

1 **Estimating the serial interval of the novel coronavirus disease**
2 **(COVID-19): A statistical analysis using the public data in Hong Kong**
3 **from January 16 to February 15, 2020**

4 Shi Zhao^{1,2,*}, Daozhou Gao³, Zian Zhuang⁴, Marc KC Chong^{1,2}, Yongli Cai⁵, Jinjun Ran⁶, Peihua Cao⁷, Kai
5 Wang⁸, Yijun Lou⁴, Weiming Wang^{5,*}, Lin Yang⁹, Daihai He^{4,*}, and Maggie H Wang^{1,2}

6
7 1 JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China

8 2 Shenzhen Research Institute of Chinese University of Hong Kong, Shenzhen, China

9 3 Department of Mathematics, Shanghai Normal University, Shanghai, China

10 4 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

11 5 School of Mathematics and Statistics, Huaiyin Normal University, Huaian, China

12 6 School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China

13 7 Clinical Research Centre, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China

14 8 Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi, 830011, China

15 9 School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

16 * Correspondence to: zhaoshi.cmsa@gmail.com (SZ), weimingwang2003@163.com (WW), and

17 daihai.he@polyu.edu.hk (DH)

18
19 **Email address of all authors**

20 SZ: zhaoshi.cmsa@gmail.com

21 DG: dzgao@shnu.edu.cn

22 ZZ: larry.zhuang@polyu.edu.hk

23 MKCC: marc@cuhk.edu.hk

24 YC: yonglicai@hytc.edu.cn

25 JR: jimran@connect.hku.hk

26 PC: cphcc@smu.edu.cn

27 KW: wangkaimath@sina.com

28 YL: yijun.lou@polyu.edu.hk

29 WW: weimingwang2003@163.com

30 LY: l.yang@polyu.edu.hk

31 DH: daihai.he@polyu.edu.hk

32 MHW: maggiew@cuhk.edu.hk

33

34 Abstract

35 **Backgrounds:** The emerging virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has
36 caused a large outbreak of novel coronavirus disease (COVID-19) in Wuhan, China since December 2019.
37 Based on the publicly available surveillance data, we identified 21 transmission chains in Hong Kong and
38 estimated the serial interval (SI) of COVID-19.

39 **Methods:** Index cases were identified and reported after symptoms onset, and contact tracing was conducted
40 to collect the data of the associated secondary cases. An interval censored likelihood framework is adopted to
41 fit a Gamma distribution function to govern the SI of COVID-19.

42 **Findings:** Assuming a Gamma distributed model, we estimated the mean of SI at 4.4 days (95% CI: 2.9–6.7)
43 and SD of SI at 3.0 days (95% CI: 1.8–5.8) by using the information of all 21 transmission chains in Hong
44 Kong.

45 **Conclusion:** The SI of COVID-19 may be shorter than the preliminary estimates in previous works. Given the
46 likelihood that SI could be shorter than the incubation period, pre-symptomatic transmission may occur, and
47 extra efforts on timely contact tracing and quarantine are recommended in combating the COVID-19 outbreak.

48 **Keywords:** novel coronavirus disease; serial interval; statistical analysis; Hong Kong; contact tracing.

49

50 Introduction

51 The novel coronavirus disease (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2
52 (SARS-CoV-2, formerly known as the ‘2019-nCoV’), which has emerged in Wuhan, China in December 2019
53 [1]. The COVID-19 cases were soon exported to other Chinese cities and overseas, and the travel-related risk
54 of disease spreading was suggested by [2, 3]. Since the first confirmed imported case in Hong Kong on
55 January 23, the local government has implemented a series of control and prevention measures for COVID-19,
56 including enhanced border screening and traffic restrictions [4, 5].

57 As of February 15, there were 56 COVID-19 cases confirmed in Hong Kong [4], and local transmission was
58 also recognized by the contact tracing investigation. Given the risk of human-to-human transmission, the
59 serial interval (SI), which refers to the time interval from illness onset in a primary case (i.e., infector) to that
60 in a secondary case (i.e., infectee) [6-9], was of interested to iterative rate of transmission generations of
61 COVID-19. SI could be used to assist strategic decision-making of public health policies and construct
62 analytical frameworks for studying the transmission dynamics of SARS-CoV-2.

63 In this study, we examined the publicly available materials released by the Centre for Health Protection (CHP)
64 of Hong Kong. Adopting the case-ascertained design [10], we identified the transmission chain from index
65 cases to secondary cases. We estimated the SI of COVID-19 based on 21 identified transmission chains from
66 the surveillance data and contact tracing data in Hong Kong.

67 Data and methods

68 As of February 15, there were 56 confirmed COVID-19 cases in Hong Kong [4], which followed the case
69 definition in official diagnostic protocol released by the World Health Organization (WHO) [11]. To identify
70 the pairs of infector (i.e., index case) and infectee (i.e., secondary case), we scanned all news press released by
71 the CHP of Hong Kong between January 16 and February 15, 2020 [5]. The exact symptoms onset dates of all
72 individual patients were released by CHP [4], which were publicly available, and used to match each
73 transmission chain. For those infectees associated with multiple infectors, we record the range of onset dates
74 of all associated infectors, i.e., lower and upper bounds. We identified 21 transmission chains, including 12
75 infectees matched with only one infector, that were used for SI estimation.

76 Following previous study [6], we adopted a Gamma distribution with mean μ and standard deviation (SD) σ ,
77 denoted by $g(\cdot|\mu, \sigma)$, to govern the distribution of SI. The interval censored likelihood, denoted by L , of SI
78 estimates is defined in Eqn (1).

$$L(\mu, \sigma | \tau, T^{\text{low}}, T^{\text{up}}) = \prod_i \left[\int_{T_i^{\text{low}}}^{T_i^{\text{up}}} h(t | T_i^{\text{low}}, T_i^{\text{up}}) g(\tau_i - t | \mu, \sigma) dt \right]. \quad (1)$$

79 The $h(\cdot)$ was the probability density function (PDF) of exposure following a uniform distribution with a range
80 from T^{low} to T^{up} . The terms T_i^{low} and T_i^{up} denoted the lower and upper bounds, respectively, for the range of
81 onset dates of multiple infectors linked to the i -th infectee. Specially, for the infectees with only one infector,
82 $T^{\text{low}} = T^{\text{up}}$, and thus $h(\cdot) = 1$. The τ_i was the observed onset date of the i -th infectee. We calculated the
83 maximum likelihood estimates of μ and σ . Their 95% confidence interval (95%CI) were calculated by using
84 the profile likelihood estimation framework with cutoff threshold determined by a Chi-square quantile [12].

85 In the dataset, the latest onset date of all infectors was on January 31 in contrast to the investigation period up
86 to February 15. Given the minimum investigation period at $(15 - 0 + 1 =) 16$ days, which is approximately
87 twice of the SI of the severe acute respiratory syndrome (SARS) [13], we ignored the right-truncated selection
88 bias, i.e., ‘infector-infectee’ pairs with longer SI, due to short investigation period. Therefore, the likelihood
89 framework defined in Eqn (1) was sufficient for the estimation in this study.

90 Results and discussion

91 The observed SIs of all 21 samples have a mean at 4.3 days, median at 4 days, interquartile range (IQR)
92 between 2 and 5, and range from 1 to 13 days. For the 12 ‘infector- infectee’ pairs, the observed SIs have a
93 mean at 3 days, median at 2 days, IQR between 2 and 4, and range from 1 to 8 days. Fig 1 shows the
94 likelihood profiles of varying SI with respect to μ and σ of SI. By using all 21 samples, we estimated the mean
95 of SI at 4.4 days (95%CI: 2.9–6.7) and SD of SI at 3.0 days (95%CI: 1.8–5.8). The fitted Gamma distribution
96 was shown in Fig 2. These estimates largely matched the results in the existing literatures [14, 15]. Limiting to
97 only consider the 12 ‘infector-infectee’ pairs, we estimated the mean of SI at 3.1 days (95%CI: 2.0–5.4) and
98 SD of SI at 1.8 days (95%CI: 1.0–4.7).

99 Comparing to the SI of SARS with mean at 8.4 days and SD at 3.4 days [13], the estimated 4.4-day SI for
100 COVID-19 indicated rapid cycles of generation replacement in the transmission chain. Hence, highly efficient
101 public health control measures, including contact tracing, isolation and screening, were strongly recommended
102 to mitigate the epidemic size. The timely supply and delivery of healthcare resources, e.g., facemasks, alcohol
103 sterilizer and manpower and equipment for treatment, were of required in response to the rapid growing
104 incidences of COVID-19.

105 As also pointed out by two recent works [14, 15], the mean of SI at 4.4 days is notably smaller than the mean
106 incubation period, roughly 5 days, estimated by many previous studies [16-19]. The pre-symptomatic
107 transmission may occur when the SI is shorter than the incubation period. If isolation can be conducted
108 immediately after the symptom onset, the pre-symptomatic transmission is likely to contribute to the most of
109 SARS-CoV-2 infections. This situation has been recognized by a recent epidemiological investigation
110 evidently [20], and implemented in the mechanistic modelling studies of COVID-19 epidemic [3, 21], where
111 the pre-symptomatic cases were contagious. As such, merely isolating the symptomatic cases will lead to a
112 considerable proportion of secondary cases, and thus contact tracing and immediately quarantine were crucial
113 to reduce the risk of infection.

114 A recently epidemiological study used 5 ‘infector-infectee’ pairs from contact tracing data in Wuhan, China
115 during the early outbreak to estimate the mean SI at 7.5 days (95%CI: 5.3–19.0) [17], which appeared larger
116 than our SI estimate at 4.4 days. Although the 95% CIs of SI estimates in this study and [17] were not
117 significantly separated, we note that the difference in SI might exist. If this difference was not due to sampling
118 chance, one of the possible explanations could be the pathogenic heterogeneity in mainland China and Hong
119 Kong. The SI estimate can be benefit from larger sample size, and the estimates in our study was backed up
120 by 21 identified transmission chains including 12 ‘infector-infectee’ pairs, and thus it is likely to be more
121 informative.

122 Accurate and consistent records on dates of illness onset were essential to the estimation of the SI. All samples
123 used in this analysis were identified in Hong Kong and collected consistently from the CHP [4, 5]. Hence, the
124 reporting criteria were most likely to be the same for all COVID-2019 cases, which potentially made our
125 findings more robust.

126 The clusters of cases can occur by person-to-person transmission within the cluster, e.g.,

- 127 • scenario (I): person A infected B, C and D, or
- 128 • scenario (II): A to B to C to D, or
- 129 • scenario (III): a mixture of (I) and (II), e.g., A to B, B to C and D, or

130 or they can occur through a common exposure to an unrecognized source of infection, e.g.,

- 131 • scenario (IV): unknown person X infected A, B, C and D; or

- 132 • scenario (V): a mixture of (IV) and (I) or (II), e.g., X to A and B, B to C and D; or

133 The lack of information in the publicly available dataset made it difficult to disentangle such complicated
134 situations. The scenarios (I) and (II) can be covered by the pair of ‘infector-infectee’ such that we could
135 identify the link between two unique consecutive infections. Under the scenario (III), we cannot clearly
136 identify the pairwise match between the infector and infectee, which means there were multiple candidate of
137 infector for one infectee. As such, we employed the PDF $h(\cdot)$ in Eqn (1) to account for the possible time of
138 exposure ranging from T^{low} to T^{up} . There is no information available on the SI for scenarios (IV) as well as (V)
139 due to the onset date of person X is unknown, and thus our analysis was limited in the scenarios (I)-(III). We
140 note that extra-cautious should be needed to interpret the clusters of cases because of this potential limitation.
141 Although we used interval censoring likelihood to deal with the multiple-infector matching issue, more
142 detailed information of the exposure history and clue on ‘who acquires infection from whom’ (WAIFW)
143 would improve our estimates.

144 Longer SI might be difficult to occur in reality due to the isolation of confirmed infections, or to identify and
145 link together due to the less accurate information associated with memory error occurred in the backward
146 contact tracing exercise. The issue associated with isolation could possibly bias the SI estimates and lead to an
147 underestimated result. Due to lack of information in the public dataset, our estimation framework could be
148 benefit from detailed records on the date of isolation of individual cases. But isolation occurs for all severe
149 infectious diseases such as SARS and COVID-19, there is no reason to believe that isolation play a bigger role
150 in one than the other, at least more evidence is needed to support that claim. It is however possible that at the
151 initial stage the SI is longer than later when strict isolation takes place. Nevertheless, a comparison of
152 estimated SI for SARS and COVID-19 in Hong Kong is still meaningful. And we found that the SI of
153 COVID-19 estimated appears shorter than that of SARS, which is the key message. It would be hard to
154 imagine that isolation is responsible for the difference. There is no reason to believe isolation is more rapid in
155 cases of COVID-19 than in SARS in Hong Kong, as well as other limitations (would have happened for both).
156 Thus, the difference we observed for COVID-19 and SARS is likely intrinsic. In conclusion, given the rapid
157 spreading of the COVID-19, effective contact tracing and quarantine/isolation were even more crucial for
158 successful control.

159

160

161 **Declarations**

162 **Ethics approval and consent to participate**

163 The follow-up data of individual patients were collected via public domain [4, 5], and thus neither ethical
164 approval nor individual consent was not applicable.

165 **Availability of materials**

166 All data used in this work were publicly available via [4, 5], and the exacted dataset was attached as a
167 supplementary files of this study.

168 **Consent for publication**

169 Not applicable.

170 **Funding**

171 DH was supported by General Research Fund (Grant Number 15205119) of the Research Grants Council
172 (RGC) of Hong Kong, China. WW was supported by National Natural Science Foundation of China (Grant
173 Number 61672013) and Huaian Key Laboratory for Infectious Diseases Control and Prevention (Grant
174 Number HAP201704), Huaian, Jiangsu, China.

175 **Acknowledgements**

176 None.

177 **Disclaimer**

178 The funding agencies had no role in the design and conduct of the study; collection, management, analysis,
179 and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the
180 manuscript for publication.

181 **Conflict of Interests**

182 The authors declared no competing interests.

183 **Authors' Contributions**

184 SZ conceived the study, carried out the analysis, and drafted the first manuscript. All authors discussed the
185 results, critically read and revised the manuscript, and gave final approval for publication.

186

187 References

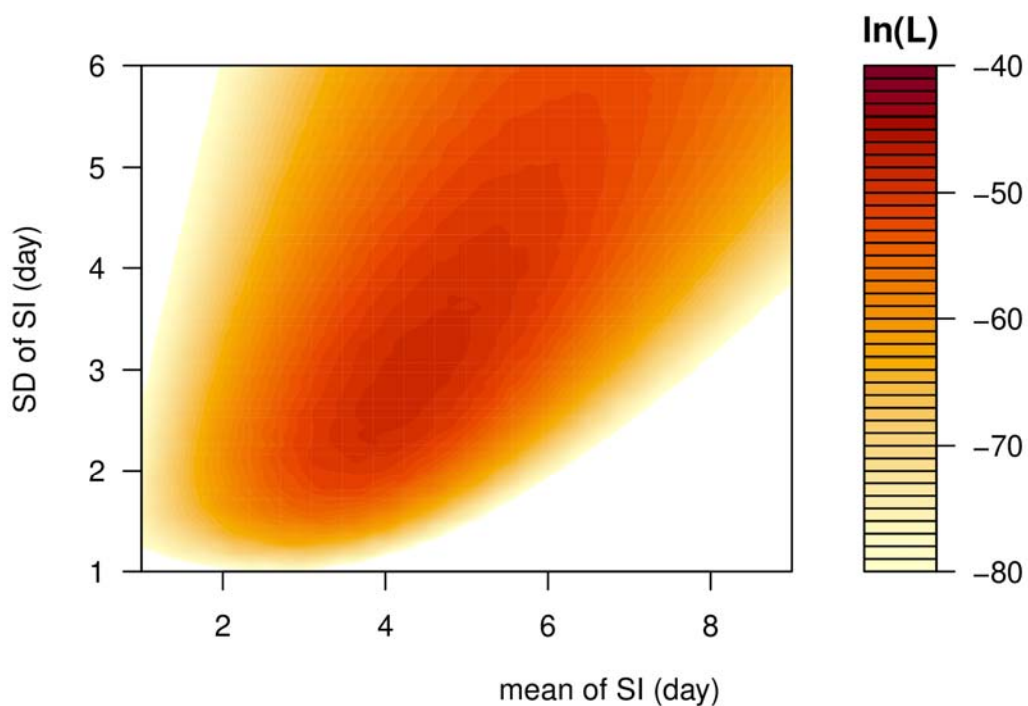
- 188 1. **'Pneumonia of unknown cause – China', Emergencies preparedness, response, Disease outbreak**
189 **news, World Health Organization (WHO)** [[https://www.who.int/csr/don/05-january-2020-](https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/)
190 [pneumonia-of-unkown-cause-china/en/](https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/)]
- 191 2. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K: **Pneumonia of unknown**
192 **etiology in Wuhan, China: potential for international spread via commercial air travel.** *Journal*
193 *of Travel Medicine* 2020;doi:10.1093/jtm/taaa1008.
- 194 3. Wu JT, Leung K, Leung GM: **Nowcasting and forecasting the potential domestic and**
195 **international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling**
196 **study.** *The Lancet* 2020.
- 197 4. **Summary of data and outbreak situation of the Severe Respiratory Disease associated with a**
198 **Novel Infectious Agent, Centre for Health Protection, the government of Hong Kong.**
199 [<https://www.chp.gov.hk/en/features/102465.html>]
- 200 5. **The collection of Press Releases by the Centre for Health Protection (CHP) of Hong Kong.**
201 [<https://www.chp.gov.hk/en/media/116/index.html>]
- 202 6. Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM: **Estimation of the serial interval of influenza.**
203 *Epidemiology (Cambridge, Mass)* 2009, **20**(3):344.
- 204 7. Fine PEM: **The Interval between Successive Cases of an Infectious Disease.** *American Journal of*
205 *Epidemiology* 2003, **158**(11):1039-1047.
- 206 8. Nishiura H, Chowell G, Heesterbeek H, Wallinga J: **The ideal reporting interval for an epidemic to**
207 **objectively interpret the epidemiological time course.** *Journal of the Royal Society Interface* 2010,
208 **7**(43):297-307.
- 209 9. Wallinga J, Lipsitch M: **How generation intervals shape the relationship between growth rates**
210 **and reproductive numbers.** *Proceedings of the Royal Society B: Biological Sciences* 2007,
211 **274**(1609):599-604.
- 212 10. Yang Y, Longini I, Halloran ME: **Design and evaluation of prophylactic interventions using**
213 **infectious disease incidence data from close contact groups.** *J R Stat Soc Ser C-Appl Stat* 2006,
214 **55**:317-330.
- 215 11. **Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases, World**
216 **Health Organization (WHO)** [[https://www.who.int/health-topics/coronavirus/laboratory-](https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus)
217 [diagnostics-for-novel-coronavirus](https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus)]
- 218 12. Fan J, Huang T: **Profile likelihood inferences on semiparametric varying-coefficient partially**
219 **linear models.** *Bernoulli* 2005, **11**(6):1031-1057.
- 220 13. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew SK, Tan CC,
221 Samore MH: **Transmission dynamics and control of severe acute respiratory syndrome.** *Science*
222 2003, **300**(5627):1966-1970.
- 223 14. Nishiura H, Linton NM, Akhmetzhanov AR: **Serial interval of novel coronavirus (2019-nCoV)**
224 **infections.** *medRxiv* 2020:2020.2002.2003.20019497.
- 225 15. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, Pang CH, Zhang Y, Chen Z, Zhou X-H: **Estimation of**
226 **the Time-Varying Reproduction Number of COVID-19 Outbreak in China.** *medRxiv*
227 2020:2020.2002.2008.20021253.
- 228 16. Backer JA, Klinkenberg D, Wallinga J: **Incubation period of 2019 novel coronavirus (2019-nCoV)**
229 **infections among travellers from Wuhan, China, 20–28 January 2020.** *Eurosurveillance* 2020,
230 **25**(5):2000062.
- 231 17. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY *et al*:
232 **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.**
233 *New England Journal of Medicine* 2020.

- 234 18. Linton MN, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov RA, Jung S-m, Yuan B, Kinoshita R,
235 Nishiura H: **Incubation Period and Other Epidemiological Characteristics of 2019 Novel**
236 **Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case**
237 **Data.** *Journal of Clinical Medicine* 2020, **9**(2).
- 238 19. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith H, Azman AS, Reich NG, Lessler J: **The**
239 **incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and**
240 **application.** *medRxiv* 2020:2020.2002.2002.20020016.
- 241 20. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, Zimmer T, Thiel V, Janke C,
242 Guggemos W *et al*: **Transmission of 2019-nCoV Infection from an Asymptomatic Contact in**
243 **Germany.** *New England Journal of Medicine* 2020.
- 244 21. Chowell G, Dhillon R, Srikrishna D: **Getting to zero quickly in the 2019-nCov epidemic with**
245 **vaccines or rapid testing.** *medRxiv* 2020:2020.2002.2003.20020271.

246

247

248 Figures



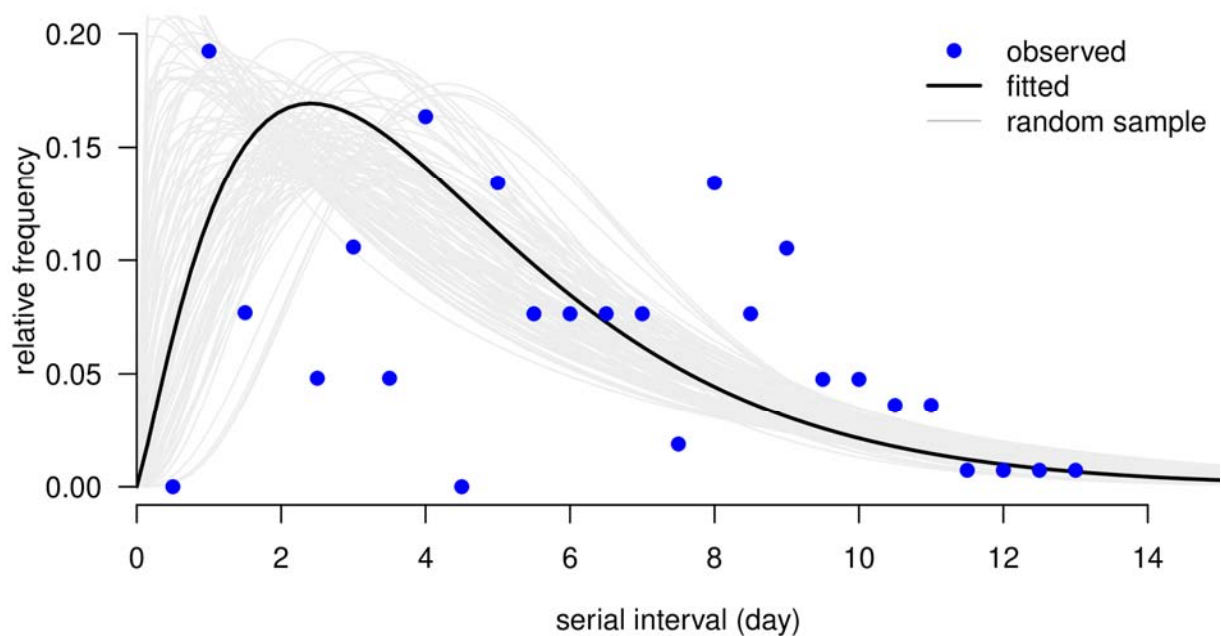
249

250 Figure 1.

251 The likelihood profile of the varying serial interval (SI) of COVID-19 by using all samples. The color scheme
252 is shown on the right-hand side, and a darker color indicates a larger log-likelihood, i.e., $\ln(L)$, value.

253

254



255

256 Figure 2.

257 The distribution of serial interval (SI) by using all samples. The blue dots are the observed distribution of SI,
258 the black bold curve is the fitted distribution of SI, and the light grey curves are 1000 simulation samples.

259

260