Journal of Water & Health

IWA PUBLISHING

Journal of Water and Health Vol 20 No 5, 816 doi: 10.2166/wh.2022.013

Modeling the relationship between SARS-CoV-2 RNA in wastewater or sludge and COVID-19 cases in three New England regions

Elyssa Anneser[a,†], Emily Riseberg[a,*,†], Yolanda M. Brooks[b], Laura Corlin[a,c] and Christina Stringer[d]

[a] Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, USA
[b] Department of Sciences, St Joseph's College of Maine, Standish, ME, USA
[c] Department of Civil and Environmental Engineering, Tufts University School of Engineering, Medford, MA, USA
[d] New England Interstate Water Pollution Control Commission, Lowell, MA, USA
*Corresponding author. E-mail: emilyriseberg@fas.harvard.edu
†These authors contributed equally to the work.

ER, 0000-0003-4024-2533

ABSTRACT

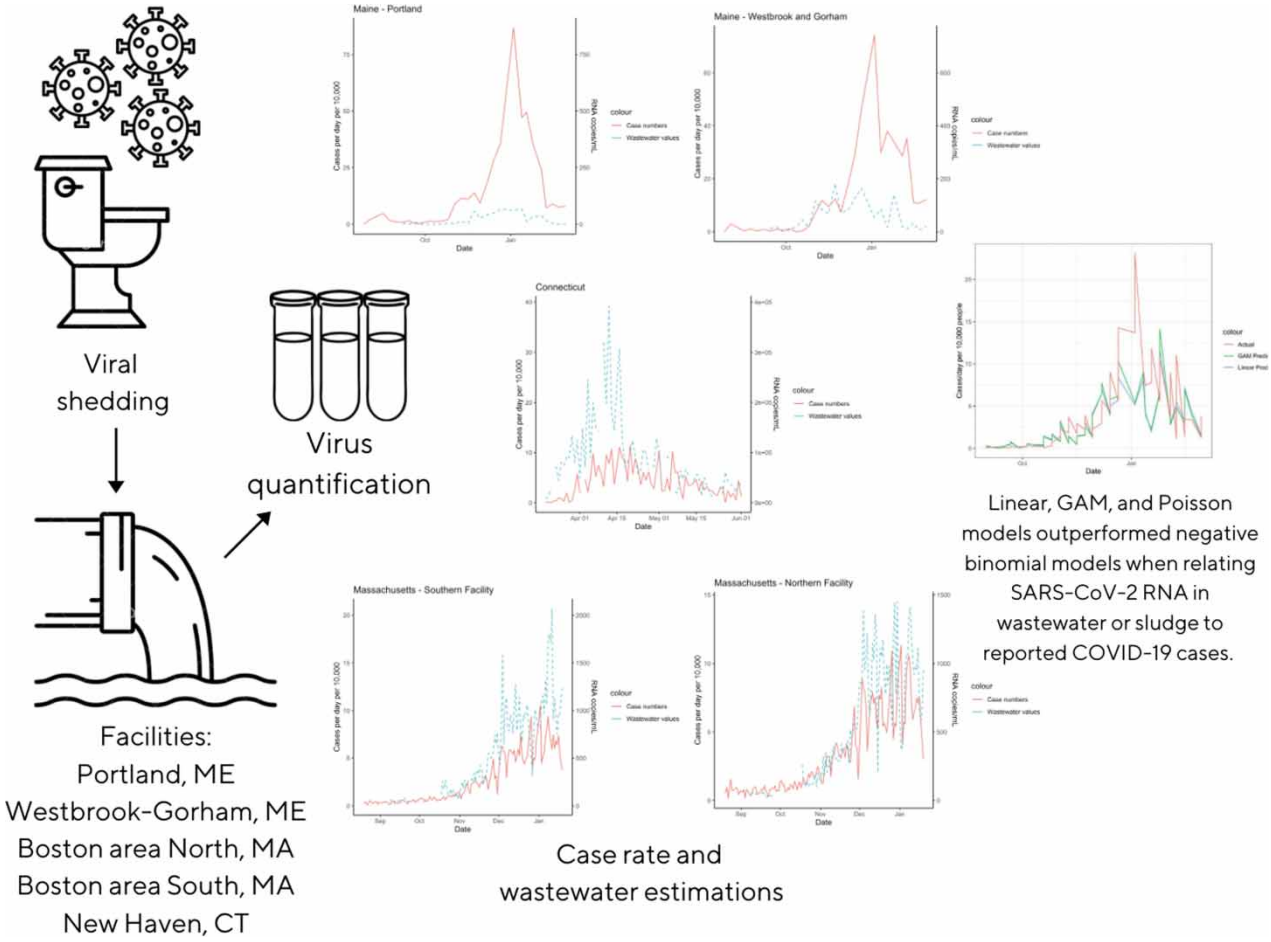
Background: We aimed to compare statistical techniques estimating the association between SARS-CoV-2 RNA in untreated wastewater and sludge and reported coronavirus disease 2019 (COVID-19) cases. Methods: SARS-CoV-2 RNA concentrations (copies/mL) were measured from 24-h composite samples of wastewater in Massachusetts (MA) (daily; 8/19/2020–1/19/2021) and Maine (ME) (weekly; 9/1/2020–3/2/2021) and sludge samples in Connecticut (CT) (daily; 3/1/2020–6/1/2020). We fit linear, generalized additive with a cubic regression spline (GAM), Poisson, and negative binomial models to estimate the association between SARS-CoV-2 RNA concentration and reported COVID-19 cases. Results: The models that fit the data best were linear [adjusted R^2=0.85 (MA), 0.16 (CT), 0.63 (ME); root-mean-square error (RMSE)=0.41 (MA), 1.14 (CT), 0.99 (ME)), GAM (adjusted R^2=0.86 (MA), 0.16 (CT) 0.65 (ME); RMSE=0.39 (MA), 1.14 (CT), 0.97 (ME)], and Poisson [pseudo R^2=0.84 (MA), 0.21 (CT), 0.52 (ME); RMSE=0.39 (MA), 0.67 (CT), 0.79 (ME)]. Conclusions: Linear, GAM, and Poisson models outperformed negative binomial models when relating SARS-CoV-2 RNA in wastewater or sludge to reported COVID-19 cases.

Key words: COVID-19, SARS-CoV-2, sludge, statistical model comparisons, wastewater, wastewater-based epidemiology

HIGHLIGHTS

- Comparisons of statistical models can help inform public health departments on how to implement wastewater-based epidemiology.
- We compared the performance of linear, generalized additive, Poisson, and negative binomial models relating SARS-CoV-2 RNA in wastewater or sludge to reported COVID-19 cases.
- Linear, generalized additive, and Poisson models performed best and negative binomial models performed worst.

GRAPHICAL ABSTRACT



INTRODUCTION

Coronavirus Disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, emerged in late 2019 and quickly escalated into a global health emergency and ongoing pandemic. The manifestation of symptoms and severity of COVID-19 vary among individuals but often include fever, shortness of breath, and loss of sense of taste or smell (CDC 2021). Asymptomatic carriers are estimated to contribute to approximately a quarter of positive cases and spread (He *et al.* 2021), and unless these individuals opt to receive a test or are part of a routine surveillance testing program, their cases will be undetected. Therefore, individual-level testing alone will generally underestimate the community prevalence of COVID-19 (Omori *et al.* 2020).

To complement individual-level testing, public health officials can take advantage of routine community-level wastewater surveillance since SARS-CoV-2 is shed fecally by symptomatic and asymptomatic carriers (La Rosa *et al.* 2020; Polo *et al.* 2020). Wastewater screening is a tool of wastewater-based epidemiology (WBE) and has been used to detect the community prevalence of other fecally shed viruses and illicit drug use (Sinclair *et al.* 2008; Zarei *et al.* 2020). Since the beginning of the COVID-19 pandemic, WBE has been increasingly used as a surveillance technique to monitor the state of the pandemic by localities throughout the world (Goulding *et al.* 2020; Feng *et al.* 2021; Shah *et al.* 2022). For example, previous studies found associations between the reported occurrence of COVID-19 and the concentration of SARS-CoV-2 ribonucleic acid (RNA) in wastewater in the Boston, Massachusetts (MA) metro area and in sludge in New Haven, Connecticut (CT) (Peccia *et al.* 2020; Wu *et al.* 2020). Studies in the Boston area suggest that the actual COVID-19 prevalence was approximately 4–200 times higher than the confirmed clinical case counts indicated (Wu *et al.* 2020). Additionally, WBE has been useful in predicting future trends in COVID-19 incidence since RNA concentrations of SARS-CoV-2 in sludge were found to lead clinical case trends by 2–10 days (Peccia *et al.* 2020).

Several statistical modeling techniques have been proposed to estimate the association between viral genetic markers in wastewater and sludge and COVID-19 cases in a community (McMahan *et al.* 2020; Medema *et al.* 2020; Peccia *et al.* 2020; Vallejo *et al.* 2020; Weidhaas *et al.* 2021). Linear regression and simple correlations have been used in multiple settings (Medema *et al.* 2020; Vallejo *et al.* 2020; Weidhaas *et al.* 2021); however, a standard framework to account for flow rate, environmental effects, and other confounding factors has not been established. Similarly, these methods fail to account for the potential for a nonlinear relationship between SARS-CoV-2 in wastewater and COVID-19 case counts. A second approach uses generalized additive models. One study analyzing data from Spain suggested that generalized additive models performed similarly to models that did not have distributional assumptions (Vallejo *et al.* 2020). A third common modeling strategy fits Poisson models to estimate the count of reported new cases. For example, these have been used in New Haven, CT to estimate trends of reported case numbers and hospitalization numbers (Peccia *et al.* 2020). One limitation of Poisson regression, however, is the assumption that the mean and variance of the outcome (i.e., case counts) are equal. Negative binomial models relax this assumption. Negative binomial models have been used to estimate COVID-19 cases in China based on proxies for water-borne SARS-CoV-2 (Wang *et al.* 2021). To our knowledge, negative binomial regression models have not yet been utilized to assess the association between SARS-CoV-2 RNA in wastewater or sludge and COVID-19 case counts. Given the building body of literature on potential models to estimate community COVID-19 prevalence from wastewater surveillance, these modeling approaches should be quantitatively compared and contextualized across multiple communities for implementation by public health officials.

In this study, we compared four statistical modeling approaches (linear, generalized additive, Poisson, negative binomial) to find an operational model for the associations between SARS-CoV-2 RNA concentrations in wastewater influent or sludge and COVID-19 for four communities in New England, USA (Boston, MA; Portland, Maine (ME); Gorham and Westbrook, ME; and New Haven, CT) serviced by distinct wastewater treatment facilities.

METHODS

SARS-CoV-2 sample collection and characteristics of sewersheds

Wastewater and sludge samples were collected from the Massachusetts Water Resource Authority [MWRA; Boston (and surrounding area), MA], East Shore Water Pollution Abatement Facility (ESWPAF; New Haven, CT), and the Portland Water District (PWD; Portland, Westbrook, and Gorham, ME) (Figure 1). Wastewater samples from MA and ME were collected via 24-h (1 L) composite samples of influent, with at least eight samples collected over the course of 24-h to represent the flow over the course of the day. In MA, samples were collected three to seven times per week both in the Northern and Southern facilities of the watershed system (MWRA – Wastewater COVID-19 Tracking n.d.). We analyzed MA data from 8/19/2020 to 1/19/2021. In ME, samples were collected weekly, and we analyzed data from 9/1/2020 to 3/2/2021. Primary sludge samples (40 mL) from CT were collected daily. We analyzed data from 3/1/2020 to 6/1/2020.

Quantification of SARS-CoV-2

The quantification methods differed slightly among the geographic regions. For MA, the laboratory methods for SARS-CoV-2 concentration estimation were reported previously (Wu *et al.* 2020). Briefly, MA used the TRIzol-chloroform approach to extract RNA and reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) to quantify SARS-CoV-2 RNA via the N1, N2, and N3 markers (Biobot Analytics, Inc. Cambridge, MA) (Wu *et al.* 2020). The TRIzol reagent was mixed with 300 μL of chloroform for 1 min and incubated at room temperature for 5 min. Then 600 μL of aqueous phase were transferred to new 1.50 mL tubes and mixed with an equal volume of isopropanol. After centrifuging for 10 min, the pellet was washed twice with 75% ethanol and then with 30 μL of diethyl pyrocarbonate (DEPC) water to recover the RNA (Wu *et al.* 2020). Values were normalized to the population using the RNA virus pepper mild mottle virus, a highly prevalent fecal marker (Wu *et al.* 2020; Feng *et al.* 2021). The limit of detection for the MA samples was 4.80 copies/mL of sewage after normalization, and samples below this limit were reported as nondetects.

For CT, the laboratory methods were also previously reported (Peccia *et al.* 2020). Total RNA was isolated using a commercial kit (RNeasey PowerSoil Total RNA Kit, Qiagen), and the RNA pellets were dissolved in 50 μL of ribonuclease-free water. Total RNA concentrations were measured using spectrophotometry (Peccia *et al.* 2020). CT evaluated the presence of the human *RP* gene marker in the sludge to establish the presence of human material (Peccia *et al.* 2020). The Bio-Rad iTaq Universal Probes One-Step kit was used, with 20 μL reactions run at 50 °C for 10 min, 95 °C for 1 min, and 40 cycles at 95 and 60 °C for 10 and 30 s, respectively, as recommended by the manufacturer. SARS-CoV-2 concentrations were calculated using

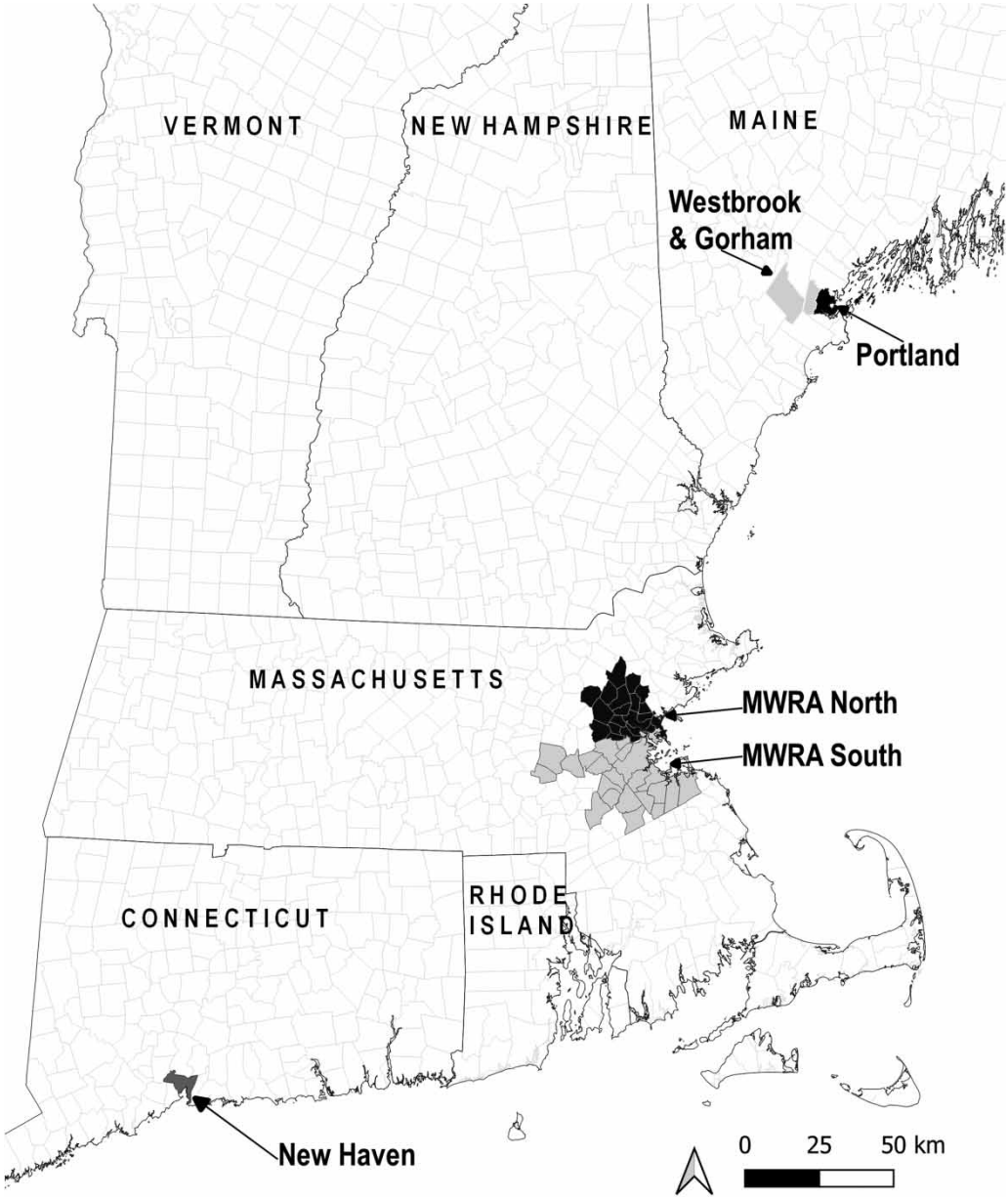


Figure 1 | Map of wastewater and sludge sampling areas in New England. Massachusetts wastewater data was collected from the Massachusetts Water Resources Association (MWRA). CT sludge data was collected from the East Shore Water Pollution Abatement Facility. Maine wastewater data was collected from the Portland Water District.

a standard curve, in which DNA from the WA1-USA strain of SARS-CoV-2 RNA was used as a template to determine transcripts of the SARS-CoV-2 *N* gene (Vogels *et al.* 2020). SARS-CoV-2 RNA copies were normalized to the total RNA in the sample (Peccia *et al.* 2020). Values below 39 copies/mL were not used.

ME utilized the SARS-CoV-2 RT-PCR test (IDEXX Laboratories, Inc. Westbrook, ME), a test of RT-qPCR. The reaction volume was 25 μL, and the template volume was 5 μL. SARS-CoV-2 RNA was quantified in all samples via the N1 and N2 markers (CDC 2020). Absolute concentrations of N1 and N2 genetic markers were calibrated with an external standard curve that was run every four runs. Each step of the standard curve on the external standard curve was run in triplicate. In each run, there was a no-template control, negative extraction control, and one step from the standard curve (2.0×10^3 or 2.0×10^4 copies per reaction). The recovery of SARS-CoV-2 was measured via spikes of bovine respiratory syncytial virus (BRSV) in the wastewater sample. Each sample was run in duplicate. More details on SARS-CoV-2 RNA quantification

have been reported previously (Brooks *et al.* 2021). ME samples were adjusted using the 2019 American Community Survey population size estimates. The limit of detection and quantification for ME samples was 0.76 copies/mL before adjustment. Values below this limit were estimated to be half the concentration of the limit.

COVID-19 case count data

In MA, daily county-level case counts were available for each of the counties served by the MWRA (Middlesex, Suffolk, Norfolk, and Plymouth) (COVID-19 Response Reporting n.d.). Since not all cities and towns in these counties are served by the MWRA, we estimated the percentage of each county's population served by the MWRA using population size estimates (*U.S. Census Bureau QuickFacts n.d.*). As Middlesex County is served by both the Northern and Southern MWRA systems, we estimated the population served by each MWRA system in this county. Supplemental Table 1 shows the percentage of each county serviced by the Northern or Southern MWRA systems. The percentage of the county population served by the MWRA was multiplied by new daily laboratory-confirmed cases in each county to obtain the estimated daily case counts in the areas served by the Northern and Southern systems (Archive of COVID-19 Cases in Massachusetts 2021). We assumed that the total number of new cases per population had constant spatial distribution across each city and town within a county. For CT, we used the number of laboratory-confirmed new cases per day in New Haven as reported by the Connecticut Department of Public Health (COVID-19 Daily Report 2020). For ME, we obtained COVID-19 case counts (number of reported cases per week in ZIP codes within Portland and Westbrook/Gorham) from the Maine Center for Disease Control and Prevention (Coronavirus Disease 2019 (COVID-19) 2021). For consistency with the MA and CT data, we estimated daily cases in ME as the weekly reported case count divided by seven. Population sizes in all cities and towns were estimated using the 2019 American Community Survey (*U.S. Census Bureau QuickFacts n.d.*).

Covariate data

Models for each geographic region were fit separately. The covariates in each model were ambient temperature (all regions), flow rate (MA and ME), and facility (Northern or Southern for MA; Portland or Westbrook/Gorham for ME; not a covariate for CT since only one facility was included). We included temperature because it can affect the ability to detect SARS-CoV-2 RNA in wastewater (Mandal *et al.* 2020). We included flow rate because this has been suggested as a potential confounding factor in previous wastewater-based models of COVID-19 (Weidhaas *et al.* 2021). Average daily temperature data were obtained from the National Oceanic and Atmospheric Administration database (National Centers for Environmental Information 2021). Daily temperature in Boston was used to estimate both Northern and Southern facility temperatures for MA. Daily temperature in Portland was used to estimate the temperature in Portland, Westbrook, and Gorham, ME. Flow rate of influent wastewater was recorded at the time of sample collection at the MWRA (MA) and PWD (ME). Flow rate did not apply to the ESWPAF because the SARS-CoV-2 RNA concentration was calculated using normalized values from sludge instead of wastewater.

Statistical analysis

We compared case counts, SARS-CoV-2 RNA concentrations, and flow rates for facilities within states using Wilcoxon matched pair sign rank tests. We evaluated the correlation between daily average temperature and copies/mL of SARS-CoV-2 RNA at each facility using Spearman rank correlation. We assessed the association between SARS-CoV-2 RNA copies/mL in wastewater (or sludge) and daily COVID-19 measures using four modeling approaches (linear, generalized additive with a cubic regression spline, Poisson, and negative binomial models). SARS-CoV-2 RNA copies/mL data were log-transformed. Weeks with zero reported cases or zero copies detected/mL were removed from the analysis (n=3 in ME; 7/27/2020, 10/12/2020, and 2/21/2021). For the linear and generalized additive models, the outcome measure was natural log-transformed new COVID-19 cases per day. For the Poisson and negative binomial models, the outcome measure was the count of reported new cases per day. Each model was estimated twice: once unadjusted for potential confounders and once adjusted for average daily temperature (degrees Celsius), flow rate (millions of gallons per day; MA and ME only), and facility (MA and ME only). Adjusted models were assessed for multicollinearity, and variance inflation factors were less than or equal to three. The adjusted models (including flow rate and facility) were computed as follows:

$$\text{Linear model: } \ln(\text{new cases/population}) = \beta_0 + \beta_1 \ln(\text{SARS} - \text{CoV} - 2) + \beta_2 \text{temperature} + \beta_3 \text{ flow rate} + \beta_4 \text{ facility}$$

$$\text{Generalized additive model: } \ln(\text{new cases/population}) = \gamma_0 + f(\ln(\text{SARS} - \text{CoV} - 2)) + \gamma_1 \text{temperature} + \gamma_2 \text{flow rate} + \gamma_3 \text{ facility}$$

$$\text{Poisson and negative binomial model: new cases} = \alpha_0 + \alpha_1 \ln(\text{SARS} - \text{CoV} - 2) + \alpha_2 \text{temperature} + \alpha_3 \text{flow rate} + \alpha_4 \text{ facility}$$

We conducted a sensitivity analysis using weekly COVID-19 outcome measures for ME only. Additionally, we incorporated time lags with the MA data to assess whether associations were stronger when case counts were measured four, five, and six days prior. For all models, we compared model fit using adjusted R^2 (linear and generalized additive models), McFadden's pseudo-R^2 (Poisson and negative binomial models), and root-mean-square error (RMSE; all models). All analyses were conducted in R 4.10 using the 'mgcv', 'DescTools', 'MASS', 'ggplot2, and 'foreign' packages. All data were obtained from publicly available sources, and the study was considered Not Human Subjects Research by the Tufts Institutional Review Board.

RESULTS

The COVID-19 case counts, SARS-CoV-2 RNA concentrations, ambient temperature, and flow rate in the communities served by the watersheds under study are presented in Table 1. The estimated populations served were 1,605,054 people in the Northern MWRA district of MA; 781,829 in the Southern MWRA district of MA; 130,250 in the New Haven district of CT; 37,052 residents in Westbrook and Gorham, ME; and 66,215 in Portland, ME. The trend in the number of reported new cases per day per 10,000 population (Supplemental Figure 1) generally reflected the trends in the SARS-CoV-2 concentration in wastewater and sludge (Figure 2). Copies/mL of SARS-CoV-2 were significantly different in Portland versus Westbrook and Gorham, and the documented number of cases was significantly different for the MWRA Northern and Southern facilities (Supplemental Table 2). The variation in copies/mL among sampling locations depends on sampling methods (wastewater in MA and ME versus sludge in CT; sampling frequency) and number of cases (more cases in MA than ME). In both MA and ME, flow rate varied significantly by facility (Supplemental Table 2). The correlation coefficients between average daily temperature and wastewater values of SARS-CoV-2 copies/mL were −0.79 (MA, Northern facility), −0.81 (MA, Southern facility), −0.68 (ME, Portland), −0.29 (ME, Westbrook/Gorham), and −0.36 (CT).

Log-transformed SARS-CoV-2 concentrations were significantly associated with case count measures in every model (Table 2). For MA and ME (but not CT), adjusting for covariates generally attenuated effect estimates for SARS-CoV-2 RNA. In the linear regression models for MA and ME, the ambient temperature was a significant covariate (Supplemental Table 3). Similarly, the ambient temperature was a significant predictor in the generalized additive model, Poisson, and negative binomial model types for MA and ME. The temperature was not a significant predictor in any of the models for CT (though CT data represented only three months, so temperatures were less variable). Flow rate was a significant predictor in Poisson models for MA and ME and in the negative binomial model for MA only. The facility was a significant predictor in the generalized additive (MA), Poisson (MA and ME), and negative binomial (MA and ME) models.

Table 3 presents a comparison of the model fit statistics for models relating SARS-CoV-2 RNA to COVID-19 cases in each region. In terms of R^2 values in each of the three states, the multivariable generalized additive model performed best for MA

Table 1 | Population-level characteristics[a] of the areas served by the three locations[b] over the study periods

	Massachusetts (8/19/2020–1/19/202)	Connecticut (3/1/2020–6/1/2020)	Maine (9/1/2020–3/2/2021)
Case counts per day	427 (180, 1265)	45.5 (26.0, 78.8)	12.8 (1.64, 26.9)
Daily new cases/10,000 people	1.79 (0.76, 5.30)	3.49 (2.00, 6.05)	1.24 (0.16, 2.60)
Copies of SARS-CoV-2 (copies/mL)	327 (85.0, 858)	70997 (39917, 115719)	21.1 (8.84, 57.2)
Flow rate (millions of gallons/day)	138 (114, 182)	–	7.87 (6.58, 8.94)
Temperature (degrees Celsius)	10.6 (3.33, 16.7)	9.22 (6.89, 13.2)	5.84 (−0.49, 15.3)

[a]Values presented are median (25th percentile, 75th percentile).
[b]Values in Massachusetts and Maine were computed as population size-weighted averages among facilities.

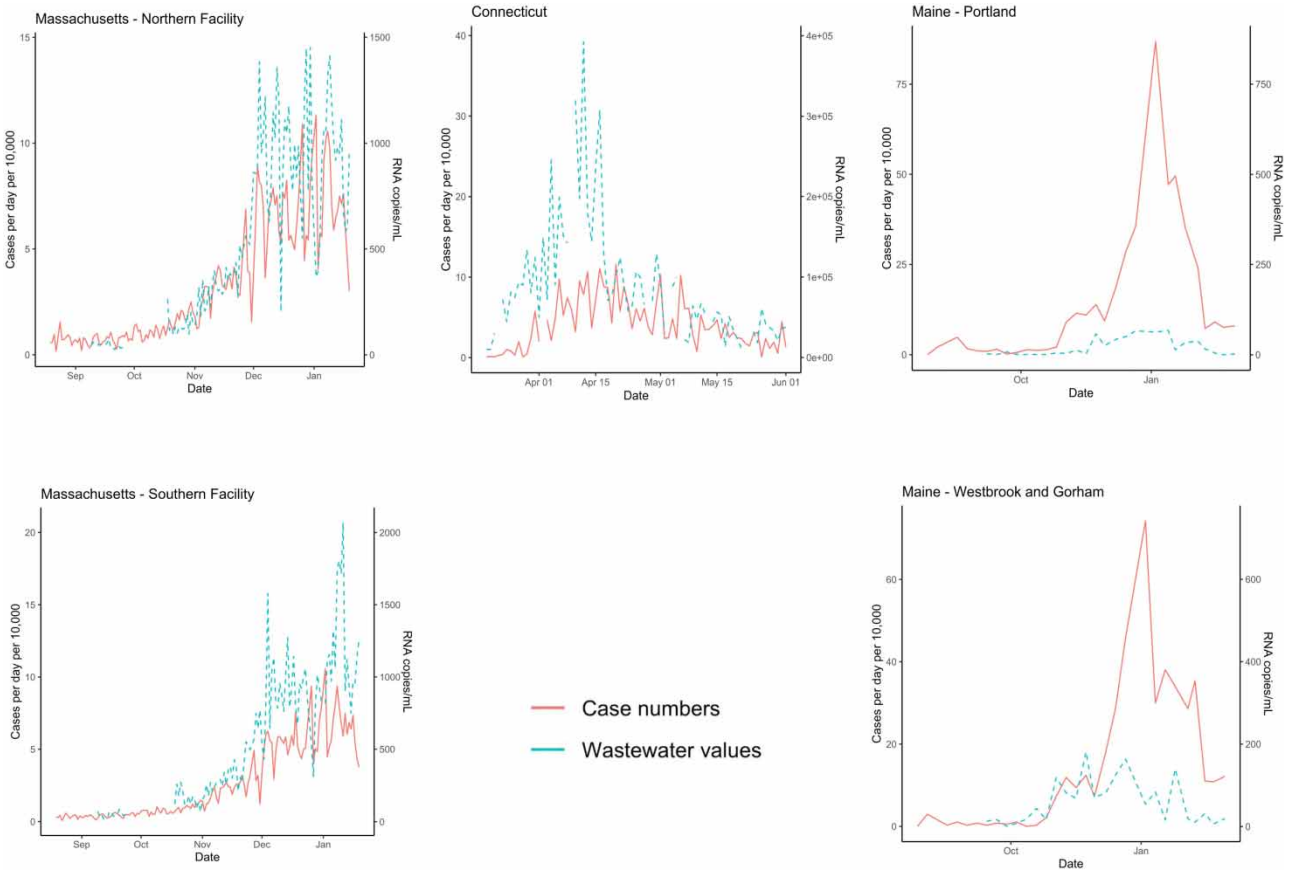


Figure 2 | Distributions of SARS-CoV-2 RNA in wastewater and sludge. The data collection time periods were 8/19/2020–1/19/2021 in Massachusetts, 3/1/2020–6/1/2020 in CT, and 9/1/2020–3/2/2021 in Maine. The vertical axis scales vary by facility.

Table 2 | Unadjusted and adjusted[a] results of statistical models relating natural log-transformed SARS-CoV-2 RNA and daily COVID-19 cases[b]

	Massachusetts	Connecticut	Maine
Unadjusted linear model; β (95% CI)	0.76 (0.71, 0.80)	0.65 (0.30, 1.00)	0.61 (0.40, 0.83)
Adjusted linear model; β (95% CI)	0.62 (0.55, 0.69)	0.71 (0.33, 1.08)	0.28 (0.05, 0.51)
Unadjusted generalized additive model; p-value	<0.01	<0.01	<0.01
Adjusted generalized additive model; p-value	<0.01	<0.01	0.03
Unadjusted Poisson model; β (95% CI)	0.67 (0.67, 0.68)	0.47 (0.43, 0.51)	0.37 (0.31, 0.44)
Adjusted Poisson model; β (95% CI)	0.59 (0.58, 0.60)	0.47 (0.43, 0.52)	0.39 (0.31, 0.47)
Unadjusted negative binomial model; β (95% CI)	0.71 (0.64, 0.77)	0.47 (0.22, 0.70)	0.49 (0.29, 0.70)
Adjusted negative binomial model; β (95% CI)	0.61 (0.55, 0.68)	0.46 (0.20, 0.72)	0.34 (0.16, 0.52)

[a]Adjusted models are adjusted for average daily temperature (degrees Celsius), flow rate (millions of gallons per day; Massachusetts and Maine only), and facility (Massachusetts and Maine only).
[b]Linear and generalized additive models estimate population-adjusted case counts. Poisson and negative binomial models estimate case numbers.

(adjusted R^2=0.86) and ME (adjusted R^2=0.65), although results were similar between generalized additive models and linear models. The multivariable Poisson model performed best for CT (pseudo-R^2=0.21 for Poisson model, adjusted R^2=0.16 for generalized additive and linear models) and also performed well for MA (pseudo-R^2=0.84). In all three states, the negative binomial model performed worst. In terms of the RMSE, the Poisson model seemed to better fit the data than the other model types for CT and ME. The RMSE values for MA were similar among the linear, generalized additive, and Poisson models.

Table 3 | Comparison of model fit for statistical models relating natural log-transformed SARS-CoV-2 RNA and daily COVID-19 cases

	Massachusetts	Connecticut	Maine
Unadjusted linear[a] model			
Adjusted R^2	0.80	0.16	0.40
Root-mean-square error	0.47	1.15	1.31
Adjusted[b] linear model			
Adjusted R^2	0.85	0.16	0.63
Root-mean-square error	0.41	1.14	0.99
Unadjusted generalized additive model			
Adjusted R^2	0.82	0.16	0.40
Root-mean-square error	0.45	1.15	1.31
Adjusted generalized additive model			
Adjusted R^2	0.86	0.16	0.65
Root-mean-square error	0.39	1.14	0.97
Unadjusted Poisson model			
Pseudo R^2	0.58	0.21	0.20
Root-mean-square error	0.66	0.67	1.02
Adjusted Poisson model			
Pseudo R^2	0.84	0.21	0.52
Root-mean-square error	0.39	0.67	0.79
Unadjusted negative binomial model			
Pseudo R^2	0.07	0.02	0.05
Root-mean-square error	0.67	0.67	1.03
Adjusted negative binomial model			
Pseudo R^2	0.15	0.02	0.16
Root-mean-square error	0.39	0.67	0.81

[a]Linear and generalized additive models estimate population-adjusted case counts. Poisson and negative binomial models estimate case numbers.
[b]Models are adjusted for average daily temperature (degrees Celsius), flow rate (millions of gallons per day; Massachusetts and Maine only), and facility (Massachusetts and Maine only).

In a sensitivity analysis using weekly case numbers for ME rather than daily numbers, results remained generally the same (Supplemental Table 4). Similar to the daily results, the adjusted generalized additive model had the highest (adjusted) R^2 (0.63), and the adjusted Poisson model had the lowest RMSE (0.78). When accounting for time lags in reporting of cases in MA, results remained similar at four-, five-, and six-day lags (Supplemental Table 5).

DISCUSSION

In this multi-state analysis of wastewater and sludge-measured SARS-CoV-2 RNA concentrations, linear, generalized additive, Poisson, and negative binomial models all indicated significant associations between SARS-CoV-2 RNA concentration and daily reported COVID-19 cases. The adjusted linear, generalized additive, and Poisson models resulted in higher R^2 and lower RMSEs across the three New England metropolitan regions. This is among the first studies to compare model fit for different types of statistical models relating SARS-CoV-2 RNA and COVID-19 cases across several locations. It is also one of the first to consider SARS-CoV-2 RNA measured in sludge. Our results could inform ongoing and future wastewater monitoring efforts to aid in pandemic preparedness and response.

Since SARS-CoV-2 was first detected in wastewater, researchers and public health practitioners have used the concentration of viral RNA to estimate COVID-19 cases in sewersheds (Ahmed *et al.* 2020). Previous research found nonparametric and parametric regression models to be useful tools when interpreting the concentration of viral particles in wastewater and sludge (McMahan *et al.* 2020; Vallejo *et al.* 2020). A study conducted by Wu *et al.* leveraged MWRA

wastewater data to suggest the number of prevalent COVID-19 cases was higher than anticipated based on individual testing, with asymptomatic cases possibly explaining the difference between clinically documented cases and observed SARS-CoV-2 shedding in wastewater (Wu *et al.* 2020). The current study builds upon these findings by using uniform modeling techniques to compare associations with reported case numbers in communities in three New England states. Our results indicate that using a linear model versus a generalized additive model versus a Poisson model does not make a large difference in terms of model fit. For example, in MA both the R^2 and the RMSEs in each of the three models only differed by a maximum of 0.02. Additionally, as the significant covariates differed by a municipality, our analysis suggests that local context (including characteristics such as flow rate) should be considered when fitting models estimating COVID-19 cases from wastewater and sludge.

The strength of association between viral RNA in wastewater or sludge and COVID-19 case counts varied by geographic region. These differences may be explained by differences in the study time period (e.g., CT data were from spring 2020, whereas MA and ME data were collected from mid to late 2020 through early 2021), number of days included in the analysis (e.g., shorter time period in CT), testing capacity (especially if testing capacity impacts COVID-19 transmission and reported case counts), and community characteristics (e.g., occupational or policy differences affecting human mobility patterns). Collectively, these differences (along with differences in sample collection methods between sludge and wastewater) may help explain why model performance was weaker in CT than in MA or ME. Other differences may be due to how viral RNA is measured in wastewater and sludge (e.g., the limits of detection for the two wastewater sites were 4.80 copies/mL in MA and 0.76 copies/mL in ME). The higher limit of detection in MA implies that some concentrations of SARS-CoV-2 RNA could have been underestimated in these facilities; however, this detection limit was constant over time and would theoretically underestimate similar values each day. Nevertheless, given the ability to detect (with >99% probability) SARS-CoV-2 RNA if there were at least one case per 6,500 people in the MA area served and given the large population served, the somewhat higher detection limit was not likely to substantively affect the analysis. Furthermore, while the negative binomial regression relaxes the assumption of equal mean and variance, it is possible that the negative binomial model was consistently outperformed by the other models because it is more sensitive to smaller sample sizes than other similar models (e.g., Poisson regression), and our analysis was based on relatively small sample sizes (especially for CT) (Piegorsch 1990; Tang 2015).

Additionally, the performance of statistical models in the COVID-19 context may depend on factors related to viral shedding that we did not account for in our analysis. The viral shedding rate is still unclear, as are differences in viral shedding due to different COVID-19 variants and based on differences by vaccination status (Wölfel *et al.* 2020; Mallavarpu Ambrose *et al.* 2021). Some work suggests that the shedding rate may vary between 10^2 and 10^8 copies/g throughout a 14–21 day shedding period, but further research is still needed to determine the duration and amount of shedding throughout the disease course (Randazzo *et al.* 2020; Wölfel *et al.* 2020). Additionally, in interpreting the output of these statistical models, public health officials should remember that the models do not account for local outbreaks or clusters within a sewershed and should instead be used to estimate changes in the number of cases within the community for which the model is being fit.

Other factors that can affect the validity of the wastewater and sludge estimates for SARS-CoV-2 RNA concentration include precipitation and flow rate. All sewer systems where data were collected had combined sewer for at least half of their sewer systems; these combined systems carry stormwater and wastewater together (Hartford | The Clean Water Project n.d.; Wastewater Maintenance | Portland ME n.d.; MWRA – Combined Sewer Overflow Control Program 2022). In our analyses, precipitation was not a significant covariate (results with precipitation not shown). In ME, sampling dates were at times changed based on rain, so this could potentially explain the nonsignificant results in ME. Flow rate can partially account for precipitation; nonetheless, wastewater studies without flow rate data should consider precipitation rates (Wu *et al.* 2020).

WBE is a useful tool that, as evidenced by our findings, can be implemented through several straightforward statistical modeling types with data collected from either sludge or wastewater. Public health departments and organizations considering wastewater monitoring should work in partnership with local water and wastewater utilities that collect samples. We suggest that wastewater treatment facilities develop a sampling plan that includes 24-h composite sampling, as was done in CT and ME. We also encourage continued data sharing of results in a timely manner to advance public health goals of predicting and responding to pandemic conditions.

We found that incorporating a time lag did not impact the results greatly, so a time lag could be used depending on the average time from test to reporting but may only marginally affect results due to day-to-day similarities in RNA concentrations and cases. We found that including facility as a covariate affected the results and interpretation, suggesting that models may need to consider local context. Accounting for local context explicitly within models will also help to inform local public health measures, such as mask mandates, mobile testing centers, and vaccination clinic targeting, and prediction and

forecasting efforts by health departments. These types of applications of WBE have been implemented globally. For example, South Africa has been using WBE to surveille the spread of COVID-19 and track new variants (Aguiar-Oliveira *et al.* 2020; Kumar *et al.* 2020; Fuschi *et al.* 2021).

There were several limitations of this study. First, the SARS-CoV-2 wastewater concentrations and number of cases in ME were measured weekly rather than daily in MA and CT. For consistency with the MA and CT data, we calculated daily values for ME by dividing the reported weekly cases by seven, assuming that there were no temporal trends over the course of the week. Our sensitivity analysis suggests that this assumption was reasonable, and future applications may prefer 7-day moving averages to smooth out the noise of individual day outliers (e.g., holiday and weekend testing differences). There could also have been differences in testing capabilities among the locations, which could have resulted in certain cities or towns having higher reported case numbers than others compared to the number of true cases. This limitation highlights the importance of WBE as a method of disease surveillance, as it can capture the total case burden, including those cases that may not be diagnosed. Second, we estimated population size for each sewershed without accounting for time-varying differences in population due to employment, education, or other trends. This could underestimate the strength of association if, for example, many more people work in the sewershed than reside there (thus contributing to the wastewater influent but not case counts). Relatedly, we assumed that everyone within the study communities contributed to public wastewater, though in certain areas some of the population may use private septic systems. Third, the timing, sampling, and analytic methods for wastewater and sludge collection and analysis varied by municipality. This could result in variable data quality and model performance. Fourth, daily viral RNA concentrations measured at the different facilities within MA and ME may not be independent of the other facility; if they were not independent, the standard errors may be artificially inflated. Relatedly, we did not account for temporal autocorrelation in any modeling strategy. Other modeling approaches, such as auto-regressive models, could address this unsatisfied assumption and could be considered in the future (Akaike 1969). Fifth, average daily temperature was, in some cases, highly correlated with SARS-CoV-2 copies/mL, which raises some concern for multicollinearity. We assessed multicollinearity using the variance inflation factor and did not find strong evidence; however, we cannot rule out that this assumption was violated. Given that temperature is an important factor in social distancing practices and evident case numbers, we felt it was important to incorporate it into the model. Sixth, we only assessed the effect of a time lag in MA, as we had data on the largest number of days in this location. CT, while sampled daily, had data on a shorter time period, and ME data were collected weekly. Nonetheless, we did not find a strong difference in the unlagged versus four-, five-, or six-day lagged models in MA.

Our study also had several strengths. We used consistent data management and statistical approaches to assess the relationship between viral RNA concentration and COVID-19 case counts in multiple metropolitan New England areas. We were able to adjust for relevant covariates such as flow rate, ambient temperature, and facility. Finally, we were able to compare four frequently used statistical modeling approaches. Our work could thus inform future community-level COVID-19 surveillance efforts leveraging wastewater-based modeling. Future studies should assess whether the statistical modeling types and/or specific model parameters we estimated vary in their ability to predict COVID-19 trends prospectively and whether the model parameters need to be tuned to local conditions for communities outside of New England.

CONCLUSIONS

We observed that SARS-CoV-2 RNA concentrations in wastewater and sludge were significantly associated with local COVID-19 case numbers in MA, CT, and ME. The linear, generalized additive, and Poisson modeling strategies outperformed the negative binomial models. Public health surveillance efforts can confidently use any one of multiple statistical approaches to estimate COVID-19 burden in communities.

ACKNOWLEDGEMENTS

We would like to thank Eilidh Sidaway, Brianna Shelley, Bailey Gryskwicz for laboratory analyses at Saint Joseph's College of Maine, Dr Jordan Peccia for providing the data from New Haven, and the Tufts Data Lab for assistance with figure generation.

FUNDING STATEMENT

Support for this project was provided by NEIWPCC (www.neiwpcc.org), a regional commission that helps the states of the Northeast preserve and advance water quality. The support was in-kind, consisting of the staff time of Dr Christina Stringer. Dr Stringer served in an advisory capacity, overseeing the work conducted by the primary authors, Elyssa Anneser and Emily Riseberg. This included guidance and oversight of the study design, as well as analysis and interpretation of data, and playing a minor role in report writing and article submission. Laura Corlin was supported by Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) grant number K12HD092535 and by the Tufts University/Tufts Medical Center Rapid Response Seed Funding Program. Portland Water District funded the project in Portland, Westbrook, and Gorham, Maine.

CONFLICTS OF INTEREST

Y. Brooks received some material supplies from IDEXX. E. Anneser, E. Riseberg, L. Corlin, and C. Stringer declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories https://github.com/emilyriseberg/COVID_WBE.

REFERENCES

Aguiar-Oliveira, M. d. L., Campos, A. R., Matos, A., Rigotto, C., Sotero-Martins, A., Teixeira, P. F. P. & Siqueira, M. M. 2020 Wastewater-based epidemiology (WBE) and viral detection in polluted surface water: a valuable tool for COVID-19 surveillance – a brief review. *International Journal of Environmental Research and Public Health* **17** (24), 9251. https://doi.org/10.3390/ijerph17249251

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**, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764.

Akaike, H. 1969 Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21** (1), 243–247. https://doi.org/10.1007/BF02532251.

Archive of COVID-19 Cases in Massachusetts 2021 Mass.Gov. Available from: https://www.mass.gov/info-details/archive-of-covid-19-cases-in-massachusetts

Brooks, Y. M., Gryskwicz, B., Sheehan, S., Piers, S., Mahale, P., McNeil, S., Chase, J., Webber, D., Borys, D., Hilton, M., Robinson, D., Sears, S., Smith, E., Lesher, E. K., Wilson, R., Goodwin, M. & Pardales, M. 2021 Detection of SARS-CoV-2 in wastewater at Residential College, Maine, USA, August–November 2020. *Emerg Infect Dis J – CDC* **27** (12). https://doi.org/10.3201/eid2712.211199.

CDC 2020 *Research Use Only 2019-Novel Coronavirus (2019-NCoV) Real-Time RT-PCR Primers and Probes*. Centers for Disease Control and Prevention. Available from: http://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html

CDC 2021 *Symptoms of COVID-19*. Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

Coronavirus Disease 2019 (COVID-19) 2021 Division of Disease Surveillance | Maine.Gov. Available from: https://www.maine.gov/dhhs/mecdc/infectious-disease/epi/airborne/coronavirus/index.shtml

COVID-19 Daily Report 2020. Available from: https://data.ct.gov/stories/s/COVID-19-Daily-Report/q5as-kyim/

COVID-19 Response Reporting n.d. Mass.Gov. Available from: https://www.mass.gov/info-details/covid-19-response-reporting (Accessed December 17 2021)

Feng, S., Roguet, A., McClary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G. & McLellan, S. L. 2021 Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. *ACS ES&T Water* **1** (8), 1955–1965. https://doi.org/10.1021/acsestwater.1c00160.

Fuschi, C., Pu, H., Negri, M., Colwell, R. & Chen, J. 2021 Wastewater-based epidemiology for managing the COVID-19 pandemic. *ACS ES&T Water* **1** (6), 1352–1362. https://doi.org/10.1021/acsestwater.1c00050.

Goulding, N., Hickman, M., Reid, M., Amundsen, E. J., Baz-Lomba, J. A., O'Brien, J. W., Tscharke, B. J., de Voogt, P., Emke, E., Kuijpers, W., Hall, W. & Jones, H. E. 2020 A comparison of trends in wastewater-based data and traditional epidemiological indicators of stimulant consumption in three locations. *Addiction (Abingdon, England)* **115** (3), 462–472. https://doi.org/10.1111/add.14852.

Hartford | The Clean Water Project n.d. Available from: https://www.thecleanwaterproject.com/in-your-town/hartford (Accessed March 19 2022).

He, J., Guo, Y., Mao, R. & Zhang, J. 2021 Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *Journal of Medical Virology* **93** (2), 820–830. https://doi.org/10.1002/jmv.26326.

Kumar, M., Patel, A. K., Shah, A. V., Raval, J., Rajpara, N., Joshi, M. & Joshi, C. G. 2020 First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2. *Science of the Total Environment* **746**, 141326. https://doi.org/10.1016/j.scitotenv.2020.141326.

La Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L. & Suffredini, E. 2020 First detection of SARS-CoV-2 in untreated wastewaters in Italy. *The Science of the Total Environment* **736**, 139652. https://doi.org/10.1016/j.scitotenv.2020.139652.

Mallavarpu Ambrose, J., Priya Veeraraghavan, V., Kullappan, M., Chellapandiyan, P., Krishna Mohan, S. & Manivel, V. A. 2021 Comparison of immunological profiles of SARS-CoV-2 variants in the COVID-19 pandemic trends: an immunoinformatics approach. *Antibiotics* **10** (5), 535. https://doi.org/10.3390/antibiotics10050535.

Mandal, P., Gupta, A. K. & Dubey, B. K. 2020 A review on presence, survival, disinfection/removal methods of coronavirus in wastewater and progress of wastewater-based epidemiology. *Journal of Environmental Chemical Engineering* **8** (5), 104317. https://doi.org/10.1016/j.jece.2020.104317.

McMahan, C. S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Deaver, J. A., Popat, S., Karanfil, T. & Freedman, D. L. 2020 *COVID-19 Wastewater Epidemiology: A Model to Estimate Infected Populations*. https://doi.org/10.1101/2020.11.05.20226738.

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. 2020 Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in The Netherlands. *Environmental Science & Technology Letters* **7** (7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357.

MWRA – Combined Sewer Overflow Control Program 2022. Available from: https://www.mwra.com/03sewer/html/sewcso.htm

MWRA – Wastewater COVID-19 Tracking n.d. Available from: https://www.mwra.com/biobot/biobotdata.htm (Accessed March 30 2022).

National Centers for Environmental Information 2021 *National Oceanic and Atmospheric Administration (NOAA)*. Available from: https://www.ncdc.noaa.gov/

Omori, R., Mizumoto, K. & Chowell, G. 2020 Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *International Journal of Infectious Diseases* **94**, 116–118. https://doi.org/10.1016/j.ijid.2020.04.021.

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. *Nature Biotechnology* **38** (10), 1164–1167. https://doi.org/10.1038/s41587-020-0684-z.

Piegorsch, W. W. 1990 Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics* **46** (3), 863–867. https://doi.org/10.2307/2532104.

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D. L., Singer, A. C., Graham, D. W. & Romalde, J. L. 2020 Making waves: wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Research* **186**, 116404. https://doi.org/10.1016/j.watres.2020.116404.

QuickFacts n.d. *United States Census Bureau*. Available from: https://www.census.gov/quickfacts/fact/table/MA,chelseacitymassachusetts/PST045219 (Accessed February 9 2021).

Randazzo, W., Cuevas-Ferrando, E., Sanjuán, R., Domingo-Calap, P. & Sánchez, G. 2020 Metropolitan wastewater analysis for COVID-19 epidemiological surveillance. *International Journal of Hygiene and Environmental Health* **230**, 113621. https://doi.org/10.1016/j.ijheh.2020.113621.

Shah, S., Gwee, S. X. W., Ng, J. Q. X., Lau, N., Koh, J. & Pang, J. 2022 Wastewater surveillance to infer COVID-19 transmission: a systematic review. *The Science of the Total Environment* **804**, 150060. https://doi.org/10.1016/j.scitotenv.2021.150060.

Sinclair, R. G., Choi, C. Y., Riley, M. R. & Gerba, C. P. 2008 Pathogen surveillance through monitoring of sewer systems. *Advances in Applied Microbiology* **65**, 249–269. https://doi.org/10.1016/S0065-2164(08)00609-6.

Tang, Y. 2015 Sample size estimation for negative binomial regression comparing rates of recurrent events with unequal follow-up time. *Journal of Biopharmaceutical Statistics* **25** (5), 1100–1113. https://doi.org/10.1080/10543406.2014.971167.

Vallejo, J. A., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B. K., Nasser, M., Cid, Á., Veiga, M. C., Acevedo, A., Lamora, C., Bou, G., Cao, R. & Poza, M. 2020 Predicting the number of people infected with SARS-COV-2 in a population using statistical models based on wastewater viral load. *MedRxiv*. https://doi.org/10.1101/2020.07.02.20144865.

Vogels, C. B. F., Brito, A. F., Wyllie, A. L., Fauver, J. R., Ott, I. M., Kalinich, C. C., Petrone, M. E., Casanovas-Massana, A., Muenker, C., Moore, A. J., Klein, J., Lu, P., Lu-Culligan, A., Jiang, X., Kim, D. J., Kudo, E., Mao, T., Moriyama, M., Oh, J. I., Park, A., Silva, J., Song, E., Takahashi, T., Taura, M., Tokuyama, M., Venkataraman, A., Weizman, O-E., Wong, P., Yang, Y., Cheemarla, N. R., White, E. B., Lapidus, S., Earnest, R., Geng, B., Vijayakumar, P., Odio, C., Fournier, J., Bermejo, S., Farhadian, S., Dela Cruz, C. S., Iwasaki, A., Ko, A. I., Landry, M. L., Foxman, E. F. & Grubaugh, N. D. 2020 Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology* **5** (10), 1299–1305. https://doi.org/10.1038/s41564-020-0761-6.

Wang, J., Li, W., Yang, B., Cheng, X., Tian, Z. & Guo, H. 2021 Impact of hydrological factors on the dynamic of COVID-19 epidemic: a multi-region study in China. *Environmental Research* **198**, 110474. https://doi.org/10.1016/j.envres.2020.110474.

Wastewater Maintenance | Portland, ME n.d. Available from: https://www.portlandmaine.gov/489/Wastewater-Maintenance (Accessed March 19 2022).

Weidhaas, J., Aanderud, Z. T., Roper, D. K., VanDerslice, J., Gaddis, E. B., Ostermiller, J., Hoffman, K., Jamal, R., Heck, P., Zhang, Y., Torgersen, K., Laan, J. V. & LaCross, N. 2021 Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *The Science of the Total Environment* **775**, 145790. https://doi.org/10.1016/j.scitotenv.2021.145790.

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. *Nature* **581** (7809), 465–469. https://doi.org/10.1038/s41586-020-2196-x.

Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W. L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A. D., Erickson, T. B., Chai, P. R., Thompson, J. & Alm, E. J. 2020 SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *MSystems* **5** (4). https://doi.org/10.1128/mSystems.00614-20.

Zarei, S., Salimi, Y., Repo, E., Daglioglu, N., Safaei, Z., Güzel, E. & Asadi, A. 2020 A global systematic review and meta-analysis on illicit drug consumption rate through wastewater-based epidemiology. *Environmental Science and Pollution Research International* **27** (29), 36037–36051. https://doi.org/10.1007/s11356-020-09818-6.

First received 14 January 2022; accepted in revised form 15 April 2022. Available online 26 April 2022