

1
2
3
4
5 **Predicting daily COVID-19 case rates from SARS-CoV-2 RNA**
6 **concentrations across a diversity of wastewater catchments**
7

8
9
10
11 Alessandro Zulli¹, Annabelle Pan¹, Stephen M. Bart^{2,3}, Forrest W. Crawford^{4,5,6,7},
12 Edward H. Kaplan^{1,7,8}, Matthew Cartter^{3,8}, Albert I. Ko⁹, Duncan Cozens¹⁰, Marcela
13 Sanchez¹, Doug E. Brackney^{8,10}, and Jordan Peccia^{1,*}
14

15
16 *¹Department of Chemical and Environmental Engineering, School of Engineering and*
17 *Applied Science, Yale University, New Haven, CT, USA*
18

19 *²Epidemic Intelligence Service, Centers for Disease Control and Prevention*
20

21 *³Connecticut Department of Public Health, Hartford, CT, USA*
22

23 *⁴Department of Biostatistics, Yale School of Public Health, Yale University, New Haven,*
24 *CT, USA*
25

26 *⁵Department of Statistics and Data Science, Yale University, New Haven, CT, USA*
27

28 *⁶Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT,*
29 *USA*
30

31 *⁷Yale School of Management, Yale University, New Haven, CT, USA*
32

33 *⁸Yale School of Public Health, Yale University, New Haven, CT, USA*
34

35 *⁹Department of Epidemiology of Microbial Disease, Yale School of Public Health, Yale*
36 *University, New Haven, CT, USA*
37

38 *¹⁰Connecticut Agricultural Experimental Station, State of Connecticut, New Haven, CT,*
39 *USA*
40

41
42
43
44
45 *Corresponding authors: Jordan Peccia, Tel: +1 (203) 432-4385; email: Jordan.Peccia@Yale.edu

46 **Abstract**

47 We assessed the relationship between municipality COVID-19 case rates and SARS-CoV-2
48 concentrations in the primary sludge of corresponding wastewater treatment facilities. Over
49 1,000 daily primary sludge samples were collected from six wastewater treatment facilities with
50 catchments serving 18 cities and towns in the State of Connecticut, USA. Samples were analyzed
51 for SARS-CoV-2 RNA concentrations during a six-month time period that overlapped with fall
52 2020 and winter 2021 COVID-19 outbreaks in each municipality. We fit a single regression
53 model to estimate reported case rates in the six municipalities from SARS-CoV-2 RNA
54 concentrations collected daily from corresponding wastewater treatment facilities. Results
55 demonstrate the ability of SARS-CoV-2 RNA concentrations in primary sludge to estimate
56 COVID-19 reported case rates across treatment facilities and wastewater catchments, with
57 coverage probabilities ranging from 0.94 to 0.96. Leave-one-out cross validation suggests that
58 the model can be broadly applied to wastewater catchments that range in more than one order of
59 magnitude in population served. Estimation of case rates from wastewater data can be useful in
60 locations with limited testing availability or testing disparities, or delays in individual COVID-19
61 testing programs.

62

63 **Keywords**

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

66

67

68

69 **Introduction**

70 Wastewater surveillance has the potential to identify and track outbreaks of human pathogens
71 that demonstrate gut tropism, including severe acute respiratory syndrome coronavirus 2 (SARS-
72 CoV-2), the causative agent of COVID-19.¹ Recent studies demonstrate that SARS-CoV-2 RNA
73 concentrations in domestic wastewater reflect the rise and fall of COVID-19 cases based on daily
74 positive tests in a community,²⁻⁴ and important epidemiological parameters such as the effective
75 reproduction number can be estimated using wastewater SARS-CoV-2 concentrations, in
76 addition to other epidemic indicators.⁵ If SARS-CoV-2 wastewater results are rapidly reported,
77 they could provide a leading indicator of community infection rates over COVID-19 case rates
78 and hospital admissions data.⁵⁻⁷

79 Estimates of infection based on COVID-19 case rates from testing of human specimens have
80 been the standard for applying community interventions aimed at decreasing morbidity and
81 mortality. Alternative approaches to COVID-19 testing are necessary to estimate the number of
82 infections in locations with disparities in testing practices⁸, limited testing resources,⁹ or during
83 outbreaks when testing capacity cannot meet demand. SARS-CoV-2 RNA concentration
84 measurements in wastewater may resolve these issues, but quantitative relationships between
85 SARS-CoV-2 concentration in wastewater and infection (or proxies of infection such as cases)
86 are not well-resolved.

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

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

93

94 **Materials and Methods**

95 **Wastewater sampling.** Primary sewage sludge samples of 40 to 45 ml in volume were
96 collected daily from six wastewater treatment plants (WWTP) in the State of Connecticut, USA.
97 **Table 1** lists the cities and towns served, and details the specific plant processes from which
98 primary sludge was produced and samples were withdrawn. Grab samples from Stamford,
99 Bridgeport, New Haven and New London were collected daily from August 3, 2020 to February
100 10, 2021. Samples from Hartford were collected daily from August 10, 2020 to February 10,
101 2021. Samples from Norwich were collected daily from August 17, 2020 to February 10, 2021.
102 All samples were collected between 8 am and 9 am, stored at -20° C before being transported to
103 Yale University laboratories on ice, and analyzed immediately upon arrival.

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

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

116 SARS-CoV-2 RNA concentration was estimated using one-step qRT-PCR with SARS-CoV-2
117 N1 and N2 primer sets.^{10, 11} CrAssphage, a ubiquitous bacteriophage that is highly concentrated
118 in the human gut,¹²⁻¹⁴ was quantified as a PCR control. All primer sets were run in separate
119 reactions. Analysis was conducted using a one-step RT-qPCR kit (BioRad iTaq™ Universal
120 Probes One-Step Kit). Triplicate 20 µL reactions using a 5x diluted template were run at 55° C
121 for 10 min and 95° C for 1 min, followed by 40 cycles consisting of 95°C for 10 seconds and 55°
122 C for 30 seconds⁶ and Ct values for triplicates were average. N1 and N2 primer set standards
123 were constructed using ten-fold dilutions (5×10^1 to 5×10^8 copies per reaction) of the N gene
124 transcripts for the SARS-CoV-2 Wuhan-Hu-1 strain. Concentrations in primary sludge samples
125 were calculated using these standard curves and are presented as SARS-CoV-2 RNA copies per
126 ng of total extracted RNA. Positive RNA and negative (no template) controls were included in
127 all qRT-PCR runs. Less than 1.5% of samples were not available from the treatment facilities. In
128 these cases, SARS-CoV-2 concentrations used in regression models were estimated by
129 interpolating between the prior and subsequent measured SARS-CoV-2 concentration.

130 **COVID-19 test data for the cities served by the treatment facilities.** The number of
131 confirmed and probable COVID-19 cases per the Council of State and Territorial
132 Epidemiologists case definition¹⁵ was provided by the Connecticut Department of Public Health
133 (CT DPH) and used in this analysis. The date for each case was assigned, in order of preference,
134 as test specimen collection date (if available), symptom onset date (if available), or date of report
135 to the CT DPH. Greater than 90% of reported case values were by date of clinical specimen
136 collection. Average test positivity rates over the study period were the following: Stamford

137 (6.2%), Bridgeport (6.1%), New Haven (3.7%), Hartford (6.0%), New London (4.6%), Norwich
138 (6.5%).

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

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

147

$$148 \quad \text{case rate}_{i,t} = \alpha_0 + \alpha_i + \sum_{j=0}^{\tau} (\beta_j + \beta_{i,j}) \text{RNA}_{i,t-j} + \sum_{d=2}^7 (\gamma_d + \gamma_{i,d}) 1\{w_t = d\} + \epsilon_{it}$$

149 (eq. 1)

150 where the reported case rate is the number of new cases reported by date of specimen collection
151 per 100,000 population, i represents municipality, t represents time (days), α_0 is the model
152 offset, α_i is the individual municipality model offset, j is the number of lagged days, τ is the
153 maximum number of lagged days, $\text{RNA}_{i,t-j}$ are RNA concentrations (SARS CoV-2 RNA copies
154 ng^{-1} RNA), w_t indicates the day of the week on day t , β_j , $\beta_{i,j}$ and γ_d and $\gamma_{i,d}$ are regression
155 coefficients, d is day of the week (Friday is the reference day), and ϵ_{it} is the residual error
156 associated with each observation. Day of the week addresses the known daily variations in
157 testing behavior.¹⁶ Adjustment for population, which shows a strong collinearity with average
158 treatment plant flow (simple linear regression: slope = $0.9 \text{ m}^3\text{d}^{-1} \text{ person}^{-1}$, $R^2=0.93$, $p=0.002$),

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

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

175

176 **Results**

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

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

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

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

204

205 Discussion

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

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

228 concentrations in primary sludge⁶. A single general regression model trained by pooled plant
229 data in a leave-one-out analysis was able to accurately estimate case rates for communities
230 served by different domestic treatment plants, suggesting that this model could be extended more
231 broadly to a variety of communities. In this case, treatment plants produced sludge using
232 different primary treatment schemes and served populations that varied by more than an order of
233 magnitude in size (**Table 1**). The approach for utilizing primary sludge instead of raw
234 wastewater allows for rapid sampling of a mixed influent stream without the use of specialized
235 sampling equipment. The concentration of SARS-CoV-2 in wastewater solids is greater than that
236 in raw wastewater³ and negates the need for concentration steps often required when using
237 untreated wastewater. The 0.5 ml sample volume considered in this study should enable
238 automation of RNA extraction. SARS-CoV-2 concentrations are likely not impacted by
239 precipitation events (infiltration and inflow) that might dilute concentrations of SARS-CoV-2 in
240 the aqueous phase of raw wastewater.

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

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

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

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

273

274

275

276 **Acknowledgements**

277 This project was supported by Cooperative Agreement no. 6NU50CK000524-01 from the
278 Centers for Disease Control and Prevention using funds from the COVID-19 Paycheck
279 Protection Program and Health Care Enhancement Act Response Activities. This activity was
280 reviewed by CDC and was conducted consistent with applicable federal law and CDC policy.
281 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
282 U.S.C. 3501 et seq. The findings and conclusions of this report are those of the author(s) and do
283 not necessarily represent the official position of the Centers for Disease Control and Prevention.

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

287

288 **Supporting Information**

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

294

295

296

297

298

299

300

301 References

302

- 303 1. Xiao, F.; Tang, M.; Zheng, X.; Liu, Y.; Li, X.; Shan, H., Evidence for gastrointestinal
304 infection of SARS-CoV-2. *Gastroenterology* **2020**, *158*, 1831-1833.e3.
- 305 2. Medema, G.; Heijnen, L.; Elsinga, G.; Italiaander, R.; Brouwer, A., Presence of SARS-
306 Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in
307 the early stage of the epidemic in The Netherlands. *Environmental Science & Technology*
308 *Letters* **2020**, *7*, 511-516.
- 309 3. Graham, K. E.; Loeb, S. K.; Wolfe, M. K.; Catoe, D.; Sinnott-Armstrong, N.; Kim, S.;
310 Yamahara, K. M.; Sassoubre, L. M.; Mendoza Grijalva, L. M.; Roldan-Hernandez, L.;
311 Langenfeld, K.; Wigginton, K. R.; Boehm, A. B., SARS-CoV-2 RNA in Wastewater
312 settled solids is associated with COVID-19 cases in a large urban sewershed.
313 *Environmental Science & Technology* **2021**, *55*, 488-498.
- 314 4. Vallejo, J. A.; Rumbo-Feal, S.; Conde-Pérez, K.; López-Oriona, Á.; Tarrío-Saavedra,
315 J.; Reif, R.; Ladra, S.; Rodiño-Janeiro, B. K.; Nasser, M.; Cid, Á.; Veiga, M. C.;
316 Acevedo, A.; Lamora, C.; Bou, G.; Cao, R.; Poza, M., Predicting the number of people
317 infected with SARS-COV-2 in a population using statistical models based on wastewater
318 viral load. *medRxiv* **2020**, 2020.07.02.20144865.
- 319 5. Kaplan, E. H.; Wang, D.; Wang, M.; Malik, A. A.; Zulli, A.; Peccia, J., Aligning
320 SARS-CoV-2 indicators via an epidemic model: application to hospital admissions and
321 RNA detection in sewage sludge. *Health Care Management Science* **2020**.
322 doi: 10.1007/s10729-020-09525-1
- 323 6. Peccia, J.; Zulli, A.; Brackney, D. E.; Grubaugh, N. D.; Kaplan, E. H.; Casanovas-
324 Massana, A.; Ko, A. I.; Malik, A. A.; Wang, D.; Wang, M.; Warren, J. L.;
325 Weinberger, D. M.; Arnold, W.; Omer, S. B., Measurement of SARS-CoV-2 RNA in
326 wastewater tracks community infection dynamics. *Nature Biotechnology* **2020**, *38*, 1164-
327 1167.
- 328 7. Nemudryi, A.; Nemudraia, A.; Wiegand, T.; Surya, K.; Buyukyoruk, M.; Cicha, C.;
329 Vanderwood, K. K.; Wilkinson, R.; Wiedenheft, B., Temporal Detection and
330 Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Reports*
331 *Medicine* **2020**, *1*, 100098.
- 332 8. Souch, J. M.; Cossman, J. S., A Commentary on Rural-Urban Disparities in COVID-19
333 Testing Rates per 100,000 and Risk Factors. *The Journal of Rural Health* **2020**, *37*, 188-
334 190.
- 335 9. Ondo, P.; Kebede, Y.; Loembe, M. M.; Bhiman, J. N.; Tessema, S. K.; Sow, A.;
336 Sall, A. A.; Nkengasong, J., COVID-19 testing in Africa: lessons learnt. *Lancet Microbe*
337 **2020**, *1*, e103-e104.
- 338 10. U.S. Centers for Diseases Control and Prevention, CDC 2019-Novel Coronavirus (2019-
339 nCoV) Real-Time RT-PCR Diagnostic Panel. **2020**.
- 340 11. Vogels, C. B. F.; Brito, A. F.; Wyllie, A. L.; Fauver, J. R.; Ott, I. M.; Kalinich, C. C.;
341 Petrone, M. E.; Casanovas-Massana, A.; Catherine Muenker, M.; Moore, A. J.; Klein,

- 342 J.; Lu, P.; Lu-Culligan, A.; Jiang, X.; Kim, D. J.; Kudo, E.; Mao, T.; Moriyama, M.;
343 Oh, J. E.; Park, A.; Silva, J.; Song, E.; Takahashi, T.; Taura, M.; Tokuyama, M.;
344 Venkataraman, A.; Weizman, O.-E.; Wong, P.; Yang, Y.; Cheemarla, N. R.; White, E.
345 B.; Lapidus, S.; Earnest, R.; Geng, B.; Vijayakumar, P.; Odio, C.; Fournier, J.;
346 Bermejo, S.; Farhadian, S.; Dela Cruz, C. S.; Iwasaki, A.; Ko, A. I.; Landry, M. L.;
347 Foxman, E. F.; Grubaugh, N. D., Analytical sensitivity and efficiency comparisons of
348 SARS-CoV-2 RT-qPCR primer-probe sets. *Nature Microbiology* **2020**, *5*, 1299-1305.
- 349 12. Crank, K.; Li, X.; North, D.; Ferraro, G. B.; Iaconelli, M.; Mancini, P.; La Rosa, G.;
350 Bibby, K., CrAssphage abundance and correlation with molecular viral markers in Italian
351 wastewater. *Water Research* **2020**, *184*, 116161.
- 352 13. Stachler, E.; Kelty, C.; Sivaganesan, M.; Li, X.; Bibby, K.; Shanks, O. C., Quantitative
353 CrAssphage PCR assays for human fecal pollution measurement. *Environmental Science*
354 *& Technology* **2017**, *51*, 9146-9154.
- 355 14. Wu, Z.; Greaves, J.; Arp, L.; Stone, D.; Bibby, K., Comparative fate of CrAssphage
356 with culturable and molecular fecal pollution indicators during activated sludge
357 wastewater treatment. *Environment International* **2020**, *136*, 105452.
- 358 15. Council of State and Territorial, Update to the standardized surveillance case definition
359 and national notification of 2019 novel coronavirus disease (COVID-19) Interim-20-ID-
360 02. **2020**.
- 361 16. Bergman, A.; Sella, Y.; Agre, P.; Casadevall, A., Oscillations in U.S. COVID-19
362 Incidence and mortality data reflect diagnostic and reporting factors. *mSystems* **2020**, *5*
363 (4), e00544-20.
- 364 17. Wurtzer, S.; Marechal, V.; Mouchel, J.-M.; Maday, Y.; Teyssou, R.; Richard, E.;
365 Almayrac, J. L.; Moulin, L., Evaluation of lockdown impact on SARS-CoV-2 dynamics
366 through viral genome quantification in Paris wastewaters. *medRxiv* **2020**,
367 2020.04.12.20062679.
- 368 18. Wu, F.; Xiao, A.; Zhang, J.; Moniz, K.; Endo, N.; Armas, F.; Bonneau, R.; Brown,
369 M. A.; Bushman, M.; Chai, P. R.; Duvallet, C.; Erickson, T. B.; Foppe, K.; Ghaeli,
370 N.; Gu, X.; Hanage, W. P.; Huang, K. H.; Lee, W. L.; Matus, M.; McElroy, K. A.;
371 Nagler, J.; Rhode, S. F.; Santillana, M.; Tucker, J. A.; Wuertz, S.; Zhao, S.;
372 Thompson, J.; Alm, E. J., SARS-CoV-2 titers in wastewater foreshadow dynamics and
373 clinical presentation of new COVID-19 cases. *medRxiv* **2020**, 2020.06.15.20117747.
- 374 19. Berchenko, Y.; Manor, Y.; Freedman, L.; Kaliner, E.; Grotto, I.; Mendelson, E.;
375 Huppert, A., Estimation of polio infection prevalence from environmental surveillance
376 data. *Science Translational Medicine* **2017**, *9*, eaaf6786.
- 377 20. He, J.; Guo, Y.; Mao, R.; Zhang, J., Proportion of asymptomatic coronavirus disease
378 2019: A systematic review and meta-analysis. *Journal of Medical Virology* **2020**, 93:820-
379 830.

380
381
382
383
384
385
386
387

388
389
390
391
392
393
394
395
396
397
398
399
400
401
402

Tables and Figures

Table 1. Wastewater treatment plant characteristics

WWTP Name	Municipalities served (total population)	Average flow rate (m ³ d ⁻¹)	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.

403
404
405
406
407
408

409
410
411
412
413
414
415
416

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

Maximum lag τ	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

417
418
419
420
421
422
423
424
425

Table 3. Leave-one-out cross validation results for the general model¹

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

426

427
428
429
430
431
432
433
434
435
436
437
438
439

440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485

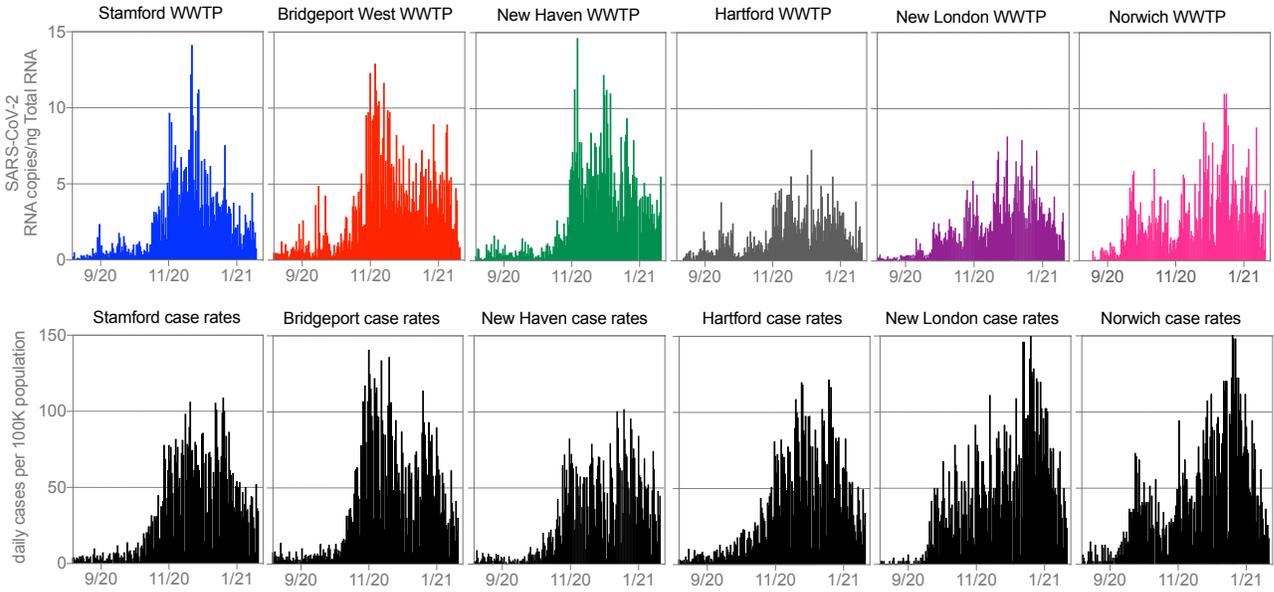
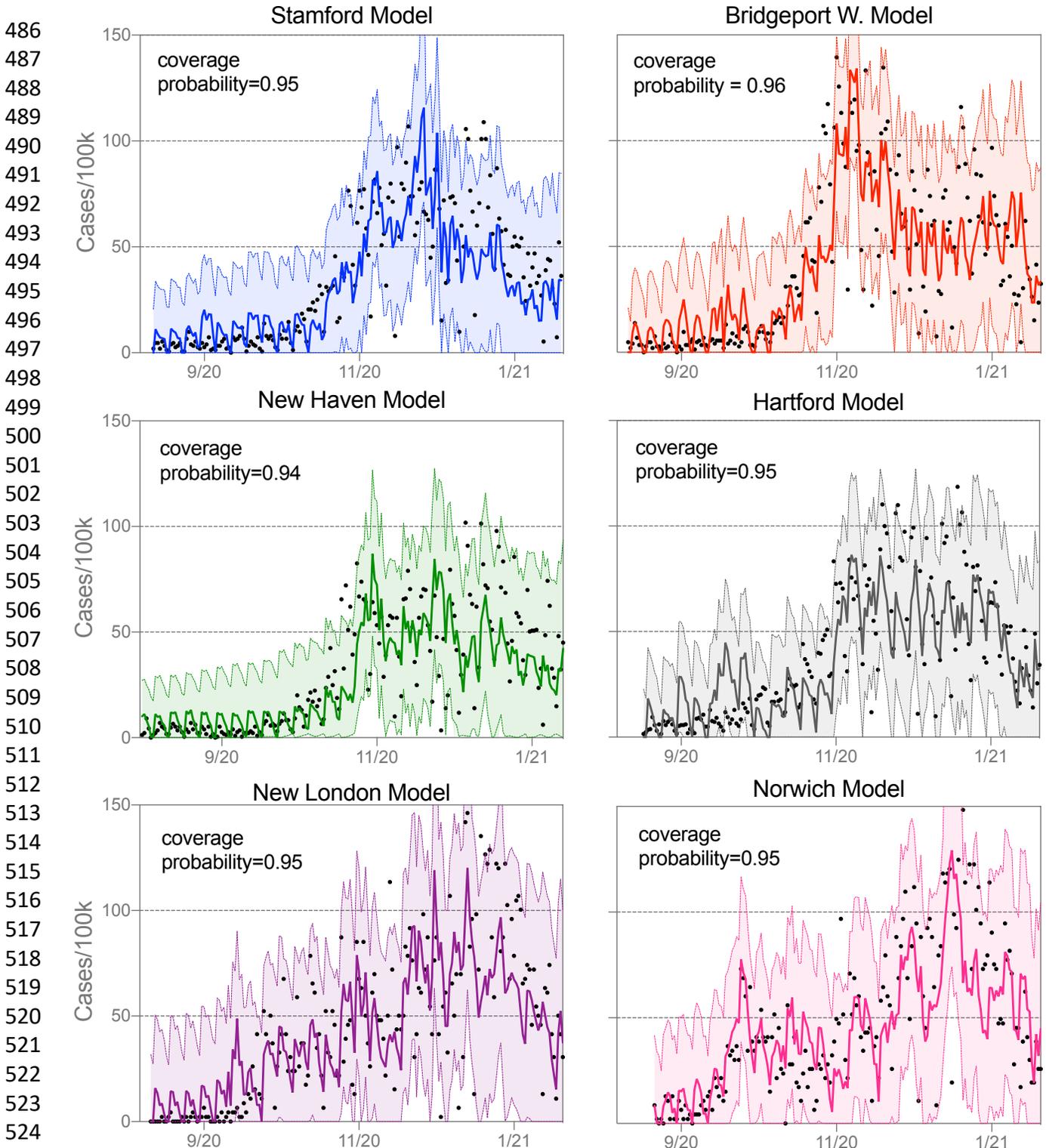


Figure 1. (top) Daily SARS-CoV-2 RNA concentrations in primary sludge of the six WWTP's considered in this study. (bottom) Daily COVID-19 cases by date of specimen collection per 100,000 residents for the cities and towns served by the above wastewater treatment plants (**see Table 1**).



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