

1    **Title: Serial interval of novel coronavirus (COVID-19) infections**

2    **Running title:** Serial interval of COVID-19

3    **Authors:**

4    Hiroshi Nishiura<sup>1,2</sup>, Natalie M. Linton<sup>1</sup>, Andrei R. Akhmetzhanov<sup>1</sup>

5

6    **Affiliations:**

7    <sup>1</sup>Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi 7 Chome,

8    Kitaku, Sapporo, 060-8638, Japan; <sup>2</sup>CREST, Japan Science and Technology

9    Agency, Honcho 4-1-8, Kawaguchi, Saitama, 332-0012 Japan;

10

11   **Correspondence to:**

12    H. Nishiura, Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi

13    7 Chome, Kitaku, Sapporo, 060-8638, Japan

14    Tel: +81 11 706 5066    Fax: +81 11 706 7819

15    E-mail: nishiurah@med.hokudai.ac.jp

16   **Article type:**

17    Short communication

18

1    **Abstract**

2    **Objective:** To estimate the serial interval of novel coronavirus (COVID-19)

3    from information on 28 infector-infectee pairs.

4    **Methods:** We collected dates of illness onset for primary cases (infectors) and

5    secondary cases (infectees) from published research articles and case

6    investigation reports. We subjectively ranked the credibility of the data and

7    performed analyses on both the full dataset ( $n=28$ ) and a subset of pairs with

8    highest certainty in reporting ( $n=18$ ). In addition, we adjusting for right

9    truncation of the data as the epidemic is still in its growth phase.

10   **Results:** Accounting for right truncation and analyzing all pairs, we estimated

11   the median serial interval at 4.0 days (95% credible interval [CrI]: 3.1, 4.9).

12   Limiting our data to only the most certain pairs, the median serial interval was

13   estimated at 4.6 days (95% CrI: 3.5, 5.9).

14   **Conclusions:** The serial interval of COVID-19 is shorter than its median

15   incubation period. This suggests that a substantial proportion of secondary

16   transmission may occur prior to illness onset. The COVID-19 serial interval is

17   also shorter than the serial interval of severe acute respiratory syndrome (SARS),

18   indicating that calculations made using the SARS serial interval may introduce

19   bias.

1    **Keywords:** coronavirus; outbreak; illness onset; generation time; statistical  
2    model; epidemiology; viruses

### 3    **Highlights**

- 4    -    The serial interval of novel coronavirus (COVID-19) infections was  
5        estimated from a total of 28 infector-infectee pairs.
- 6    -    The median serial interval is shorter than the median incubation period,  
7        suggesting a substantial proportion of pre-symptomatic transmission.
- 8    -    A short serial interval makes it difficult to trace contacts due to the rapid  
9        turnover of case generations.

10

### 11   **Introduction**

12        The epidemic of novel coronavirus (COVID-19) infections that began in  
13    China in late 2019 has rapidly grown and cases have been reported worldwide.  
14    An empirical estimate of the serial interval—the time from illness onset in a  
15    primary case (infector) to illness onset in a secondary case (infectee)—is needed  
16    to understand the turnover of case generations and transmissibility of the disease  
17    [1]. Estimates of the serial interval can only be obtained by linking dates of onset  
18    for infector-infectee pairs, and these links are not easily established. A recently  
19    published epidemiological study used contact tracing data from cases reported in  
20    Hubei Province early in the epidemic to estimate the mean serial interval at 7.5  
21    days [2], which is consistent with the 8.4-day mean serial interval reported for  
22    severe acute respiratory syndrome (SARS) from Singaporean household contact  
23    data [3]. However, there were only six infector-infectee pairs in this dataset, and  
24    sampling bias may have been introduced to the variance and mean. To further

1 assess the serial interval of COVID-19 infections we compiled a dataset of 28  
2 publicly shared infector-infectee pairs and calculated the serial interval from  
3 these data.

#### 4 **Materials and Methods**

5 We scanned publicly available information published in research articles  
6 and quoted from official reports of outbreak investigations to obtain our dataset.  
7 The date of illness onset was defined as the date on which a symptom relevant to  
8 COVID-19 infection appeared and was determined by the reporting  
9 governmental body. We subjectively ranked the credibility of the ascertained  
10 pairs into “certain” and “probable,” where the former was used for pairs and  
11 dates of illness onset were clearly defined in an academic article and the latter  
12 was applied to pairs and dates of illness onset that were clearly defined but  
13 quoted from outbreak investigation reports. Estimates were obtained for certain  
14 and probable pairs combined ( $n=28$ ) as well as for the certain pairs alone ( $n=18$ ).

15 The interval censored data were handled in units of days. We employed  
16 a Bayesian approach with doubly interval censored likelihood to obtain estimates  
17 of the serial interval [4]:

$$L(\theta_g; \mathbf{D}) = \prod_i \int_{E_{L,i}}^{E_{R,i}} \int_{S_{L,i}}^{S_{R,i}} g(e) f(s - e) ds de, \quad (1)$$

18 where  $i$  represents the identity of each pair,  $E(R,L)$  is the interval for symptom  
19 onset of the infector and  $S(R,L)$  is the interval for symptom onset of the infectee.  
20 Here,  $g(\cdot)$  is the probability density function (p.d.f.) of exposure following a  
21 uniform distribution and  $f(\cdot)$  is the p.d.f. of the serial interval, assumed to be  
22 governed by three different distributions—lognormal, gamma, and Weibull. We

1 sampled the posterior distributions using CmdStan version 2.22.1  
2 (<http://github.com/aakhmetz/nCoVSerialInteval2020>).

3 As the epidemic will continue to grow beyond our data collection cutoff  
4 point of 12 February 2020, it is possible that the naïve likelihood (1)  
5 underestimates the serial interval as sampling during the early stage of the  
6 epidemic preferentially excludes infector-infectee pairs with longer serial  
7 intervals. We adjusted for this selection bias—called right truncation—in our  
8 model. The alternative p.d.f. that accounts for right truncation during the  
9 exponential growth phase of the epidemic is written as:

$$f'(s - e, e) = \frac{f(s - e)}{\int_0^{T-e} \frac{r \exp(-ru)}{1 - \exp(-ru)} F(T - e - u) du} \quad (2)$$

10 where  $r$  is the exponential growth rate estimated at 0.14 [5] and  $T$  is the latest  
11 time of observation (12 February 2020). The widely applicable information  
12 criterion (WAIC) was used to compare between distributions and the model with  
13 the minimal WAIC value was selected as the best-fit model for each set of  
14 estimates with and without right truncation.

## 15 **Results**

16 We were able to obtain data on 28 infector-infectee pairs (see  
17 Supplementary Table). Of these, 12 pairs were family clusters. Accounting for  
18 right truncation and analyzing all pairs, the model using the lognormal  
19 distribution was selected as the best-fit model (WAIC=224.0) The median serial  
20 interval was estimated at 4.0 days (95% credible interval [CrI]: 3.1, 4.9) while  
21 the mean and standard deviation (SD) of the serial interval were estimated at 4.7  
22 days (95% CrI: 3.7, 6.0) and 2.9 days (95% CrI: 1.9, 4.9), respectively. Without

1 truncation, the model using the lognormal distribution was also the best-fit model  
2 (WAIC=128.0) with the median serial interval was estimated at 3.9 days (95%  
3 CrI: 3.1, 4.8).

4 Limiting our dataset to only certain observations, the median serial  
5 interval of the best-fit Weibull distribution model was estimated at 4.6 days (95%  
6 CI: 3.5, 5.9) with a mean and SD of 4.8 days (95% CrI: 3.8, 6.1) and 2.3 days  
7 (95% CrI: 1.6, 3.5), respectively. Without truncation, the best-fit model used the  
8 lognormal distribution and estimated the median serial interval at 4.1 days (95%  
9 CrI: 3.2, 5.0). Figure 1 shows the best-fit distributions overlaid with a published  
10 distribution of the SARS serial interval [4].

## 11 **Discussion**

12 Our estimate of the median serial interval as 4.0 days indicates that  
13 COVID-19 infection leads to rapid cycles of transmission from one generation of  
14 cases to the next. The shorter serial interval compared to SARS implies that  
15 contact tracing methods must compete against the rapid replacement of case  
16 generations, and the number of contacts may soon exceed what available  
17 healthcare and public health workers are able to handle. The difference between  
18 these distributions suggests that using serial intervals estimates from SARS data  
19 will result in overestimation of the COVID-19 basic reproduction number.

20 More importantly, the estimated median serial interval is shorter than the  
21 preliminary estimates of the mean incubation period (approximately 5 days)  
22 [3,6]. As illustrated in Figure 2, when the serial interval is shorter than the  
23 incubation period, pre-symptomatic transmission is likely to have taken place  
24 and may even occur more frequently than symptomatic transmission. A

1 substantial proportion of secondary transmission occurring before illness onset  
2 indicates that many transmissions cannot be prevented solely through isolation of  
3 symptomatic cases, as by the time contacts are traced they may have already  
4 become infectious themselves and generated secondary cases [7].

5 Correct ascertainment of dates of illness onset is critical to the calculation of  
6 the serial interval. Considering the overall mild nature of the infection [8] it is  
7 possible that different reporting jurisdictions have different criteria for  
8 determining what qualifies as illness onset for COVID-2019 cases, which is a  
9 potential bias we are unable to account for. However, the present study addresses  
10 the issue of data quality of the reported pairs in two ways. First, our data include  
11 the updated information from a recent report of pre-symptomatic transmission in  
12 Germany [9] where it was later found that the primary case was already  
13 symptomatic while in contact with persons who later became infected  
14 (Supplementary Material in [9]). Second, classification of the credibility of the  
15 data and comparing analyses including and excluding less certain (but  
16 nonetheless highly probable) pairs allowed us to determine that our results using  
17 all pairs (and therefore a greater sample size) did not differ significantly from the  
18 results using only the most credible data.

19 In conclusion, we have estimated the median serial interval of COVID-19 at  
20 4.0 days, which is shorter than the disease's median incubation period indicating  
21 that rapid cycles of transmission and substantial pre-symptomatic transmissions  
22 are occurring. Thus, containment via case isolation alone is likely to be very  
23 challenging.

24

## 1 **Acknowledgments**

2 H.N. received funding support from Japan Agency for Medical Research and  
3 Development [grant number: JP18fk0108050] the Japan Society for the  
4 Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI  
5 in Japanese abbreviation) grant nos. 17H04701, 17H05808, 18H04895 and  
6 19H01074, and the Japan Science and Technology Agency (JST) Core Research  
7 for Evolutional Science and Technology (CREST) program [grant number:  
8 JPMJCR1413]. NML received a graduate study scholarship from the Ministry of  
9 Education, Culture, Sports, Science and Technology, Japan. The funders had no  
10 role in study design, data collection and analysis, decision to publish, or  
11 preparation of the manuscript.

## 12 **Conflict of interest**

13 The authors declare no conflicts of interest.

## 14 **References**

- 15 1. Fine PE. The interval between successive cases of an infectious disease. *Am*  
16 *J Epidemiol* 2003;158:1039-1047.
- 17 2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau  
18 EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J,  
19 Li M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo  
20 Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi  
21 G, Lam TTY, Wu JTK, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z.  
22 Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-  
23 Infected Pneumonia. *N Engl J Med*. 2020; in press. doi:  
24 10.1056/NEJMoa2001316



- 1 3. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna  
2 G, Chew SK, Tan CC, Samore MH, Fisman D, Murray M. Transmission  
3 dynamics and control of severe acute respiratory syndrome. *Science*.  
4 2003;300(5627):1966-70.
- 5 4. Reich NG, Lessler J, Cummings DA, Brookmeyer R. Estimating incubation  
6 period distributions with coarse data. *Stat Med* 2009;28:2769-2784.
- 7 5. Jung S, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B,  
8 Kobayashi T, Kinoshita R, Nishiura H. Real time estimation of the risk of  
9 death from novel coronavirus (2019-nCoV) infection: Inference using  
10 exported cases. medRxiv, [http://dx.doi.org/ 10.1101/2020.01.29.20019547v1](http://dx.doi.org/10.1101/2020.01.29.20019547v1)
- 11 6. Linton NM, Kobayashi T, Yang Y, Hayashi K, Andrei, AR, Jung S, Yuan B,  
12 Kinoshita R, Nishiura H. (2020). Epidemiological characteristics of novel  
13 coronavirus infection: A statistical analysis of publicly available case data.  
14 medRxiv. <http://dx.doi.org/10.1101/2020.01.26.20018754>
- 15 7. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an  
16 infectious disease outbreak controllable. *Proceedings of the National*  
17 *Academy of Sciences of the United States of America*. 2004;101(16):6146–  
18 6151.
- 19 8. Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R,  
20 Linton NM, Jung S, Yuan B, Suzuki A, Akhmetzhanov AR. The rate of  
21 underascertainment of novel coronavirus (2019-nCoV) infection: Estimation  
22 using Japanese passengers data on evacuation flights. *J Clin Med*. 2020; in  
23 press. doi:10.3390/jcm9020419.
- 24 9. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C,  
25 Zimmer T, Thiel V, Janke C, Guggemos W, Seilmaier M, Drosten C,  
26 Vollmar P, Zwirgmaier K, Zange S, Wölfel R. Transmission of 2019-nCoV

1 infection from an asymptomatic contact in Germany. N Eng J Med. 2020; in  
2 press. doi:10.1056/NEJMc2001468.  
3

1 **Figure Legends**

2 **Figure 1. Serial interval of novel coronavirus (COVID-19) infections.**

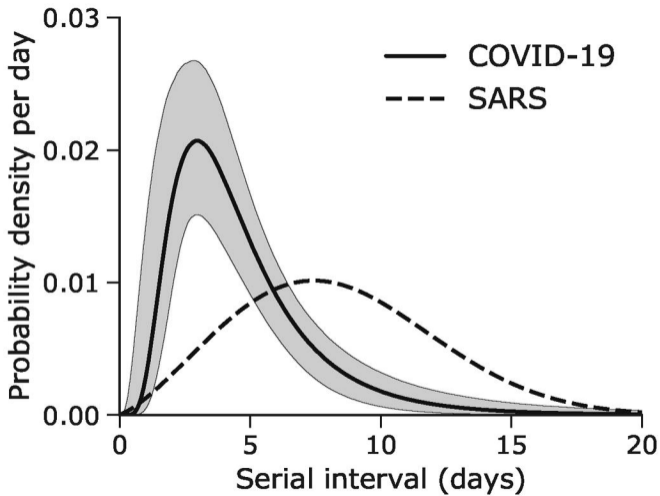
3 The solid line shows the estimated serial interval distribution of COVID-19  
4 infections using the best-fit lognormal distribution with right truncation. A  
5 distribution based on a published estimate of the serial interval for severe acute  
6 respiratory syndrome [3] is overlaid as a dashed line for comparison.

7

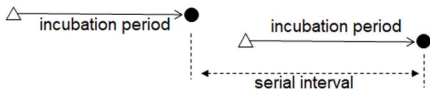
8 **Figure 2. The relationship between the incubation period and serial interval.**

9 If the transmission takes place during the symptomatic period of the primary case,  
10 the serial interval is longer than the incubation period. However, this relationship  
11 can be reversed when pre-symptomatic transmission takes place (the secondary  
12 case may even experience illness onset prior to onset in their infector).

13

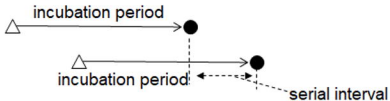


Symptomatic transmission (incubation period  $\leq$  serial interval)



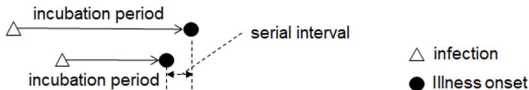
Pre-symptomatic transmission

(incubation period  $>$  serial interval & serial interval  $>$  0)



Pre-symptomatic transmission

(incubation period  $>$  serial interval & serial interval  $\leq$  0)



Time