

Correlation between wastewater and COVID-19 case incidence rates in major California sewersheds across three variant periods

Angela Rabe^{a,*†}, Sindhu Ravuri^{a,†}, Elisabeth Burnor^a, Joshua A. Steele^b, Rose S. Kantor^c, Samuel Choi^d, Stanislav Forman^e, Ryan Batjiaka^f, Seema Jain^a, Tomás M. León^a, Duc J. Vugia^a and Alexander T. Yu^a

^a California Department of Public Health COVID-19 Detection, Investigation, Surveillance, Clinical, and Outbreak Response, California Department of Public Health, Richmond and Sacramento, CA, USA

^b Southern California Coastal Water Research Project (SCCWRP), Department of Microbiology, Costa Mesa, CA, USA

^c Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

^d Orange County Sanitation District, Fountain Valley, CA, USA

^e Zymo Research Corp. Department of Sample Collection and Nucleic Acid Purification, Zymo Research Corp., Irvine, CA, USA

^f San Francisco Public Utilities Commission, San Francisco, CA, USA

*Corresponding author. E-mail: angela.rabe@cdph.ca.gov

†These first authors contributed equally to this manuscript.

 AR, 0009-0000-3440-1345

ABSTRACT

Monitoring for COVID-19 through wastewater has been used for adjunctive public health surveillance, with SARS-CoV-2 viral concentrations in wastewater correlating with incident cases in the same sewershed. However, the generalizability of these findings across sewersheds, laboratory methods, and time periods with changing variants and underlying population immunity has not been well described. The California Department of Public Health partnered with six wastewater treatment plants starting in January 2021 to monitor wastewater for SARS-CoV-2, with analyses performed at four laboratories. Using reported PCR-confirmed COVID-19 cases within each sewershed, the relationship between case incidence rates and wastewater concentrations collected over 14 months was evaluated using Spearman's correlation and linear regression. Strong correlations were observed when wastewater concentrations and incidence rates were averaged (10- and 7-day moving window for wastewater and cases, respectively, $\rho = 0.73\text{--}0.98$ for N1 gene target). Correlations remained strong across three time periods with distinct circulating variants and vaccination rates (winter 2020–2021/Alpha, summer 2021/Delta, and winter 2021–2022/Omicron). Linear regression revealed that slopes of associations varied by the dominant variant of concern, sewershed, and laboratory ($\beta = 0.45\text{--}1.94$). These findings support wastewater surveillance as an adjunctive public health tool to monitor SARS-CoV-2 community trends.

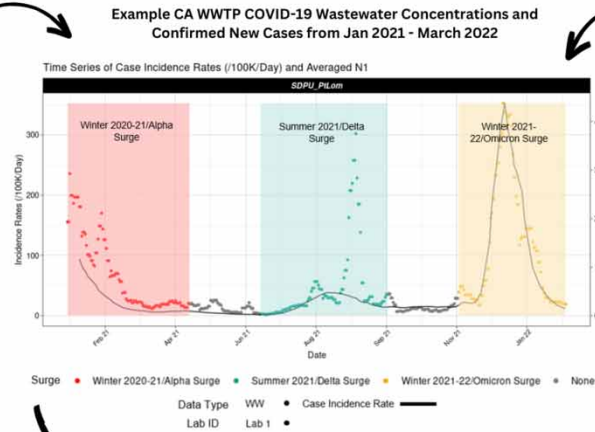
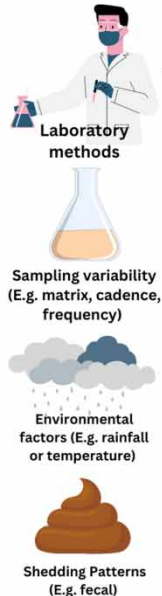
Key words: COVID-19 epidemiology, COVID-19 surveillance, public health surveillance, variant(s) of concern, wastewater surveillance

HIGHLIGHTS

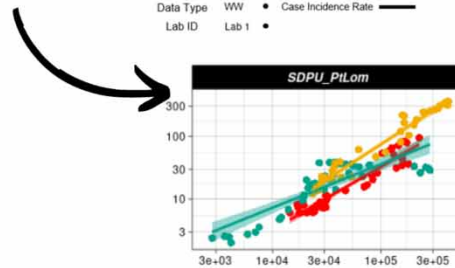
- Wastewater and incidence rate correlations were strong throughout three variant periods of the COVID-19 pandemic and spanning vaccine introduction to widespread uptake.
- Correlations persisted across different regions of California with four unique labs.
- Slopes of associations between wastewater concentration and case varied by the variant, sewershed, and laboratory.
- Wastewater is a complementary COVID-19 public health tool.

GRAPHICAL ABSTRACT

1. Potential factors that can influence wastewater concentrations overtime



2. Potential factors that can influence epidemiological case data overtime



3) Strong and positive correlations existed across three major COVID-19 variant periods, despite potential influencing factors, including varied sewersheds and laboratories, and changes in predominance and population immunity. (Spearman's $\rho > 0.78$)

INTRODUCTION

Public health surveillance of coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, relies on results from clinical testing done to identify cases, primarily through nucleic acid amplification tests for SARS-CoV-2 RNA. Limitations of such case-based surveillance include heterogeneous testing capacity and the lack of at-home tests reported to public health agencies. Case-based surveillance is also largely dependent on test-seeking behavior, missing most asymptomatic individuals and many symptomatic individuals who do not seek or cannot access testing. Since SARS-CoV-2 virus is excreted in the feces of about half of infected individuals, whether symptomatic or asymptomatic, monitoring of SARS-CoV-2 RNA concentrations in wastewater as a proxy for infected individuals has been deployed as a complimentary public health surveillance method (Chen *et al.* 2020; Medema *et al.* 2020; Park *et al.* 2021; Natarajan *et al.* 2022). This has been possible because of moderate and positive nonparametric correlations between wastewater concentrations to case-based surveillance data, as has been reported in the literature across different geographies, variant periods, laboratory techniques, and time (Feng *et al.* 2021; Greenwald *et al.* 2021; Weidhaas *et al.* 2021; Al-Faliti *et al.* 2022; Duvallet *et al.* 2022; Hegazy *et al.* 2022; Zheng *et al.* 2022; Ando *et al.* 2023). In addition, wastewater surveillance may also provide earlier warnings of trends compared with clinical surveillance as it is not limited by delays in test seeking and result reporting to public health (Aguar-Oliveira *et al.* 2020; Kirby *et al.* 2021; Olesen *et al.* 2021; Wolfe *et al.* 2021a, 2021b; Morvan *et al.* 2022; Wu *et al.* 2022).

These studies support the framing of wastewater surveillance as an informative tool for public health situational awareness and decision-making. However, the duration of most studies is short (usually a few months) and generally focuses on data from early in the pandemic prior to large shifts in population immunity, relies on data generated from a single laboratory, and/or spans only one to two variant periods. Changes in underlying population immunity and in dominant circulating SARS-CoV-2 variants over time may impact viral fecal shedding rates. Such dynamics could impact the relationship between wastewater concentrations and cases enough to change interpretations (or usability) of wastewater surveillance results (D'Aoust *et al.* 2022; Schill *et al.* 2023). Examining the correlation and relationship between wastewater with case-based data over longer periods of time, varied geographic settings, changing dominant variants, evolving population immunity, and differing laboratory methods remain an informative investigation to provide reassurance and insight into generalizability.

and robustness of wastewater surveillance for public health surveillance of SARS-CoV-2. Additionally, questions remain about whether normalization of wastewater concentrations to human fecal markers such as pepper mild mottled virus (PMMoV) improves correlations, or whether an ideal averaging window of which to interpret wastewater data exists.

In this study, we investigate the correlation between wastewater SARS-CoV-2 concentrations from the California Department of Public Health (CDPH) National Wastewater Surveillance System (NWSS) and COVID-19 polymerase chain reaction (PCR)-confirmed cases reported to CDPH and geocoded to corresponding sewersheds. Correlations are compared using raw and normalized wastewater data as well as across different averaging time windows. Wastewater data are derived from heterogeneous wastewater analysis laboratories and sewersheds during a 14-month period spanning three major SARS-CoV-2 variant periods and the time from COVID-19 vaccine introduction through widespread uptake.

METHODS

Wastewater sample collection

In January 2021, with funding from the U.S. Centers for Disease Control and Prevention (CDC) to join the NWSS program, CDPH partnered with the State Water Resources Control Board (SWRCB) and five sanitation agencies to pilot monitoring of wastewater at six wastewater treatment plants (WWTPs) servicing large (populations > 200,000) metropolitan areas of the state. These six WWTPs serve over 30% of California's population and include two utilities located in Los Angeles County (Los Angeles Sanitation [LASAN Hyp] and Los Angeles County Sanitation District [LACSD Jnt]), two in San Francisco County (San Francisco Public Utilities Commission, Oceanside [SFPUC Ocean] and Southeast [SFPUC SE]), one in San Diego County (San Diego Public Utilities [SDPU Pt Lom]), and one in Orange County (Orange County Sanitation District [OC San]). Sample collection and analysis for the data included in this study occurred from 1 January 2021 to 2 March 2022.

Raw wastewater influent (24-h composite, flow-, or time-weighted) samples were collected at a minimum of three times a week at the main headworks of six California WWTPs. The cadence and days of sampling varied between the treatment plants and over the course of the study period. Since each sanitation district was allowed to use a laboratory of their choice, four different laboratories were included, each with distinct methods for viral RNA extraction and quantification using either quantitative PCR (qPCR) or digital droplet PCR (ddPCR). Samples were quantified for specific SARS-CoV-2 gene fragments (e. g., N1 and N2) of the nucleocapsid N protein. Each laboratory determined a unique limit of detection (LOD) for the minimum concentration required to quantify the virus per sample. Quantification measurements were performed for the N1 gene by all laboratories for all samples collected; N2 and PMMoV, an indicator of human fecal material, were measured at only a subset of laboratories, at the laboratories' discretion. PMMoV is plant virus commonly found in pepper-based foods and is one of the most abundantly excreted viruses found in human fecal samples (Kitajima *et al.* 2018). It has been used throughout the COVID-19 pandemic by laboratories to normalize raw SARS-CoV-2 RNA concentrations with the goal of adjusting for differing fecal strength in wastewater samples and differing viral recovery between samples (Wolfe *et al.* 2021a, 2021b).

Laboratory methodology

Laboratory 1: Southern California Coastal Water Research Project (SCCWRP)

Aliquots of primary influent composite samples were collected in sterile 1 L bottles, stored at 4°C, transported to the laboratory, and processed within 96 h. Sample processing and RNA extraction were performed following the methods described in Steele *et al.* (2021) and Kim *et al.* (2022). Prior to processing, samples were spiked with known concentrations of Bovine Coronavirus (BcoV) to assess extraction efficiency. Then, viruses were adsorbed to 0.45-µm pore size mixed cellulose ester HA filters (Millipore Sigma, USA), and filters were extracted using Zymo BashingBead beads (Zymo Research, USA) with the NucliSENS nucleic acid extraction kit (BioMerieux, USA). One-step digital RT-PCR with the BioRad QX200 ddPCR system and QuantaSoft Analysis Pro software (BioRad, USA) was utilized to amplify and quantify nucleic acid targets for SARS-CoV-2 (N1 and N2), Bovine Coronavirus (BCoV), and PMMoV. The concentration per reaction and 95% confidence intervals were converted to copies per volume of wastewater using dimensional analysis. Quality control steps were performed as described in Steele *et al.* (2021) and Kim *et al.* (2022).

Laboratory 2: Los Angeles County Sanitation District San Jose Creek Water Quality Laboratory

Influent wastewater samples were stabilized after collection with DNA/RNA Shield (Zymo Research, USA) at a ratio of 250 µL of influent wastewater to 750 µL of DNA/RNA Shield and transported to the laboratory. If samples could not be

processed immediately upon arrival, they were stored at -80°C . Details about sample extraction and quantification can be found in Supplemental Methods. Briefly, RNA extraction was performed in triplicate for each sample using the Zymo Quick-RNA Fecal/Soil Microbe Microprep kit (Zymo Research, USA), with modifications. Approximately 1×10^5 copies of an *in vitro* transcribed RNA processing control were added to each extract prior to quantification. Quantification of SARS-CoV-2 RNA (N1) was performed via TaqMan RT-qPCR on the LightCycler 480 instrument (Roche Diagnostics, Germany). Analysis of the RT-qPCR data was performed using the second derivative method on the LightCycler480[®] software, version 1.5.1.62 (Roche Diagnostics, Germany). Further details about the RNA processing control, the RT-qPCR protocol primers, probes, and cycling conditions, and quality control steps can be found in Supplemental Methods.

Laboratory 3: Zymo Research Corporation

Influent wastewater samples were stabilized after collection with DNA/RNA Shield (Zymo Research, USA). Triplicates of 5 mL of stabilized wastewater were spiked with known concentrations of HeLa cells to gauge extraction efficiency and estimate inhibition. Samples were then combined with two volumes of Viral RNA Buffer (Zymo Research, USA) before the total volume was bound directly to a column. The resulting crude extractions were cleaned and concentrated via the RNA Clean & Concentrator kit (Zymo Research, USA) before treatment with Zymo-Spin III-HRC Filters (Zymo Research, USA) to remove inhibitors. Final dilutions were quantified via RT-qPCR and quantification cycle values were converted to gene copies using an average standard curve.

Laboratory 4: University of California, Berkeley (UC Berkeley)

Influent wastewater samples were treated with NaCl, Tris, and EDTA to lyse and preserve SARS-CoV-2 RNA and shipped overnight to the laboratory at UC Berkeley. Samples were processed, extracted, and quantified as described in [Kantor et al. \(2022\)](#). Briefly, samples were extracted in duplicate 40 mL aliquots, according to the 4S method for total RNA extraction ([Whitney et al. 2021](#)). RT-qPCR was used to quantify SARS-CoV-2 (N1), PMMoV, and Bovine Coronavirus (to assess extraction efficiency) in each sample ([Greenwald et al. 2021](#)) using Luna 2 \times master mix (New England Biolabs, USA) on a QuantStudio 3 instrument (Applied Biosystems, USA). Each RT-qPCR reaction was performed in triplicate alongside no-template controls and 7-point standard curves. Quantification cycle (C_q) values were converted to gene copies using the standard curve. Further details about sample processing steps to ensure quality control standards are described in [Kantor et al. \(2022\)](#).

Data analysis methodology

Confirmed COVID-19 cases with positive PCR test results reported to CDPH with a residential address within each treatment plant service area (also referred to as a sewershed) were aggregated using shapefiles of each WWTP's service area boundaries (ArcGIS Pro [Version 2.7.3] and R/Rstudio [R version 4.0.4]). To assign a date of occurrence for each case, episode date was used (the earliest available report of either symptom onset, test collection, test result, or death). Incidence rates were calculated by the number of new PCR-confirmed COVID-19 cases per 100,000 population within each sewershed; sewershed service populations were self-reported by each sanitation district ([Figure 1](#)).

Prior to data analyses, laboratory staff preprocessed ddPCR or qPCR data. Because each laboratory performed data processing prior to data sharing, laboratory-specific choices were made about reporting extraction replicates, averaging of data from those extraction replicates, LOD determination, and outlier removal prior to downstream analyses. Extraction replicates from Laboratory 4 were averaged using a geometric mean and extraction replicates from Laboratory 2 were averaged using an arithmetic mean. Any measurements reported as below the laboratory's LOD were included at a concentration of one half of that laboratory's estimated LOD. Because distributions of wastewater concentrations and case incidence rates were right-skewed, wastewater concentrations, case incidence rates, and moving averages of both variables were log-transformed (base 10) prior to further analyses.

To investigate the relationships between SARS-CoV-2 RNA concentrations in wastewater (both raw or PMMoV-normalized) and sewershed-bounded case incidence rates of COVID-19, Spearman's rank correlation tests were utilized. Spearman correlations were performed separately for raw wastewater concentrations, 10-day moving averages of raw wastewater concentrations, PMMoV-normalized wastewater concentrations, and 10-day moving averages of PMMoV-normalized wastewater concentrations. Some laboratories measured both N1 and N2 genes to quantify SARS-CoV-2 RNA concentrations, and analyses were performed separately for N1 and N2 targets. Due to the inherent variability in wastewater data, it is common practice to use rolling, moving averages, or other smoothing techniques to reduce noise ([Rauch et al. 2022](#)). For wastewater measurements, 10-day moving averages can be useful to accommodate varying sampling schedules

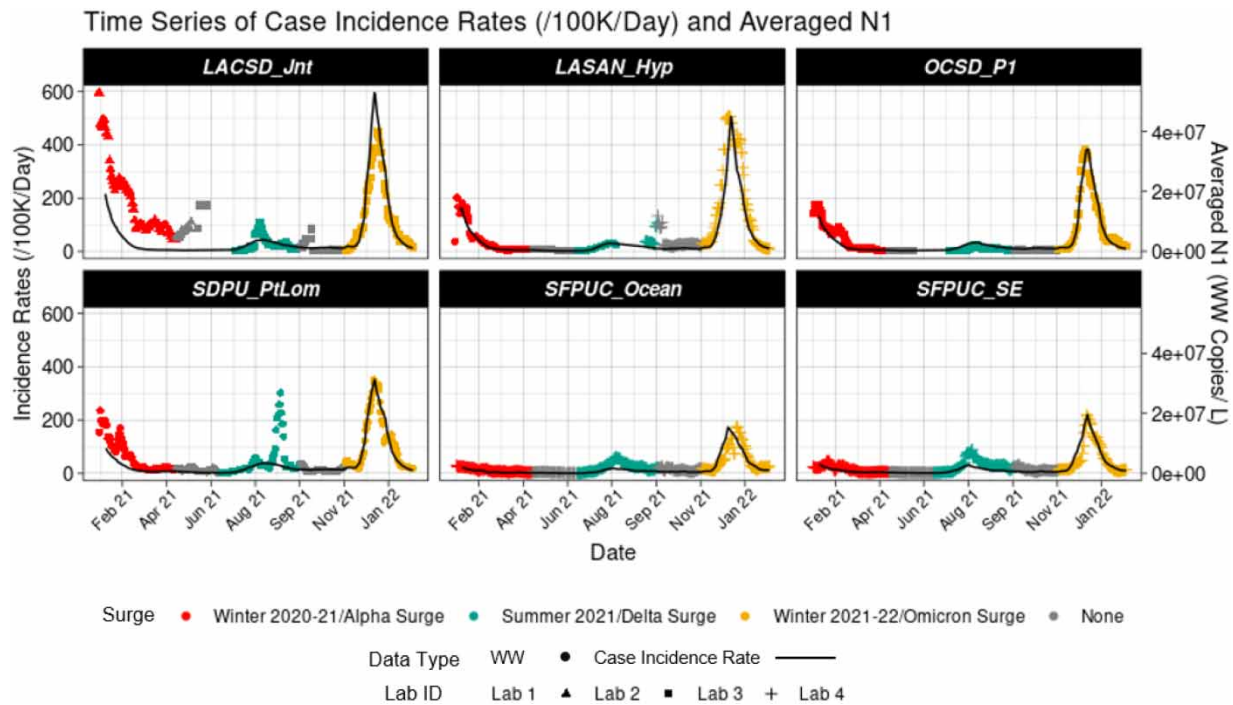


Figure 1 | Time series of 10-day averaged N1 gene concentrations (points) and 7-day averaged COVID-19 incidence rates (black solid lines) in six California sewersheds from four counties: Los Angeles County (Los Angeles Sanitation [LASAN Hyp] and Los Angeles County Sanitation District [LACSD Jnt]), San Francisco County (San Francisco Public Utilities Commission, Oceanside [SFPUC Ocean] and Southeast [SFPUC SE]), San Diego County (San Diego Public Utilities [SDPU Pt Lom]), and Orange County (Orange County Sanitation District [OC San]). Data points from each unique laboratory (Laboratory 1: SCCWRP, Laboratory 2: LACSD Laboratory, Laboratory 3: Zymo, Laboratory 4: UCB) are indicated by a different shape. Colors distinguish between COVID-19 variant-specific periods. LACSD Jnt and LASAN Hyp changed laboratories during the study period.

at different sewersheds (which can range from one to seven samples per week). For case incidence rates, 7-day moving averages are useful to accommodate variable reporting by day of the week. For the primary analysis of this study, the 10-day moving average was utilized for wastewater data and the 7-day moving average was applied for case incidence rates. A sensitivity analysis compared Spearman's correlation coefficients between N1 concentrations and case incidence rates across different moving average windows (7-, 10-, 14-, and 21-days) was applied to investigate if correlations remained similar.

To investigate how the associations between wastewater SARS-CoV-2 RNA and case incidence rates compared between COVID-19 variant periods, both Spearman correlations and simple linear regression models were fit using subset data from the winter 2020–2021/Alpha (B.1.1.7 and Q lineages) period, the summer 2021/Delta (B.1.617.2 and descendent lineages) period, and finally the winter 2021–2022/Omicron (B.1.1.529 and descendent lineages) period. The dates of these variant periods were defined using statewide case counts, vaccination rates, and prevalence of circulating VOCs. Variant data made available to CDPH were generated by the network of California laboratories known as 'California COVIDNet' (a major source of case-based genomic surveillance in California, composed of local and CDPH public health laboratories, academic partners, and commercial laboratories) and joined with case data via the state's reportable diseases system, CalREDIE.

For these analyses, the winter 2020–2021/Alpha period was defined as 1 January–15 April 2021. April 15 was selected as the end date because case counts had plateaued at a low level, vaccination rates (fully vaccinated) of eligible persons (18 and older) in California were lower at 31–44% in each sewershed but increasing, and, finally, regional circulation of the emerging Delta variant was estimated to be <2% of all circulating variants while Alpha remained the predominate lineage (>50%) (California, *Vaccination data*). Fully vaccinated individuals were defined as those who had received two doses of the mRNA vaccines (Pfizer and Moderna) or one dose of the Janssen vaccine. The summer 2021/Delta period was defined as 15 June–15 August 2021. June 15 was selected as the start date because case counts were still low but starting to increase,

sewershed vaccination rates had reached 59–76%, and Delta had taken over, accounting for approximately 50% of isolates sequenced and rapidly increasing. Finally, the winter 2021–2022/Omicron period was defined as 1 December 2021–2 March 2022. Based on sewershed averages, Delta remained the predominant variant with regards to COVID cases at the start date of this period, but Omicron rapidly and almost completely replaced Delta within weeks. Vaccination rates during this period were high, between 76 and 88% (California, State of, 2022).

The significance of associations between wastewater concentrations and case incidence rates from simple linear regression models were assessed using Newey West standard errors, to account for autocorrelation in model residuals, which arose from using moving averages as the predictor and outcome in these models (Newey & West 1987). It was hypothesized that the associations between wastewater and case incidence rates were unlikely to be linear for all models, so models with quadratic terms and cubic terms were also fit for each sub-dataset and compared against linear models using ANOVA. Finally, slope coefficients from each period were compared using the nonparametric Kruskal–Wallis test and using a permutation-based ANOVA.

RESULTS

Spearman's correlation analyses showed strong, positive, and statistically significant associations between wastewater SARS-CoV-2 concentrations and corresponding sewershed case incidence rates for all six sewersheds during the entire study period (Table 1, Figure 2). The strength of Spearman correlations varied by sewershed, gene target, and adjustments for viral RNA concentrations (e.g., normalization by PMMoV and applying moving averages) (Table 1). Generally, the weakest correlations were between unadjusted wastewater concentrations and case incidence rates ($\rho = 0.62$ – 0.88 and 0.79 – 0.92 for N1 and N2, respectively), with the exception of LACSD_Jnt – Laboratory 3 (Table 1). After applying moving averages with 10-day windows to wastewater concentrations and 7-day windows to incidence rates, correlations for the N1 and N2 genes strengthened at most sites ($\rho = 0.73$ – 0.93 and $\rho = 0.85$ – 0.98 for N1 and N2, respectively). Sensitivity analyses comparing 7-, 10-, 14-, and 21-day averaging windows for N1 gene wastewater concentrations and sewershed-restricted incidence rates revealed small overall changes in correlation, which varied by sewershed and laboratory. Strongest correlations were generally observed using 21-day moving averages for both wastewater and cases ($\rho = 0.78$ – 0.96), which may be due to an over-smoothing effect (Table 2).

Table 1 | Spearman's correlation coefficient rho (ρ) comparing wastewater SARS-CoV-2 RNA and COVID-19 clinical incidence rates at six different wastewater treatment plants

Plant	Laboratory ID	Target	Raw WW samples		Raw 10D MA WW		PMMoV-normalized WW samples		PMMoV-normalized 10D MA WW		Total samples
			ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value	
LACSD_Jnt	Laboratory 2	N1	0.816	<0.005	0.882	<0.005	NA	NA	NA	NA	38
LACSD_Jnt	Laboratory 3	N1	0.863	<0.005	0.734	<0.005	NA	NA	NA	NA	87
LASAN_Hyp	Laboratory 1	N1	0.876	<0.005	0.931	<0.005	0.914	<0.005	0.97	<0.005	88
		N2	0.866	<0.005	0.951	<0.005	0.892	<0.005	0.976	<0.005	88
LASAN_Hyp	Laboratory 4	N1	0.623	<0.005	0.733	<0.005	0.487	<0.005	0.488	<0.005	67
OCSD_P1	Laboratory 3	N1	0.769	<0.005	0.826	<0.005	NA	NA	NA	NA	135
		N2	0.915	<0.005	0.983	<0.005	NA	NA	NA	NA	25
SDPU_PtLom	Laboratory 1	N1	0.764	<0.005	0.828	<0.005	0.768	<0.005	0.813	<0.005	167
		N2	0.79	<0.005	0.848	<0.005	0.809	<0.005	0.864	<0.005	164
SFPUC_Ocean	Laboratory 4	N1	0.815	<0.005	0.923	<0.005	0.719	<0.005	0.817	<0.005	192
SFPUC_SE	Laboratory 4	N1	0.883	<0.005	0.909	<0.005	0.814	<0.005	0.908	<0.005	218

Incidence rates (7-day moving averages) correlated with (from left to right) raw N1 and N2 gene copies/L, 10-day moving averages (10D MA) of N1 and N2, raw N1/PMMoV and N2/PMMoV, and 10-day moving averages (10D MA) of N1/PMMoV and N2/PMMoV. LACSD Jnt switched laboratories during the study period: Laboratory 2 was utilized between 1 January 2021 and 28 April 2021, and Laboratory 3 between 1 May 2021 and 2 March 2022. LASAN Hyp switched laboratories as well: Laboratory 1 was utilized between 1 January 2021 and 28 July 2021, and Laboratory 4 between 6 September 2021 and 2 March 2022. For these two plants, separate correlation coefficients were calculated for each laboratory.

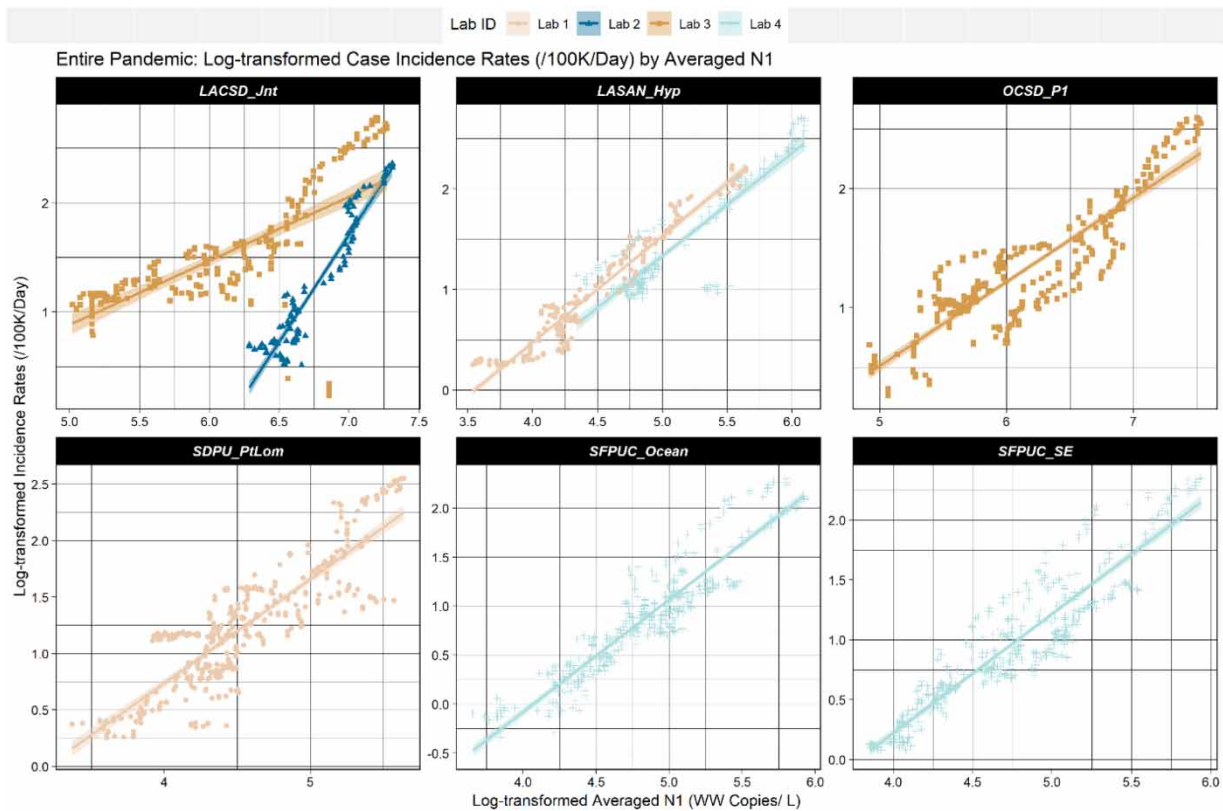


Figure 2 | Fitted linear regression models of 10-day averaged N1 gene concentrations and 7-day averaged COVID-19 incidence rates in six California sewersheds from four counties: Los Angeles County (Los Angeles Sanitation [LASAN Hyp] and Los Angeles County Sanitation District [LACSD Jnt]), San Francisco County (San Francisco Public Utilities Commission, Oceanside [SFPUC Ocean] and Southeast [SFPUC SE]), San Diego County (San Diego Public Utilities [SDPU PtLom]), and Orange County (Orange County Sanitation District [OC San]). Colors distinguish between different laboratories (Laboratory 1: SCCWRP, Laboratory 2: LACSD Laboratory, Laboratory 3: Zymo, Laboratory 4: UCB). LACSD Jnt and LASAN Hyp changed laboratories during the study period.

For sites where PMMoV measurements were available, normalizing wastewater concentrations for estimated human fecal strength weakened correlations for three sewersheds, all using the same laboratory (SFPUC_Ocean, SFPUC_SE, and LASAN_Hyp – Laboratory 4), and strengthened correlations for the remaining two sites (SDPU_PtLom and LASAN_Hyp – Laboratory 1) (Table 1). For the sewersheds where the N2 gene was quantified, correlations were strong, positive ($\rho = 0.85\text{--}0.98$) and improved with averaging. For the two sewersheds where both N2 and PMMoV were quantified, normalization of N2 by PMMoV improved correlations (Table 1).

During each variant period, Spearman correlations between 10-day moving averages of wastewater concentrations and 7-day moving averages of incidence rates remained strong ($\rho \geq 0.78$) and positive (Table 3, Figure 3). For all sewershed/laboratory combinations with more than 10 samples analyzed during the variant period, correlations between 10-day N1 gene concentrations with 7-day case incidence rates were similar across all three variant periods (for Alpha $\rho = 0.87\text{--}0.97$, Delta $\rho = 0.78\text{--}0.97$, Omicron $\rho = 0.85\text{--}0.97$).

As a complement to Spearman correlations, linear regression models were fit to assess the associations between wastewater concentrations and incidence rates during different variant periods. Errors for these models were normally distributed and homoscedastic but were autocorrelated due to the use of moving averages in both predictors and outcomes. Heteroskedasticity- and autocorrelation-consistent standard errors were used to estimate standard errors and p -values for these regression models. There was also visual evidence that the relationships between wastewater concentrations and case incidence rates were not linear for every sewershed and variant period (Figure 3). We tested for evidence of nonlinear relationships (comparing linear, quadratic, and cubic models) between SARS-CoV-2 concentrations and corresponding sewershed case incidence rates during all variant periods. This analysis suggested that a linear model provided the best fit for only

Table 2 | Sensitivity Analysis (Spearman's ρ) by treatment plant (Los Angeles Sanitation [LASAN Hyp] and Los Angeles County Sanitation District [LACSD Jnt], San Francisco Public Utilities Commission, Oceanside [SFPUC Ocean] and Southeast [SFPUC SE], San Diego Public Utilities [SDPU Pt Lom], and Orange County Sanitation District [OC San]) and laboratory (Laboratory 1: SCCWRP, Laboratory 2: LACSD Laboratory, Laboratory 3: Zymo, Laboratory 4: UCB), comparing wastewater (N1) and incidence rate averaging windows

Plant	Laboratory ID	WW 10D + IR 7D	WW 7D + IR 7D	WW 14D + IR 7D	WW 21D + IR 7D	WW 10D + IR 10D	WW 7D + IR 10D	WW 14D + IR 10D	WW 21D + IR 10D	WW 10D + IR 14D	WW 7D + IR 14D	WW 14D + IR 14D	WW 21D + IR 14D	WW 10D + IR 21D	WW 7D + IR 21D	WW 14D + IR 21D	WW 21D + IR 21D
LACSD_Jnt	Laboratory 2	0.882	0.854	0.932	0.954	0.873	0.841	0.927	0.95	0.859	0.823	0.918	0.944	0.828	0.792	0.901	0.932
LACSD_Jnt	Laboratory 3	0.734	0.793	0.79	0.833	0.747	0.805	0.809	0.857	0.76	0.814	0.826	0.88	0.764	0.806	0.837	0.902
LASAN_Hyp	Laboratory 1	0.931	0.925	0.932	0.933	0.931	0.922	0.934	0.939	0.931	0.917	0.938	0.949	0.923	0.902	0.936	0.956
LASAN_Hyp	Laboratory 4	0.733	0.746	0.709	0.71	0.737	0.735	0.719	0.722	0.733	0.735	0.739	0.744	0.724	0.706	0.754	0.783
OCSD_P1	Laboratory 3	0.826	0.822	0.804	0.776	0.841	0.835	0.822	0.797	0.863	0.854	0.849	0.829	0.885	0.871	0.877	0.869
SDPU_PtLom	Laboratory 1	0.828	0.808	0.839	0.823	0.835	0.81	0.849	0.842	0.841	0.815	0.861	0.862	0.841	0.808	0.869	0.882
SFPUC_Ocean	Laboratory 4	0.923	0.901	0.932	0.936	0.925	0.9	0.938	0.945	0.92	0.891	0.938	0.951	0.906	0.871	0.931	0.955
SFPUC_SE	Laboratory 4	0.909	0.912	0.906	0.89	0.92	0.917	0.919	0.908	0.925	0.919	0.931	0.928	0.92	0.906	0.935	0.946

Averaging windows varied from 7 to 21 days for both wastewater (WW) and case incidence rate (IR).

Table 3 | Correlation and regression analysis of each sewershed and laboratory combination across different variant periods

Plant (N1)	Laboratory ID	Winter 2020–2021/Alpha Surge							Summer 2021/Delta Surge							Winter 2021–2022/Omicron Surge						
		<i>n</i>	<i>w</i>	<i>P</i>	<i>p</i> [<i>ρ</i>]	Slope	Robust <i>p</i> (slope)		<i>n</i>	<i>w</i>	<i>ρ</i>	<i>p</i> [<i>ρ</i>]	Slope	Robust <i>p</i> (slope)		<i>n</i>	<i>w</i>	<i>ρ</i>	<i>p</i> [<i>ρ</i>]	Slope	Robust <i>p</i> (slope)	
LACSD_Jnt*	Laboratory 2	93	3.58	0.91	< 0.005	1.94	< 0.005		NA	NA	NA	NA	NA	NA		NA	NA	NA	NA	NA	NA	
LACSD_Jnt	Laboratory 3	NA	NA	NA	NA	NA	NA		79	3.41	0.84	< 0.005	0.45	< 0.005		92	4.3	0.97	< 0.005	1.01	< 0.005	
LASAN_Hyp*	Laboratory 1	99	4.11	0.87	< 0.005	1.05	< 0.005		54	4.02	0.97	< 0.005	1.20	< 0.005		NA	NA	NA	NA	NA	NA	
LASAN_Hyp	Laboratory 4	NA	NA	NA	NA	NA	NA		NA	NA	NA	NA	NA	NA		92	3.8	0.96	< 0.005	0.92	< 0.005	
OCSD_P1	Laboratory 3	85	3.59	0.97	< 0.005	0.78	< 0.005		88	4.06	0.81	< 0.005	0.57	< 0.005		92	4.28	0.95	< 0.005	0.94	< 0.005	
SDPU_PtLom	Laboratory 1	99	4.06	0.90	< 0.005	1.00	< 0.005		109	4.2	0.78	< 0.005	0.68	0.16		92	3.64	0.93	< 0.005	1.08	< 0.005	
SFPUC_Ocean	Laboratory 4	99	4.17	0.91	< 0.005	1.14	< 0.005		107	5.63	0.96	< 0.005	0.93	< 0.005		91	3.87	0.85	< 0.005	1.10	< 0.005	
SFPUC_SE	Laboratory 4	98	4.79	0.87	< 0.005	0.81	< 0.005		109	6.13	0.91	< 0.005	0.81	< 0.005		92	4.15	0.89	< 0.005	1.00	< 0.005	

Total number of data points included in period-specific simple linear regression (*n*); average number of wastewater samples collected per 10-day window during time period (*w*). Spearman's rho (*ρ*) and associated *p*-value (*p* [*ρ*]); regression model slopes (slope) and corresponding *p*-values of log10-transformed COVID-19 incidence (7-day average) with SARS-CoV-2 RNA (10-day averaged N1) by variant period (Winter 2020–2021/Alpha, Summer 2021/Delta, or Winter 2021–2022/Omicron). The slope robust *p*-values (Robust *P* [slope]) indicate the results of *t*-tests against the null hypotheses that each slope coefficient is zero (e.g., there is no relationship between wastewater concentrations and incident rates). *LACSD Jnt switched laboratories during the period of interest; quantifications from the Winter 2020–2021 period are derived using data from Laboratory 2, while the remaining two period correspond to data from Laboratory 3. Similarly, LASAN Hyp switched from Laboratory 1 to Laboratory 4 on 29 July 2021, and calculations were performed separately for each laboratory. (Laboratory 1: SCCWRP, Laboratory 2: LACSD Laboratory, Laboratory 3: Zymo, Laboratory 4: UCB).

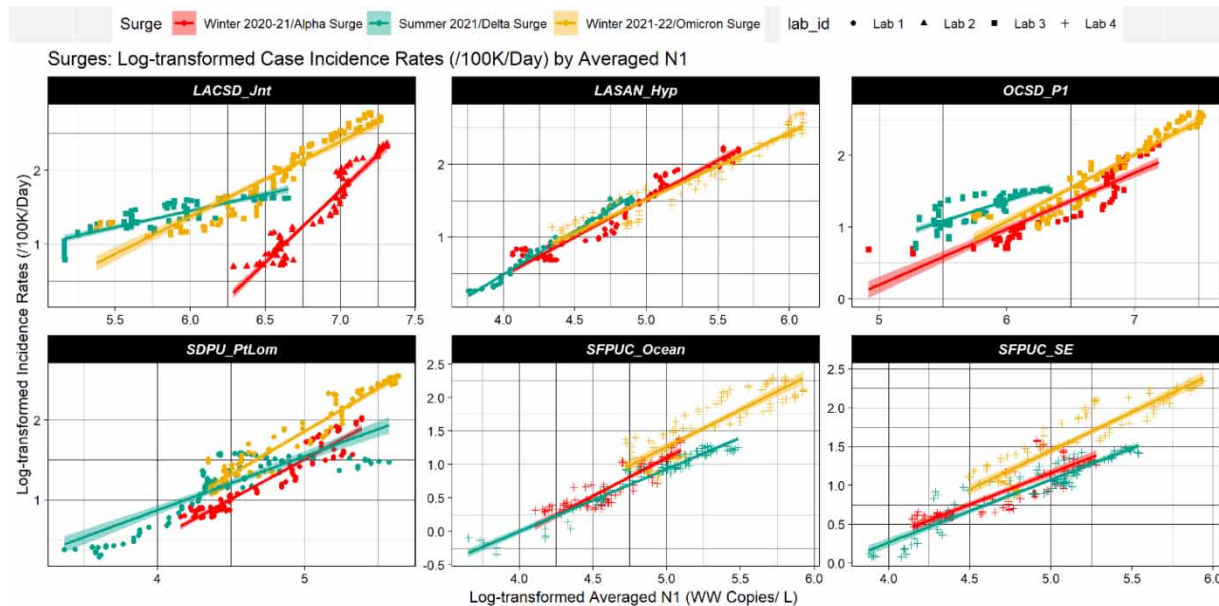


Figure 3 | Fitted linear regression models between 10-day averaged N1 gene concentrations and 7-day averaged COVID-19 incidence rates in different sewersheds, separated by winter 2020–2021/Alpha, summer 2021/Delta, and winter 2021–2022/Omicron surges. LACSD Jnt and LASAN Hyp changed laboratories during the pandemic, and therefore have separate linear regression models corresponding to each data source. Data points from unique laboratories are indicated by different shapes (Laboratory 1: SCCWRP, Laboratory 2: LACSD Laboratory, Laboratory 3: Zymo, Laboratory 4: UCB).

five of the sewershed-laboratory-period combinations. For all other subset datasets, quadratic or cubic models improved model fit (Table 4). Despite this evidence of nonlinearity in associations in many of the sewershed-laboratory-period combinations, the linear beta coefficients are presented in Table 3 and plotted in Figure 3 to allow for a more straightforward interpretation of coefficients and for direct comparisons across variant periods, sewersheds, and laboratories.

Except for SDPU_PtLom, Laboratory 3, during the Delta period, all sewershed laboratory-period combinations yielded statistically significant associations between wastewater and incidence rates ($p < 0.05$). Estimated slope coefficients ranged from 0.45 to 1.94, suggesting that the slopes of the associations varied by subset data (Table 3). A comparison of slope coefficients across time periods using a nonparametric Kruskal–Wallis test (while not considering laboratory or sewershed) did not reveal a statistically significant difference in slopes (0.074). A permutation-based, multivariate ANOVA, considering all three grouping variables, did reveal statistically significant differences in slope by sewershed ($p = 0.015$), laboratory (0.001), and variant period (0.004).

DISCUSSION

We found strong and positive correlations between wastewater SARS-CoV-2 concentrations and case incidence rates in six sewersheds that collectively represent 30% of the California population (Table 1, Figures 1 and 2). These correlations were statistically significant across four different laboratories and throughout the study period. Generally, 10-day moving average smoothed wastewater concentrations correlated better with 7-day moving average smoothed case incidence rates, as compared to non-averaged wastewater concentrations and daily incidence rates. This suggests that smoothing both wastewater and case data prior to downstream analyses is useful. Correlations were similarly strong and statistically significant during each of the Alpha, Delta, and Omicron periods spanning 2021–2022 (Table 3). The slopes of linear associations between wastewater and case incidence rates across three distinct time periods revealed that slope magnitudes varied by sewershed, laboratory, and variant period, suggesting that while highly correlated, the relationship between wastewater concentration to case incidence may vary according to these factors (Figure 3 and Table 3). Our results support wastewater surveillance as an adjunctive tool to complement traditional COVID-19 clinical surveillance, despite heterogeneous laboratory methodologies and sewersheds, as well as time-varying population immunity and variant predominance.

Table 4 | Investigation of nonlinearity in the associations between SARS-CoV-2 concentrations and incidence rates across all sewersheds, labs, and variant periods

Plant	Laboratory ID	Nonlinearity Testing Results		
		Winter 2020–2021/Alpha Surge	Summer 2021/Delta Surge	Winter 2021–2022/Omicron Surge
LACSD_Jnt	Laboratory 2	Nonlinear	NA	NA
LACSD_Jnt	Laboratory 3	NA	Nonlinear	Nonlinear
LASAN_Hyp	Laboratory 1	Nonlinear	Nonlinear	NA
LASAN_Hyp	Laboratory 4	NA	NA	Nonlinear
OCSD_P1	Laboratory 3	Nonlinear	Linear	Nonlinear
SDPU_PtLom	Laboratory 1	Nonlinear	Nonlinear	Linear
SFPUC_Ocean	Laboratory 4	Nonlinear	Nonlinear	Linear
SFPUC_SE	Laboratory 4	Linear	Nonlinear	Linear

For each model, a linear model was compared against quadratic and cubic models by an ANOVA test. Where evidence of nonlinearity was found by ANOVA ($p < 0.05$), the best-fit model is indicated as 'nonlinear' since an exhaustive analysis of different nonlinear modeling forms was not done.

These results are consistent with earlier analyses showing a correlation between wastewater and clinical surveillance data (Greenwald *et al.* 2021; Weidhaas *et al.* 2021; Duvall *et al.* 2022; Hegazy *et al.* 2022; Ando *et al.* 2023). Most correlations-based analyses investigated wastewater and case surveillance data from earlier in the COVID-19 pandemic or included only the Alpha and/or Delta periods. One of the few Omicron-inclusive studies by Hegazy *et al.* (2022) revealed a weak Spearman's correlation during the Delta period, as compared to Alpha and Omicron. We found similar correlation strength across all three variant periods, suggesting the reliability of wastewater to serve as a proxy for community disease trends over time and changing population immunity and variant dynamics.

The weakest correlations were observed between raw, non-averaged, non-normalized wastewater concentrations and non-averaged case incidence rates. Correlations of COVID-19 wastewater concentration moving averages and case incidence rate moving averages, both averaged over different time windows (7-, 10-, 14-, and 21-days), produced stronger correlations. A likely reason for this is the smoothing of expected environmental variability of wastewater samples, which contain a complex matrix of both organic and inorganic materials. Samples taken for analysis are relatively small volumes, and even when collected and composited over the duration of 24-h, cannot perfectly represent the millions of gallons of heterogeneous wastewater matrix flowing through a treatment plant each day. External influences on the samples, such as the amount of rainfall, industrial input, chemical makeup, temperature, and sewage travel time, can also impact measured concentrations and add variability (Martins *et al.* 2022; Li *et al.* 2023). Another cause of variability is in the day-to-day population dynamics of a sewershed. For example, if a sewershed represents more of a commercial or commuter population traveling to and from the sewershed each day, concentrations on weekends could be lower than through the work week. Similarly, case counts on different days of the week can be variable due to testing access and utilization (e.g., fewer testing sites open on weekends).

Since averaging windows are routinely used and reported in public health surveillance to attenuate day-to-day variability and improve the interpretability of both case and wastewater data, we performed a sensitivity analysis to help determine if an ideal averaging window existed. Wastewater concentrations averaged over longer periods of time (10-, 14-, and 21-days) paired with correspondingly longer case averaging periods appeared to have somewhat stronger correlations compared to 7-day averages, as might be expected when smoothing data. Despite small improvements in correlation offered by these increased averaging windows, we selected a 10-day moving average for wastewater (capturing on average 3–5 data points) and a 7-day moving average for cases (allowing aggregation of weekday and weekend case data) to dampen noise for each data source while maintaining more temporally granular and recent information on trends.

Correlations differed between sewershed, gene target, and with PMMoV normalization. At the three sites that measured both N1 and N2, correlations between case incidence rates and N2 were comparable to correlations for N1 (Table 2, Supplemental Figure 1). For two of the five sewershed – laboratory combinations (LASAN Hyp – Laboratory 1, SDPU Pt Lom – Laboratory 1) that measured PMMoV, normalization improved correlations; at the three remaining sewersheds (SFPUC Ocean, SFPUC SE, and LASAN Hyp – Laboratory 4), normalization to PMMoV weakened correlations. This may have been related to laboratory methodology; samples from all three sewersheds where decreased correlations were

observed were analyzed by the same laboratory (Laboratory 4). Additionally, samples taken from one site (LASAN Hyp) had an improved correlation when analyzed by Laboratory 1 and weaker correlation after normalization when analyzed by Laboratory 4. Another potential influence on the varied impact of PMMoV normalization could be a result of nonrandom differences in diets or pepper consumption behaviors in the populations served by the treatment plants at different times. These findings are consistent with other studies where PMMoV normalization did not consistently show a significant improvement to the correlations between clinical data and wastewater data (Greenwald *et al.* 2021; Li *et al.* 2022; Zheng *et al.* 2022; Maal-Bared *et al.* 2023). It is important to note that there is no clear consensus in the literature about optimal methods for data (e.g., averaging, time-shifting, and normalization) of wastewater and case-based surveillance data for the purpose of correlations-based analysis. This suggests that correlations can vary between sites and with each laboratory methodology and that individualized analyses can be useful to identify the best ways to preprocess and transform data at each site. However, in our study across disparate sites, time periods, and laboratory methods, averaging of data outperformed non-averaged data correlations, and unnormalized wastewater data performed similarly to PMMoV-normalized data. These results suggest that raw, non-normalized wastewater data, smoothed by applying moving averages, would be a good starting point for comparing wastewater data to clinical data in various settings.

Excluding comparison between variant periods at sites where the analysis laboratory changed (LACSD Jnt, LASAN Hyp), correlations remained similar and strong at all six sites throughout the three different variant periods. While the strength of correlations (ρ) remained similar, non-significant differences in the slope of linear associations between wastewater and cases were noted across time periods, with the highest slopes generally observed during the Omicron period (Table 3). However, given a relatively small sample size (six sewersheds), multiple within-sewershed laboratory changes, and the decision to not directly compare wastewater data produced from different laboratory methods, it was difficult to parse out exactly how the underlying associations between wastewater and case incidence rates may have changed during variant periods. Influencing factors may include differences in laboratory methods, different underlying populations, environmental variables between sewersheds, changes in testing access or utilization, as well as environmental and population changes between the variant periods. Differences in fecal viral shedding of infected persons with different variants of concern (VOCs) and vaccination status may also be important contributing factors. Future studies with long-term wastewater data utilizing a single laboratory covering multiple variant-dominant time periods and analyses incorporating other clinical metrics (such as test positivity rates and hospitalization rates) will add clarity to the question of whether associations between wastewater and clinical data change as dominant COVID-19 variants and population immunity change.

Our analysis of linearity in associations between wastewater and case incidence rates suggested that, depending on the underlying data (e.g., which sewershed, laboratory, or variant period is used), nonlinear models provide improved model fit, over linear models (Table 4). Thus, some of the linear models presented in Figure 3 and Table 3 would likely not have strong predictive value. Despite this, presenting linear models in this analysis allows for direct visual and statistical comparison of linear associations across different variant periods and between sewersheds and laboratories. In this study, only quadratic and cubic models were compared against linear models, and thus conclusions cannot be drawn about the best nonlinear fit for these data. Future investigations of optimal nonlinear modeling approaches will improve our understanding of these associations and allow for better predictive modeling of wastewater-derived incidence rates.

Each utility was allowed to use a laboratory of their choice, resulting in data from four labs. This introduced an additional source of variability in the data from four distinct protocols of concentration, extraction, and analysis. Overall correlations remained strong regardless of laboratory, suggesting that different laboratory methods can be reliable for wastewater surveillance. However, there were important differences between laboratories. Notably, normalization with PMMoV weakened correlation at one laboratory while improving correlation at three laboratories. Additionally, while correlations were strong regardless of laboratory used, the slope of linear association between laboratories was not always the same, even for analyses done in the same sewershed but utilizing different laboratories (Figure 2). This suggests that review and evaluation regarding correlations are important for interpreting results, especially when laboratories and methods change over time. Additional caution should be used when directly comparing wastewater data from different laboratories.

CONCLUSIONS

During a 14-month period that included changing dominance of three different VOCs (Alpha, Delta, and Omicron) and the increasing, widespread adoption of newly introduced vaccines, strong and significant Spearman correlations ($\rho \geq 0.73$;

$p < 0.05$) were observed between 10-day averaged wastewater viral RNA concentrations and 7-day averaged case incidence rates for all six sewersheds in California. These correlations were strong and significant regardless of the laboratory choice (four different laboratories with distinct protocols). The slopes of associations between wastewater concentrations and case incidence rates varied by the dominant variant, sewershed, and laboratory, suggesting further work is needed to understand how these different factors may impact the relationship between wastewater and case incidence rates over time. These results build confidence that trends in wastewater reflect underlying community disease activity and support the use of wastewater surveillance as an adjunctive tool for COVID-19 public health surveillance across heterogeneous sewersheds and laboratory methods.

ACKNOWLEDGEMENTS

This study was supported in part by the Epidemiology and Laboratory Capacity for Infectious Diseases Cooperative Agreement (no. 6NU50CK000539-03-02) from CDC. We thank the staff at the utilities for the hard and extra work to collect and ship these samples in efforts to benefit the public health and COVID-19 pandemic response. We also thank key California State Water Resources Control Board (SWRCB) staff from the Division of Water Quality and Office of Information Management and Analysis who supported and facilitated the implementation of this project. A special thank you to the staff at the utilities involved including Margil Jimenez and the Orange county Sanitation Team, the City of San Diego scientists from the Marine Microbiology and Environmental Chemistry Services laboratories, and operations crew from the Point Loma Wastewater Treatment Plant and Pump Station 2, the Los Angeles County Sanitation Districts Laboratories and Research Sections, the Catena Foundation and San Francisco Public Utilities Commission for supporting the Nelson Laboratory at the University of California, Berkeley and the Department of Microbiome Research and Bioinformatics at the Zymo Research Corp., particularly Michael Lisek, Kenneth Day, Liya Zhu, Kristopher Locken, Shuiquan Tang, and Xiaoxiao Cheng. Graphical abstract was created and designed using Canva Pro (web version).

DISCLAIMER

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views or opinions of the California Department of Public Health or the California Health and Human Services Agency.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

Dr. Rose Kantor carried out some work as a consultant for the Rockefeller Foundation in 2022 related to data analysis for sequencing SARS-CoV-2 in wastewater.

REFERENCES

- Aguiar-Oliveira, M. D., Campos, A. R., Matos, A., Rigotto, C., Sotero-Martins, A., Teixeira, P. F. & 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>.
- Al-Faliti, M., Kotlarz, N., McCall, C., Harris, A. R., Smith, A. L., Stadler, L. B., de los Reyes, F. L. & Delgado Vela, J. 2022 *Comparing rates of change in SARS-COV-2 wastewater load and clinical cases in 19 sewersheds across four major metropolitan areas in the United States. ACS ES&T Water* **2** (11), 2233–2242. <https://doi.org/10.1021/acsestwater.2c00106>.
- Ando, H., Murakami, M., Ahmed, W., Iwamoto, R., Okabe, S. & Kitajima, M. 2023 *Wastewater-based prediction of COVID-19 cases using a highly sensitive SARS-COV-2 RNA detection method combined with mathematical modeling. Environment International* **173**, 107743. <https://doi.org/10.1016/j.envint.2023.107743>.
- California, State of. 2022 *Vaccination Data. Coronavirus COVID-19 Response*. Available from: <https://covid19.ca.gov/vaccination-progress-data/>
- Chen, C., Gao, G., Xu, Y., Pu, L., Wang, Q., Wang, L., Wang, W., Song, Y., Chen, M., Wang, L., Yu, F., Yang, S., Tang, Y., Zhao, L., Wang, H., Wang, Y., Zeng, H. & Zhang, F. 2020 *SARS-COV-2-positive sputum and feces after conversion of pharyngeal samples in patients with COVID-19. Annals of Internal Medicine* **172** (12), 832–834. <https://doi.org/10.7326/m20-0991>.
- D'Aoust, P. M., Tian, X., Towhid, S. T., Xiao, A., Mercier, E., Hegazy, N., Jia, J.-J., Wan, S., Kabir, M. P., Fang, W., Fuzzen, M., Hasing, M., Yang, M. I., Sun, J., Plaza-Diaz, J., Zhang, Z., Cowan, A., Eid, W., Stephenson, S., Servos, M. R. & Delatolla, R. 2022 *Wastewater to*

- clinical case (WC) ratio of COVID-19 identifies insufficient clinical testing, onset of new variants of concern and population immunity in urban communities. *Science of the Total Environment* **853**, 158547. <https://doi.org/10.1016/j.scitotenv.2022.158547>.
- Duvallet, C., Wu, F., McElroy, K. A., Imakaev, M., Endo, N., Xiao, A., Zhang, J., Floyd-O'Sullivan, R., Powell, M. M., Mendola, S., Wilson, S. T., Cruz, F., Melman, T., Sathyanarayana, C. L., Olesen, S. W., Erickson, T. B., Ghaeli, N., Chai, P., Alm, E. J. & Matus, M. 2022 Nationwide trends in COVID-19 cases and SARS-COV-2 RNA wastewater concentrations in the United States. *ACS ES&T Water* **2** (11), 1899–1909. <https://doi.org/10.1021/acsestwater.1c00434>.
- 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>.
- Greenwald, H. D., Kennedy, L. C., Hinkle, A., Whitney, O. N., Fan, V. B., Crits-Christoph, A., Harris-Lovett, S., Flamholz, A. I., Al-Shayeb, B., Liao, L. D., Beyers, M., Brown, D., Chakrabarti, A. R., Dow, J., Frost, D., Koekemoer, M., Lynch, C., Sarkar, P., White, E., Kantor, R. & Nelson, K. L. 2021 Tools for interpretation of wastewater SARS-COV-2 temporal and spatial trends demonstrated with data collected in the San Francisco Bay area. *Water Research X* **12**, 100111. <https://doi.org/10.1016/j.wroa.2021.100111>.
- Hegazy, N., Cowan, A., D'Aoust, P. M., Mercier, É., Towhid, S. T., Jia, J.-J., Wan, S., Zhang, Z., Kabir, M. P., Fang, W., Graber, T. E., MacKenzie, A. E., Guilherme, S. & Delatolla, R. 2022 Understanding the dynamic relation between wastewater SARS-COV-2 signal and clinical metrics throughout the pandemic. *Science of the Total Environment* **853**, 158458. <https://doi.org/10.1016/j.scitotenv.2022.158458>.
- Kantor, R. S., Greenwald, H. D., Kennedy, L. C., Hinkle, A., Harris-Lovett, S., Metzger, M., Thornton, M. M., Paluba, J. M. & Nelson, K. L. 2022 Operationalizing a routine wastewater monitoring laboratory for SARS-COV-2. *PLOS Water* **1** (2). <https://doi.org/10.1371/journal.pwat.0000007>.
- Kim, S., Kennedy, L. C., Wolfe, M. K., Criddle, C. S., Duong, D. H., Topol, A., White, B. J., Kantor, R. S., Nelson, K. L., Steele, J. A., Langlois, K., Griffith, J. F., Zimmer-Faust, A. G., McLellan, S. L., Schussman, M. K., Ammerman, M., Wigginton, K. R., Bakker, K. M. & Boehm, A. B. 2022 SARS-COV-2 RNA is enriched by orders of magnitude in primary settled solids relative to liquid wastewater at publicly owned treatment works. *Environmental Science: Water Research & Technology* **8** (4), 757–770. <https://doi.org/10.1039/d1ew00826a>.
- Kirby, A. E., Walters, M. S., Jennings, W. C., Fugitt, R., LaCross, N., Mattioli, M., Marsh, Z. A., Roberts, V. A., Mercante, J. W., Yoder, J. & Hill, V. R. 2021 Using wastewater surveillance data to support the COVID-19 response – United States, 2020–2021. *MMWR. Morbidity and Mortality Weekly Report* **70** (36), 1242–1244. <https://doi.org/10.15585/mmwr.mm7036a2>.
- Kitajima, M., Sassi, H. P. & Torrey, J. R. 2018 Pepper mild mottle virus as a water quality indicator. *Npj Clean Water* **1** (1). <https://doi.org/10.1038/s41545-018-0019-5>.
- Li, L., Mazurowski, L., Dewan, A., Carine, M., Haak, L., Guarin, T. C., Dastjerdi, N. G., Gerrity, D., Mentzer, C. & Pagilla, K. R. 2022 Longitudinal monitoring of SARS-COV-2 in wastewater using viral genetic markers and the estimation of unconfirmed COVID-19 cases. *Science of The Total Environment* **817**, 152958. <https://doi.org/10.1016/j.scitotenv.2022.152958>.
- Li, X., Zhang, S., Sherchan, S., Orive, G., Lertxundi, U., Haramoto, E., Honda, R., Kumar, M., Arora, S., Kitajima, M. & Jiang, G. 2023 Correlation between SARS-COV-2 RNA concentration in wastewater and COVID-19 cases in community: a systematic review and meta-analysis. *Journal of Hazardous Materials* **441**, 129848. <https://doi.org/10.1016/j.jhazmat.2022.129848>.
- Maal-Bared, R., Qiu, Y., Li, Q., Gao, T., Hrudey, S. E., Bhavanam, S., Ruecker, N. J., Ellehoj, E., Lee, B. E. & Pang, X. 2023 Does normalization of SARS-COV-2 concentrations by pepper mild mottle virus improve correlations and lead time between wastewater surveillance and clinical data in Alberta (Canada): comparing twelve SARS-COV-2 normalization approaches. *Science of the Total Environment* **856**, 158964. <https://doi.org/10.1016/j.scitotenv.2022.158964>.
- Martins, R. M., Carvalho, T., Bittar, C., Quevedo, D. M., Miceli, R. N., Nogueira, M. L., Ferreira, H. L., Costa, P. I., Araújo, J. P., Spilki, F. R., Rahal, P. & Calmon, M. F. 2022 Long-term wastewater surveillance for SARS-COV-2: one-year study in Brazil. *Viruses* **14** (11), 2333. <https://doi.org/10.3390/v14112333>.
- 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>.
- Morvan, M., Jacomo, A. L., Souque, C., Wade, M. J., Hoffmann, T., Pouwels, K., Lilley, C., Singer, A. C., Porter, J., Evens, N. P., Walker, D. I., Bunce, J. T., Engeli, A., Grimsley, J., O'Reilly, K. M. & Danon, L. 2022 An analysis of 45 large-scale wastewater sites in England to estimate SARS-COV-2 community prevalence. *Nature Communications* **13** (1). doi:10.1038/s41467-022-31753-y.
- Natarajan, A., Zlitni, S., Brooks, E. F., Vance, S. E., Dahlen, A., Hedlin, H., Park, R. M., Han, A., Schmidtke, D. T., Verma, R., Jacobson, K. B., Parsonnet, J., Bonilla, H. F., Singh, U., Pinsky, B. A., Andrews, J. R., Jagannathan, P. & Bhatt, A. S. 2022 Gastrointestinal symptoms and fecal shedding of SARS-COV-2 RNA suggest prolonged gastrointestinal infection. *Med* **3** (6). <https://doi.org/10.1016/j.medj.2022.04.001>.
- Newey, W. & West, K. 1987 A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** (3), 703–708.
- Olesen, S. W., Imakaev, M. & Duvallet, C. 2021 Making waves: defining the lead time of wastewater-based epidemiology for COVID-19. *Water Research* **202**, 117433. <https://doi.org/10.1016/j.watres.2021.117433>.

- Park, S., Lee, C.-W., Park, D.-I., Woo, H.-Y., Cheong, H. S., Shin, H. C., Ahn, K., Kwon, M.-J. & Joo, E.-J. 2021 **Detection of SARS-COV-2 in fecal samples from patients with asymptomatic and mild COVID-19 in Korea.** *Clinical Gastroenterology and Hepatology* **19** (7). <https://doi.org/10.1016/j.cgh.2020.06.005>.
- Rauch, W., Schenk, H., Insam, H., Markt, R. & Kreuzinger, N. 2022 **Data modelling recipes for SARS-CoV-2 wastewater-based epidemiology.** *Environmental Research* **214** (1), 113809. <https://doi.org/10.1016/j.envres.2022.113809>.
- Schill, R., Nelson, K. L., Harris-Lovett, S. & Kantor, R. S. 2023 **The dynamic relationship between COVID-19 cases and SARS-COV-2 wastewater concentrations across time and space: considerations for model training data sets.** *Science of the Total Environment* **871**, 162069. <https://doi.org/10.1016/j.scitotenv.2023.162069>.
- Steele, J. A., Zimmer-Faust, A. G., Griffith, J. F. & Weisberg, S. B. 2021 **Sources of variability in methods for processing, storing, and concentrating SARS-COV-2 in influent from urban wastewater treatment plants.** <https://doi.org/10.1101/2021.06.16.21259063>
- 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.** *Science of The Total Environment* **775**, 145790. <https://doi.org/10.1016/j.scitotenv.2021.145790>.
- Whitney, O. N., Kennedy, L. C., Fan, V. B., Hinkle, A., Kantor, R., Greenwald, H., Crits-Christoph, A., Al-Shayeb, B., Chaplin, M., Maurer, A. C., Tjian, R. & Nelson, K. L. 2021 **Sewage, salt, silica, and SARS-COV-2 (4s): an economical kit-free method for direct capture of SARS-COV-2 RNA from wastewater.** *Environmental Science & Technology* **55** (8), 4880–4888. <https://doi.org/10.1021/acs.est.0c08129>.
- Wolfe, M. K., Archana, A., Catoe, D., Coffman, M. M., Dorevich, S., Graham, K. E., Kim, S., Grijalva, L. M., Roldan-Hernandez, L., Silverman, A. I., Sinnott-Armstrong, N., Vugia, D. J., Yu, A. T., Zambrana, W., Wigginton, K. R. & Boehm, A. B. 2021a **Scaling of SARS-COV-2 RNA in settled solids from multiple wastewater treatment plants to compare incidence rates of laboratory-confirmed COVID-19 in their sewersheds.** *Environmental Science & Technology Letters* **8** (5), 398–404. <https://doi.org/10.1021/acs.estlett.1c00184>.
- Wolfe, M. K., Topol, A., Knudson, A., Simpson, A., White, B., Vugia, D. J., Yu, A. T., Li, L., Balliet, M., Stoddard, P., Han, G. S., Wigginton, K. R. & Boehm, A. B. 2021b **High-frequency, high-throughput quantification of SARS-COV-2 RNA in wastewater settled solids at eight publicly owned treatment works in northern California shows strong association with COVID-19 incidence.** *mSystems* **6** (5). <https://doi.org/10.1128/msystems.00829-21>.
- 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. & Alm, E. J. 2022 **SARS-COV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases.** *Science of the Total Environment* **805**, 150121. <https://doi.org/10.1016/j.scitotenv.2021.150121>.
- Zheng, X., Li, S., Deng, Y., Xu, X., Ding, J., Lau, F. T. K., In Yau, C., Poon, L. L. M., Tun, H. M. & Zhang, T. 2022 **Quantification of SARS-COV-2 RNA in wastewater treatment plants mirrors the pandemic trend in Hong Kong.** *Science of the Total Environment* **844**, 157121. <https://doi.org/10.1016/j.scitotenv.2022.157121>.

First received 15 June 2023; accepted in revised form 28 August 2023. Available online 11 September 2023