

1 Classification: Biological Sciences, BIOLOGICAL SCIENCES

2 **Title:** Estimation of the incubation period of COVID-19 using viral load data

3 (Short title) Estimation of the incubation period of COVID-19 using viral load data

4 **Authors**

5 Keisuke Ejima<sup>1,‡,\*</sup>, Kwang Su Kim<sup>2,‡</sup>, Christina Ludema<sup>1</sup>, Ana I. Bento<sup>1</sup>, Shoya Iwanami<sup>2</sup>, Yasuhisa Fujita<sup>2</sup>,  
6 Hirofumi Ohashi<sup>3</sup>, Yoshiki Koizumi<sup>4</sup>, Koichi Watashi<sup>3,5,6,7</sup>, Kazuyuki Aihara<sup>8</sup>, Hiroshi Nishiura<sup>9</sup>, and  
7 Shingo Iwami<sup>2,6,10,11,12,\*</sup>

8 **Affiliations**

9 <sup>1</sup>Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington,  
10 IN, USA.

11 <sup>2</sup>Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, Japan.

12 <sup>3</sup>Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan.

13 <sup>4</sup>National Center for Global Health and Medicine, Tokyo, Japan.

14 <sup>5</sup>Department of Applied Biological Science, Tokyo University of Science, Noda, Japan.

15 <sup>6</sup>MIRAI, JST, Saitama, Japan.

16 <sup>7</sup>Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan.

17 <sup>8</sup>International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced  
18 Study, The University of Tokyo, Tokyo, Japan.

19 <sup>9</sup>Graduate School of Medicine, Hokkaido University, Hokkaido, Japan

20 <sup>10</sup>Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan.

21 <sup>11</sup>NEXT-Ganken Program, Japanese Foundation for Cancer Research (JFCR), Tokyo, Japan.

22 <sup>12</sup>Science Groove Inc., Fukuoka, Japan.

23 \* Corresponding authors: [kejima@iu.edu](mailto:kejima@iu.edu) (KE) [siwami@kyushu-u.org](mailto:siwami@kyushu-u.org) (SI)

24 ‡ These authors contributed equally to this work.

- 25 **Author contributions:** Conceived and designed the study: KE HN SI. Analysed the data: KE KSK SI.
- 26 Wrote the paper: KE KSK CL AIB SI YF YI HO YL KW KA HS SI. All authors read and approved the
- 27 final manuscript.

28 **Abstract (144/250)**

29 The incubation period, or the time from infection to symptom onset of COVID-19 has been usually estimated  
30 using data collected through interviews with cases and their contacts. However, this estimation is influenced  
31 by uncertainty in recalling effort of exposure time. We propose a novel method that uses viral load data  
32 collected over time since hospitalization, hindcasting the timing of infection with a mathematical model for  
33 viral dynamics. As an example, we used the reported viral load data from multiple countries (Singapore, China,  
34 Germany, France, and Korea) and estimated the incubation period. The median, 2.5, and 97.5 percentiles of  
35 the incubation period were 5.23 days (95% CI: 5.17, 5.25), 3.29 days (3.25, 3.37), and 8.22 days (8.02, 8.46),  
36 respectively, which are comparable to the values estimated in previous studies. Using viral load to estimate  
37 the incubation period might be a useful approach especially when impractical to directly observe the infection  
38 event.

39

40 **Keywords:**

41 SARS-CoV-2, COVID-19, mathematical model, infectious disease epidemiology, incubation period

42

## 43 **Text (1461/1600)**

### 44 **Introduction**

45 The current COVID-19 outbreak is characterized by a longer incubation period (i.e., time from  
46 infection to symptom onset) than that of influenza and other respiratory viruses. The median incubation period  
47 of COVID-19 is estimated as 5 to 6 days (1-4), while that of influenza A and B and SARS-CoV-1 are 1.4, 0.6  
48 and 4.0 days, respectively (5).

49 Estimating the incubation period is challenging, because we rarely directly observe the time of  
50 infection or the time of symptom onset (examples to the contrary in HIV infection show the intense follow up  
51 needed to observe these events (6, 7)). The first study estimating the incubation period of SARS-CoV-2 was  
52 Li (4), where they fit a log-binomial model to a subset of cases where detailed information about their exposure  
53 to another case was available. Another set of early studies used information from cases that were identified  
54 outside of Hubei province (1-3) to estimate the incubation period. In these studies, the time of exposure was  
55 inferred using the duration of travel to Wuhan. Bi et al (4) added considerably to this literature by estimating  
56 the incubation period from contact-based surveillance in which all the contacts of identified cases were tested  
57 prospectively and a more complete chain of transmission could be documented.

58 However, even with meticulous contact tracing effort, directly observing infector-infectee pairs is a  
59 time-consuming process, especially when the incubation period is lengthy. Measuring the incubation period  
60 through contact tracing is more difficult if the infector-infectee pair had a lot of contact with each other,  
61 leading to a wide range of tracing among suspected individuals. Indeed, Bi et al, demonstrated large  
62 uncertainty (the interval of exposure was more than 10 days for about 25% of the cases) on the timing of  
63 infection for COVID-19 in China (4). Although a majority of these studies (1, 2, 4) use a statistical modeling  
64 technique that accounts for uncertainty both in the reports of exposure time and the time of symptom onset  
65 (8), they had to inherently use a heuristic weight function for the censored information.

66 Here we propose another approach to estimate the incubation period, where we use longitudinal data  
67 on viral load and hindcast the point of initial infection. Viral load data were collected at the early stage of the  
68 epidemic for clinical purposes (e.g., understanding the aetiology and the pathophysiology of COVID-19) and  
69 to ensure patients were no longer shedding virus (or more precisely, viral fragments) before hospital discharge.

70 The data were analysed using a mathematical model that describes the viral dynamics, which typically draw  
71 a bell-shaped curve (i.e., viral load increases exponentially first until the peak where it starts declining).  
72 Although the data are available only after the onset of symptoms, the timing of infection can be estimated by  
73 hindcasting the model for each case.

## 75 **Results**

76 We extracted the viral load data reported in five papers. **Figure 1** shows the timing of infection for  
77 each case estimated from viral load data using the virus dynamics model. The peak of viral load appears after  
78 2-3 days from symptom onset. The AICs of the three models (log-normal, gamma, the Weibull distributions)  
79 were, 10014.3, 10410.5, 12071.0, respectively. Thus, the lognormal distribution was preferred. **Figure 2**  
80 summarized the estimated cumulative distribution function and probability density function of incubation  
81 period for COVID-19. The median, 2.5, and 97.5 percentile of incubation period was 5.23 days (95%CI: 5.17  
82 to 5.25), 3.29 days (3.25 to 3.37), and 8.22 days (8.02 to 8.46), respectively.

## 84 **Discussion**

85 Inferring the timing of infection is challenging in general. Given asymptomatic and pre-symptomatic  
86 transmission and the long incubation period, not all cases were aware of an exposure or the specific time of  
87 exposure. Thus, we propose using viral load data, which is externally measured and independent from recalling  
88 effort. The median of the estimated incubation period was about 5 days, and 97.5% of cases develop symptoms  
89 in about 8 days. These estimations are consistent with published estimates(1-4).

90 The strength of this approach is that it can complement limitations that classical interview-based  
91 approach has pertained to ascertain the exposure event. Our proposed approach may be applicable not only to  
92 the human infectious disease and zoonoses such as influenza and COVID-19, but to animal/livestock  
93 infectious diseases such as foot and mouth disease when contact recall is not possible.

94 We note that there are several studies proposed statistical approach to estimate the incubation period  
95 using observed biomarkers, especially for HIV/AIDS. Shi et al. and Geskus used CD4 counts to estimate the  
96 incubation period as well as residual time (i.e., time from AIDS diagnosis to current time)(9, 10). The

97 uniqueness of our approach compared with these examples is that we have used conventional viral dynamics  
98 model for respiratory diseases with acute course of illness.

99 Limitations should be noted. Our approach did not account for any uncertainty in reporting of  
100 symptoms, which was accounted in the previous approach by Reich et al. (8). Combining ours with Reich et  
101 al. might reduce uncertainty surrounding the precise reporting of exposure and illness onset events. The model  
102 we used in this study did not include detailed immune response or antiviral effects given limited information.  
103 The proposed approach requires collection of viral loads over time since symptom onset, which might not be  
104 feasible for all patients or in resource limited contexts. Additional diagnostic testing methods are presently  
105 developed to measure SARS-CoV-2 viral load in saliva, which would ease the process and mitigate the risk  
106 of infection of those involved in collecting the samples(11, 12).

107 Valid estimation of incubation period is essential to mitigate risk by simplifying the process of contact  
108 tracing and understanding the role pre-symptomatic infection. Unifying the proposed approach with existing  
109 epidemiological methods, precise determination of the length of quarantine will be achieved.

## 111 **Materials and Methods**

### 112 **Data**

113 The viral load data from five previously published papers among hospitalized COVID-19 patients were  
114 used (13-17). All cases used in our analysis presented symptoms before or after hospitalization. For  
115 consistency, the viral load data from upper respiratory specimens were used in the analysis. The cases treated  
116 with antivirals or with less than two data points were excluded. For all the studies from which we extracted  
117 data, ethics approval was obtained from the ethics committee at each institute. Written informed consent was  
118 obtained from the cases or their next of kin in the original studies. We summarized the data in **Table 1**.

### 120 **A mathematical model for virus dynamics to estimate the day of infection establishment**

121 The virus and its target cell dynamics are described by a mathematical model previously proposed  
122 in(13, 18, 19).

123 
$$\frac{df(t)}{dt} = -\beta f(t)V(t), \quad \frac{dV(t)}{dt} = \gamma f(t)V(t) - \delta V(t),$$

124 where  $f(t)$  and  $V(t)$  are the ratio of uninfected target cells and the amount of virus, respectively. The  
125 parameters  $\beta$ ,  $\gamma$ , and  $\delta$  are the rate constant for virus infection, the maximum rate constant for viral replication  
126 and the death rate of infected cells, respectively. The viral load data from the five different papers were fitted  
127 to the model with mixed effects, which assumed that the parameters for each individual follow normal  
128 distributions with the same population mean. Further, the day of SARS-CoV-2 infection establishment, in  
129 other words, the start of the exponential growth phase of viral loads(19), was estimated using the estimated  
130 parameters for each individual case. The time of infection event,  $T_{\text{inf}}$ , was identified by means of back-  
131 calculation, using dataset when the viral load reaches the threshold. The viral load threshold for infection  
132 establishment was estimated using the data of three cases whose primary cases and exposure history is known  
133 (thus the day of infection establishment is known)(15), in which the start day of exposure is assumed to be  
134 equal to the day of infection event. To address the uncertainty of the estimation, we resampled 100 parameter  
135 sets for each individual and obtained corresponding 100 of  $T_{\text{inf}}$ .

### 136 137 **Incubation period estimation**

138 The estimated incubation periods,  $T_{\text{inf}}$ , were fitted to three parametric distributions: Weibull, gamma, and log-  
139 normal distributions. Comparing the Akaike Information Criteria (AIC) for those three distributions, the best  
140 model (i.e., with lowest AIC) was used for further analyses. The parametric bootstrap method was employed  
141 to assess the parameter uncertainty. Specifically, the bootstrap sample was generated by resampling with  
142 replacement from the all estimated  $T_{\text{inf}}$ . The proposed parametric models (i.e., Weibull, gamma, and log-  
143 normal distributions) were fitted to the bootstrapped data for parameter inference. We repeated this process  
144 1000 times and obtained 1000 parameter sets, and the median, 2.5, and 97.5 percentiles of the distribution are  
145 computed.

### 146 147 **Viral load threshold for infection establishment**

The viral load threshold for infection establishment was estimated using the data from the three cases with known primary cases reported in China (i.e., Patients D, H and L)(15): a case (Patient E) from Wuhan visited Patient D and Patient L in Zhuhai on January 17. Patients D and L developed symptoms on January 23 and 20, respectively (thus their primary case is Patient E). Two cases (Patient I and P) from Wuhan visited their daughter, Patient H, in Zhuhai on January 11. Patient H developed fever on January 17 (thus her primary cases are Patient I and P). Assuming that infection was established on the day when primary and secondary cases first met, the viral load on the day of infection establishment was computed by hindcasting the mathematical model with the estimated parameters, which we defined as the infection establishment threshold:  $10^{-5.77} - 10^{-4.32}$ ,  $10^{-3.81} - 10^{-2.63}$ , and  $10^{-1.32} - 10^{0.11}$  for Patients D, H and L, respectively. We used the middle value ( $10^{-2.83}$ ) as the threshold.

## Acknowledgments

This study was supported in part by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education 2019R1A6A3A12031316 (to K.S.K.); Grants-in-Aid for JSPS Scientific Research (KAKENHI) Scientific Research B 17H04085 (to K.W.), 18KT0018 (to S.I.), 18H01139 (to S.I.), 16H04845 (to S.I.), Scientific Research S 15H05707 (to K.A.), Scientific Research in Innovative Areas 20H05042 (to S.I.), 19H04839 (to S.I.), 18H05103 (to S.I.); AMED JP20dm0307009 (to K.A.); AMED CREST 19gm1310002 (to S.I.); AMED J-PRIDE 19fm0208019j0003 (to K.W.), 19fm0208006s0103 (to S.I.), 19fm0208014h0003 (to S.I.), 19fm0208019h0103 (to S.I.); AMED Research Program on HIV/AIDS 19fk0410023s0101 (to S.I.); Research Program on Emerging and Re-emerging Infectious Diseases 19fk0108050h0003 (to S.I.); Program for Basic and Clinical Research on Hepatitis 19fk0210036j0002 (to K.W.), 19fk0210036h0502 (to S.I.); Program on the Innovative Development and the Application of New Drugs for Hepatitis B 19fk0310114j0003 (to K.W.), 19fk0310101j1003 (to K.W.), 19fk0310103j0203 (to K.W.), 19fk0310114h0103 (to S.I.); JST PRESTO (to S.N.); JST MIRAI (to K.W. and S.I.); The Yasuda Medical Foundation (to K.W.); Smoking Research Foundation (to K.W.); Takeda Science Foundation (to K.W.); Mochida Memorial Foundation for Medical and Pharmaceutical Research (to K.W.); Mitsui Life Social Welfare Foundation (to S.I. and K.W.); Shin-Nihon of Advanced Medical Research (to

175 S.I.); Suzuken Memorial Foundation (to S.I.); Life Science Foundation of Japan (to S.I.); SECOM Science  
176 and Technology Foundation (to S.I.); The Japan Prize Foundation (to S.I.); Toyota Physical and Chemical  
177 Research Institute (to S.I.); Fukuoka Financial Group, Inc. (to S.I.); Kyusyu Industrial Advancement Center  
178 Gapfund Program (to S.I.); Foundation for the Fusion Of Science and Technology (to S.I.).

179 **Competing Interest Statement**

180 The authors declare that they have no competing interests.

181

182

## 183 References

- 184 1. Lauer SA, *et al.* (2020) The Incubation Period of Coronavirus Disease 2019 (COVID-19) From  
185 Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*.
- 186 2. Linton NM, *et al.* (2020) Incubation Period and Other Epidemiological Characteristics of 2019  
187 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available  
188 Case Data. *Journal of clinical medicine* 9(2).
- 189 3. Backer JA, Klinkenberg D, & Wallinga J (2020) Incubation period of 2019 novel coronavirus (2019-  
190 nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill* 25(5).
- 191 4. Bi Q, *et al.* (2020) Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their  
192 close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*.
- 193 5. Lessler J, *et al.* (2009) Incubation periods of acute respiratory viral infections: a systematic review.  
194 *The Lancet Infectious Diseases* 9(5):291-300.
- 195 6. Robb ML, *et al.* (2016) Prospective Study of Acute HIV-1 Infection in Adults in East Africa and  
196 Thailand. 374(22):2120-2130.
- 197 7. Rolland M, *et al.* (2020) Molecular dating and viral load growth rates suggested that the eclipse  
198 phase lasted about a week in HIV-1 infected adults in East Africa and Thailand. *PLoS pathogens*  
199 16(2):e1008179.
- 200 8. Reich NG, Lessler J, Cummings DAT, & Brookmeyer R (2009) Estimating incubation period  
201 distributions with coarse data. 28(22):2769-2784.
- 202 9. Geskus RB (2000) On the inclusion of prevalent cases in HIV/AIDS natural history studies through a  
203 marker-based estimate of time since seroconversion. *Stat Med* 19(13):1753-1769.
- 204 10. Shi M, Taylor JM, & Muñoz A (1996) Models for residual time to AIDS. *Lifetime data analysis*  
205 2(1):31-49.
- 206 11. Azzi L, *et al.* (2020) Saliva is a reliable tool to detect SARS-CoV-2. *Journal of Infection*.
- 207 12. Wyllie AL, *et al.* (2020) Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients  
208 than nasopharyngeal swabs.2020.2004.2016.20067835.
- 209 13. Kim KS, *et al.* (2020) Modelling SARS-CoV-2 Dynamics: Implications for Therapy.  
210 *medRxiv*:2020.2003.2023.20040493.
- 211 14. Wölfel R, *et al.* (2020) Virological assessment of hospitalized patients with COVID-2019. *Nature*.
- 212 15. Zou L, *et al.* (2020) SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.  
213 *N Engl J Med* 382(12):1177-1179.
- 214 16. Young BE, *et al.* (2020) Epidemiologic Features and Clinical Course of Patients Infected With  
215 SARS-CoV-2 in Singapore. *Jama*.
- 216 17. Lescure F-X, *et al.* (2020) Clinical and virological data of the first cases of COVID-19 in Europe: a  
217 case series. *The Lancet Infectious Diseases*.
- 218 18. Ikeda H, *et al.* (2016) Quantifying the effect of Vpu on the promotion of HIV-1 replication in the  
219 humanized mouse model. *Retrovirology* 13(1):23.
- 220 19. Perelson AS (2002) Modelling viral and immune system dynamics. *Nature reviews. Immunology*  
221 2(1):28-36.
- 222

223

**Table 1. Summary of data**

<b>Papers</b>	<b>country</b>	<b>number of included cases (number of excluded cases)</b>	<b>site of viral load data used for this analysis</b>
Young et al.(16)	Singapore	12 (6)	nasopharyngeal swab
Zou et al.(15)	China	8 (8)	nasal swab
Kim et al.(13)	Korea	1 (0)	upper respiratory specimens (nasopharyngeal and oropharyngeal swab)
Wölfel et al.(14)	Germany	8 (1)	pharyngeal swab
Lescure et al.(17)	France	2 (3)	nasopharyngeal swab

224

## 225 **Figure legends**

226 **Fig. 1.** The estimated day of infection establishment for each case using day from symptom onset as a time  
227 scale. The dots and the bars are the median, 2.5, and 97.5 percentiles of the estimated day of infection  
228 establishment for each case.

229 **Fig. 2.** The estimated incubation period. (A) The cumulative distribution function. The grey lines were drawn  
230 based on the different bootstrap samples. The horizontal bars are 95%CI at 2.5%, 50%, and 97.5% of the  
231 distribution. The solid red curve corresponds to the curve estimated using all the sample. (B) The probability  
232 density function.



