

# Predicting daily COVID-19 case rates from SARS-CoV-2 RNA concentrations across a diversity of wastewater catchments

Alessandro Zulli<sup>1</sup>, Annabelle Pan<sup>1</sup>, Stephen M. Bart<sup>2,3</sup>, Forrest W. Crawford<sup>4,5,6,7</sup>, Edward H. Kaplan<sup>1,7,8</sup>, Matthew Cartter<sup>3,8</sup>, Albert I. Ko<sup>9</sup>, Duncan Cozens<sup>10</sup>, Marcela Sanchez<sup>1</sup>, Doug E. Brackney<sup>8,10</sup>, and Jordan Peccia<sup>1,\*</sup>

<sup>1</sup>*Department of Chemical and Environmental Engineering, School of Engineering and Applied Science, Yale University, New Haven, CT, USA*

<sup>2</sup>*Epidemic Intelligence Service, Centers for Disease Control and Prevention*

<sup>3</sup>*Connecticut Department of Public Health, Hartford, CT, USA*

<sup>4</sup>*Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, USA*

<sup>5</sup>*Department of Statistics and Data Science, Yale University, New Haven, CT, USA*

<sup>6</sup>*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA*

<sup>7</sup>*Yale School of Management, Yale University, New Haven, CT, USA*

<sup>8</sup>*Yale School of Public Health, Yale University, New Haven, CT, USA*

<sup>9</sup>*Department of Epidemiology of Microbial Disease, Yale School of Public Health, Yale University, New Haven, CT, USA*

<sup>10</sup>*Connecticut Agricultural Experimental Station, State of Connecticut, New Haven, CT, USA*

\*Corresponding authors: Jordan Peccia, Tel: +1 (203) 432-4385; email: [Jordan.Peccia@Yale.edu](mailto:Jordan.Peccia@Yale.edu)

## Abstract

We assessed the relationship between municipality COVID-19 case rates and SARS-CoV-2 concentrations in the primary sludge of corresponding wastewater treatment facilities. Over 1,000 daily primary sludge samples were collected from six wastewater treatment facilities with catchments serving 18 cities and towns in the State of Connecticut, USA. Samples were analyzed for SARS-CoV-2 RNA concentrations during a six-month time period that overlapped with fall 2020 and winter 2021 COVID-19 outbreaks in each municipality. We fit a single regression model to estimate reported case rates in the six municipalities from SARS-CoV-2 RNA concentrations collected daily from corresponding wastewater treatment facilities. Results demonstrate the ability of SARS-CoV-2 RNA concentrations in primary sludge to estimate COVID-19 reported case rates across treatment facilities and wastewater catchments, with coverage probabilities ranging from 0.94 to 0.96. Leave-one-out cross validation suggests that the model can be broadly applied to wastewater catchments that range in more than one order of magnitude in population served. Estimation of case rates from wastewater data can be useful in locations with limited testing availability or testing disparities, or delays in individual COVID-19 testing programs.

## Keywords

SARS-CoV-2, wastewater-based epidemiology, primary sludge, regression analysis, case rate estimation, quantitative PCR

## Introduction

Wastewater surveillance has the potential to identify and track outbreaks of human pathogens that demonstrate gut tropism, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19.<sup>1</sup> Recent studies demonstrate that SARS-CoV-2 RNA concentrations in domestic wastewater reflect the rise and fall of COVID-19 cases based on daily positive tests in a community,<sup>2-4</sup> and important epidemiological parameters such as the effective reproduction number can be estimated using wastewater SARS-CoV-2 concentrations, in addition to other epidemic indicators.<sup>5</sup> If SARS-CoV-2 wastewater results are rapidly reported, they could provide a leading indicator of community infection rates over COVID-19 case rates and hospital admissions data.<sup>5-7</sup>

Estimates of infection based on COVID-19 case rates from testing of human specimens have been the standard for applying community interventions aimed at decreasing morbidity and mortality. Alternative approaches to COVID-19 testing are necessary to estimate the number of infections in locations with disparities in testing practices<sup>8</sup>, limited testing resources,<sup>9</sup> or during outbreaks when testing capacity cannot meet demand. SARS-CoV-2 RNA concentration measurements in wastewater may resolve these issues, but quantitative relationships between SARS-CoV-2 concentration in wastewater and infection (or proxies of infection such as cases) are not well-resolved.

The following manuscript reports 192 days of monitoring daily primary sludge for SARS-CoV-2 concentrations across six different U.S. wastewater treatment facilities ranging in over one order of magnitude in flow rate, serving 18 municipalities and approximately 1 million residents in the State of Connecticut, USA. Regression models were evaluated to estimate

community COVID-19 reported case rates from a time-course of primary sludge SARS-CoV-2 RNA concentrations.

## Materials and Methods

**Wastewater sampling.** Primary sewage sludge samples of 40 to 45 ml in volume were collected daily from six wastewater treatment plants (WWTP) in the State of Connecticut, USA. **Table 1** lists the cities and towns served, and details the specific plant processes from which primary sludge was produced and samples were withdrawn. Grab samples from Stamford, Bridgeport, New Haven and New London were collected daily from August 3, 2020 to February 10, 2021. Samples from Hartford were collected daily from August 10, 2020 to February 10, 2021. Samples from Norwich were collected daily from August 17, 2020 to February 10, 2021. All samples were collected between 8 am and 9 am, stored at -20° C before being transported to Yale University laboratories on ice, and analyzed immediately upon arrival.

**SARS-CoV-2 RNA primary sludge concentrations.** To quantify SARS-CoV-2 RNA concentrations, the 40 to 45 ml sample was mixed on a vortexer for 1 minute and 0.5 ml of primary sludge was added to a commercial extraction kit optimized for the isolation of total RNA from raw wastewater (Zymo, *Quick-RNA Fecal/Soil Microbe Microprep*, wastewater protocol). Modifications to the extraction protocol included the addition of 0.1 ml of phenol/chloroform/isoamyl alcohol (25:24:1) in the initial bead beating step and eluting isolated RNA into 50 µl of ribonuclease-free water. Before September 20, 2020 and prior to the rapid rise in case rates in early October, primary sludge RNA extraction was accomplished using the RNeasy PowerSoil Total RNA kit (Qiagen) as previously described.<sup>6</sup> For all RNA extracts, total RNA was measured by spectrophotometry and normalized to 200 ng µL<sup>-1</sup> (NanoDrop, Thermo

Fisher Scientific) prior to quantitative PCR to account for daily variations in sludge solids content.

SARS-CoV-2 RNA concentration was estimated using one-step qRT-PCR with SARS-CoV-2 N1 and N2 primer sets.<sup>10, 11</sup> CrAssphage, a ubiquitous bacteriophage that is highly concentrated in the human gut,<sup>12-14</sup> was quantified as a PCR control. All primer sets were run in separate reactions. Analysis was conducted using a one-step RT-qPCR kit (BioRad iTaq™ Universal Probes One-Step Kit). Triplicate 20 µL reactions using a 5x diluted template were run at 55° C for 10 min and 95° C for 1 min, followed by 40 cycles consisting of 95°C for 10 seconds and 55° C for 30 seconds<sup>6</sup> and Ct values for triplicates were average. N1 and N2 primer set standards were constructed using ten-fold dilutions ( $5 \times 10^1$  to  $5 \times 10^8$  copies per reaction) of the N gene transcripts for the SARS-CoV-2 Wuhan-Hu-1 strain. Concentrations in primary sludge samples were calculated using these standard curves and are presented as SARS-CoV-2 RNA copies per ng of total extracted RNA. Positive RNA and negative (no template) controls were included in all qRT-PCR runs. Less than 1.5% of samples were not available from the treatment facilities. In these cases, SARS-CoV-2 concentrations used in regression models were estimated by interpolating between the prior and subsequent measured SARS-CoV-2 concentration.

**COVID-19 test data for the cities served by the treatment facilities.** The number of confirmed and probable COVID-19 cases per the Council of State and Territorial Epidemiologists case definition<sup>15</sup> was provided by the Connecticut Department of Public Health (CT DPH) and used in this analysis. The date for each case was assigned, in order of preference, as test specimen collection date (if available), symptom onset date (if available), or date of report to the CT DPH. Greater than 90% of reported case values were by date of clinical specimen collection. Average test positivity rates over the study period were the following: Stamford

(6.2%), Bridgeport (6.1%), New Haven (3.7%), Hartford (6.0%), New London (4.6%), Norwich (6.5%).

Cases were compiled from CT DPH data from the individual towns served by each wastewater treatment plant (**Table 1**) and adjusted per 100,000 population based on town census size. Under the assumption that each member of a population contributes an equivalent volumetric flow of raw wastewater and primary sludge across the cities and towns considered, the SARS-CoV-2 primary sludge concentrations reported here represent a per capita basis.

**Model development.** Multiple regression analysis was used to fit COVID-19 reported cases per 100,000 population on the day of specimen collection to observed values of SARS-CoV-2 RNA copies per ng of total RNA in primary sludge. The model structure is

$$case\ rate_{i,t} = \alpha_0 + \alpha_i + \sum_{j=0}^{\tau} (\beta_j + \beta_{i,j}) RNA_{i,t-j} + \sum_{d=2}^7 (\gamma_d + \gamma_{i,d}) 1\{w_t = d\} + \epsilon_{it}$$

(eq. 1)

where the reported case rate is the number of new cases reported by date of specimen collection per 100,000 population,  $i$  represents municipality,  $t$  represents time (days),  $\alpha_0$  is the model offset,  $\alpha_i$  is the individual municipality model offset,  $j$  is the number of lagged days,  $\tau$  is the maximum number of lagged days,  $RNA_{i,t-j}$  are RNA concentrations (SARS CoV-2 RNA copies ng<sup>-1</sup> RNA),  $w_t$  indicates the day of the week on day  $t$ ,  $\beta_j$ ,  $\beta_{i,j}$  and  $\gamma_d$  and  $\gamma_{i,d}$  are regression coefficients,  $d$  is day of the week (Friday is the reference day), and  $\epsilon_{it}$  is the residual error associated with each observation. Day of the week addresses the known daily variations in testing behavior.<sup>16</sup> Adjustment for population, which shows a strong collinearity with average treatment plant flow (simple linear regression: slope = 0.9 m<sup>3</sup>d<sup>-1</sup> person<sup>-1</sup>, R<sup>2</sup>=0.93, p=0.002),

was accomplished by considering the case rate per 100,000 residents as the dependent variable. A weighted least squares approach was used to weight (by reciprocal of the variance) case rate contributions to the model and reduce heteroskedasticity. A single model covering all treatment facilities was fit using the entire RNA concentration and case rate data set from all six treatment facilities from August 3, 2020 to February 10, 2021.

The statistical significance for predictive parameters was analyzed through a two-tailed t-test performed on the regression coefficients for all RNA concentration lags, and treatment plant and day of the week offsets. Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) F test, and coverage probability (calculated as the proportion of times that measured case rates fell within the 95% prediction interval of the model-estimated case rates), were used to select the maximum number of lagged RNA concentrations included in the model. A leave-one-out cross validation was utilized to test the model's ability to estimate COVID-19 reported case rates for other wastewater catchments. Six regression models, similar to **Equation 1**, with the exception that municipality was not included, were trained by leaving out data from one of the six treatment facilities. These models were then utilized to estimate case rates in the cities covered by the treatment facility excluded from the model training set.

## Results

**SARS-CoV-2 concentrations.** The SARS-CoV-2 RNA concentrations in primary sludge for the six treatment facilities are shown in **Figure 1** from August 3, 2020 to February 10, 2021 and reflect fall COVID-19 outbreaks in each city. The bottom row of **Figure 1** displays the reported case rates over the same time period in the municipalities served by the six treatment plants. The timing of SARS-CoV-2 RNA concentrations in primary sludge visually tracks the dynamics in

reported case rates, with observed increases in primary sludge RNA concentrations at each plant coinciding with the start of the fall 2020 outbreaks for the corresponding cities.

**Model selection and validation.** This study estimated COVID-19 case rates using the regression model presented in equation 1 which included lagged SARS-CoV-2 RNA wastewater concentrations, municipality, and day of the week as variables. An initial analysis was conducted to determine the maximum number of prior days of wastewater RNA concentrations ( $RNA_{i,t-j}$ , lagged days) to use. **Table 2** presents Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), F statistics with p values, and coverage probabilities for maximum lags ranging from 0 to 6 days and indicates optimal model fit with minimized model complexity when using lagged RNA concentrations from 0 to 4 days. Graphical comparisons of measured reported case rates versus model-estimated case rates are provided in **Figure 2** for each WWTP using models with lagged RNA concentrations from 0 to 4 days. When applied to the six individual municipalities, resulting coverage probabilities range was 0.93 to 0.96, with a root mean standard error of 13.1 cases per 100,000 population for the entire model (see **Figure S1** for model confidence intervals). **Tables S1-7** provide regression coefficients, intercepts, and their significance for the general linear models in **Equation 1**. Regression coefficients were typically greatest at lags 0 and 1 day.

Results of the leave-one-out cross validation analysis revealed coverage probabilities (95% prediction intervals) of the estimated case rates that ranged from 0.90 to 0.98 for the six different municipalities considered, suggesting that the model can provide estimates of measured cases from sludge SARS-CoV-2 RNA concentrations for a variety of wastewater catchments that contain treatment facilities that produce primary sludge (**Table 3**).



## Discussion

Estimating the number of infections directly from wastewater pathogen concentrations is a central goal of wastewater surveillance practice. The gut tropism of coronaviruses and widespread COVID-19 testing provide a unique opportunity to develop these tools. We monitored daily SARS-CoV-2 RNA concentrations in the primary sludge of six wastewater treatment plants that covered 18 U.S. municipalities and explored regression analyses to estimate COVID-19 cases rates on the day of specimen collection. Our results demonstrate the feasibility, utility, and simplicity of estimating COVID-19 case rates across a variety of different wastewater treatment catchments.

Prior studies have noted a concordance in wastewater SARS-CoV-2 RNA concentrations with other indicators of infection. Early work in the 2020 COVID-19 outbreak observed similar behavior between reported daily positive COVID-19 testing results in a community and the SARS-CoV-2 RNA concentrations in that community's wastewater,<sup>17</sup> and relationships between wastewater RNA concentrations and COVID-19 reported cases have been reported for SARS-CoV-2 and polio virus.<sup>2, 7, 18, 19</sup> Several features of this study are unique to these prior studies and mark advances in the science underpinning wastewater-based epidemiology. We estimate COVID-19 case rates solely from current and prior days of primary sludge RNA concentrations and day of the week across treatment plants that range in more than one order of magnitude in size as measured by average flow rate and population served. Inclusion of concentration data from prior days yielded a model that can accounts for the previously observed offsets<sup>6, 7, 18</sup> between wastewater RNA concentrations and reported case data. The regression model used revealed that concentration lags of 0 to 1 day best predict case rates and confirms the previously described 0 to 2 day lag between case rates by date of specimen collection and SARS-CoV-2

concentrations in primary sludge<sup>6</sup>. A single general regression model trained by pooled plant data in a leave-one-out analysis was able to accurately estimate case rates for communities served by different domestic treatment plants, suggesting that this model could be extended more broadly to a variety of communities. In this case, treatment plants produced sludge using different primary treatment schemes and served populations that varied by more than an order of magnitude in size (**Table 1**). The approach for utilizing primary sludge instead of raw wastewater allows for rapid sampling of a mixed influent stream without the use of specialized sampling equipment. The concentration of SARS-CoV-2 in wastewater solids is greater than that in raw wastewater<sup>3</sup> and negates the need for concentration steps often required when using untreated wastewater. The 0.5 ml sample volume considered in this study should enable automation of RNA extraction. SARS-CoV-2 concentrations are likely not impacted by precipitation events (infiltration and inflow) that might dilute concentrations of SARS-CoV-2 in the aqueous phase of raw wastewater.

The ability to estimate case rates from etiological agent concentrations in wastewater can be of significant epidemiological value. Case rates are commonly used as a proxy for changes in community infection, and have been used as the standard for implementing non-pharmaceutical interventions and policies to reduce COVID-19 transmission and associated hospitalizations and deaths. In jurisdictions where testing is limited or does not exist, these models can be used as an independent estimate of infection rates. Wastewater RNA concentrations can be reported the same day the sample is collected, thus the statistical models used herein can be utilized to estimate up-to-date reported case rates in a community when COVID-19 testing data lags.

**Limitations.** This modeling approach relies on statistical relationships between wastewater primary sludge RNA concentrations and COVID-19 case rates, which are primarily based on

diagnostic test results. The reported prediction intervals reflect the variability in these reported case rates. While commonly used as a proxy, reported cases are believed to underestimate infection due to asymptomatic COVID-19 infections.<sup>20</sup> Clinical testing volumes in the municipalities considered were dynamic over the study period; responding to events such as school openings, holidays, and shifting of resources to locations with increasing case rates. The measured RNA concentrations in wastewater are not subject to such variation in testing policy and should therefore exhibit a more direct relationship to the unobservable changes in SARS-CoV-2 infection in the community. While clinical testing data is considered an imperfect measure of infection, even in locations with strong testing programs, we note that the data resulting from testing programs has been indispensable in understanding the progression of outbreaks and initiating action to stem the spread of COVID-19 throughout the world. That case results can be estimated from SARS-CoV-2 concentrations in sewage sludge provides an added measure of confidence in the use of reported cases to monitor the epidemic.

**Summary.** Measuring the concentration of pathogens in domestic wastewater can be a useful indicator of infection trends within a population. This study demonstrated that a single regression model populated by daily lagged SARS-CoV-2 sewage sludge concentrations could estimate COVID-19 case rates across communities served by six different wastewater treatment facilities. Cross-validation by leave-out analysis suggests the regression model can provide estimates of COVID-19 case rates for a broad variety of treatment facilities that produce primary sludge. Estimating case rates from wastewater pathogen concentrations can be useful in locations with limited or delayed COVID-19 testing programs or for infectious diseases where individual testing programs are not well-developed.

## Acknowledgements

This project was supported by Cooperative Agreement no. 6NU50CK000524-01 from the Centers for Disease Control and Prevention using funds from the COVID-19 Paycheck Protection Program and Health Care Enhancement Act Response Activities. This activity was reviewed by CDC and was conducted consistent with applicable federal law and CDC policy. See e.g., 45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. 241(d); 5 U.S.C. 552a; 44 U.S.C. 3501 et seq. The findings and conclusions of this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

We wish to thank the Stamford, Bridgeport West, New Haven, Hartford South Meadows, New London, and Norwich, CT wastewater treatment facilities for assistance in sampling and participation in the study.

## Supporting Information

General offset and coefficients for the regression model (Table S1), Stamford specific offset and coefficients for the regression model (Table S2), New Haven specific offset and coefficients for the regression model (Table S3), Hartford specific offset and coefficients for the regression model (Table S4), New London specific offset and coefficients for the regression model (Table S5), Norwich specific offset and coefficients for the regression model (Table S6).

# References

1. Xiao, F.; Tang, M.; Zheng, X.; Liu, Y.; Li, X.; Shan, H., Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* **2020**, *158*, 1831-1833.e3.
2. Medema, G.; Heijnen, L.; Elsinga, G.; Italiaander, R.; Brouwer, A., 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* **2020**, *7*, 511-516.
3. Graham, K. E.; Loeb, S. K.; Wolfe, M. K.; Catoe, D.; Sinnott-Armstrong, N.; Kim, S.; Yamahara, K. M.; Sassoubre, L. M.; Mendoza Grijalva, L. M.; Roldan-Hernandez, L.; Langenfeld, K.; Wigginton, K. R.; Boehm, A. B., SARS-CoV-2 RNA in Wastewater settled solids is associated with COVID-19 cases in a large urban sewershed. *Environmental Science & Technology* **2021**, *55*, 488-498.
4. 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., Predicting the number of people infected with SARS-COV-2 in a population using statistical models based on wastewater viral load. *medRxiv* **2020**, 2020.07.02.20144865.
5. Kaplan, E. H.; Wang, D.; Wang, M.; Malik, A. A.; Zulli, A.; Peccia, J., Aligning SARS-CoV-2 indicators via an epidemic model: application to hospital admissions and RNA detection in sewage sludge. *Health Care Management Science* **2020**. doi: 10.1007/s10729-020-09525-1
6. 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., Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology* **2020**, *38*, 1164-1167.
7. Nemudryi, A.; Nemudraia, A.; Wiegand, T.; Surya, K.; Buyukyoruk, M.; Cicha, C.; Vanderwood, K. K.; Wilkinson, R.; Wiedenheft, B., Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Reports Medicine* **2020**, *1*, 100098.
8. Souch, J. M.; Cossman, J. S., A Commentary on Rural-Urban Disparities in COVID-19 Testing Rates per 100,000 and Risk Factors. *The Journal of Rural Health* **2020**, *37*, 188-190.
9. Ondoa, P.; Kebede, Y.; Loembe, M. M.; Bhiman, J. N.; Tessema, S. K.; Sow, A.; Sall, A. A.; Nkengasong, J., COVID-19 testing in Africa: lessons learnt. *Lancet Microbe* **2020**, *1*, e103-e104.
10. U.S. Centers for Disease Control and Prevention, CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. **2020**.
11. 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.; Catherine Muenker, M.; Moore, A. J.; Klein,

- J.; Lu, P.; Lu-Culligan, A.; Jiang, X.; Kim, D. J.; Kudo, E.; Mao, T.; Moriyama, M.; Oh, J. E.; 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., Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. *Nature Microbiology* **2020**, *5*, 1299-1305.
12. Crank, K.; Li, X.; North, D.; Ferraro, G. B.; Iaconelli, M.; Mancini, P.; La Rosa, G.; Bibby, K., CrAssphage abundance and correlation with molecular viral markers in Italian wastewater. *Water Research* **2020**, *184*, 116161.
13. Stachler, E.; Kelty, C.; Sivaganesan, M.; Li, X.; Bibby, K.; Shanks, O. C., Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environmental Science & Technology* **2017**, *51*, 9146-9154.
14. Wu, Z.; Greaves, J.; Arp, L.; Stone, D.; Bibby, K., Comparative fate of CrAssphage with culturable and molecular fecal pollution indicators during activated sludge wastewater treatment. *Environment International* **2020**, *136*, 105452.
15. Council of State and Territorial, Update to the standardized surveillance case definition and national notification of 2019 novel coronavirus disease (COVID-19) Interim-20-ID-02. **2020**.
16. Bergman, A.; Sella, Y.; Agre, P.; Casadevall, A., Oscillations in U.S. COVID-19 Incidence and mortality data reflect diagnostic and reporting factors. *mSystems* **2020**, *5* (4), e00544-20.
17. Wurtzer, S.; Marechal, V.; Mouchel, J.-M.; Maday, Y.; Teyssou, R.; Richard, E.; Almayrac, J. L.; Moulin, L., Evaluation of lockdown impact on SARS-CoV-2 dynamics through viral genome quantification in Paris wastewaters. *medRxiv* **2020**, 2020.04.12.20062679.
18. Wu, F.; Xiao, A.; Zhang, J.; Moniz, K.; Endo, N.; Armas, F.; Bonneau, R.; Brown, M. A.; Bushman, M.; Chai, P. R.; Duvallet, C.; Erickson, T. B.; Foppe, K.; Ghaeli, N.; Gu, X.; Hanage, W. P.; Huang, K. H.; Lee, W. L.; Matus, M.; McElroy, K. A.; Nagler, J.; Rhode, S. F.; Santillana, M.; Tucker, J. A.; Wuertz, S.; Zhao, S.; Thompson, J.; Alm, E. J., SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *medRxiv* **2020**, 2020.06.15.20117747.
19. Berchenko, Y.; Manor, Y.; Freedman, L.; Kaliner, E.; Grotto, I.; Mendelson, E.; Huppert, A., Estimation of polio infection prevalence from environmental surveillance data. *Science Translational Medicine* **2017**, *9*, eaaf6786.
20. He, J.; Guo, Y.; Mao, R.; Zhang, J., Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of Medical Virology* **2020**, 93:820-830.

## Tables and Figures

**Table 1.** Wastewater treatment plant characteristics

| WWTP Name               | Municipalities served (total population)                                   | Average flow rate (m <sup>3</sup> d <sup>-1</sup> ) | Population served | Primary sludge process details   |
|-------------------------|--|---|-------------------|--|
| Stamford                | Stamford and Darien, CT (151,528)  | 91,000  | 140,000           | Bar screens→primary clarifier→ de-gritting of primary sludge with hydrocyclones. Sample of thickened sludge collected after hydrocyclones. |
| Bridgeport West         | Bridgeport, CT (144,900)   | 120,000   | 105,000           | Bar screens→gravity thickener→primary clarifier. Sample collected at primary solids effluent pump.   |
| New Haven               | New Haven, Hamden, East Haven, Woodbridge, CT (228,862)                    | 151,000   | 200,000           | Bar screen→grit chamber→ primary clarifier→gravity thickener. Sample collected at effluent pump of gravity thickener.                      |
| Hartford, South Meadows | Hartford, West Hartford, Newington, Bloomfield, Wethersfield, CT (263,021) | 300,000   | 300,000           | Bar screen→grit chamber→ primary clarifier→gravity thickener. Sample collected at effluent pump of gravity thickener.                      |
| New London              | New London and Waterford, CT (45,826)                                      | 23,000  | 40,000            | Bar screen grit chamber→ primary sedimentation. Sample collected at primary solids pump.   |
| Norwich                 | Norwich, Sprague, Franklin, Bozrah (46,495)                                | 32,000  | 22,500            | Bar screen→grit chamber→primary clarifier→gravity thickener. Sample collected at effluent pump from gravity thickener.                     |

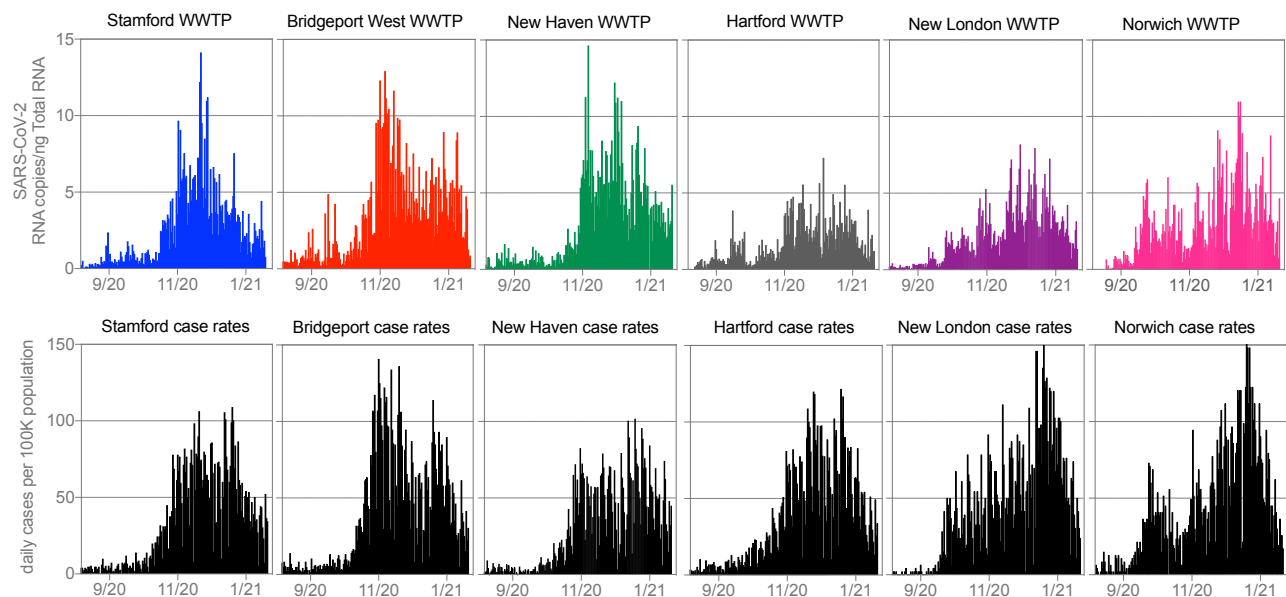
**Table 2.** Resulting parameters for model selection to determine the maximum lagged RNA concentration.

| Maximum lag $\tau$ | AIC  | BIC   | F statistic (p value) | Coverage Probability |
|--------------------|------|-------|-----------------------|----------------------|
| 0                  | 9835 | 10180 |                       | 0.94                 |
| 1                  | 9663 | 10048 | 35.6 (<0.01)          | 0.95                 |
| 2                  | 9613 | 10020 | 12.8 (<0.01)          | 0.95                 |
| 3                  | 9557 | 10013 | 8.5 (<0.01)           | 0.95                 |
| 4                  | 9543 | 9999  | 9.2 (<0.01)           | 0.95                 |
| 5                  | 9521 | 10018 | 3.6 (<0.01)           | 0.95                 |
| 6                  | 9536 | 10034 | 4.1 (<0.01)           | 0.95                 |

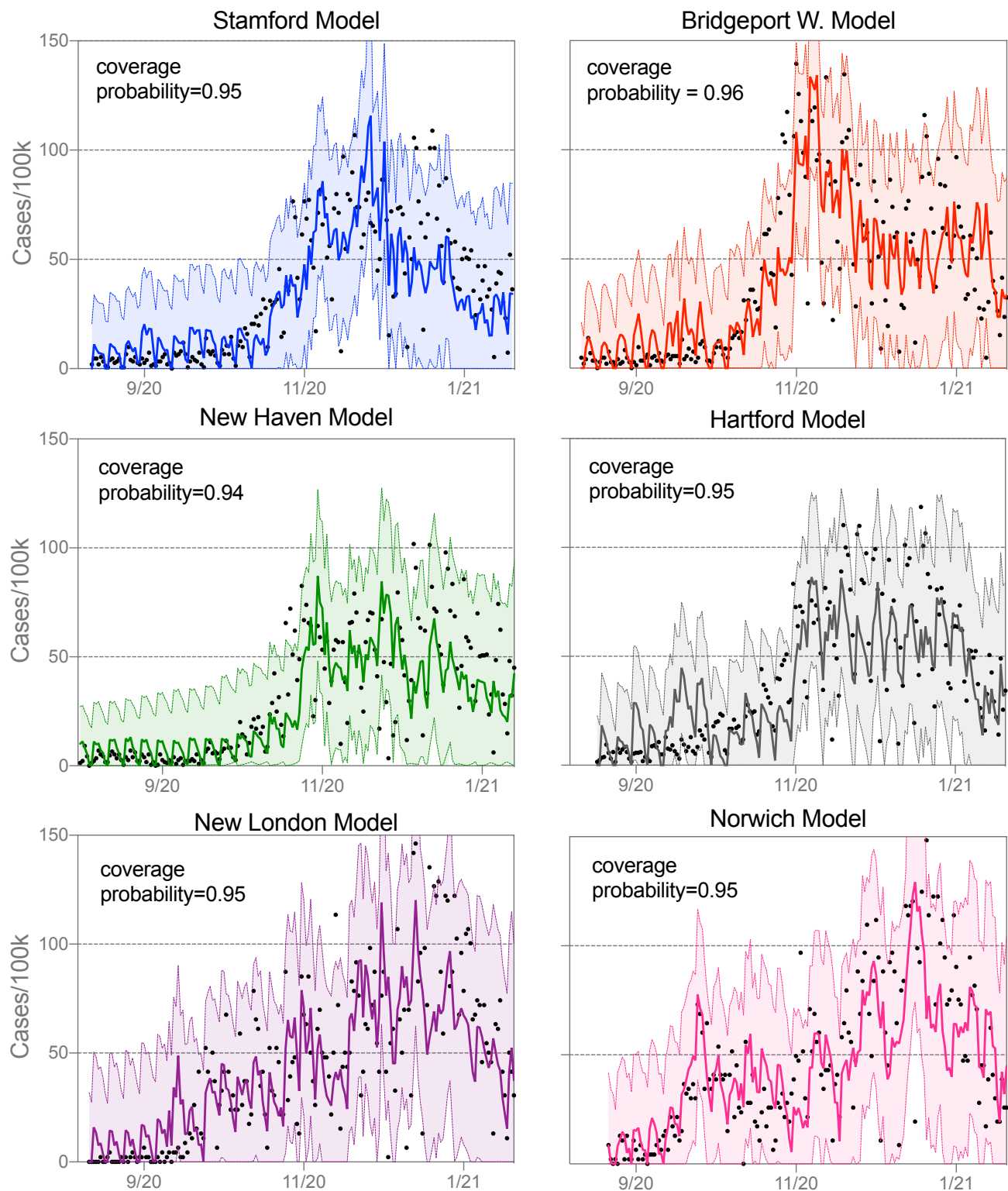
**Table 3.** Leave-one-out cross validation results for the general model<sup>1</sup>

| Town Omitted    | F-statistic (p value) | Coverage Probability |
|-----------------|-----------------------|----------------------|
| Stamford        | 285.6 (p<0.001)       | 0.98                 |
| Bridgeport West | 273.4 (p<0.001)       | 0.96                 |
| New Haven       | 339 (p<0.001)         | 0.91                 |
| Hartford        | 303.3 (p<0.001)       | 0.91                 |
| New London      | 315.6 (p<0.001)       | 0.90                 |
| Norwich         | 291.5 (p<0.001)       | 0.94                 |





**Figure 1.** (top) Daily SARS-CoV-2 RNA concentrations in primary sludge of the six WWTP's considered in this study. (bottom) Daily COVID-19 cases by date of specimen collection per 100,000 residents for the cities and towns served by the above wastewater treatment plants (**see Table 1**).



**Figure 2.** Comparison between COVID-19 case rates by day of clinical specimen collection and estimated case rates using RNA concentrations lagged from 0 to 4 days. The model is depicted by a solid line, measured cases are black dots, and 95% prediction intervals are shown by shading.  $F=37.42$ ,  $p<0.001$ .