Bayesian sequential approach to monitor COVID-19 variants through positivity rate from wastewater

J. Cricelio Montesinos-López[1, +], Maria L. Daza–Torres[1, +], Yury E. García[1], César Herrera[2], C. Winston Bess[3], Heather N. Bischel[3], and Miriam Nuño[1,*]

[1] *Department of Public Health Sciences, University of California Davis, California 95616, United States*
[2] *Department of Mathematics, Purdue University, Indiana 47907, United States*
[3] *Department of Civil and Environmental Engineering, University of California Davis, Davis, California 95616, United States*
[+] *co-first author*

[*] *corresponding author: mnuno@ucdavis.edu*

Abstract

Trends in COVID-19 infection have changed throughout the pandemic due to myriad factors, including changes in transmission driven by social behavior, vaccine development and uptake, mutations in the virus genome, and public health policies. Mass testing was an essential control measure for curtailing the burden of COVID-19 and monitoring the magnitude of the pandemic during its multiple phases. However, as the pandemic progressed, new preventive and surveillance mechanisms emerged. Implementing vaccine programs, wastewater (WW) surveillance, and at-home COVID-19 tests reduced the demand for mass severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing. This paper proposes a sequential Bayesian approach to estimate the COVID-19 positivity rate (PR) using SARS-CoV-2 RNA concentrations measured in WW through an adaptive scheme incorporating changes in virus dynamics. PR estimates are used to compute thresholds for WW data using the CDC thresholds for low, substantial, and high transmission. The effective reproductive number estimates are calculated using PR estimates from the WW data. This approach provides insights into the dynamics of the virus evolution and an analytical framework that combines different data sources to continue monitoring the COVID-19 trends. These results can provide public health guidance to reduce the burden of future outbreaks as new variants continue to emerge. The proposed modeling framework was applied to the City of Davis and the campus of the University of California Davis.

Keywords: SARS-CoV-2; COVID–19; Positivity rate; Wastewater-based epidemiology (WBE); Effective reproductive number; Bayesian sequential data assimilation.

1 Introduction

Effectively monitoring the evolution of the COVID-19 pandemic and controlling the spread of disease remains a major public health challenge. The rapid evolution of SARS-CoV-2, the virus that causes COVID-19, and variable community responses to public health interventions involving sustained behavior change, stay-at-home restrictions, face coverings, testing, and vaccination. Statistical and mathematical models are an important component of effective monitoring systems to track COVID-19 cases, hospitalizations, and deaths (Capistran et al., 2021; Giordano et al., 2020). However, classic models in epidemiology are limited. It is necessary to develop monitoring systems that capture the complex dynamics of the evolving pandemic (Daza-Torres et al., 2022; Engbert et al., 2020). Good data availability and quality, which have also changed over time, are essential in mathematical and statistical model development (Jewell et al., 2020; Shinde et al., 2020). Mass testing was an essential

control measure for curtailing the burden of COVID-19, particularly during its early phases. As the pandemic progressed, new interventions and monitoring strategies surged, including vaccine programs, wastewater (WW) surveillance (Ahmed et al., 2020; Huisman et al., 2022), and at-home COVID-19 tests (Rader et al., 2022).

Mass testing campaigns, contact tracing, isolation, and mobility restrictions made it possible to estimate the burden of disease during the early phases of the pandemic (Bassi et al., 2021; Yang et al., 2020). However, the return to "normal" was accompanied by a decrease in COVID-19 clinical testing programs and an increase in at-home diagnostic tests. These changes created limits on the utility of individual case data for public health decision-making. Using the number of confirmed cases to determine the prevalence of disease in a community can provide biased information since observed cases depend on the tests conducted, strategies to sample people, and the timing of case detection (Böttcher et al., 2022; Fenton et al., 2020). Therefore, without more specifications, the number of positive cases may not be an appropriate indicator to monitor disease burden. Positivity rate (PR) has been shown to be a better indicator for disease spread than confirmed cases because it considers both tests conducted and cases detected (Dallal et al., 2021; Montesinos-López et al., 2021).

During the COVID-19 pandemic, public health officials commonly used the PR to infer the adequacy of population-level testing and the extent of SARS-CoV-2 transmission in a population (Dallal et al., 2021; Fenga and Gaspari, 2021; Furuse et al., 2021). A low PR indicated low viral prevalence and a testing program with sufficient surveillance capacity. In contrast, a higher PR suggested a low number of tests being conducted, suggesting that many infected individuals were undetected (Dowdy and D'Souza, 2020; Fernandez-Cassi et al., 2021; Mercer and Salit, 2021).

At the beginning of the pandemic, the World Health Organization (WHO) recommended a PR threshold of 5% to declare COVID-19 transmission under control (WHO, 2020). However, as the number of people being tested for COVID-19 changed over time, the PR alone was determined insufficient to assess community-level transmission. Limited levels of testing meant that public health authorities focused on passive case-finding (i.e., only those considered most likely to be infected due to symptoms or contacts were tested). As a result, PR tended to be artificially high, and models using PR as input tended to overestimate the proportion of people infected. This is contrary to models based only on observed cases, which may underestimate the COVID-19 prevalence (Daza-Torres et al., 2023). Despite these limitations, PR can still provide a reasonable estimate of the extent of an outbreak if PR is combined with additional information.

Public health authorities are turning to wastewater-based epidemiology (WBE) as an alternative strategy for less-biased population-level surveillance of SARS-CoV-2 RNA. WBE uses biomarkers in WW to monitor trends in community-level health indices. WBE methods have been used to detect changes in drug consumption (Zuccato et al., 2005), dietary patterns (Choi et al., 2019), and the circulation of pathogens like poliovirus and norovirus (Asghar et al., 2014). SARS-CoV-2 RNA in WW correlates well with reported cases of COVID-19 (Ai et al., 2021; Daza-Torres et al., 2023; Huisman et al., 2022). However, some studies have shown that the relationship between WW and COVID-19 clinical cases varies over time (D'Aoust et al., 2022; Xiao et al., 2022). This relationship is affected by testing practices and the emergence of new variants (Frampton et al., 2021; Graham et al., 2021; Xiao et al., 2022). The work of (Xiao et al., 2022) showed that increments in the ratio of WW signal and reported daily new positive clinical tests could be attributed to insufficient testing in the general population (i.e., insufficient testing to capture the exponential growth of actual COVID-19 cases). These findings suggest that models assuming observed cases as input will tend to underestimate the prevalence of COVID-19, as was noted in (Daza-Torres et al., 2023).

In this paper, we correlate WW with PR to monitor COVID-19 trends and to help overcome the limitations of relying only on clinical cases. We pose an adaptive scheme to model the non-

autonomous nature of the prolonged COVID-19 pandemic. The PR is modeled through a sequential Bayesian approach (Daza-Torres et al., 2022) with a Beta regression using the WW viral loads as a covariable. The resulting model allows us to compute a PR based on WW measurements and incorporates changes in viral transmission dynamics through an adaptive scheme. The PR estimate is used to calculate values for WW data that correspond to PR thresholds using criteria proposed by the United States Centers for Disease Control (CDC) in 2021 Christie et al. (2021). The PR thresholds indicated low transmission for a $PR \leq 0.05$, substantial transmission for a $PR \in (0.08, 0.1)$, and high transmission for a $PR \geq 0.1$. Due to uncertainties in relating WW data to COVID-19 case counts, public health authorities typically evaluate trends in WW data to asses changes in infection rates rather than absolute thresholds. However, it may be useful for public health authorities to interpret WW data in terms of PR. Our modeling approach provides insights into the evolution of virus transmission dynamics and a methodology that combines different sources of information to continue monitoring the COVID-19 trends.

The paper is organized as follows: Section 2 describes the data and the modeling approach. Section 4 presents the results 3, and Section 4 discusses our modeling approach's conclusions and limitations.

2 Materials and methods

The analytical framework was developed using data from the City of Davis (Davis) and replicated for the campus of the University of California Davis (UC Davis). The Davis and UC Davis WW collection areas (commonly referred to as sewersheds) are geographically adjacent. The analysis includes laboratory-confirmed incident COVID-19 cases and WW data from July 1, 2021, to July 1, 2022, for Davis and UC Davis. Daily COVID-19 cases and tests for Davis were provided by Healthy Davis Together (HDT) (HDT, 2020). HDT is the community pandemic response program launched in Davis from September 2020 to June 30, 2022, to mitigate the spread of COVID-19. HDT involved a broad set of interventions, including free saliva-based asymptomatic testing with high throughput methods to process large volumes of tests. Testing and cases for UC Davis come from the campus community COVID-19 screening program, which included mandatory completion of bi-weekly asymptomatic tests to access campus facilities. The study site is thus unique compared to most WW surveillance regions in that there was an extraordinarily high number of clinical tests performed in both sewersheds during the study period (Figure 1) for a relatively small population size (425,314 tests were performed over the study period by Davis and 835,785 for UC Davis for approximately 66,799 residents in the combined surveillance regions). The high number of tests resulted in $PR \leq 0.05$ for the majority of the study period in both locations (Figure 2). These conditions provide a useful context to estimate WW thresholds that correspond to the CDC-defined PR thresholds. The WHO contends that disease dynamics based on case data can be confidently tracked when $PR \leq 0.05$ (WHO, 2020). Otherwise, when $PR \geq 0.05$, WW measurements offer a more robust measure of true disease dynamics (Fernandez-Cassi et al., 2021)

2.1 Wastewater settled solids methods

Wastewater settled solids for the Davis wastewater treatment plant (WWTP) were collected daily from the primary clarifier, transported on the same day of collection to the analytical laboratory, and processed within 24 h as previously described (Wolfe et al., 2021). Wastewater settled solids were obtained from daily composite influent samples from the UC Davis WWTP. Composite influent samples were collected using a refrigerated autosampler (Hach Sigma 900 MAX) located at

the WWTP headworks and programmed to collect flow-weighted influent sample volumes every 20 minutes for a total volume of 19 L in 24 hours. Composite influent samples were then transferred to one or two 4L low-density polyethylene containers (LDPE Cubitainers^{TM}, Thermo Scientific^{TM} I-Chem^{TM}) and stored at 4C prior to settling (up to 6 days of storage). Each 4L sample was pasteurized in a 60C water bath for 45 min immediately prior to settling. Pasteurized influent samples were inverted to mix, poured into a 3-gallon high-density polyethylene conical vessel equipped with a sampling port (FF3G, FastFerment^{TM}), and left to settle for 2 hours. Settled solids were obtained from either 4L or 8L of influent from a single day (two 4L samples were combined into one settling vessel when 8L were used). Settled solids were collected by dispensing from the bottom of the settling vessel into one or two 50mL polypropylene centrifuge tubes (for 4L or 8L initial volume, respectively). If two tubes of settled solids were obtained, the supernatants were carefully decanted from each, and remaining settled solids were combined. Between sampling episodes, the settling tanks were emptied and the tank and sampling port valves were bleached (10% commercial bleach for 1 hr), rinsed with deionized water, and left to air dry. Samples of settled solids were stored at 4C and subsequently transported on ice in a cooler by a courier to the laboratory.

Samples processing was completed within 8 days of initial sample collection. Sample RNA extraction, purification, and droplet digital reverse transcriptase PCR (ddRT-PCR) followed the same protocol as for the Davis samples. These protocols are described in detail elsewhere (Topol et al., 2021, 2022).

We normalize the SARS-CoV-2 RNA concentration determined (N gene copies per gram dry weight solids) by the concentration of mild pepper mottle virus (PMMoV gene copies per gram dry weight solids) to yield the dimensionless metric, N/PMMoV. The N gene is present in all variants of the virus. PMMoV is a highly abundant RNA virus detected broadly in WW (Rothman et al., 2021; Symonds et al., 2018). PMMoV serves as a process control such that normalization of N by PMMoV for each sample helps correct SARS-CoV-2 concentrations for virus extraction efficiency. PMMoV is also often used to account for variations population size, rainfall, and water usage between different WW collection areas. We expect these latter factors to have less of an effect on N gene concentrations determined herein because water is removed from the WW settled solids samples prior to sample analysis and concentrations are reported in terms of the dry weight of the dewatered solids. Figure 1 shows the normalized N gene concentration.

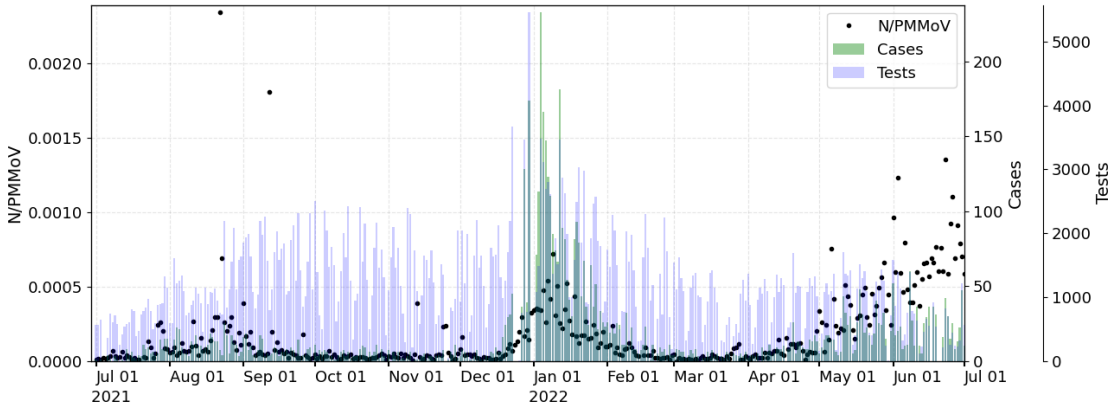
2.1.1 Smoothed wastewater signal

Given that WW signals are often noise-corrupted, we applied a 7-day trimmed mean for daily WW data to reduce uncertainty and minimize daily fluctuations. The smoothed data is later correlated to raw PR (Figure 2).

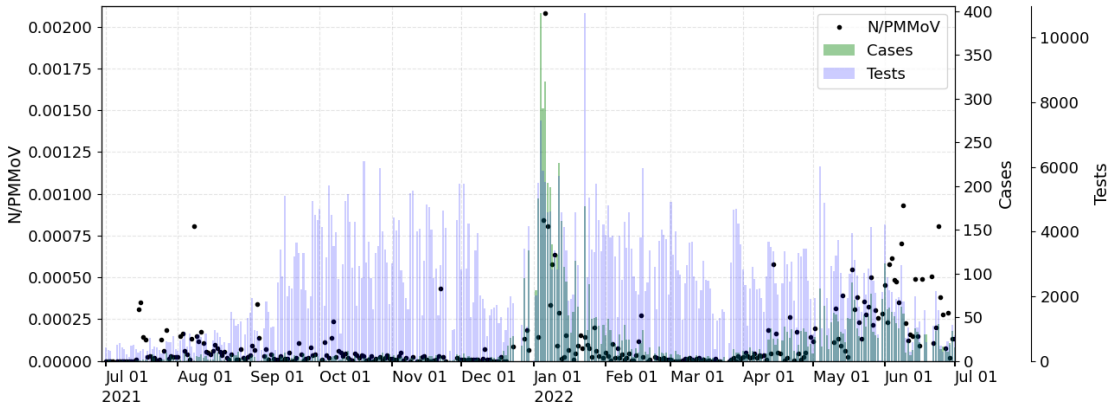
2.2 Statistical model

We model PR as a Beta distribution using WW viral loads as a covariate, assuming a Bayesian approach. We assume Y_i as the PR at day i, defined as the ratio of the number of new positive cases among the number of tests performed at day i. Beta regression is a good choice of model for continuous data with response variables expressed as proportions.

The Beta distribution is reparametrized as $Y_i \sim \mathcal{B}(\mu_i, \phi)$, with mean μ_i and variance $\sigma_i^2 = \mu_i(1-\mu_i)/(1+\phi)$, ϕ is known as the precision parameter since, for fixed μ_i, the larger ϕ the smaller the variance of Y_i; ϕ^{-1} is a dispersion parameter (Ferrari and Cribari-Neto (2004)). The mean μ_i can be expressed as a function of the linear predictor $\eta_i = \beta^T \mathbf{x}_i$, where β is a $(p+1)$-dimensional vector of unknown regression coefficients (including the intercept), and $\mathbf{x}_i$ is the vector of covariates



(A)



(B)

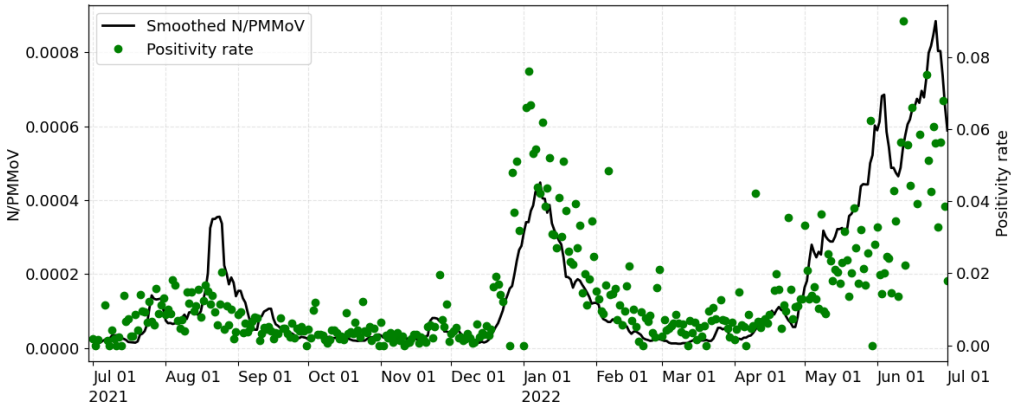
Figure 1: Normalized wastewater data (N/PMMoV), number of COVID-19 tests conducted (Tests), and positive cases (Cases) for (A) Davis and (B) UC Davis.

plus a one for the intercept. In this study, only the 7-day trimmed mean of the wastewater data (Smoothed N/PMMoV), denoted as C_i, is included in the linear predictor. Thus, the linear predictor is given by $\eta_i = \beta_0 + \beta_1 C_i$, and the logit link, the inverse of the logistic function, is used in the Beta regression (i.e., $logit(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \eta_i$).

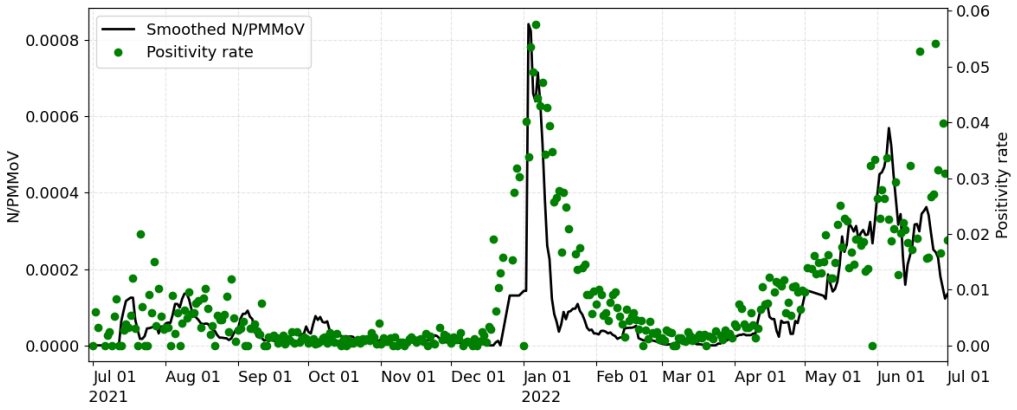
The aim of this inference problem is to estimate $\theta = (\beta_0, \beta_1)$ from measurements of WW data, $\mathbf{C} = (C_1, C_2, \ldots, C_n)$, and COVID-19 PR, $\mathbf{Y} = (y_1, y_2, \ldots, y_n)$. Thus, the likelihood function for the previous model is given by:

$$L(\theta|\mathbf{C}, \mathbf{Y}) = \prod_{i=1}^{n} \frac{1}{\mathrm{B}(a_i, b_i)} y_i^{a_i - 1}(1 - y_i)^{b_i - 1},$$

where $a_i = \mu_i\phi$, $b_i = \phi - a_i$, $\mu_i = 1/(1 + \exp(-\eta_i)$ is the mean (logistic function), $\eta_i = \beta_0 + \beta_1 C_i$ is the linear predictor, and ϕ is the dispersion parameter.



(A)



(B)

Figure 2: 7-day trimmed mean of WW data (Smoothed N/PMMoV) and daily PR for (A) Davis and (B) UC Davis (July 1, 2021 – July 01, 2022).

2.2.1 Bayesian statistical approach

We adopt a Bayesian statistical approach, which is well suited to model multiple sources of uncertainty and allows for incorporating background knowledge of the model's parameters. In this framework, a prior distribution, $\pi_\Theta(\theta)$, is required to account for unknown parameter θ in order to obtain the posterior distribution. Having specified the likelihood and the prior, we use Bayes' rule to calculate the posterior distribution,

$$\pi_{\Theta|\mathbf{C},\mathbf{Y}}(\theta|\mathbf{C},\mathbf{Y}) = \frac{\pi_\Theta(\theta)L(\theta|\mathbf{C},\mathbf{Y})}{Z(\mathbf{Y})},$$

where $Z(\mathbf{Y}) = \int \pi_\Theta(\theta)L(\theta|\mathbf{C},\mathbf{Y})d\theta$ is the normalization constant. The posterior distribution is simulated using an existing Markov chain Monte Carlo (MCMC) method, the t-walk algorithm (Christen et al., 2010).

2.2.2 Bayesian sequential method

We adapt the sequential approach proposed in Daza-Torres et al. (2022) to our model to update forecasts over time. The aim is to train the model using only a subset of the most recent data. The forecast is then updated sequentially in a sliding window of data.

We let L be the length in days of the period used to train the model. The data window is then moved forward every n day as new data becomes available. We set t_0 as the first initial time to start the analysis and the subsequent initial times as $t_{k+1} = t_k + n$. The training period is taken as $[t_k, t_k + L]$ and the forecasting period as $[t_k + L, t_k + L + F]$, see Figure 3.

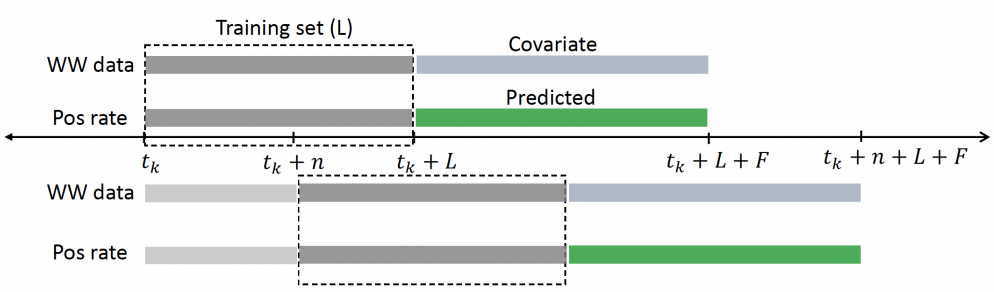


Figure 3: The model is fitted with data from the training period (grey). Then, the estimated parameters are used to predict the forecasting period (green). The training window is then moved n days forward. When new data becomes available, we update all forecasts; the latest posterior becomes the newest prior to the next training period.

We denote $\theta_k = (\beta_0^{(k)}, \beta_1^{(k)})$ the model parameters to be inferred and the vectors of data as $\mathbf{C}_{k,n} = (C_k, \ldots, C_{k+n}), \mathbf{Y}_{k,n} = (y_k, \ldots, y_{k+n})$ at period k. Note that, from the beginning, θ_k is assumed to change in time within each forecast window. If $k = 0$, we postulate a prior distribution $\pi_{\Theta_k}(\theta_k)$ and a likelihood $L(\theta_k|\mathbf{C}_{k,n}, Y_{k,n})$ previously described in Section 2.2. The probabilistic prediction of y_t, in the forecasting period $t \in [t_k + L, t_k + L + F]$, is obtained by using the estimated parameters through the MCMC method and the WW data concentration $(\mathbf{C}_{k+L,F})$.

Afterward, the forecasting window is updated by setting $t_{k+1} = t_k + n$, with n the number of days until the next forecast. In the new training window $[t_{k+1}, t_{k+1} + L]$, we propose a new prior distribution $\pi_{\Theta_{k+1}}(\theta_{k+1})$ for the model parameters θ_{k+1} using samples from the posterior distribution obtained in the previous forecast. Finally, we set $k = k + 1$ and repeat the process described above to create a new forecast, see Daza-Torres et al. (2022) for implementation details.

Regarding the elicitation of the parameters' prior distribution for the first forecast, at $k = 0$, we assume a Normal distribution for the parameters $\beta_0^{(k)}$, $\beta_1^{(k)}$, with mean and standard deviation $(0, 0)$ and $(1000, 1000)$, respectively (i.e. noninformative priors). We set L to twice the length from symptoms onset to mild disease clinical outcome, namely 30 days, and F is chosen to be 10 days. The forecasting was updated every 10 days.

2.3 Action thresholds for WW concentration

We let Y be the PR, and C corresponds to the SARS-CoV-2 RNA concentrations measured in wastewater. Then, the cumulative distribution function of Y given C is defined as $F_{Y|C}(y|c) := P(Y \leq y|C = c)$, which represents the probability that PR is less than or equal to y given that the SARS-CoV-2 RNA concentration in wastewater is c. Henceforth for simplicity, we will use $F(y|c)$

instead of $F_{Y|C}(y|c)$ without loss of generality. The quantile function F^{-1} of Y given $C = c$ is defined[1] by

$$F_c^{-1}(p) := \inf\{y \in \mathbb{R} : F(y|c) \geq p\}, \quad p \in (0, 1).$$

The p-quantile of a data set is defined as the value where a p fraction of the data is below that value and $(1 - p)$ fraction of the data is above that value (e.g., the 0.5 quantile is the median).

Using the CDC thresholds for PR values corresponding to low ($Y \leq 0.05$), moderate ($Y \in (0.05, 0.08)$), substantial ($Y \in (0.08, 0.1)$), and high ($Y \geq 0.1$) transmission and the parameter estimates from the assumed beta regression model, we propose a methodology for estimating WW concentrations associated with PR thresholds at a given point in time. We find the value of WW concentrations, c, such that with probability $1 - \alpha$, the PR is less than the CDC threshold $y \in \{0.5, 0.8, 0.1\}$ (that is, to find c such that $F_c^{-1}(\alpha) = y$). Note that α is the precision we set to estimate the threshold. With $\alpha = 0.05$ we are being conservative and choose lower bounds of WW concentration values associated with PR thresholds.

The proposed method for finding these thresholds, assuming that we have simulations of the posterior distribution for the parameter $\theta = (\beta_0, \beta_1)$, is given in Algorithm 1. First, we suggest a search grid for the WW viral load concentration. Then, for each concentration, we simulate the predicted posterior distribution of the PR. Lastly, we calculate the α-quantile for each concentration level and find the concentration level that yields the quantile closest to the value of the desired PR threshold.

Algorithm 1: Finding threhsold for WW using PR.

Input : A sample, $\theta_1, \theta_2, \ldots, \theta_N$, from the posterior distribution of Θ, where $\theta_i = (\beta_{0,i}, \beta_{1,i}), \quad i = 1, \ldots, N$. The precision parameter ϕ, the threshold for the PR, T_{pr}, and the probability α;

Output: Threshold for WW concentration T_{ww};

Step 1. Generate a grid of WW concentration, $c_1, c_2, \ldots, c_L$;

for *l in 1* **to** L **do**

  for *l in 1* **to** N **do**

    Step 2. Simulate $Y_{l,i} \sim Beta(\mu_{l,i}, \phi)$, where $\mu_{l,i} = \frac{1}{1+\exp(-\eta_{l,i})}$ and $\eta_{l,i} = \beta_{0,i} + \beta_{1,i} c_l$;

  Step 2. Compute the α-quantile of $\mathbf{Y}_l = (Y_{l,1}, \ldots, Y_{l,N})$, namely Q_l,;

Step 3. Find l^* such that $l^* = \underset{l}{\operatorname{argmin}} \quad |Q_l - T_{pr}|$;

Step 4. Set $T_{ww} = c_{l^*}$

2.4 Effective reproductive number

The number of people in a population who are susceptible to infection by an infected individual at any particular time is denoted by R_e, the effective reproductive number. This dimensionless quantity is sensitive to time-dependent variation due to reductions in susceptible individuals, changes in population immunity, and other factors. R_e can be estimated by the ratio of the number of new

[1]The definition says that for given p, we are looking for some y (PR), such that $F(y|c) \geq p$. However, since multiple values of y could meet this condition, we take the smallest y of those.

infections (I_t) generated at time t, to the total infectious individuals at time t, given by $\sum_{s=1}^{t} I_{t-s} w_s$, the sum of new infections up to time step $t-1$, weighted by the infectivity function w_s. We implement Cori et al. (2013)'s approach to calculate the R_e from the PR estimated with the WW data. Note that the PR is an estimation of the proportion of infected persons. Therefore, if we multiply the PR by a T value, representing the total number of tests carried out in the study period, we will have an estimate of the incidence, which we can use to compute the R_e. Since R_e is a scale-free metric, we should get similar results for different T values. We use T as the average of the tests carried out in Davis or UC Davis.

3 Results

The analytical framework was developed using the City of Davis data and replicated for the University of California, Davis campus.

To display changes in the relation between PR and signal WW across variants, we estimate the PR using the sequential model proposed in Section 2.2.2 and compute the posterior distribution for the parameter β_1 over time (Figures 4-5 top-panel). The results show how the PR and WW signal relation changed, mainly when a new variant emerged. These changes remain similar during the period where the variant is dominant.

Using the algorithm to compute the thresholds for WW data with the estimated PR in each time window. We did not calculate thresholds for the period of September-December 2021 since the corresponding posterior distribution for β_1 contained zero. In other words, the probability that the β_1 parameter will be zero was positive, suggesting that there is no significant association between WW data and PR (Figures 4-5). We illustrate the estimated thresholds for WW corresponding to low, moderate, substantial, and high transmission thresholds proposed by the CDC (Figure 6). High variability in the threshold estimation coincides with the emergence of new variants. This variability is reduced in the period where the variant is dominant.

Figure 7 illustrates R_e, which was computed assuming the median of the predicted PR multiplied by the average number of tests performed per day for Davis and UC Davis in the study period, $T = 1,198$ and $T = 2,381$ respectively. We also compare with the R_e computed with the observed cases. R_e trends determined from WW are similar in magnitude and depict similar trends for R_e calculated using observed cases during the periods analyzed.

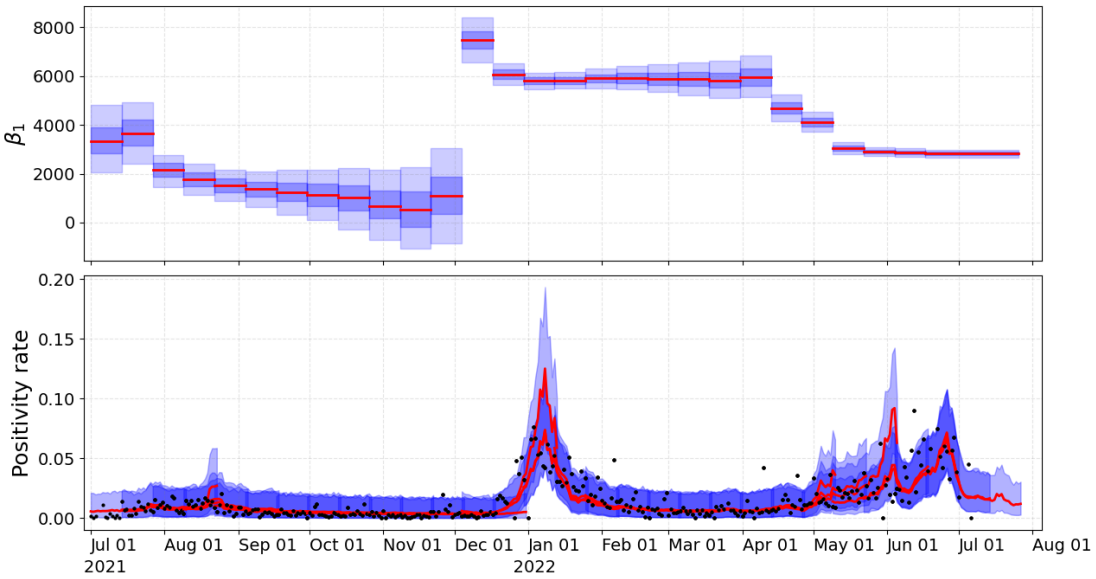


Figure 4: Estimated PR for the City of Davis (lower-panel) and posterior distribution of β_1 (upper-panel). Red-solid line and blue-shadow area describe the mean and SD, respectively.

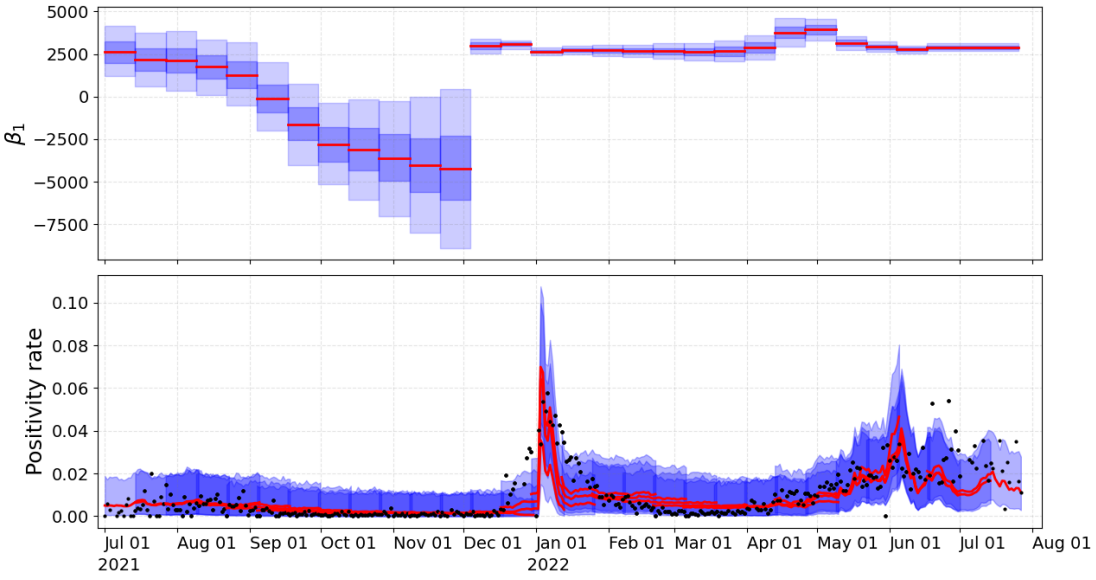
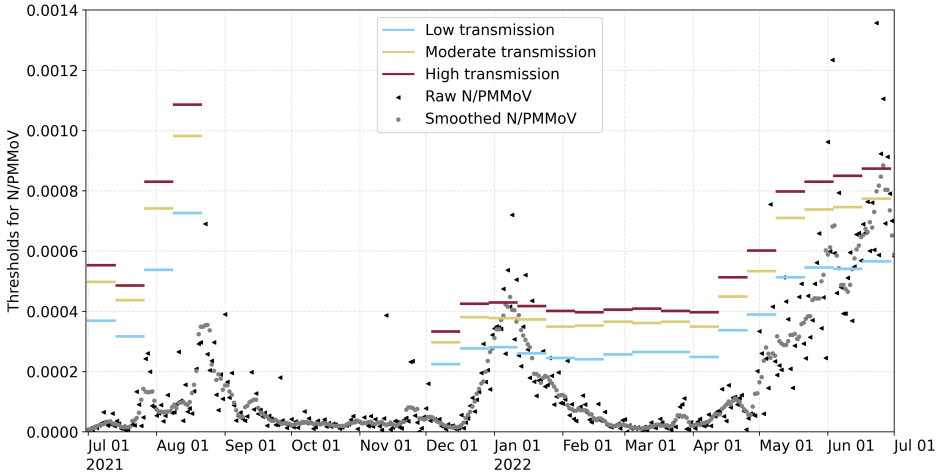
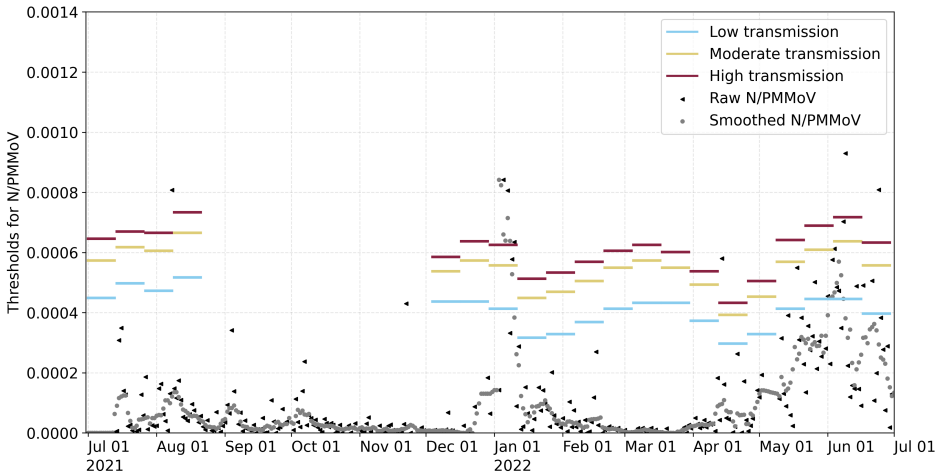


Figure 5: Estimated PR for UC Davis (lower-panel) and posterior distribution of β_1 (upper-panel). Red-solid line and blue-shadow area describe the mean and SD, respectively.

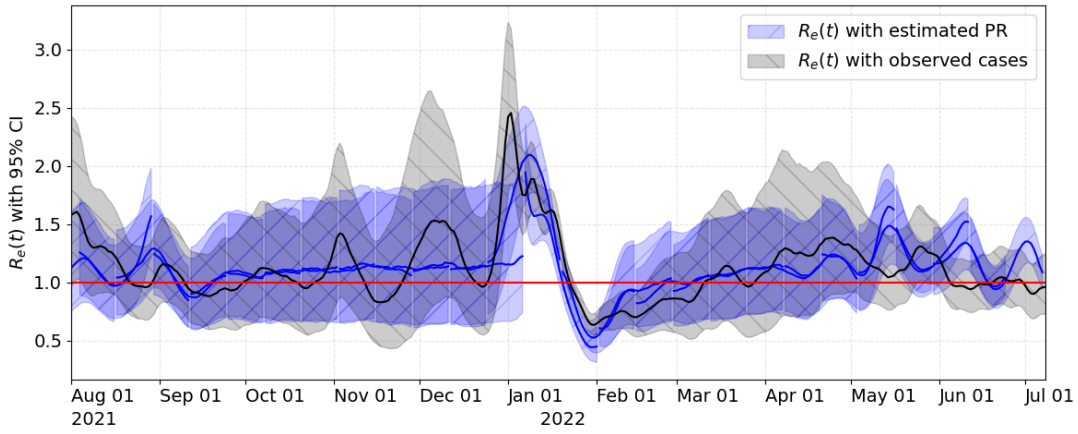


(A)

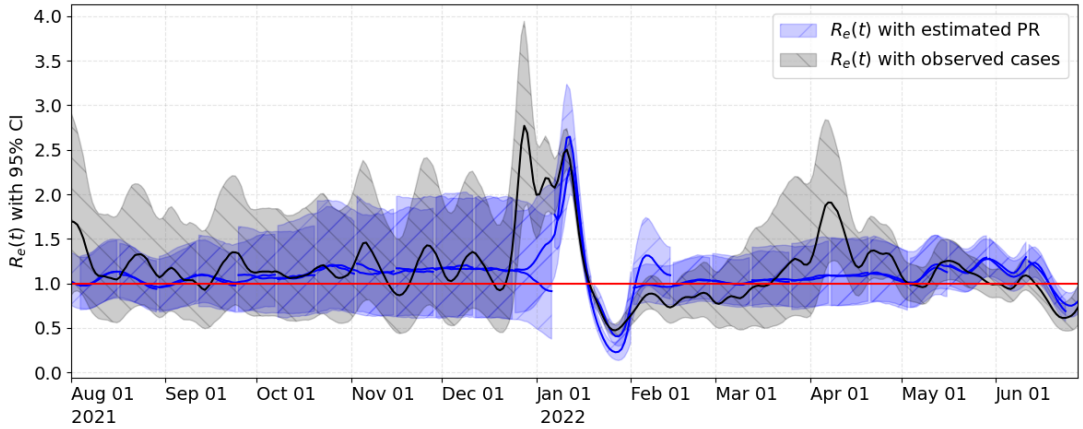


(B)

Figure 6: Wastewater thresholds over time for (A) the City of Davis and (B) UC Davis. Blue, yellow, and red horizontal lines correspond to low, moderate, and high transmission thresholds, respectively. Raw (black triangles) and Smoothed (grey dots) N/PMMoV WW data.



(A)



(B)

Figure 7: Effective R_e of (A) the City of Davis and (B) UC Davis computed with observed cases and the median of the predicted PR multiplied by the average number of tests performed for Davis in the study period. Solid lines and shaded regions illustrate the median and 95% prediction intervals, respectively.

4 Discussion

Substantial changes in test availability and test seeking behavior may confound estimates of case counts in a community (Daza-Torres et al., 2023). Positivity rate (PR) may provide a more accurate reflection of the state of the epidemic. PR is an important metric because it indicates how widespread an outbreak is within a particular area where testing is being conducted and whether current levels of testing are sufficient to accurately capture levels of disease transmission. Increases in PR can indicate that it may be a good time to incorporate restrictions to slow the spread of disease (Dowdy and D'Souza, 2020).

This study proposes a sequential Bayesian framework to model the COVID-19 PR via wastewater data for near real-time monitoring of the COVID-19 pandemic. The PR is modeled as a reparametrized Beta regression, and the parameters are estimated using a Bayesian approach. The

changing dynamics of the virus and the data availability impose a challenge in developing helpful mathematical modeling for the surveillance and monitoring of COVID-19. Here, we propose an adaptive modeling framework as an alternative to overcome some of the limitations of traditional models. The adaptive capacity of the model is well suited to capture the variability in virus trends over time by leveraging knowledge gained.

We then use the model developed to offer a retrospective estimate of WW thresholds (for the settled solids analytical method performed) that corresponded to PR thresholds recommended by the CDC. The WW thresholds determined for UC Davis appeared more stable through time and through waves of different COVID-19 variants than WW thresholds estimated for the City of Davis. This can be explained by the fact that the UC Davis PR was nearly always less than 5%, while the City of Davis PR exceeded 5% periodically over the study period. Confidence in disease dynamics breaks down as PR rises above 5% (WHO, 2020). The WW thresholds estimated for the City of Davis were similar to those for UC Davis when PR remained low but increased dramatically at the end of the study period when clinical testing rates declined in Davis, usage of at-home test kits increased, and PR surpassed 5%. Mandatory asymptomatic testing continued for UC Davis through the end of the study period but has since been eliminated. In the absence of strong clinical testing programs, the relatively more stable WW thresholds determined over this study period may serve as a future reference to assess relative COVID-19 infection dynamics for these sewersheds. We caution against the direct translation of the estimated WW thresholds to other sewersheds and analytical methods for WW. Further research is needed to investigate the application of this framework to other sewersheds, for alternative WW analytical methods, and other respiratory and enteric pathogens present in WW.

One of the limitations of this framework is that it requires access to both WW and test data for continued adaptation, which implies continuous community monitoring through testing. The capacity of current testing programs has decreased significantly with recent transitions to a new normal and the implementation of prevention and surveillance mechanisms such as vaccines, at-home tests, and WW surveillance. New limited testing may lead to passive case-finding (i.e., only those most likely to be infected are tested). The PR may thus overestimate the current burden of the disease under these conditions. It is important to highlight that reduction in information poses new challenges and limitations in COVID-19 monitoring. With reductions in testing, public health authorities must decide between an indicator that overestimates the burden of the disease (PR) or a projection that underestimates the burden of the disease (case counts). We propose to use the PR in combination with WW measurements to reduce bias in assessments of disease dynamics.

In our study, the relationship between wastewater concentrations and PR changed when PR increased. This was likely due to changes in test-seeking behavior and test availability. Viral shedding may also change through time due to changes in vaccination status, acquired immunity, and changes in transmission patterns for different viral variants - although there is limited evidence available from fecal shedding studies. WW thresholds were estimated herein based on PR for a community where testing rates were extraordinarily high, given the population size. While it is not feasible at this time to establish *a priori* public health thresholds based on WW concentrations to estimate the burden of the disease through time, a record of historical values and thresholds provides meaningful context to guide public health authorities as new waves of infection arise.

Availability Statement

All code and underlying data are publicly available through the GitHub repository (Montesinos-López and Daza-Torres, 2023). Analyses were carried out using Python version 3.9.

Human Participant Protection

This study was determined to be exempt from institutional review board review by the UC Davis Office of Research.

Credit authorship contribution statement

J.C.M.L. and M.L.D.T: Conceptualization, Methodology, Software, Formal analysis, Writing - Original draft preparation, Writing - Reviewing & Editing. **Y.E.G. and C.H.:** Literature review, Writing - Reviewing & Editing. **C.W.B.** Laboratory sample processing, Writing - methods. **H.N.B.:** Project conception, funding, research oversight, collaborator coordination, Writing - methods, Writing - Reviewing & Editing. **M.N.:** Methodology, research oversight/supervision, Writing - Reviewing & Editing.

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.

Acknowledgments

This research was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant number UL1 TR001860. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This research was also supported by Healthy Davis Together (HDT) programs at the University of California, Davis, and the Centers for Disease Control and Prevention Foundation. Additionally, we thank Colleen C. Naughton, Alexandria B. Boehm, and Marlene K. Wolfe for their valuable comments. All authors reviewed and approved the final manuscript. We declare no conflicts of interest.

References

W. Ahmed, N. Angel, J. Edson, K. Bibby, A. Bivins, J. W. O'Brien, P. M. Choi, M. Kitajima, S. L. Simpson, J. Li, B. Tscharke, R. Verhagen, W. J. Smith, J. Zaugg, L. Dierens, P. Hugenholtz, K. V. Thomas, and J. F. Mueller. 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*, 728:138764, 2020. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2020.138764.

Y. Ai, A. Davis, D. Jones, S. Lemeshow, H. Tu, F. He, P. Ru, X. Pan, Z. Bohrerova, and J. Lee. Wastewater sars-cov-2 monitoring as a community-level covid-19 trend tracker and variants in ohio, united states. *Science of The Total Environment*, 801:149757, 2021. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2021.149757. URL https://www.sciencedirect.com/science/article/pii/S0048969721048324.

H. Asghar, O. M. Diop, G. Weldegebriel, F. Malik, S. Shetty, L. El Bassioni, A. O. Akande, E. Al Maamoun, S. Zaidi, A. J. Adeniji, et al. Environmental surveillance for polioviruses in the

global polio eradication initiative. *The Journal of Infectious Diseases*, 210(suppl_1):S294–S303, 2014. doi: 10.1093/infdis/jiu384.

F. Bassi, G. Arbia, and P. Falorsi. Observed and estimated prevalence of covid-19 in italy: How to estimate the total cases from medical swabs data. *Science of The Total Environment*, 764:142799, 2021. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2020.142799. URL https://www.sciencedirect.com/science/article/pii/S0048969720363282.

L. Böttcher, M. R. D'Orsogna, and T. Chou. A statistical model of covid-19 testing in populations: effects of sampling bias and testing errors. *Philosophical Transactions of the Royal Society A*, 380(2214):20210121, 2022. doi: 10.1098/rsta.2021.0121. URL https://doi.org/10.1098/rsta.2021.0121.

M. A. Capistran, A. Capella, and J. A. Christen. Forecasting hospital demand in metropolitan areas during the current covid-19 pandemic and estimates of lockdown-induced 2nd waves. *PLOS ONE*, 16(1):1–16, 01 2021. doi: 10.1371/journal.pone.0245669. URL https://doi.org/10.1371/journal.pone.0245669.

P. M. Choi, B. Tscharke, S. Samanipour, W. D. Hall, C. E. Gartner, J. F. Mueller, K. V. Thomas, and J. W. O'Brien. Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. *Proceedings of the National Academy of Sciences*, 116(43):21864–21873, 2019. doi: 10.1073/pnas.1910242116.

J. A. Christen, C. Fox, et al. A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–281, 2010. doi: 10.1214/10-BA603.

A. Christie, J. T. Brooks, L. A. Hicks, E. K. Sauber-Schatz, J. S. Yoder, M. A. Honein, C. COVID, and R. Team. Guidance for implementing covid-19 prevention strategies in the context of varying community transmission levels and vaccination coverage. *Morbidity and mortality weekly report*, 70(30):1044, 2021. doi: http://dx.doi.org/10.15585/mmwr.mm7030e2.

A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178 (9):1505–1512, 2013. doi: 10.1093/aje/kwt133.

A. A. Dallal, U. AlDallal, and J. A. Dallal. Positivity rate: an indicator for the spread of covid-19. *Current Medical Research and Opinion*, 37(12):2067–2076, 2021. doi: 10.1080/03007995.2021.1980868. URL https://doi.org/10.1080/03007995.2021.1980868. PMID: 34517740.

P. M. D'Aoust, X. Tian, S. T. Towhid, A. Xiao, E. Mercier, N. Hegazy, J.-J. Jia, S. Wan, M. P. Kabir, W. Fang, M. Fuzzen, M. Hasing, M. I. Yang, J. Sun, J. Plaza-Diaz, Z. Zhang, A. Cowan, W. Eid, S. Stephenson, M. R. Servos, M. J. Wade, A. E. MacKenzie, H. Peng, E. A. Edwards, X.-L. Pang, E. J. Alm, T. E. Graber, and R. Delatolla. Wastewater to clinical case (wc) ratio of covid-19 identifies insufficient clinical testing, onset of new variants of concern and population immunity in urban communities. *Science of The Total Environment*, 853:158547, 2022. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2022.158547. URL https://www.sciencedirect.com/science/article/pii/S0048969722056467.

M. L. Daza-Torres, M. A. Capistrán, A. Capella, and J. A. Christen. Bayesian sequential data assimilation for covid-19 forecasting. *Epidemics*, 39:100564, 2022. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2022.100564. URL https://www.sciencedirect.com/science/article/pii/S1755436522000196.

M. L. Daza-Torres, J. C. Montesinos-López, M. Kim, R. Olson, C. W. Bess, L. Rueda, M. Susa, L. Tucker, Y. E. García, A. J. Schmidt, C. C. Naughton, B. H. Pollock, K. Shapiro, M. Nuño, and H. N. Bischel. Model training periods impact estimation of covid-19 incidence from wastewater viral loads. *Science of The Total Environment*, 858:159680, 2023. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2022.159680.

D. Dowdy and G. D'Souza. Covid-19 testing: understanding the "percent positive". *Johns Hopkins Bloomberg School of Public Health*, 2020. URL https://publichealth.jhu.edu/2020/covid-19-testing-understanding-the-percent-positive.

R. Engbert, M. M. Rabe, R. Kliegl, and S. Reich. Sequential Data Assimilation of the Stochastic SEIR Epidemic Model for Regional COVID-19 Dynamics. *Bulletin of Mathematical Biology*, 83 (1):1–16, 2020. ISSN 1522-9602. URL https://doi.org/10.1007/s11538-020-00834-8.

L. Fenga and M. Gaspari. Predictive capacity of covid-19 test positivity rate. *Sensors*, 21(7), 2021. ISSN 1424-8220. doi: 10.3390/s21072435. URL https://www.mdpi.com/1424-8220/21/7/2435.

N. E. Fenton, M. Neil, M. Osman, and S. McLachlan. Covid-19 infection and death rates: the need to incorporate causal explanations for the data and avoid bias in testing. *Journal of Risk Research*, 23(7-8):862–865, 2020. doi: 10.1080/13669877.2020.1756381. URL https://doi.org/10.1080/13669877.2020.1756381.

X. Fernandez-Cassi, A. Scheidegger, C. Bänziger, F. Cariti, A. Tuñas Corzon, P. Ganesanandamoorthy, J. C. Lemaitre, C. Ort, T. R. Julian, and T. Kohn. Wastewater monitoring outperforms case numbers as a tool to track covid-19 incidence dynamics when test positivity rates are high. *Water Research*, 200:117252, 2021. ISSN 0043-1354. doi: https://doi.org/10.1016/j.watres.2021.117252. URL https://www.sciencedirect.com/science/article/pii/S0043135421004504.

S. Ferrari and F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004. doi: 10.1080/0266476042000214501.

D. Frampton, T. Rampling, A. Cross, H. Bailey, J. Heaney, M. Byott, R. Scott, R. Sconza, J. Price, M. Margaritis, et al. Genomic characteristics and clinical effect of the emergent sars-cov-2 b.1.1.7 lineage in london, uk: a whole-genome sequencing and hospital-based cohort study. *The Lancet Infectious Diseases*, 21(9):1246–1256, 2021. ISSN 1473-3099. doi: https://doi.org/10.1016/S1473-3099(21)00170-5.

Y. Furuse, Y. K. Ko, K. Ninomiya, M. Suzuki, and H. Oshitani. Relationship of test positivity rates with covid-19 epidemic dynamics. *International Journal of Environmental Research and Public Health*, 18(9), 2021. ISSN 1660-4601. doi: 10.3390/ijerph18094655. URL https://www.mdpi.com/1660-4601/18/9/4655.

G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, 26(6):855–860, 2020. ISSN 1546-170X. doi: https://doi.org/10.1038/s41591-020-0883-7.

M. S. Graham, C. H. Sudre, A. May, M. Antonelli, B. Murray, T. Varsavsky, K. Kläser, L. S. Canas, E. Molteni, M. Modat, et al. Changes in symptomatology, reinfection, and transmissibility associated with the sars-cov-2 variant b. 1.1. 7: an ecological study. *The Lancet Public Health*, 6 (5):e335–e345, 2021. ISSN 2468-2667. doi: https://doi.org/10.1016/S2468-2667(21)00055-4.

HDT. Healthy Davis Together. https://healthydavistogether.org/testing-data/, 2020. (Accessed on 06/08/2022).

J. S. Huisman, J. Scire, L. Caduff, X. Fernandez-Cassi, P. Ganesanandamoorthy, A. Kull, A. Scheidegger, E. Stachler, A. B. Boehm, B. Hughes, A. Knudson, A. Topol, K. R. Wigginton, M. K. Wolfe, T. Kohn, C. Ort, T. Stadler, and T. R. Julian. Wastewater-based estimation of the effective reproductive number of sars-cov-2. *Environmental Health Perspectives*, 130(5):057011, 2022. doi: 10.1289/EHP10050. URL https://ehp.niehs.nih.gov/doi/abs/10.1289/EHP10050.

N. P. Jewell, J. A. Lewnard, and B. L. Jewell. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA*, 323(19):1893–1894, 05 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.6585. URL https://doi.org/10.1001/jama.2020.6585.

T. R. Mercer and M. Salit. Testing at scale during the covid-19 pandemic. *Nature Reviews Genetics*, 22(7):415–426, 2021. doi: 10.1038/s41576-021-00360-w. URL https://doi.org/10.1038/s41576-021-00360-w.

J. C. Montesinos-López and M. L. Daza-Torres. Bayesian sequential approach to monitor covid-19 variants through positivity rate from wastewater. https://github.com/mdazatorres/Bayesian_sequential_approach_PR_WW, 2023.

J. C. Montesinos-López, M. L. Daza-Torres, Y. E. García, L. A. Barboza, F. Sanchez, A. J. Schmidt, B. H. Pollock, and M. Nuño. The role of sars-cov-2 testing on hospitalizations in california. *Life*, 11:1336, 2021. doi: 10.3390/life11121336.

B. Rader, A. Gertz, A. D. Iuliano, M. Gilmer, L. Wronski, C. M. Astley, K. Sewalk, T. J. Varrelman, J. Cohen, R. Parikh, et al. Use of at-home covid-19 tests—united states, august 23, 2021–march 12, 2022. *Morbidity and Mortality Weekly Report*, 71(13):489–494, 2022. doi: 10.15585/mmwr.mm7113e1.

J. A. Rothman, T. B. Loveless, J. Kapcia III, E. D. Adams, J. A. Steele, A. G. Zimmer-Faust, K. Langlois, D. Wanless, M. Griffith, L. Mao, et al. Rna viromics of southern california wastewater and detection of sars-cov-2 single-nucleotide variants. *Applied and environmental microbiology*, 87(23):e01448–21, 2021. doi: 10.1128/AEM.01448-21.

G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien. Forecasting models for coronavirus disease (covid-19): a survey of the state-of-the-art. *SN Computer Science*, 1(4):1–15, 2020. doi: 10.1007/s42979-020-00209-9.

E. Symonds, K. H. Nguyen, V. Harwood, and M. Breitbart. Pepper mild mottle virus: a plant pathogen with a greater purpose in (waste) water treatment development and public health management. *Water research*, 144:1–12, 2018. doi: https://doi.org/10.1016/j.watres.2018.06.066.

A. Topol, M. Wolfe, B. White, K. Wigginton, and A. BB Boehm. High throughput pre-analytical processing of wastewater settled solids for sars-cov-2 rna analyses v2. *ZappyLab, Inc.*, 2021. doi: dx.doi.org/10.17504/protocols.io.b2kmqcu6. URL https://www.protocols.io/view/high-throughput-pre-analytical-processing-of-waste-kxygxpod4l8j/v2. (Accessed on 06/08/2022).

A. Topol, M. Wolfe, B. White, K. Wigginton, and A. BB Boehm. High throughput sars-cov-2, pmmov, and bcov quantification in settled solids using digital rt-pcr. protocols. io, 2022. URL https://www.protocols.io/view/high-throughput-sars-cov-2-pmmov-and-bcov-quantifi-e6nvw5orwvmk/v5?version_warning=no. (Accessed on 06/08/2022).

WHO. Public health criteria to adjust public health and social measures in the context of covid-19: annex to considerations in adjusting public health and social measures in the context of covid-19, 12 may 2020. Technical documents, World Health Organization, 2020. URL https://apps.who.int/iris/handle/10665/332073.

M. K. Wolfe, A. Topol, A. Knudson, A. Simpson, B. White, D. J. Vugia, A. T. Yu, L. Li, M. Balliet, P. Stoddard, et al. High-frequency, high-throughput quantification of sars-cov-2 rna in wastewater settled solids at eight publicly owned treatment works in northern california shows strong association with covid-19 incidence. *Msystems*, 6(5):e00829–21, 2021. doi: https://doi.org/10.1128/mSystems.00829-21.

A. Xiao, F. Wu, M. Bushman, J. Zhang, M. Imakaev, P. R. Chai, C. Duvallet, N. Endo, T. B. Erickson, F. Armas, B. Arnold, H. Chen, F. Chandra, N. Ghaeli, X. Gu, W. P. Hanage, W. L. Lee, M. Matus, K. A. McElroy, K. Moniz, S. F. Rhode, J. Thompson, and E. J. Alm. Metrics to relate covid-19 wastewater data to clinical testing dynamics. *Water Research*, 212:118070, 2022. ISSN 0043-1354. doi: https://doi.org/10.1016/j.watres.2022.118070. URL https://www.sciencedirect.com/science/article/pii/S0043135422000331.

J. Yang, X. Chen, X. Deng, Z. Chen, H. Gong, H. Yan, Q. Wu, H. Shi, S. Lai, M. Ajelli, et al. Disease burden and clinical severity of the first pandemic wave of covid-19 in wuhan, china. *Nature communications*, 11(1):1–10, 2020.

E. Zuccato, C. Chiabrando, S. Castiglioni, D. Calamari, R. Bagnati, S. Schiarea, and R. Fanelli. Cocaine in surface waters: a new evidence-based tool to monitor community drug abuse. *Environmental Health*, 4(1):1–7, 2005. doi: 10.1186/1476-069X-4-14.