

1 **Age-dependent association between SARS-CoV-2 cases reported by passive surveillance and**
2 **viral load in wastewater**

3 Ryosuke Omori^{a,*}, Fuminari Miura^b, Masaaki Kitajima^c,

4

5 ^aDivision of Bioinformatics, International Institute for Zoonosis Control, Hokkaido University, North 20
6 West 10, Kita-Ku, Sapporo, Hokkaido, 001-0020, Japan

7 ^bCenter for Marine Environmental Studies (CMES), Ehime University, 3 Bunkyo, Matsuyama,
8 Ehime, 790-8577, Japan

9 ^cDivision of Environmental Engineering, Faculty of Engineering, Hokkaido University, North 13 West 8,
10 Kita-ku, Sapporo, Hokkaido 060-8628, Japan

11 [□]Corresponding author: International Institute for Zoonosis Control, Hokkaido University, North 20 West 10,
12 Kita-Ku, Sapporo, Hokkaido, 001-0020, Japan.

13 E-mail address: omori@czc.hokudai.ac.jp (R. Omori).

14

15

16 **Abstract**

17 The true number of individuals infected with severe acute respiratory syndrome coronavirus 2
18 (SARS-CoV-2) is difficult to estimate using a case-reporting system (i.e., passive surveillance) alone
19 because of asymptomatic infection. While wastewater-based epidemiology has been implemented as an
20 alternative/additional monitoring tool to reduce reporting bias, the relationship between passive and
21 wastewater surveillance data has yet to be explicitly examined. Since there is strong age dependency in the
22 symptomatic ratio of SARS-CoV-2 infections, this study aimed to estimate i) an age-dependent association
23 between the number of reported cases and the viral load in wastewater and ii) the time lag between those
24 time series. The viral load in wastewater was modeled as a combination of contributions from different age
25 groups' virus shedding, incorporating the delay, and fitted with daily case count data collected from the
26 Massachusetts Department of Public Health and SARS-CoV-2 RNA concentrations in wastewater collected
27 by the Massachusetts Water Resources Authority. The estimated lag between the time series of viral loads in
28 wastewater and of reported cases was 10.8 days (95% confidence interval =[10.2, 11.6]) for wastewater
29 treatment plant's northern area and 8.8 days [8.4, 9.1] for southern area. The estimated contribution rate of a
30 reported case to the viral load in wastewater in the 0–19 yr age group was 0.38 [0.35, 0.41] for northern area
31 and 0.40 [0.37, 0.43] for southern area, that in the 80+ yr age group was 0.67 [0.65, 0.69] for northern area
32 and 0.51 [0.49, 0.52] for southern area. The estimated lag between those time series suggested the
33 predictability of reported cases ten days later using viral loads in wastewater. The contribution of a reported
34 case in passive surveillance to the viral load in wastewater differed by age, suggesting a large variation in
35 viral shedding kinetics among age groups.

36

37 **Keywords:** SARS-CoV-2, COVID-19, wastewater, wastewater-based epidemiology, mathematical modeling.

38 Abbreviations: SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; WBE, wastewater-based
39 epidemiology; PMMoV, pepper mild mottle virus

40

41

42 **1. Introduction**

43 Following the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in
44 2019 (1), the epidemic has resulted in a large number of deaths in many countries. The development of
45 effective pharmaceutical interventions for eradication is ongoing, and non-pharmaceutical interventions (i.e.,
46 restriction of movement) are still required to reduce the risk of mortality. For the latter to be effective, an
47 understanding of the current state of the epidemic is needed. In infectious disease epidemiology, the time
48 series of the number of confirmed cases is monitored (so-called passive surveillance) and used to understand
49 the current state of an epidemic. However, in the context of the COVID-19 pandemic, such data have large
50 biases because the testing frequency and propensity are not constant over time, so the sampling process is
51 non-random (2). For instance, the number of confirmed cases strongly depends on the number of people who
52 recognize their infection or suspicious contact with infected individuals. SARS-CoV-2 is known to cause
53 asymptomatic infections and the rate of symptomatic infection depends on age (3, 4). These characteristics
54 make it more difficult to estimate the true number of incidences from the number of confirmed cases,
55 requiring an alternative approach to identifying these potentially unreported transmissions.

56 In this context, WBE has drawn attention from scientists (5) and public health authorities (6, 7).
57 Even during the initial phase of the epidemic, SARS-CoV-2 RNA detection in wastewater was reported in
58 various countries such as Australia (8), Japan (9), the United States (10), and many countries in Europe (11,
59 12). The Netherlands has already implemented a nationwide wastewater monitoring system (13), and more
60 than 40 states in the United States have utilized WBE (14, 15). While this has potential as an early warning
61 system, considerable challenges remain (16). For instance, although several studies have conducted lagged
62 correlation analysis (17-19), there exists no explicit modeling method that can relate the observed viral load
63 in wastewater to the number of infected individuals.

64 To capture trends in ongoing transmissions from the viral load in wastewater, it is essential to
65 understand their quantitative relationship. While the detection rate by passive surveillance depends strongly
66 on age, resulting in a bias in the number of confirmed cases, in principle the viral load in wastewater reflects
67 actual infection levels in communities as long as the wastewater monitoring methods (i.e., condition of
68 influent samples, timing of wastewater sampling, virus detection protocols, etc.) are consistent over time.

69 Therefore, a comparison between passive surveillance and wastewater surveillance could be expected to
70 quantitatively measure age-dependent biases in passive surveillance.

71 The present study aimed to estimate the age-dependent association between the number of reported
72 cases in passive surveillance and the viral load in wastewater to compare passive and wastewater
73 surveillance methods. We constructed a statistical model describing the association between incidence per
74 age group and the viral load in wastewater, then estimated the lag of detection timing between passive
75 surveillance and wastewater surveillance and age dependency in the contribution rate of incidences to the
76 viral load in wastewater.

77 **2. Methods**

78 **2.1. Data collection**

79 Daily case count data were collected from situation reports produced by the Massachusetts
80 Department of Public Health (20). The definition of a “confirmed” case was a patient who met confirmatory
81 laboratory criteria (i.e., detection of SARS-CoV-2 RNA in a clinical or autopsy specimen using a molecular
82 amplification test) (21). Since age-specific data were available only until August 11, 2020, we analyzed the
83 case series from March 22 through that date.

84 For wastewater, we used data from the Massachusetts Water Resources Authority, in which SARS-
85 CoV-2 RNA concentrations were measured by Biobot Analytics (22). The wastewater treatment facility used
86 has two major influent streams (“northern” and “southern”), and wastewater samplings were conducted 3–7
87 times per week for each influent stream. The concentration of SARS-CoV-2 was normalized to the
88 concentration of a human fecal indicator (PMMoV) in order to compensate for variations in fecal load in
89 wastewater, which can be affected by factors such as fecal excretion patterns and dilution. Details of the
90 sampling and molecular methods were described in a previous study (19).

91 **2.2. Model**

92 To understand the heterogeneity in the contribution of incidence to the viral load in wastewater by
93 age, age-group-specific incidences were compared to the viral load in wastewater. Assuming that the lag

94 between the detection timing from wastewater and the reporting by passive surveillance is similar between
95 age groups, the viral load in wastewater at time t , $w(t)$, can be written as:

$$96 \quad w(t) = \sum_a k_a i_a(t - \tau), \quad (1)$$

97 where τ denotes the lag between the detection timing from wastewater and reporting by passive surveillance,
98 i_a denotes the incidence in age group a , and k_a denotes the contribution rate of an incidence in age group a to
99 the viral load in wastewater. This model was fitted to the data for viral load in wastewater $w_{data}(t)$ by
100 maximizing the likelihood function considering the Poisson sampling process, L :

$$101 \quad L = \prod_t \text{pmf}(\text{Poisson}(w(t)), w_{data}(t)). \quad (2)$$

102 We then calculated profile likelihood-based confidence intervals for each parameter. In the model, $i_a(t)$ was
103 defined as the change in continuous time, but the sampling times of real data were discrete. To obtain this
104 variable in continuous time, we determined the two-week moving average of the reported cases from passive
105 surveillance by cubic spline interpolation.

106 To understand the heterogeneity in the contribution of age-group-specific reported cases, we
107 conducted a multiple regression analysis between the number of reported cases by passive surveillance and
108 the viral load in wastewater. The coefficients in the model, k_a s, were compared to evaluate the contribution
109 of each age group. The relative evaluation of coefficients in multiple regression is biased by multicollinearity
110 if the explanatory variables (in our study, the number of reported cases per age group) are strongly correlated.
111 To reduce this bias, we calculated the Pearson correlation coefficients between pairs of explanatory variables
112 in the model (i_a in Equation (1)). If a correlation coefficient > 0.7 was found, one of two explanatory
113 variables was excluded until all correlation coefficients were < 0.7 . All of the above computations were
114 implemented in Mathematica ver. 12.0.0.0.

115 **3. Results and discussion**

116 **3.1. Multicollinearity in the regression model of viral load in wastewater**

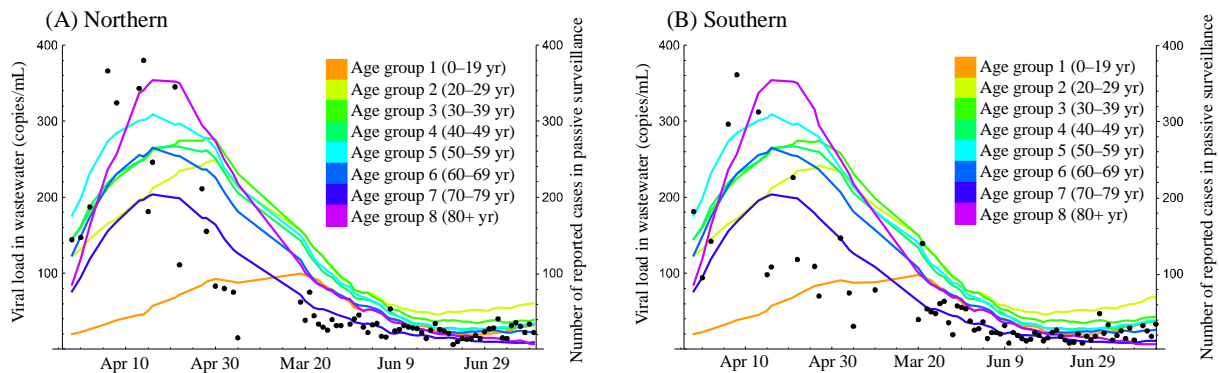
117 Strong correlations in the number of reported cases by passive surveillance between age groups were
118 found (Table 1), which expected a large bias in the multiple regression analysis. As shown in Fig. 1, the
119 time-series trends in the number of reported cases between the close age groups were similar. This similarity

120 can be considered to induce multicollinearity in the regression model. The time series of reported cases by
 121 passive surveillance in age groups 0–19 years and 80+ years were chosen as the explanatory variables to
 122 satisfy the criteria. In the selected model, all correlation coefficients between explanatory variables were <
 123 0.7.

124 **Table 1**
 125 Pearson correlation coefficients of the time series of reported cases by passive surveillance between different
 126 age group pairs.
 127

	Age group (yr)						
	20–29	30–39	40–49	50–59	60–69	70–79	80+
0–19	0.78	0.71	0.67	0.63	0.60	0.45	0.44
20–29		0.96	0.94	0.94	0.91	0.77	0.70
30–39			0.99	0.97	0.97	0.87	0.82
Age group 40–49				0.98	0.98	0.89	0.84
50–59					0.98	0.89	0.83
60–69						0.93	0.88
70–79							0.97

128
 129



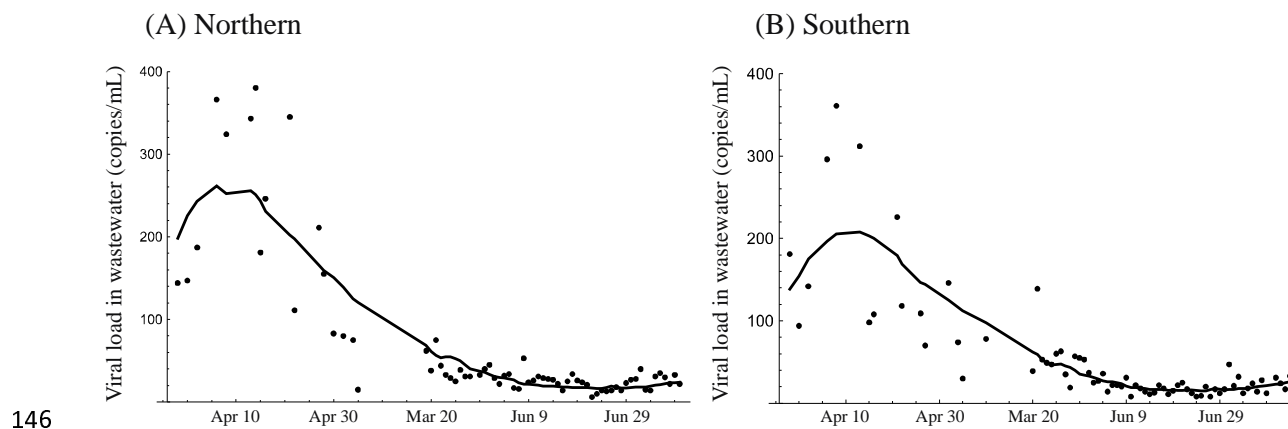
130

131 **Fig. 1.** Comparison between viral load in wastewater and the number of reported cases per age group in
 132 passive surveillance. Dots show viral load in wastewater sampled in the treatment plant's (A) northern and
 133 (B) southern area. Solid lines show the two-week moving average of the number of reported cases per age
 134 group in passive surveillance, with color representing age group.
 135

136 3.2. Model fittings

137 We used the model described in Equation (1) to estimate the lag between the detection timing from
 138 wastewater and the reporting from passive surveillance, τ , and the contribution rate of an incidence in age
 139 group a to the viral load in wastewater, k_a . The maximum likelihood estimates of τ were 10.8 days (95%

140 confidence interval [95%CI] = [10.2, 11.6]) for the northern area and 8.8 days (95%CI = [8.4, 9.1]) for the
141 southern. The maximum likelihood estimates of k_a for the age group 0–19 yr were 0.38 (95%CI = [0.35,
142 0.41]) for the northern area and 0.40 [95%CI = [0.37, 0.43]) for the southern. The maximum likelihood
143 estimates of k_a for the age group 80+ yr were 0.67 (95%CI = [0.65, 0.69]) for the northern area and 0.51
144 (95%CI = [0.49, 0.52]) for the southern. The model's estimated parameters reasonably explained the data for
145 viral load in wastewater sampled in both the northern and southern areas (Fig. 2).



147 **Fig. 2.** Model fittings to the viral load in wastewater sampled in the (A) northern and (B) southern areas.
148 Solid lines show the model prediction and dots show the viral load in wastewater.

149 Our estimated length of the lag between the detection timing from wastewater and the reporting by
150 passive surveillance suggested that the former can occur earlier than the latter, agreeing with a previous
151 study (23). Compared to the incubation period of SARS-CoV-2 (24-26), our estimate of τ suggests that the
152 virus is discharged from the body to sewers at early stages of infection. These results also suggest the
153 potential success of early outbreak detection using wastewater. The estimated lag length was similar between
154 both areas. The lag length was related to the natural history of the pathogen (i.e., the timing of virus
155 discharge from body to sewer, and the timing of symptom onset, which triggers testing in passive
156 surveillance) and testing capacity. These factors were likely similar between both areas because of
157 geographical proximity.

158 The contribution of reported cases from the passive survey in age group 80+ yr to the viral load in
159 wastewater was larger than that of 0–19 yr, a trend that was similar between both areas. This estimate

160 suggests that the amount of virus shedding to wastewater among young individuals was smaller than that
161 among elderly individuals. Furthermore, since the rate of asymptomatic infection among young individuals is
162 higher than that among the elderly (4), the amount of virus shedding to wastewater among individuals with
163 asymptomatic infection may be smaller than those with symptomatic infection. The number of confirmed
164 cases from passive surveillance largely depends on the rate of symptomatic infection. If the amount of virus
165 shedding to wastewater depends on the severity of symptoms, wastewater surveillance will also be affected
166 by the rate of symptomatic infection.

167 The goodness of fit was better for the decreasing parts of the time series (Fig. 2). This might be
168 simply because the beginning of the time series contained too much uncertainty, but another possible
169 interpretation is that the estimability of parameters is ensured by an observation period that contains rich
170 information on the clearance of viruses from infected individuals. In such a case, the difference in shedding
171 kinetics among age groups determines the temporal changes in the viral load in wastewater. Although
172 previous studies have investigated the duration of shedding or its time course with mainly symptomatic
173 patient data (27-29), further studies on the determinants of the excretion characteristics of SARS-CoV-2 are
174 needed.

175 **4. Conclusion**

176 The estimated lag between the time series of viral loads in wastewater and of reported cases in
177 passive surveillance suggested the predictability of reported cases ten days later using viral loads in
178 wastewater. Our modeling of the association between passive and wastewater surveillance showed that the
179 contribution of a reported case in passive surveillance to the viral load in wastewater differed by age,
180 emphasizing the importance of understanding age-dependent shedding kinetics. Further insights into the
181 relationship between the amount of virus discharge to wastewater, age, and severity of symptoms would
182 improve the predictive ability of WBE and make it more robust as a surveillance system.

183

184 **Acknowledgment**

185 This work was supported by JST, CREST (Grant Number JPMJCR20H1), JSPS KAKENHI (Grant Number
186 20J00793), Grants-in-Aid for Cross departmental Young Researcher Grants of Hokkaido University, Japan.

187 The authors would like to thank Editage for the English language review.

188

189 **CRedit author contribution statement**

190 Ryosuke Omori: Conceptualization, Methodology, Formal analysis, Writing - Original draft, Funding
191 acquisition, Fuminari Miura: Methodology, Data curation, Writing - Original draft, Masaaki Kitajima:
192 Writing - Original draft, Funding acquisition

193

194 **Declaration of competing interest**

195 The authors declare that they have no conflict of interest.

196

197

198 **References**

- 199 1. Huang CL, Wang YM, Li XW, Ren LL, Zhao JP, Hu Y, et al. Clinical features of patients infected with
200 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
- 201 2. Omori R, Mizumoto K, Chowell G. Changes in testing rates could mask the novel coronavirus
202 disease (COVID-19) growth rate. *Int J Infect Dis*. 2020;94:116-8.
- 203 3. Omori R, Mizumoto K, Nishiura H. Ascertainment rate of novel coronavirus disease (COVID-19) in
204 Japan. *Int J Infect Dis*. 2020;96:673-5.
- 205 4. Kronbichler A, Kresse D, Yoon S, Lee KH, Effenberger M, Shin JI. Asymptomatic patients as a source
206 of COVID-19 infections: A systematic review and meta-analysis. *Int J Infect Dis*. 2020;98:180-6.
- 207 5. Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, et al. Wastewater-Based Epidemiology:
208 Global Collaborative to Maximize Contributions in the Fight Against COVID-19. *Environ Sci Technol*.
209 2020;54(13):7754-7.
- 210 6. COVID-19 WBE Collaborative [Available from: <https://www.covid19wbec.org/>].
- 211 7. Naughton CC, Roman FA, Alvarado AGF, Tariqi AQ, Deeming MA, Bibby K, et al. Show us the Data:
212 Global COVID-19 Wastewater Monitoring Efforts, Equity, and Gaps. *medRxiv*. 2021.
- 213 8. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, et al. First confirmed detection of SARS-
214 CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-
215 19 in the community. *Sci Total Environ*. 2020;728:138764.
- 216 9. Haramoto E, Malla B, Thakali O, Kitajima M. First environmental surveillance for the presence of
217 SARS-CoV-2 RNA in wastewater and river water in Japan. *Sci Total Environ*. 2020;737:140405.
- 218 10. Sherchan SP, Shahin S, Ward LM, Tandukar S, Aw TG, Schmitz B, et al. First detection of SARS-CoV-2
219 RNA in wastewater in North America: A study in Louisiana, USA. *Sci Total Environ*. 2020;743:140621.
- 220 11. La Rosa G, Iaconelli M, Mancini P, Bonanno Ferraro G, Veneri C, Bonadonna L, et al. First detection
221 of SARS-CoV-2 in untreated wastewaters in Italy. *Sci Total Environ*. 2020;736:139652.
- 222 12. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in
223 Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The
224 Netherlands. *Environ Sci Tech Let*. 2020;7(7):511-6.
- 225 13. (RIVM) TDNifPHatE. Coronavirus monitoring in sewage research 2020 [Available from:
226 <https://www.rivm.nl/en/covid-19/sewage>].
- 227 14. Prevention CfDca. National Wastewater Surveillance System (NWSS) [Available from:
228 <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/wastewater-surveillance.html>].
- 229 15. Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, et al. Wastewater Surveillance of SARS-CoV-2
230 across 40 US states. *medRxiv*. 2021.
- 231 16. Zhu Y, Oishi W, Maruo C, Saito M, Chen R, Kitajima M, et al. Early warning of COVID-19 via
232 wastewater-based epidemiology: potential and bottlenecks. *Sci Total Environ*. 2021;767:145124.
- 233 17. Nemudryi A, Nemudraia A, Wiegand T, Surya K, Buyukyoruk M, Cicha C, et al. Temporal Detection
234 and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Rep Med*. 2020;1(6):100098.
- 235 18. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al. Measurement
236 of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat Biotechnol*. 2020;38(10):1164-
237 7.
- 238 19. Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, et al. SARS-CoV-2 titers in wastewater foreshadow
239 dynamics and clinical presentation of new COVID-19 cases. *Medrxiv*. 2020.
- 240 20. Health tMDoP. Archive of COVID-19 cases in Massachusetts [Available from:
241 <https://www.mass.gov/info-details/archive-of-covid-19-cases-in-massachusetts>].
- 242 21. Edpidemiologists CoSaT. Update to the standardized surveillance case definition and national
243 notification for 2019 novel coronavirus disease (COVID-19) [Available from:
244 [https://cdn.ymaws.com/www.cste.org/resource/resmgr/ps/positionstatement2020/Interim-20-ID-
245 02_COVID-19.pdf](https://cdn.ymaws.com/www.cste.org/resource/resmgr/ps/positionstatement2020/Interim-20-ID-02_COVID-19.pdf)].
- 246 22. Authority MWR. Wastewater COVID-19 Tracking [Available from:
247 <https://www.mwra.com/biobot/biobotdata.htm>].

- 248 23. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, et al. Measurement
249 of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology*.
250 2020;38(10):1164-7.
- 251 24. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical characteristics of coronavirus
252 disease 2019 in China. *New England journal of medicine*. 2020;382(18):1708-20.
- 253 25. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of
254 Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020;382(13):1199-207.
- 255 26. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of
256 Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application.
257 *Ann Intern Med*. 2020;172(9):577-82.
- 258 27. Wolfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Muller MA, et al. Virological assessment
259 of hospitalized patients with COVID-2019. *Nature*. 2020;581(7809):465-9.
- 260 28. Han MS, Seong MW, Kim N, Shin S, Cho SI, Park H, et al. Viral RNA Load in Mildly Symptomatic and
261 Asymptomatic Children with COVID-19, Seoul, South Korea. *Emerg Infect Dis*. 2020;26(10):2497-9.
- 262 29. Miura F, Kitajima M, Omori R. Duration of SARS-CoV-2 viral shedding in faeces as a parameter for
263 wastewater-based epidemiology: Re-analysis of patient data using a shedding dynamics model. *Science of*
264 *The Total Environment*. 2021;769:144549.

265
266