v))^{\theta-1} w^{\frac{2}{\theta}-2} \left(1 + (\theta-1)w^{\frac{1}{\theta}-2}\right), \quad (12)$$

168 where  $w = (-\log(u))^\theta + (-\log(v))^\theta$ . Consequently, the log-likelihood function<sup>30</sup> is defined by

$$\begin{aligned} \log(L(\theta|u, v)) &= w^{\frac{1}{\theta}} - \log(uv) + (\theta-1)\log(\log(u)\log(v)) \\ &+ \log\left(w^{\frac{2}{\theta}-2}\right) \log\left[(\theta-1)w^{\frac{1}{\theta}-2}\right]. \end{aligned} \quad (13)$$

169 For the bivariate Clayton copula, the generator is  $\varphi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$  and inverse generator  $\varphi(s)^{-\theta} =$   
 170  $\max\{(1 + \theta s)^{-1/\theta}, 0\}$  for  $\theta \in [-1, \infty) \setminus \{0\}$ . The bivariate Clayton copula is given as

$$C(u, v) = \max\left(\left(u^{-\theta} + v^{-\theta} - 1\right)^{\frac{1}{\theta}}, 0\right), \quad (14)$$

171 where full dependence is reached as  $\theta \rightarrow \infty$ , while independence is reached as  $\theta \rightarrow 0$ .

172 The density is given by

$$c(u, v; \theta) = \frac{(\theta+1)(uv)^\theta}{(u^\theta v^\theta (uv)^\theta)^{\frac{1}{\theta}+2}} \quad (15)$$

173 The log-likelihood function for the Clayton copula is defined by

$$\begin{aligned} \log(L(\theta|u, v)) &= \log(1) + \theta - (\theta+1)(\log(u) + \log(v)) \\ &- \left(\frac{1+2\theta}{\theta}\right) \log(u^{-\theta} + v^{-\theta} + 1) \end{aligned} \quad (16)$$

174 The independence copula is a special case of several Archimedean copulas, as well as a special case of  
 175 the Gaussian copula with a correlation matrix equal to the identity matrix. It has no correlation (copula)  
 176

177 parameter and no tail dependence. For the bivariate independence copula, the generator is  $\varphi(t) =$   
 178  $\exp(-t)$ . We applied the independence copula as the special case of the Gumbel copula ( $\theta = 1$ ).  
 179

180 *Measuring dependence*

181 We use Kendall's tau to assess dependence between the generation interval and incubation period.  
 182 Rank correlations such as Kendall's tau only depend on the unique copula of the joint distribution and are  
 183 therefore invariant to monotone transformations of the marginals.<sup>29</sup> The relationship between Kendall's  
 184 tau and the various copula parameters are listed in Supplementary Table 4.  
 185

186 **Supplementary Table 4.** Bivariate copulas and their properties

Copula	Copula parameter	Independence	Kendall's tau	Range of tau	Lower tail dependence	Upper tail dependence
Gaussian	$\theta \in (-1,1)$	$\theta = 0$	$\frac{2}{\pi} \sin^{-1}(\theta)$	$[-1,1]$	0	0
Gumbel	$\theta \in [1, \infty)$	$\theta = 1$	$1 - \theta^{-1}$	$[0,1]$	0	$2 - \frac{1}{\theta}$
Clayton	$\theta \in [-1, \infty) \setminus \{0\}$	$\theta \rightarrow 0$	$\frac{\theta}{\theta + 2}$	$[0,1]$	$2^{-\frac{1}{\theta}}$	0
Independence	None	Always	0	0	0	0

187  
 188 Of the three copula that allow for dependence between the parameters (Gaussian, Gumbel, Clayton),  
 189 only the Gaussian copula considers negative correlation.  
 190

191 *Mixture model*

192 We used a Bayesian mixture model to determine the best-fit combination of various copulas and  
 193 marginal distributions. For all possible combinations of copula and marginal distributions we assign  
 194 individual likelihoods as the sum of component contributions and formulated the model in terms of latent  
 195 variables. We considered the four copulas described above, and the gamma, lognormal, and Weibull  
 196 distributions for the generation interval and incubation period for a total of  $M = 4 \times 3 \times 3 = 36$   
 197 combinations. These  $M$  combinations mix in proportion  $\lambda$ , where  $\lambda_m \geq 0$  and  $\sum_m^M \lambda_m = 1$ .

198 The outcome is drawn from one of these combinations, the identity of which is controlled by a  
 199 categorical mixing distribution  $z \sim \text{Categorical}(\lambda)$ . We fixed Kendall's tau and the means and SDs for the  
 200 generation interval and incubation period across all possible combinations  $M$ . We used informative priors  
 201 for the means of the generation interval and incubation period, obtained from previous publications.<sup>13,31</sup>  
 202  
 203  
 204

205 **Author contributions**

206 NML and NH conceived the study. Data were collected by NML. Data were analyzed by NML and ARA, and  
207 interpreted by NML, ARA, and HN. NML wrote the first draft of the manuscript, which was revised by NML,  
208 ARA, and HN. All authors approved the final version of the manuscript.

210 **Competing interests**

211 The authors declare no competing interests.

213 **Acknowledgements**

214 The authors thank the reporting jurisdictions in Japan for conducting thorough investigations and  
215 publishing information about cases and transmission settings of public health import, Atsuna Tokumoto for  
216 help with initial data collection on transmission links for two prefectures, and the Nishiura lab for additional  
217 support with data collection on case characteristics. NML received a Japanese Ministry of Education,  
218 Culture, Sports, Science, and Technology (MEXT) graduate scholarship. HN received funding from a Health  
219 and Labor Sciences Research Grant (20CA2024 and 20HA2007), the Japan Agency for Medical Research and  
220 Development (AMED; JP20fk0108140 and JP20fk0108535s0101), the Japan Society for the  
221 Promotion of Science (JSPS)KAKENHI (21H03198), and the Japan Science and Technology Agency  
222 (JST) SICORP (e-ASIA) program (JPMJSC20U3).

224 **References**

- 225 1. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic  
226 control with digital contact tracing. *Science (80- )*. 2020;368(6491):eabb6936.  
227 doi:10.1126/science.abb6936
- 228 2. Ganyani T, Kremer C, Chen D, et al. Estimating the generation interval for coronavirus disease  
229 (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*. 2020;25(17):1-8.  
230 doi:10.2807/1560-7917.ES.2020.25.17.2000257
- 231 3. Tindale LC, Stockdale JE, Coombe M, et al. Evidence for transmission of covid-19 prior to symptom  
232 onset. *Elife*. 2020;9:1-34. doi:10.7554/eLife.57149
- 233 4. Li M, Liu K, Song Y, Wang M, Wu J. Serial interval and generation interval for imported and local  
234 infectors, respectively, estimated using reported contact-tracing data of COVID-19 in China. *Front*  
235 *Public Heal*. 2021;8(577431). doi:10.3389/fpubh.2020.577431
- 236 5. Te Beest DE, Wallinga J, Donker T, Van Boven M. Estimating the generation interval of influenza a  
237 (H1N1) in a range of social settings. *Epidemiology*. 2013;24(2):244-250.  
238 doi:10.1097/EDE.0b013e31827f50e8
- 239 6. Bushman M, Worby C, Chang HH, Kraemer M, Hanage WP. Transmission of SARS-CoV-2 before and  
240 after symptom onset: Impact of nonpharmaceutical interventions in China. *Eur J Epidemiol*.

- 241 2021;36(4):429-439. doi:10.1101/2020.12.16.20214106
- 242 7. He X, Lau EH, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat*  
243 *Med.* 2020;26:672–675. doi:10.1101/2020.03.15.20036707
- 244 8. Xu X-K, Liu XF, Wu Y, et al. Reconstruction of transmission pairs for novel coronavirus disease 2019  
245 (COVID-19) in Mainland China: Estimation of superspreading events, serial interval, and hazard of  
246 infection. *Clin Infect Dis.* 2020;2019(Xx):1-5. doi:10.1093/cid/ciaa790
- 247 9. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly  
248 reported confirmed cases. *Emerg Infect Dis.* 2020;26(6):1341-1343. doi:10.3201/eid2606.200357
- 249 10. Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of  
250 coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet*  
251 *Infect Dis.* 2020;3099(20):1-10. doi:10.1016/s1473-3099(20)30230-9
- 252 11. Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating incubation period distributions with  
253 coarse data. *Stat Med.* 2009;28:2769-2784. doi:10.1002/sim.3659
- 254 12. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV)  
255 infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance.*  
256 2020;(January):1-6.
- 257 13. Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of  
258 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available  
259 case data. *J Clin Med.* 2020;9(538). doi:10.3390/jcm9020538
- 260 14. Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID-19 symptom  
261 onset highlights the importance of continued contact tracing. *Elife.* 2021;10:e65534.
- 262 15. Cheng, Hao-Yuan, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H. Contact tracing assessment of  
263 COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after  
264 symptom onset. *JAMA Intern Med.* 2020;180(9):1156-1163. doi:10.1001/jamainternmed.2020.2020
- 265 16. Zhang W. Estimating the presymptomatic transmission of COVID19 using incubation period and  
266 serial interval data. *medRxiv.* Published online 2020.
- 267 17. Xia W, Liao J, Li C, Li Y. Transmission of coronavirus disease 2019 during the incubation period may  
268 lead to a quarantine loophole. *medRxiv.* Published online 2020.
- 269 18. Hart WS, Abbott S, Endo A, et al. Inference of SARS-CoV-2 generation times using UK household  
270 data. 2021;(March 2020):1-32.
- 271 19. Zhao S, Tang B, Musa SS, et al. Estimating the generation interval and inferring the latent period of  
272 COVID-19 from the contact tracing data. *Epidemics.* 2021;36:100482.  
273 doi:10.1016/j.epidem.2021.100482
- 274 20. Lau YC, Tsang TK, Kennedy-shaffer L, et al. Joint estimation of the generation time and incubation  
275 period for coronavirus disease (COVID-19). *J Infect Dis.* Published online 2021.
- 276 21. Ren X, Li Y, Yang X, et al. Evidence for pre-symptomatic transmission of coronavirus disease 2019

- 277 (COVID-19) in China. *Influenza Other Respi Viruses*. 2021;15(1):19-26. doi:10.1111/irv.12787
- 278 22. Lehtinen S, Ashcroft P, Bonhoeffer S. On the relationship between serial interval, infectiousness  
279 profile and generation time. *J R Soc Interface*. 2021;18(174):20200756. doi:10.1098/rsif.2020.0756
- 280 23. Challen R, Brooks-pollock E, Tsaneva-atanasova K, Danon L. Meta-analysis of the SARS-CoV-2 serial  
281 interval and the impact of parameter uncertainty on the COVID-19 reproduction number. Published  
282 online 2020:1-27.
- 283 24. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from  
284 publicly reported confirmed cases: Estimation and application. *Ann Intern Med*. 2020;172(9):577-  
285 582. doi:10.7326/M20-0504
- 286 25. McAloon C, Collins Á, Hunt K, et al. Incubation period of COVID-19: A rapid systematic review and  
287 meta-analysis of observational research. *BMJ Open*. 2020;10(8):1-9. doi:10.1136/bmjopen-2020-  
288 039652
- 289 26. Klinkenberg D, Nishiura H. The correlation between infectivity and incubation period of measles,  
290 estimated from households with two cases. *J Theor Biol*. 2011;284(1):52–60.  
291 doi:10.1016/j.jtbi.2011.06.015
- 292 27. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DAT. Incubation periods of acute  
293 respiratory viral infections: a systematic review. *Lancet Infect Dis*. 2015;9(5):291-300.  
294 doi:10.1016/S1473-3099(09)70069-6.Incubation
- 295 28. Sklar A. Fonctions de Répartition à n Dimensions et Leurs Marges. *Publ L'Institut Stat L'Universit{é}*  
296 *Paris*. Published online 1959.
- 297 29. Czado C. *Analyzing Dependent Data with Vine Copulas: A Practical Guide with R*. Vol 222.; 2019.  
298 doi:10.1007/978-3-030-13785-4\_1
- 299 30. Correia Martins André LMB. Copula models for dependence: Comparing classical and Bayesian  
300 approaches. Published online 2019.
- 301 31. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections.  
302 *Int J Infect Dis*. Published online 2020:113332. doi:10.1016/j.ijid.2020.02.060
- 303 32. Geismar C, Fragaszy E, Nguyen V, et al. Serial interval of COVID-19 and the effect of Variant B.1.1.7:  
304 analyses from a prospective community cohort study. *medRxiv*. Published online 2021:1-11.
- 305 33. Ryu S, Lim J-S, Ali ST, Cowling BJ. Changes in the serial interval and transmission dynamics associated  
306 with the SARS- CoV-2 Delta variant in South Korea. *medRxiv*. Published online 2021:iii-iv.  
307 doi:10.1051/978-2-7598-2485-4-001
- 308 34. Pung R, Mak TM, Kucharski AJ, Lee VJ. Serial intervals in SARS-CoV-2 B.1.617.2 variant cases. *Lancet*.  
309 2021;6736(C):10-11. doi:10.1016/s0140-6736(21)01697-4
- 310