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

Running title: Serial interval of COVID-19

Authors:

Hiroshi Nishiura[1,2], Natalie M. Linton[1], Andrei R. Akhmetzhanov[1]

Affiliations:

[1]Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi 7 Chome, Kitaku, Sapporo, 060-8638, Japan; [2]CREST, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama, 332-0012 Japan;

Correspondence to:

H. Nishiura, Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi 7 Chome, Kitaku, Sapporo, 060-8638, Japan

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

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

Article type:

Short communication

Abstract

Objective: To estimate the serial interval of novel coronavirus (COVID-19) from information on 28 infector-infectee pairs.

Methods: We collected dates of illness onset for primary cases (infectors) and secondary cases (infectees) from published research articles and case investigation reports. We subjectively ranked the credibility of the data and performed analyses on both the full dataset (*n*=28) and a subset of pairs with highest certainty in reporting (*n*=18). In addition, we adjusting for right truncation of the data as the epidemic is still in its growth phase.

Results: Accounting for right truncation and analyzing all pairs, we estimated the median serial interval at 4.0 days (95% credible interval [CrI]: 3.1, 4.9). Limiting our data to only the most certain pairs, the median serial interval was estimated at 4.6 days (95% CrI: 3.5, 5.9).

Conclusions: The serial interval of COVID-19 is shorter than its median incubation period. This suggests that a substantial proportion of secondary transmission may occur prior to illness onset. The COVID-19 serial interval is also shorter than the serial interval of severe acute respiratory syndrome (SARS), indicating that calculations made using the SARS serial interval may introduce bias.

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

Highlights

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

Introduction

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

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

Materials and Methods

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

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

$$L\left(\Theta_g; \boldsymbol{D}\right) = \prod_i \int_{E_{L.i}}^{E_{R.i}} \int_{S_{L.i}}^{S_{R.i}} g(e) f(s-e) \mathrm{d}s \mathrm{d}e, \quad (1)$$

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

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

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

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

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

Results

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

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

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

Discussion

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

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

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

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

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

Acknowledgments

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

Conflict of interest

The authors declare no conflicts of interest.

References

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

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

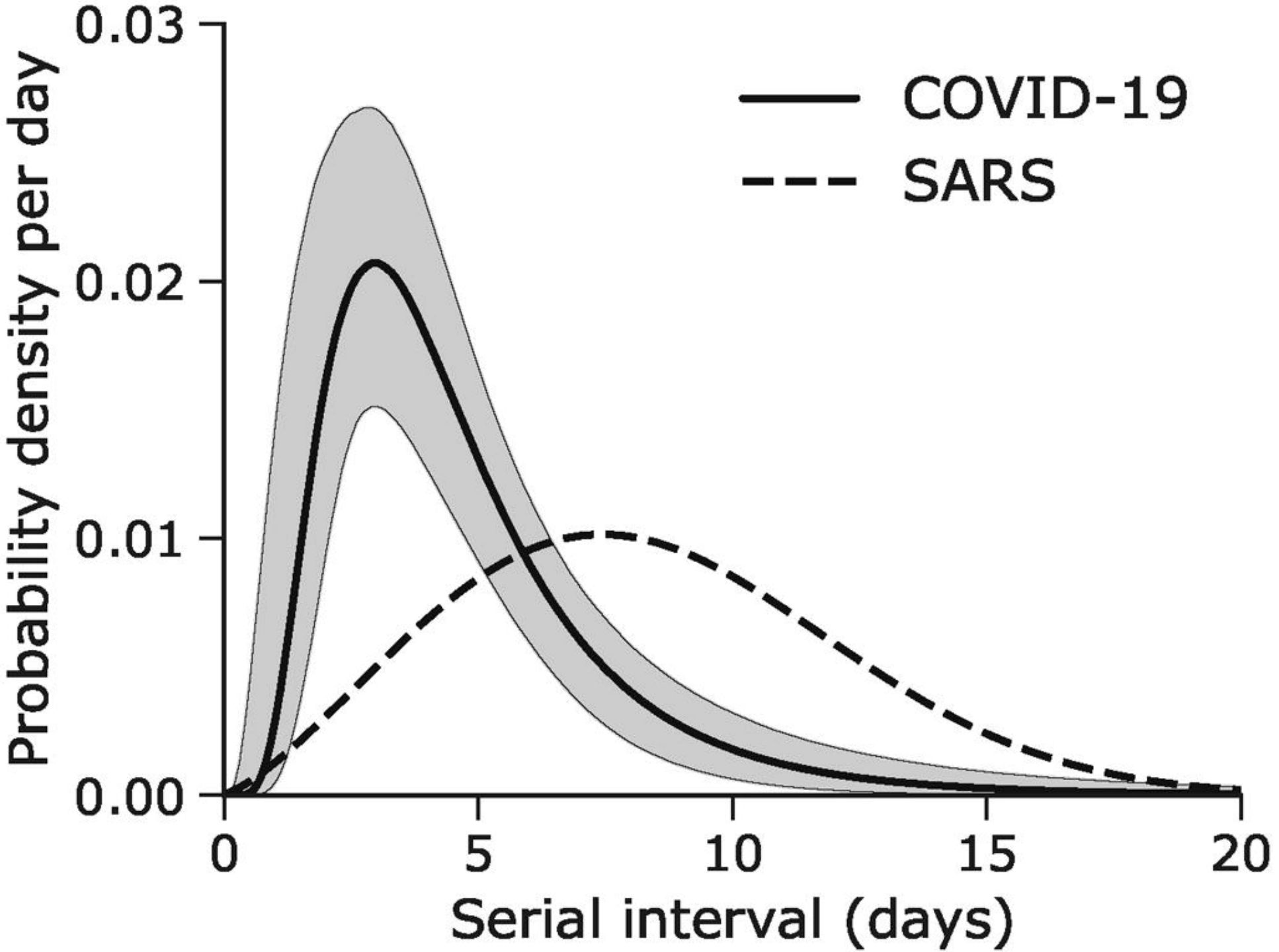
Figure Legends

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

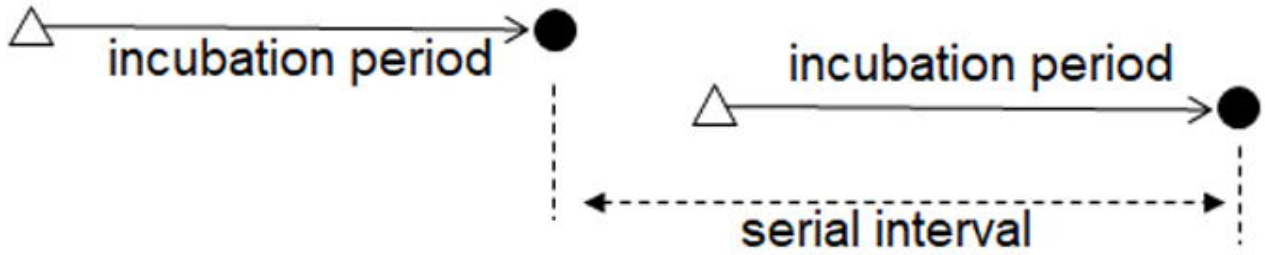
The solid line shows the estimated serial interval distribution of COVID-19 infections using the best-fit lognormal distribution with right truncation. A distribution based on a published estimate of the serial interval for severe acute respiratory syndrome [3] is overlaid as a dashed line for comparison.

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

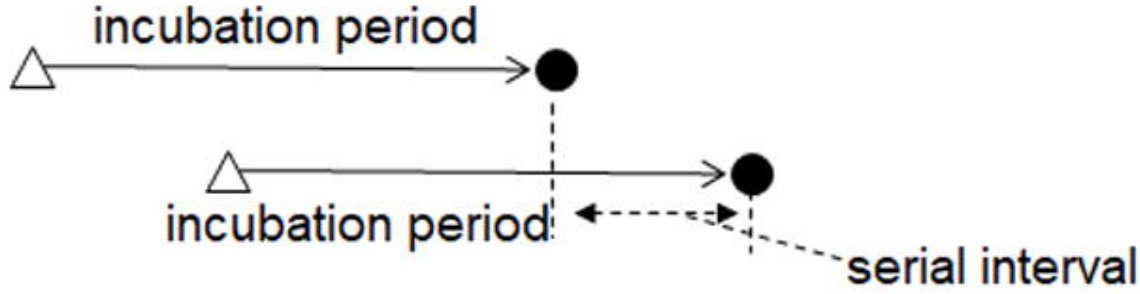
If the transmission takes place during the symptomatic period of the primary case, the serial interval is longer than the incubation period. However, this relationship can be reversed when pre-symptomatic transmission takes place (the secondary case may even experience illness onset prior to onset in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


0.03
0.02
0.01
0.00
Probability density per day
COVID-19
SARS
0
5
10
15
20
Serial interval (days)

Symptomatic transmission (incubation period ≤ serial interval)



Pre-symptomatic transmission
(incubation period > serial interval & serial interval > 0)



Pre-symptomatic transmission
(incubation period > serial interval & serial interval ≤ 0)