Infectious Disease Modelling 9 (2024) 645–656

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Estimating the effective reproduction number of COVID-19 from population-wide wastewater data: An application in Kagawa, Japan

Yuta Okada , Hiroshi Nishiura *

Kyoto University School of Public Health, Yoshida-Konoe, Sakyo, Kyoto, 606-8601, Japan

ARTICLE INFO

Article history:
Received 16 February 2024
Received in revised form 27 March 2024
Accepted 27 March 2024
Available online 3 April 2024
Handling editor: Daihai He

Keywords:
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Wastewater-based epidemiology
Shedding load distribution
Mathematical model

ABSTRACT

Although epidemiological surveillance of COVID-19 has been gradually downgraded globally, the transmission of COVID-19 continues. It is critical to quantify the transmission dynamics of COVID-19 using multiple datasets including wastewater virus concentration data. Herein, we propose a comprehensive method for estimating the effective reproduction number using wastewater data. The wastewater virus concentration data, which were collected twice a week, were analyzed using daily COVID-19 incidence data obtained from Takamatsu, Japan between January 2022 and September 2022. We estimated the shedding load distribution (SLD) as a function of time since the date of infection, using a model employing the delay distribution, which is assumed to follow a gamma distribution, multiplied by a scaling factor. We also examined models that accounted for the temporal smoothness of viral load measurement data. The model that smoothed temporal patterns of viral load was the best fit model (WAIC = 2795.8), which yielded a mean estimated distribution of SLD of 3.46 days (95% CrI: 3.01–3.95 days). Using this SLD, we reconstructed the daily incidence, which enabled computation of the effective reproduction number. Using the best fit posterior draws of parameters directly, or as a prior distribution for subsequent analyses, we first used a model that assumed temporal smoothness of viral load concentrations in wastewater, as well as infection counts by date of infection. In the subsequent approach, we examined models that also incorporated weekly reported case counts as a proxy for weekly incidence reporting. Both approaches enabled estimations of the epidemic curve as well as the effective reproduction number from twice-weekly wastewater viral load data. Adding weekly case count data reduced the uncertainty of the effective reproduction number. We conclude that wastewater data are still a valuable source of information for inferring the transmission dynamics of COVID-19, and that inferential performance is enhanced when those data are combined with weekly incidence data.

* Corresponding author.
E-mail addresses: okada.yuta.4y@kyoto-u.ac.jp (Y. Okada), nishiura.hiroshi.5r@kyoto-u.ac.jp (H. Nishiura).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.03.006

Abbreviations

LOD	limit of detection
LOOIC	Leave One Out Information Criterion
LOQ	limit of quantification
MCMC	Markov Chain Monte Carlo
PHSM	public health and social measures
PMMoV	pepper mild mottle virus
SLD	shedding load distribution
WAIC	Widely Applicable Information Criterion

1. Introduction

Since the emergence of COVID-19 in Wuhan, China in 2019, the world has continued to suffer from the morbidity and mortality associated with the disease, even after a series of vaccination campaigns that began in late 2020. Despite the ongoing impact of COVID-19, efforts to contain it via public health and social measures (PHSM) and close monitoring of the epidemic dynamics have been downgraded at both regional and national levels. In Japan, universal case reporting of COVID-19 cases under Infectious Disease Law was ceased on May 8th, 2023. Subsequently, only data from the weekly surveillance of confirmed cases by sentinel healthcare facilities has been available. The sentinel surveillance system involves a reporting delay of approximately 10 days, which hampers tracking of the actual size of the epidemic.

With the diminishing effort with regard to COVID-19 tracking by epidemiological surveillance, indicators that facilitate real-time understanding of the epidemic dynamics have inevitable been deployed since the early stage of the pandemic, especially for real-time forecasting of the societal and public health burden. In this context, wastewater surveillance could fill the gap between the public health need for COVID-19 transmission monitoring and the diminishing effort spent on data collection. Based on the knowledge that SARS-CoV-2 is shed in the gastrointestinal tract (Vaselli et al., 2021; D. Wang et al., 2020; W. Wang et al., 2020a; Wu et al., 2020; Xu et al., 2020), its usefulness as a leading epidemiological indicator and as a tool for estimating the actual transmission dynamics of the epidemic have been demonstrated in published studies (Fernandez-Cassi et al., 2021; Huisman et al., 2022; Kitajima et al., 2022; Mattei et al., 2023; Peccia et al., 2020; Peng et al., 2023). However, the relationship between wastewater viral load and epidemic dynamics is not consistent and can even vary depending on the epidemiological settings of transmission. Such variability may be attributable to the different systems used to measure viral load and to the differences in sewage systems. It may also be due to variation in biological characteristics (such as the volume of gastrointestinal viral shedding or the viral strain) and to meteorological or environmental factors that may induce random effects, including the dilution or degeneration of viruses in the sewage system (Wade et al., 2022).

In practice, another problem with wastewater surveillance is the difficulty in establishing routine daily observation owing to limited financial and human resources (Ando et al., 2022; LaTurner et al., 2021). Therefore, the usefulness of low-frequency wastewater sampling is a critical issue with regard to epidemiological investigations. Huisman et al. (2022) demonstrated that a frequency of less than three times a week may result in unreliable estimates of the effective reproduction number based on wastewater surveillance data alone. Whether their finding also holds in other settings is of interest, and any methodology that enables better quantification of the transmission dynamics with even lower sampling frequency is a critical subject for exploration.

Motivated by the factors described above, we herein propose a modelling framework for estimating the relationship between infection counts by date of infection and viral concentrations in wastewater. We also describe the reconstruction of the epidemic dynamics of COVID-19 using wastewater viral concentration data. The estimations are attained with or without weekly incidence data. To demonstrate the practical usefulness of our method, we applied our model to COVID-19 case count data and wastewater viral concentration data from Takamatsu city, Kagawa Prefecture, Japan.

2. Materials and methods

2.1. Epidemiological data

2.1.1. Viral load concentration of SARS-CoV-2 in wastewater

We obtained the wastewater viral load concentration data from Takamatsu City East Sewage Treatment Plant in Takamatsu city (Kagawa Prefectural Government, 2023). We selected Takamatsu city, which is the capital city of Kagawa Prefecture because the relationship between its wastewater drainage basin and its residential population is consistent. The wastewater viral load concentration was continuously reported twice a week on average from October 2021 to September 2022 (Kagawa Prefectural Government, 2023). The measurements were obtained according to the EPISENS-S method (Ando et al., 2022), and

the viral concentrations of SARS-CoV-2 in the wastewater were publicly reported in units of copies per liter (copies/L), with a limit for detection of 93 copies/L and a limit for quantification of 463 copies/L. The same viral load concentration data for July 4th to September 26th, 2022, are also partly available from the website of the Cabinet Secretariat of the Japanese Government (Cabinet Secretariat of the Japanese Government, 2023), which also includes data on the concentration of pepper mild mottle virus (PMMoV) RNA. However, because the PMMoV data were only available for approximately 4 months, we opted not to use them. We also limited our analysis on wastewater data from the Kagawa prefectural government's website from January 2022 to September 2022, and excluded data from 2021 to focus on analysis of the time period during which the Omicron variant (B.1.1.529) was dominant and widespread in the population.

Some viral concentration values were "below the limit of quantification (LOQ)" or "below the limit of detection (LOD)." To overcome this issue, we used arbitrary values of 463 or 93 copies/L, respectively. (Using either a half or a quarter of these values did not significantly affect the subsequent estimations. See Supplementary Table 2 for details).

2.1.2. Confirmed COVID-19 incidence data

We used the daily incidence data of confirmed COVID-19 cases registered in the Health Center Real-Time Information-sharing System on COVID-19 (HER-SYS) (Ministry of Health Labour and Welfare of Japan., 2022). The HER-SYS data were chosen because of its daily availability, in contrast to the published infection count data and wastewater data announced by Kagawa Prefecture (Kagawa Prefectural Government, 2023), which were only available for days on which wastewater viral measurements were obtained.

2.1.3. Data for estimation of the time delay from symptom onset to reporting of COVID-19

The daily COVID-19 incidence data from HER-SYS consists of incidence counts by date of report, but did not include information on the date of onset. To estimate the delay from symptom onset to the time of reporting, we used the incidence data provided by the Tokyo Metropolitan government, which are publicly available and include symptom onset dates for symptomatic COVID-19 cases (Bureau of Social Welfare and Public Health: Tokyo Metropolitan Government, 2023). We extracted all COVID-19 case records reported in Tokyo from January 1st, 2022 to September 26th, 2022 that included onset date information. This information enabled us to make statistical estimations of the time lag from the date of onset to the date of reporting.

2.1.4. Incubation period and generation interval of COVID-19

We retrieved the incubation period data from a publicized preliminary report on SARS-CoV-2 Omicron subvariants (National Institute of Infectious Diseases, 2022). We assumed that the generation interval of COVID-19 caused by Omicron subvariants remained almost unchanged throughout 2022, and follows the report that was originally shared during the early period of the spread of the Omicron variant (Alex Selby, 2022).

2.2. Models for estimating COVID-19 dynamics using wastewater viral load data

2.2.1. Quantifying the lag from infection to report in COVID-19 incidence data

First, we estimated the time lag from infection to report using the COVID-19 incidence data from Tokyo because actual data on Takamatsu were unavailable. Let this dataset from Tokyo be D with information on date of report, $t_{i,report}$, and date of onset $t_{i,onset}$ for case $i \in D$. Let the cumulative distribution function of a reporting delay of up to τ days after onset be $CDF_{O\to R}(\tau|\theta)$, where θ is the set of parameters that define $CDF_{O\to R}$. Then, the likelihood of observing the data on the report and onset dates for all individuals in the dataset is described as follows:

$$L(\theta|D) = \prod_{i\in D} \frac{CDF_{O\to R}(t_{i,report} - t_{i,onset} + 1|\theta) - CDF_{O\to R}(t_{i,report} - t_{i,onset}|\theta)}{CDF_{O\to R}(T - t_{i,onset}|\theta)} \tag{1}$$

where T is the final date of data D, which is September 26th, 2022 in our application. We assumed that the cumulative distribution function $CDF_{O\to R}$ followed the Weibull distribution and, employing a maximum likelihood method, we obtained the point estimate of θ, i.e., $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta})$ where $\widehat{\alpha}$ and $\widehat{\beta}$ are the point estimates of the shape and scale parameters, respectively, that define the Weibull distribution of interest. (See Supplementary Methods for details).

Regarding the incubation period, we used the probability distribution function over time after exposure derived from the empirical data obtained from the National Institute of Infectious Diseases (2022). Combining the incubation period data with the estimated delay from onset to reporting, the relationship between the daily reported incidence, $Y(t)$, and the incidence by infection date, $C(t)$, can be described as follows:

$$Y(t) = \sum_{\mathrm{T}=0}^{\infty} C(t-\mathrm{T}) \sum_{\tau=0}^{\mathrm{T}} p_{incubation}(\tau) p_{report}(\mathrm{T}-\tau), \tag{2}$$

where p_{report} is a discretized probability distribution over time, obtained from $CDF_{O\to R}$ as described above:

$$p_{report}(\mathrm{T}-\tau) = [CDF_{O\to R}(\mathrm{T}-\tau+1|\widehat{\theta}) - CDF_{O\to R}(\mathrm{T}-\tau|\widehat{\theta})] \tag{3}$$

2.2.2. Estimating the relationship between COVID-19 cases by infection date and viral load in wastewater

As a preparation for projecting COVID-19 dynamics from the wastewater results, we first estimated the relationship between case counts by infection data and the viral load in the wastewater, which is also mentioned as "shedding load distribution (SLD)" in published studies (Fernandez-Cassi et al., 2021; Huisman et al., 2022). The model that describes the viral load concentration with the reported case counts in Takamatsu city can be formulated as:

$$Y(t) \sim NegBin(\mu_Y(t), \varphi) \tag{4}$$

$$log[VL(t)] \sim Normal\left(log[\mu_{VL}(t)], \sigma^2_{wastewater}\right), t \in t_{VL\ observation} \tag{5}$$

$$\mu_Y(t) = \sum_{\tau=1}^{\infty} g_{rep}(\tau)\mu_I(t-\tau) \tag{6}$$

$$\mu_{VL}(t) = k\sum_{\tau=1}^{\infty} g_{VL}(\tau)\mu_I(t-\tau) \tag{7}$$

$$log[\mu_{VL}(t+1)] \sim Normal\left(log[\mu_{VL}(t)], \sigma^2_{VL}\right) \tag{8}$$

Here, in Eq. (4), $Y(t)$ represents the daily reported case counts in Takamatsu city, which are assumed to follow a negative binomial distribution with mean $\mu_Y(t)$ and overdispersion parameter φ. In Eq. (5), we modelled the natural logarithm of $VL(t)$, $log[VL(t)]$, as a realization of the natural logarithm of the underlying expected values, $\mu_{VL}(t)$, which follow normal distributions with observation error $\sigma_{wastewater}$. Eq. (6) links the expected values of infections by date of report, $\mu_Y(t)$, to the expected infection counts by date of infection, μ_I, by the convolution of μ_I and g_{rep}, the delay function from infection to reporting as in Eq. (2). In the same manner, the expected values of viral load measurements, $\mu_{VL}(t)$, are linked to μ_I by an SLD consisting of the convolution of μ_I and g_{VL} together with a scaling parameter k. We modelled g_{VL} as a probability distribution function following a Gamma distribution in a time-discretized form that is parameterized by the shape and scale parameters. We also applied temporal smoothing to the natural logarithm of $\mu_{VL}(t)$, $log[\mu_{VL}(t)]$, as in Eq. (8) to account for the sparsity of viral load measurement data and the inherent volatility of the measurements.

The parameters that govern $\mu_Y(t)$, $\mu_I(t)$, $\mu_{VL}(t)$, φ, $\sigma_{wastewater}$, σ_{VL}, k, and the two parameters that give g_{VL} are estimated by the Markov Chain Monte Carlo (MCMC) method, using informative priors for some parameters to be estimated (See Supplementary Methods for details.) We excluded the first 14 days of reported case counts for the likelihood calculation because possible infectors that were infected before the start of our data were not included or considered in our analysis.

Furthermore, we tested models that incorporate

$$log[\mu_I(t+1)] \sim Normal\left(log[\mu_I(t)], \sigma^2_I\right) \tag{9}$$

which, in the same manner as for μ_{VL}, introduces temporal smoothness of expected infections by date of infection, and selected the best fit model using Widely Applicable Information Criterion (WAIC) and Leave One Out Information Criterion (LOOIC).

Through the aforementioned estimation process, the effective reproduction number as a function of time t (day), R_t, was obtained using the renewal equation:

$$R_t = \frac{\mu_I(t)}{\sum_{\tau=1}^{\infty} g_{GI}(\tau)\mu_I(t-\tau)} \tag{10}$$

where g_{GI} is the generation interval as a function of time since the infection date (Fraser, 2007; Nishiura, 2007; Nishiura & Chowell, 2009).

2.2.3. Reconstruction of the epidemic dynamics from the viral load data

Using parameter sets from the best fit model obtained as previously described, we utilized the resulting SLD to reconstruct an epidemic curve and the reproduction number using two approaches based on the models described by Eqs. (5) and (7)–(9).

2.2.3.1. Approach 1. In the first approach, we used 1000 pairs of posterior draws $\sigma_{wastewater}$, σ_{VL}, k, and the shape and scale parameters of g_{VL} given by Gamma distribution from a single chain of the estimated model, as described in the previous section. (For computational reasons, we used the 1000 pairs of parameters from the first chain, instead of the whole 4000 pairs of parameters). We estimated the parameter σ_I in every run. Thus, μ_I was estimated by the MCMC method using each of the 1000 pairs of parameters, resulting in 1,000,000 trajectories of μ_I.

We also performed a similar analysis with the addition of once-a-week proxy weekly aggregate data of infection case counts throughout the period according to the following equation:

$$Y_{weekly}(t) \sim NegBin(\mu_Y(t), \varphi) \tag{15}$$

2.2.3.2. Approach 2. Rather than directly applying posterior draws of parameters from the SLD estimation process as in Approach 1, in the second approach we fitted a multivariate normal distribution to the posterior draw of the natural logarithm of pairs of parameters $\log(shape)$ (of Gamma distribution), $\log(scale)$ (of Gamma distribution), $\log(k)$, $\log(\sigma_{wastewater})$, and $\log(\sigma_{VL})$ using the R package "mclust". Using this fitted multivariate normal distribution as priors in the following estimation (See Supplementary Methods), we estimated μ_I and, consequently, the effective reproduction number, R_t, in the same model as in Approach 1. Note that, contrary to Approach 1, posterior draws were used not as given data but as priors, which enabled the re-estimation of the parameters of interest.

For both approaches, we evaluated the scale ratio of expected cases to reported case count throughout the study period as follows:

$$\rho = \frac{\sum_t \mu_Y(t)}{\sum_t Y(t)} \tag{16}$$

where a value of ρ closer to 1 means a closer fit to the observed epidemic scale.

2.3. Software

All analyses were conducted in R (version 4.2.2) and Stan via CmdStan (version 2.31.0). (CmdStanR, 2023; R Core Team, 2023; Stan Development Team, 2023). For all the estimations described above, we obtained 1000 MCMC samples from each of four chains, with 500 warm-up iterations, which were omitted from the results. We confirmed that all the chains converged by an R-hat function value of below 1.01, and that there were no divergent chains.

3. Results

3.1. Descriptive analysis of COVID-19 case counts and viral load concentration in wastewater

Fig. 1 shows the time series of daily reported incidence of COVID-19 (gray dots) in Takamatsu City, Kagawa, and the viral load measurement data (copies/L) that are available twice a week (red). For comparative purposes, we plotted a blue line representing the estimated case counts by date of infection using the R package "surveillance" function (Salmon et al., 2016). Note that the surge in reported case counts in January 2022 and June 2022 accord with the surge in viral load concentration in the wastewater. However, interpreting the correlation between these two different time series is difficult owing to the sparsity and volatility of the wastewater viral concentration data.

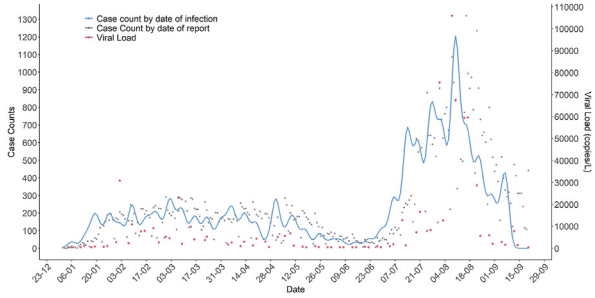

Fig. 1. Time series plot of COVID-19 case counts and wastewater viral load concentration data from Jan to Sept 2022, Takamatsu City, Kagawa, Japan.
Daily reported case counts are represented by gray dots, whereas the blue line shows the estimated case counts by date of infection using the R package "surveillance" function (Salmon et al., 2016). The red dots represent the viral load concentration time series data from the wastewater samples obtained twice a week.

3.2. Estimation of the relationship between wastewater viral concentration and case counts

To quantify and visualize the SLD, which represents the relationship between wastewater viral concentration and COVID-19 case counts by date of infection, we performed statistical analysis on the models as described by Eqs. (4)–(9). Based on WAIC and LOOIC (See Supplementary Table 1), we chose the model that accounts for temporal smoothing for expected viral load concentration values, μ_{VL}, as described in Eq. (8). A summary of parameters from the best fit models is shown in Table 1, and the Gamma distribution characterizing g_{VL} is shown in Fig. 2, with a mean of 3.66 days (95% credible interval (CrI): 3.28–4.05 days).

Fig. 3 presents the time series estimates for COVID-19 case counts by infection date (A), the effective reproduction number (B), and the expected viral load concentration in the wastewater (C). The first and last 14 days are shaded in light blue because, owing to the nature of our model, underestimation of either the infector population or the infected population occurs in the initial or final periods. In panel (A), the expected infection counts by date of infection, μ_I (with 95% CrIs), are compared with the reported daily infection counts (red dots). Panel (B) is a plot showing the variation in the effective reproduction number R_t (with 95% CrIs) over time. It captures the COVID-19 dynamics by date of infection based on the expected infection counts by date of infection in (A). In panel (C), the expected viral load concentration in the wastewater, μ_{VL}, is compared with the measured SARS-CoV-2 viral concentration in the wastewater (red dots). Although the trends in measured and estimated viral load concentrations agreed well, there were apparent discrepancies between these two time series when the numbers of cases were not sharply increasing or decreasing. These discrepancies may be attributable to random effects, including measurement errors, and the plot of μ_{VL} in (C) demonstrates how our model smoothed out these effects.

3.3. Reconstruction of COVID-19 epidemic curve from the wastewater data

Fig. 4 shows the results from the reconstruction of the COVID-19 infection counts by date of infection. In each plot, the red, green, and blue shading represent 50%, 80%, and 95% credible intervals, respectively. Panel (A) represents the results from Approach 1 without weekly reported case counts. Panel (B) shows the results with weekly counts. A comparison of the two reveals that the scale of the epidemic curve is better estimated with weekly case count data, although the overall time trends are similar. Similarly, the results from Approach 2 (C) without and (D) with weekly case count data are shown side by side in the second row of Fig. 4. (See also Supplementary Figure on estimates of viral concentration from the same estimation process, which also reveals scaling issues, as discussed here).

As is visibly suggested by Fig. 4, in both approaches the uncertainty diminishes when proxy weekly case count data are used in the estimation process. Also, it can be observed that the overestimation of the epidemic scale around the epidemic peak of Aug 2022 was mitigated by using proxy weekly case counts. This can also be seen in Table 2, which shows the estimated total infection counts relative to the reported total case counts throughout the study period. The uncertainty of $\mu_I(t)$ was smaller when proxy weekly case data were used, especially when Approach 2 was employed (Fig. 4). Moreover, the scale ratio shown in Table 2 was smallest when the same model was used.

Fig. 5 shows the effective reproduction numbers. They were estimated using the same process adopted to produce Fig. 4. The estimated R_t values from the case data alone agreed well with the wastewater-based R_t estimates. In both Approach 1 (first row, panels A and B) and Approach 2 (second row, panels C and D), the trends from the estimation with or without weekly case counts are generally in line with each other.

4. Discussion

Herein, we propose a general mathematical and statistical framework for the reconstruction of the transmission dynamics of COVID-19 based on twice-weekly wastewater sampling at a population level. Our framework can be easily applied to locations where the basin for sampled wastewater and the residential population of interest are clearly linked. This was of key importance to our study. The first step of this framework enables the estimation of SLD, i.e., the time course of SARS-CoV-2

Table 1
Parameters that describe the relationship between COVID-19 infection and SARS-CoV-2 viral concentration in wastewater.

Parameters	Median	95% Credible Interval	
		Lower	Upper
Gamma Mean	3.46	3.01	3.95
Gamma SD	3.00	2.65	3.40
k	18.8	15.2	23.6
φ	3.77	3.21	4.38
$\sigma_{wastewater}$	0.935	0.804	1.11
σ_{VL}	0.116	0.104	0.130

The parameters of Gamma distribution that give g_{VL} in Eq. (7), as well as the scaling parameter k, the overdispersion parameter φ, the observation error term $\sigma_{wastewater}$, the standard deviation for gaussian smoothing for the natural logarithm of viral load concentration, σ_{VL}, are shown with 95% credible intervals derived from posterior distribution.

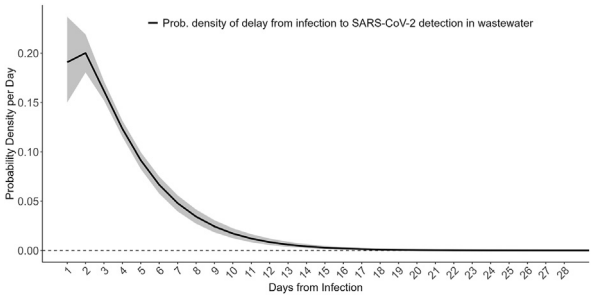

Fig. 2. Probability density of time from infection to SARS-CoV-2 detection in the wastewater.
The black solid line represents the median of the delay distribution as a probability function of the time since the date of infection; the gray shading indicates the 95% credible interval.

concentration in wastewater from the date of infection together with scaling factors from infection counts to viral concentration in the wastewater. Using the SLD estimates from the first step, we also demonstrated that the estimated SLD can be used to reconstruct the COVID-19 dynamics based on twice-weekly wastewater viral concentration measurements. This is very encouraging because it suggests that the scale of the COVID-19 epidemic can be inferred even without individual-based surveillance data that stem from universal case reporting.

Our study demonstrated that the SLD can be estimated in a similar manner to that reported by Mattei et al. (2023), in which several factors such as degradation, dilution, and volume of viral excretion into feces are summarized by a single scaling factor. This approach differs from that reported by Huisman et al. (2022), which follows a bottom-up approach based on viral shedding data from previous studies with some assumptions about viral measurements. We chose a comparatively simple approach that considers the variability and uncertainty of biological, environmental, and technical factors regarding viral excretion and measurement in wastewater, as discussed elsewhere (Ahmed et al., 2020; Cheung et al., 2020; Mattei et al., 2023; W. Wang et al., 2020b; Wölfel et al., 2020). Those studies that handled such variability and uncertainty were applied to SARS-CoV-2 variants prior to the emergence of Omicron variant B.1.1.529.

In addition to the issue of SLD, the consideration of measurement errors regarding SARS-CoV-2 concentrations in wastewater is a key advancement proposed herein. An estimation of SLD cannot avoid the issue of random effects on viral concentration data due to various factors including measurement errors (Wade et al., 2022). Therefore, we combined the SLD with equations that account for such random factors, which were partly taken into account by Peccia et al. (2020). We trust that our framework adds new statistical insight into the use of SLD to a series of published studies that have reconstructed COVID-19 dynamics from wastewater data using mathematical models (Fernandez-Cassi et al., 2021; Huisman et al., 2022; Mattei et al., 2023).

Another important finding was that the best fit model for estimating SLD assumed temporal smoothness of viral load concentration data. Owing to the scarcity of available data, the estimation of missing wastewater measurements is also necessary, especially when considering the reconstruction of COVID-19 dynamics based solely on wastewater data. Therefore, we assumed that temporal smoothness poses constraints that may help the estimation of a time series by providing missing values. In addition, some random factors other than the measurement of wastewater viral load may not be controlled by Eq. (8). Therefore, these factors may instead be accounted for at least partially by temporal smoothness, which may have improved our estimates of SLD.

The present study demonstrated that SLD estimates enable reconstruction of both the trend and scale of the COVID-19 epidemic curve, based on twice-weekly wastewater sampling. Despite the need for temporal smoothness assumption for infection counts by date of infection, possibly owing to data sparseness, our findings indicate that wastewater sampling will be useful, even when conducted twice weekly. In both approaches, combining the analysis with weekly reported case count data improved the estimation of the epidemic scale, although adding epidemiological (case) data did not drastically alter the effective reproduction number estimates.

Comparing the two approaches in the reconstruction step, the first approach, which aggregates the estimates from each run for all posterior draws from the first step is straightforward but computationally expensive, whereas the second approach can be computed readily. In addition, in terms of the estimated scale of total infection counts, Approach 2 tends not to overestimate the total burden compared with Approach 1. Any decision regarding which of these two approaches is best should be based on future validation. However, considering possible future changes to the virological characteristics of SARS-CoV-2, the latter approach might be more adaptive to future epidemiological data because parameters that provide SLD are re-estimated based on the given data using priors obtained from previous steps.

There are several limitations to our study. First, as mentioned in the method section, we did not normalize the wastewater viral concentration data using environmental viruses such as PMMoV. However, process control methodologies using PMMoV or other viruses such as bovine coronavirus have yet to be established (LaTurner et al., 2021), and we believe that the error term for viral concentration measurements in our model plays a similar role in smoothing out random variability in measurements or environmental factors. Our framework can be readily adapted to wastewater data normalized by environmental

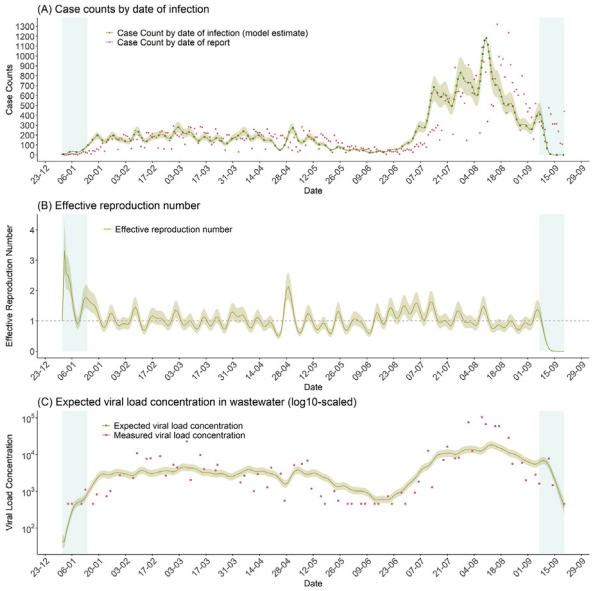

Fig. 3. Estimates of case counts by date of infection, effective reproduction number, and expected viral concentration in wastewater samples obtained between Jan and Sept 2022 from Takamatsu City, Kagawa, Japan.
(A) Estimates of case counts of COVID-19 by infection date. The yellow line (median) and the shading (95% credible interval) represent the estimated infection counts by date of infection, μ_I, which precede the reported daily infection counts (gray dots).
(B) Effective reproduction number R_t values, which were estimated based on Eq. (10), are represented by the yellow line (median) and shading (95% credible interval).
(C) Expected viral load concentration in wastewater (μ_I) values are represented by the yellow line (median) and shading (95% credible interval). For comparison, the measured values of viral load are represented by red dots.
For all panels, the first and last 14 days are represented in light blue because the number of infectors is underestimated in the former period, whereas the number of infected is underestimated in the latter period owing to the nature of our model.

viruses, so future validation is warranted when relevant data become available. Another issue with the wastewater data is that we filled the missing values for measurements below the limit of quantification or detection. However, the arbitrary values that we chose to fill missing values did not significantly affect our estimates of shedding delay function. (See Supplementary Table 2).

The second limitation is the limited statistical power and limited geographical location of the COVID-19 incidence data as well as the frequency of wastewater data used in our study. As of June 1st, 2023, the population of Takamatsu city is 411,876 (Takamatsu City, 2023), which is relatively small compared with other megacities in Japan or in other countries. With regard to measurement frequency, viral concentration data that were measured at different frequencies were examined in a previous study (Huisman et al., 2022). These data limitations may have limited the statistical power of our analysis. In addition, it would have been better if we had been able to test our model on data from different geographical locations because geographical comparisons may lead to the discovery of previously unrecognized technical issues. This is the scope for our future research, especially after efforts to monitor SARS-CoV-2 in wastewater have continued for a substantial period at several locations.

Another limitation with regard to data is the presumable weekly reported case counts used in the reconstruction process. The sentinel surveillance of COVID-19 is not designed to correctly capture the age structure of the epidemic dynamic. Therefore, owing to the nature of their selection, most sentinel hospitals or clinics tend to focus on pediatric healthcare. Thus, our results from the examination using the weekly aggregated case count data as a proxy for total infection counts may not be readily applicable to the current official weekly reports of COVID-19 in Japan.

Another limitation is the assumption that the virological characteristics or biological reactions such as gastrointestinal shedding did not significantly change throughout the period. It is not clear whether such factors differ among various sub-variants of Omicron, so our assumption here might have led to incorrect estimates. To take this point into account, further research on the characteristics of Omicron variants as well as more abundant wastewater data in terms of frequency and geography are needed.

Technically, the assumptions of temporal smoothness for viral load concentration or infection counts are key limitations to our study. Smoothing techniques of various forms, such as effective reproduction numbers, are applied to the estimation of epidemic dynamics indicators (Abbott et al., 2020; Parag, 2021). In the present study, we proposed an approach to control temporal variation of viral concentration or infection counts by date of infection. This was not only because such assumptions at least statistically improved the SLD estimation, but also because they were essential for the efficient reconstruction of the COVID-19 dynamics from our wastewater data. Omission of the smoothness assumption should be considered in the future

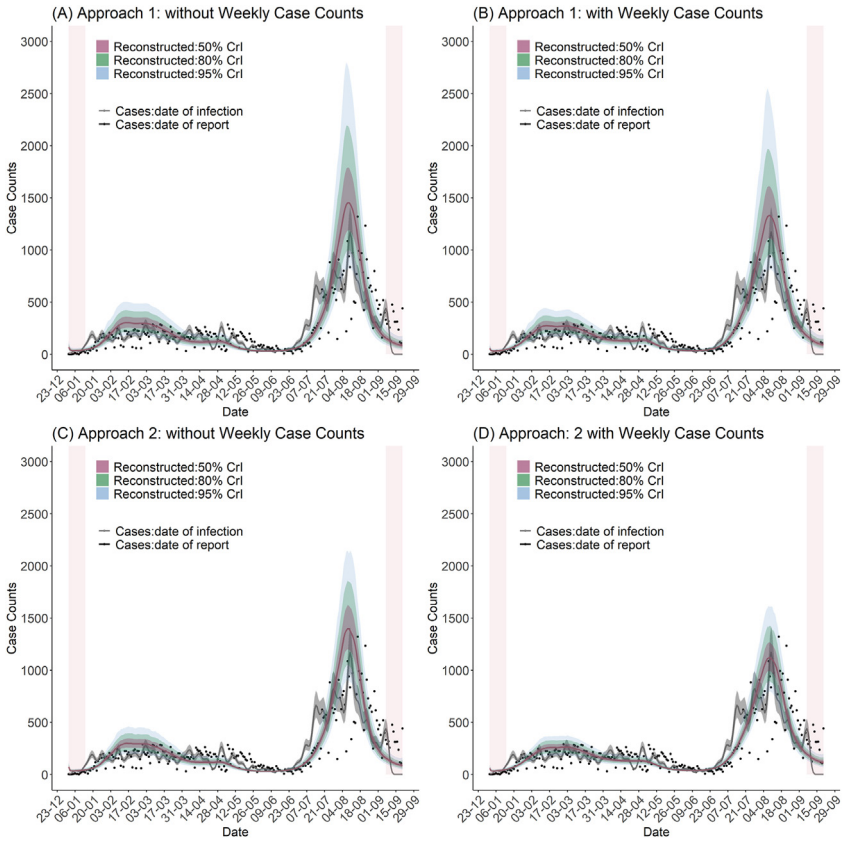

Fig. 4. Reconstruction of the time series of infection counts by date of infection using wastewater viral concentration data.
In each panel, the green line represents the median time series, whereas the red, green, and blue shading indicates 50%, 80%, and 95% credible intervals, respectively. (Gray line with gray shades in each panel represents estimated infection counts by date of infection that is also shown in panel A of Fig. 3.)
(A) Infection counts estimated from the reconstruction model assuming temporal smoothness for the expected viral load concentrations in the wastewater and the infection counts by date of infection, using Approach 1.
(B) Infection counts estimated from the same model as in (A) using weekly case count data.
(C) Infection counts estimated from the reconstruction model assuming temporal smoothness for the expected viral load concentrations in the wastewater and infection counts by date of infection, using Approach 2.
(D) Infection counts estimated from the same model as in (C) using weekly case count data.

when daily wastewater data for designated residential populations are available. In addition, out-of-sample validation is warranted to further legitimize our framework, which will be possible when data from subsequent COVID-19 waves are available.

Even considering the limitations described above, our general framework enhances the usefulness of wastewater surveillance against COVID-19 because it enables estimations of COVID-19 dynamics even in resource-limited scenarios. Our models demonstrated a framework to estimate SLD with consideration for errors in the measurement of wastewater viral concentration data. Our framework also enabled the reconstruction of COVID-19 dynamics from twice-weekly wastewater data with assumptions about temporal smoothness of wastewater viral concentrations and infection counts by date of infection. Our results may encourage public health agencies to increase monitoring of COVID-19 case counts, and to conduct more rigorous wastewater surveillance in the future.

Table 2
Ratio of estimated infection counts to the observed case counts by each reconstruction model of COVID-19 dynamics throughout the study period.

Approach	Use of Weekly Case Count Data	Scale ratio (%)	95% Credible interval	
			Lower	Upper
1	No	111.3	82.58	150.4
	Yes	109.2	88.27	141.0
2	No	108.0	92.08	128.7
	Yes	107.6	94.11	123.7

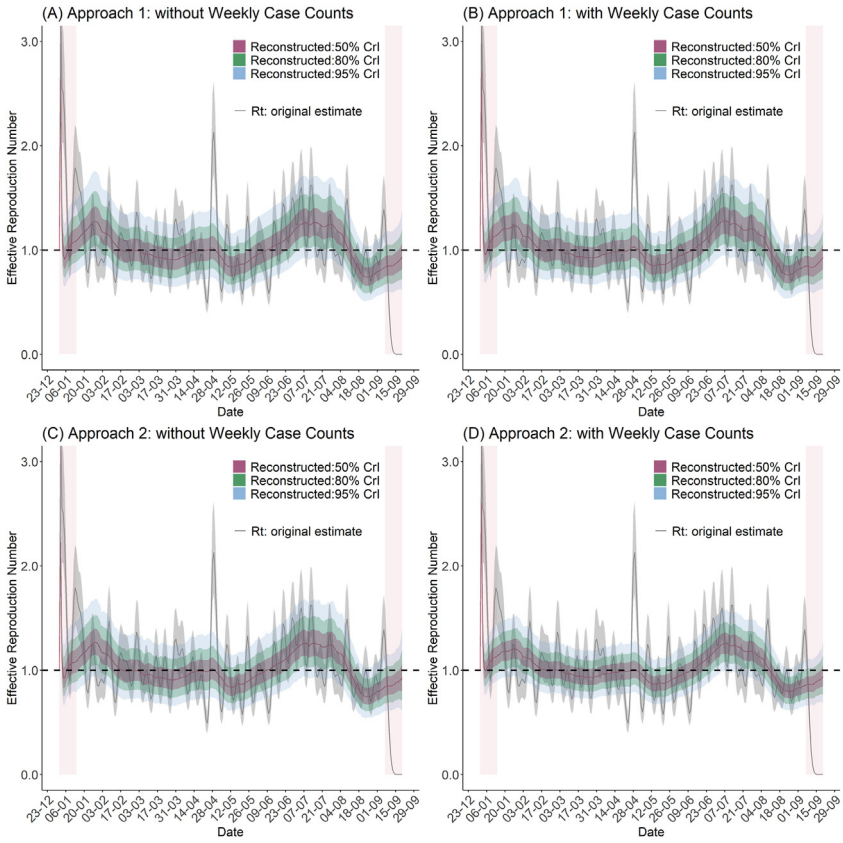

Fig. 5. Reconstructing the effective reproduction number over time by date of infection using wastewater viral concentration data.
In each panel, the green line represents the median time series, whereas the red, green, and blue shading indicates 50%, 80%, and 95% credible intervals, respectively. (Gray line with gray shades in each panel is the same as from panel B in Fig. 3, which is shown for comparison.)
(A) Effective reproduction numbers estimated from the reconstruction model assuming temporal smoothness for the expected viral load concentration in the wastewater and infection counts by date of infection, using Approach 1.
(B) Effective reproduction numbers estimated from the same model as in (A) using weekly case count data.
(C) Effective reproduction numbers estimated from the reconstruction model assuming temporal smoothness for the expected viral load concentration in the wastewater and infection counts by date of infection, using Approach 2.
(D) Effective reproduction numbers estimated from the same model as in (C) using weekly case count data.

Conflicts of interest

All authors declare no conflicts of interest with regard to this paper.

Ethical approval statement

The epidemiological data from the Ministry of Health, Labour and Welfare were received after de-identification. The wastewater viral concentration data are publicly available from the website of Kagawa Prefecture, Japan. Because the present study used only openly accessible data, ethical approval was not required owing to the nature of our study.

Funding sources

Y.O. received funding from the SECOM Science and Technology Foundation, The Kyoto University Foundation, and Fujiwara Memorial Foundation. H.N. received funding from Health and Labour Sciences Research Grants (grant numbers 20CA2024, 21HB1002, 21HA2016, and 23HA2005), the Japan Agency for Medical Research and Development (grant numbers JP23fk0108612 and JP23fk0108685), JSPS KAKENHI (grant numbers 21H03198 and 22K19670), the Environment Research and Technology Development Fund (grant number JPMEERF20S11804) of the Environmental Restoration and Conservation Agency of Japan, Kao Health Science Research, the Daikin GAP Fund of Kyoto University, the Japan Science and Technology Agency SICORP program (grant numbers JPMJSC20U3 and JPMJSC2105), and the RISTEX program for Science, Technology, and

Innovation Policy (grant number JPMJRS22B4). The funders had no role in the study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

CRediT authorship contribution statement

Yuta Okada: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Hiroshi Nishiura:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the local governments, public health centers, and institutes for surveillance, laboratory testing, epidemiological investigation, and data collection. We thank Frank Kitching, MSc from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript. The funders had no role in the study design, data collection and analysis, the decision to publish, or preparation of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2024.03.006.

Relative percentage of total infection counts estimated by each approach in comparison with the total of 62,468 reported infection counts within our study period.

References

Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I., Munday, J. D., Meakin, S., Doughty, E. L., Chun, J. Y., Chan, Y.-W. D., Finger, F., Campbell, P., Endo, A., Pearson, C. A. B., Gimma, A., Russell, T., Flasche, S., Kucharski, A. J., Eggo, R. M., & Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res, 5*, 112. https://doi.org/10.12688/WELLCOMEOPENRES.16006.2

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment, 728*, Article 138764. https://doi.org/10.1016/J.SCITOTENV.2020.138764

Alex Selby. (2022). In *Estimating generation time of Omicron - covid-19* [WWW Document]. URL https://sonorouschocolate.com/covid19/index.php?title=Estimating_Generation_Time_Of_Omicron, 6.21.23.

Ando, H., Iwamoto, R., Kobayashi, H., Okabe, S., & Kitajima, M. (2022). The efficient and practical virus identification system with ENhanced sensitivity for solids (EPISENS-S): A rapid and cost-effective SARS-CoV-2 RNA detection method for routine wastewater surveillance. *Science of The Total Environment, 843*, Article 157101. https://doi.org/10.1016/J.SCITOTENV.2022.157101

Bureau of Social Welfare and Public Health: Tokyo Metropolitan Government. (2023). In *Details of announcement of new coronavirus-positive patients in Tokyo* [WWW Document]. URL https://catalog.data.metro.tokyo.lg.jp/dataset/t000010d0000000068, 6.26.23.

Cabinet Secretariat of the Japanese Government. (2023). In *Wastewater surveillance | Cabinet Secretariat promotion office for novel coronavirus infectious diseases* [WWW Document]. URL https://corona.go.jp/surveillance/, 6.26.23.

Cheung, K. S., Hung, I. F. N., Chan, P. P. Y., Lung, K. C., Tso, E., Liu, R., Ng, Y. Y., Chu, M. Y., Chung, T. W. H., Tam, A. R., Yip, C. C. Y., Leung, K. H., Fung, A. Y. F., Zhang, R. R., Lin, Y., Cheng, H. M., Zhang, A. J. X., To, K. K. W., Chan, K. H., Yuen, K. Y., & Leung, W. K. (2020). Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: Systematic review and meta-analysis. *Gastroenterology, 159*, 81–95. https://doi.org/10.1053/J.GASTRO.2020.03.065

CmdStanR. (2023). In *R interface to CmdStan • cmdstanr* [WWW Document]. URL https://mc-stan.org/cmdstanr/, 6.25.23.

Fernandez-Cassi, X., Scheidegger, A., Bänziger, C., Cariti, F., Tuñas Corzon, A., Ganesanandamoorthy, P., Lemaitre, J. C., Ort, C., Julian, T. R., & Kohn, T. (2021). Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high. *Water Research, 200*, Article 117252. https://doi.org/10.1016/J.WATRES.2021.117252

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One, 2*, e758. https://doi.org/10.1371/JOURNAL.PONE.0000758

Huisman, J. S., Scire, J., Caduff, L., Fernandez-Cassi, X., Ganesanandamoorthy, P., Kull, A., Scheidegger, A., Stachler, E., Boehm, A. B., Hughes, B., Knudson, A., Topol, A., Wigginton, K. R., Wolfe, M. K., Kohn, T., Ort, C., Stadler, T., & Julian, T. R. (2022). Wastewater-based estimation of the effective reproductive number of SARS-CoV-2. *Environmental Health Perspectives, 130*. https://doi.org/10.1289/EHP10050

Kagawa Prefectural Government. (2023). Wastewater epidemiological surveillance. *Kagawa Prefecture* (in Japanese).[WWW Document].URL https://www.pref.kagawa.lg.jp/gesuido/gesuido/ekigaku.html, 6.25.23.

Kitajima, M., Murakami, M., Kadoya, S. S., Ando, H., Kuroita, T., Katayama, H., & Imoto, S. (2022). Association of SARS-CoV-2 load in wastewater with reported COVID-19 cases in the Tokyo 2020 olympic and paralympic village from July to September 2021. *JAMA Network Open, 5*, Article e2226822. https://doi.org/10.1001/JAMANETWORKOPEN.2022.26822

LaTurner, Z. W., Zong, D. M., Kalvapalle, P., Gamas, K. R., Terwilliger, A., Crosby, T., Ali, P., Avadhanula, V., Santos, H. H., Weesner, K., Hopkins, L., Piedra, P. A., Maresso, A. W., & Stadler, L. B. (2021). Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. *Water Research, 197*, Article 117043. https://doi.org/10.1016/J.WATRES.2021.117043

Mattei, M., Pinto, R. M., Guix, S., Bosch, A., & Arenas, A. (2023). Analysis of SARS-CoV-2 in wastewater for prevalence estimation and investigating clinical diagnostic test biases. *Water Research, 120223*. https://doi.org/10.1016/J.WATRES.2023.120223

Ministry of Health Labour and Welfare of Japan. (2022), 2022. In *Report on the number of tests for new coronavirus infection* [WWW Document].URL https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00129.html, 6.19.23.

National Institute of Infectious Diseases. (2022). In *Estimation of incubation period for SARS-CoV-2 mutant B.1.1.529 strain (Omicron strain): Preliminary report* [WWW Document]. URL https://www.niid.go.jp/niid/ja/2019-ncov/2551-cepr/10903-b11529-period.html, 6.19.23.

Nishiura, H. (2007). Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918-19. *Theoretical Biology and Medical Modelling, 4*, 1–9. https://doi.org/10.1186/1742-4682-4-20/FIGURES/5

Nishiura, H., & Chowell, G. (2009). In *The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. Mathematical and Statistical Estimation Approaches in Epidemiology*. https://doi.org/10.1007/978-90-481-2313-1_5/COVER, 103–121.

Parag, K. V. (2021). Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology, 17*, Article e1009347. https://doi.org/10.1371/JOURNAL.PCBI.1009347

Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., Weinberger, D. M., Arnold, W., & Omer, S. B. (2020). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics, 2020 *Nature Biotechnology, 38*, 1164–1167. https://doi.org/10.1038/s41587-020-0684-z, 10 38.

Peng, K. K., Renouf, E. M., Dean, C. B., Hu, X. J., Delatolla, R., & Manuel, D. G. (2023). An exploration of the relationship between wastewater viral signals and COVID-19 hospitalizations in Ottawa, Canada. *Infect Dis Model, 8*, 617–631. https://doi.org/10.1016/J.IDM.2023.05.011

R Core Team. (2023). In *R: The R project for statistical computing [WWW document]*. URL https://www.r-project.org/, 6.25.23.

Salmon, M., Schumacher, D., & Höhle, M. (2016). Monitoring count time series in R: Aberration detection in public health surveillance. *Journal of Statistical Software, 70*, 1–35. https://doi.org/10.18637/JSS.V070.I10

Stan Development Team. (2023). In *Stan modeling language users guide and reference manual, VERSION 2.32* [WWW Document]. URL https://mc-stan.org/, 6. 25.23.

Takamatsu City. (2023). In *[Statistics] estimated population (preliminary population report: Population census base)* [WWW Document]. URL http://www.city.takamatsu.kagawa.jp/kurashi/shinotorikumi/tokei/jinko/suikei/r05.html, 7.3.23.

Vaselli, N. M., Setiabudi, W., Subramaniam, K., Adams, E. R., Turtle, L., Iturriza-Gómara, M., Solomon, T., Cunliffe, N. A., French, N., Hungerford, D., Hungerford, D., Vivancos, R., Gabbay, M., Buchan, I., Carrol, E. D., Gamble, C., Crossley, L., Joseph, N., Wilton, M., … Gkioni, E. (2021). Investigation of SARS-CoV-2 faecal shedding in the community: A prospective household cohort study (COVID-LIV) in the UK. *BMC Infectious Diseases, 21*, 1–9. https://doi.org/10.1186/S12879-021-06443-7/TABLES/2

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., Niemeyer, D., Jones, T. C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirglmaier, K., Drosten, C., & Wendtner, C. (2020). Virological assessment of hospitalized patients with COVID-2019, 2020 *Nature, 581*, 465–469. https://doi.org/10.1038/s41586-020-2196-x, 7809 581.

Wade, M. J., Lo Jacomo, A., Armenise, E., Brown, M. R., Bunce, J. T., Cameron, G. J., Fang, Z., Farkas, K., Gilpin, D. F., Graham, D. W., Grimsley, J. M. S., Hart, A., Hoffmann, T., Jackson, K. J., Jones, D. L., Lilley, C. J., McGrath, J. W., McKinley, J. M., McSparron, C., … Kasprzyk-Hordern, B. (2022). Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *Journal of Hazardous Materials, 424*, Article 127456. https://doi.org/10.1016/J.JHAZMAT.2021.127456

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, China. *JAMA, 323*, 1061–1069. https://doi.org/10.1001/JAMA.2020.1585

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020a). Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA, 323*, 1843–1844. https://doi.org/10.1001/JAMA.2020.3786

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020b). Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA, 323*, 1843–1844. https://doi.org/10.1001/JAMA.2020.3786

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G., & Huang, X. (2020). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet Gastroenterol Hepatol, 5*, 434–435. https://doi.org/10.1016/S2468-1253(20)30083-2

Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., Zhang, H., Liu, H., Xia, H., Tang, J., Zhang, K., & Gong, S. (2020). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding, 2020 *Nature Medicine, 26*, 502–505. https://doi.org/10.1038/s41591-020-0817-4, 4 26.