Supporting Information for

**Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves**

**Authors:** Loren Hopkins1,2, David Persse1,3,4, Kelsey Caton1, Katherine Ensor2,Rebecca Schneider1, Camille McCall5, and Lauren B. Stadler5,\*

**Affiliations:**

1Houston Health Department, 8000 N. Stadium Dr., Houston, TX

2Department of Statistics, Rice University, 6100 Main Street MS 138, Houston, TX

3Department of Medicine and Surgery, Baylor College of Medicine, Houston, TX

4City of Houston Emergency Medical Services, Houston, TX

5Department of Civil and Environmental Engineering, Rice University, 6100 Main Street MS-519, Houston, TX

\*Corresponding author: [lauren.stadler@rice.edu](mailto:lauren.stadler@rice.edu)

**1. Materials and Methods**

Table S1. Wastewater treatment plants sampled, average flow rates, service populations, and geographic service areas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Wastewater treatment plant** | **Abbreviation** | **Flowrate, MGD (AVG ± SD)** | **Population** | **Average gal/cap/day** | **Area, square miles** |
| 69th Street | 69 | 80.03 ± 21.77 | 551,150 | 145 | 96.72 |
| Almeda Sims | AS | 13.76 ± 11.93 | 117,968 | 117 | 54.93 |
| Beltway | BW | 6.93 ± 4.70 | 70,900 | 98 | 9.76 |
| Cedar Bayou | CD | 0.78 ± 0.42 | 1,722 | 453 | 3.27 |
| Chocolate Bayou | CB | 4.03 ± 4.19 | 37,359 | 108 | 14.61 |
| Clinton Park | CP | 0.69 ± 0.81 | 3,825 | 180 | 4.14 |
| Easthaven | EH | 1.89 ± 1.85 | 16,030 | 118 | 4.78 |
| FWSD#23 | 23 | 3.09 ± 2.52 | 40,689 | 76 | 15.14 |
| Forest Cove | FC | 0.28 ± 0.13 | 4,170 | 67 | 2.73 |
| Greenridge | GR | 3.06 ± 3.06 | 28,742 | 106 | 6.87 |
| Homestead | HO | 1.58 ± 1.53 | 9,375 | 169 | 6.12 |
| Imperial Valley | IV | 1.75 ± 0.70 | 16,804 | 104 | 2.22 |
| Intercontinental Airport | IA | 1.91 ± 0.63 | 2,408 | 793 | 38.73 |
| Keegans Bayou | KB | 14.25 ± 10.31 | 124,000 | 115 | 13.78 |
| Kingwood Central | KW | 3.49 ± 1.46 | 52,055 | 67 | 23.04 |
| Kingwood West | MG | 0.61 ± 0.20 | 2,589 | 236 | 2.6 |
| MUD#203 | 203 | 0.38 ± 0.12 | 4,010 | 95 | 2.57 |
| Metro Central | MC | 1.99 ± 1.64 | 20,161 | 99 | 9.86 |
| Northbelt | NO | 2.37 ± 1.49 | 12,892 | 184 | 15.79 |
| Northeast | NE | 3.88 ± 4.25 | 33,102 | 117 | 14.41 |
| Northgate | NG | 2.75 ± 1.03 | 19,867 | 138 | 3.6 |
| Northwest | NW | 9.99 ± 5.84 | 95,600 | 104 | 22.62 |
| Park Ten | PT | 0.62 ± 0.31 | 5,497 | 113 | 2.19 |
| Sagemont | SG | 4.52 ± 3.49 | 20,608 | 219 | 5.9 |
| Sims Bayou South\* | SS | 23.93 ± 18.22 | 109,414 | 219 | 47.84 |
| Sims Bayou North\* | SB | 8.22 ± 7.54 | 109,414 | 75 | 47.84 |
| Southeast | SE | 4.88 ± 4.85 | 32,485 | 150 | 9.06 |
| Southwest | SW | 37.59 ± 26.39 | 293,227 | 128 | 38.72 |
| Tidwell Timbers | TT | 0.11 ± 0.06 | 1,133 | 97 | 0.57 |
| Turkey Creek | TC | 7.00 ± 4.85 | 59,188 | 118 | 10.46 |
| Upper Brays | UB | 10.33 ± 7.44 | 97,918 | 105 | 12.81 |
| WCID#111 | 111 | 2.24 ± 0.28 | 20,920 | 107 | 3.35 |
| WCID#47 | 47 | 3.36 ± 2.28 | 33,645 | 100 | 6.27 |
| WCID#76 | 76 | 0.37 ± 0.22 | 976 | 379 | 0.5 |
| West District | WD | 10.06 ± 6.62 | 85,129 | 118 | 17.86 |
| West Lake | WL | 0.20 ± 0.07 | 600 | 333 | 0.53 |
| Westway | WW | 0.40 ± 0.18 | 3,623 | 110 | 0.99 |
| White Oak | WO | 1.84 ± 0.91 | 20,758 | 89 | 3.31 |
| Willowbrook | WB | 1.28 ± 0.52 | 8,610 | 149 | 3.01 |
| **TOTAL** |  | **272.23 ± 159.04** | **2,168,563** | **162 ± 133 (AVG ± STDEV)** | **532** |

\*Sims Bayou South and North have overlapping geographic service areas.

The protocols for wastewater sample processing, concentration, RNA extraction, RT-ddPCR, and sequencing are described as below. In developing and reporting the procedures, we followed the Environmental Microbiology Minimum Information (EMMI) guidelines1.

*Rice Methods.*

Concentration. Concentration was performed in duplicate for each wastewater sample by first aliquoting 50 mL of sample into 50mL conical tubes and centrifuging for 10 minutes at 4,100 g and 4 C to remove solids. Samples were then transferred to a pre-DI-washed 0.45 μM electronegative HA filter (HAWG047S6, MilliporeSigma) followed by 1 mL of 1.25M MgCl2 for a final concentration of 25 mM. Samples were stirred with a sterile pipette tip and allowed to sit for 5 minutes before filtration. For extraction with Maxwell (details describe below), sample filters were placed in a bead tube containing 0.1 mm glass beads and stored at -80 until bead beating and extraction. For extraction with chemagic (details describe below), sample filters were placed in bead tube containing 0.1 mm glass bead and 1 mL of lysis buffer followed by bead beating and extraction.

Extraction. Extraction was initially performed on the Maxwell 48 RSC automated platform (AS8500, Promega) using a modified protocol for the Maxwell RSC PureFood GMO and Authentication Kit (AS1600, Promega). Briefly, 700 µL of CTAB was added to each bead tube containing filters and bead beaten at max speed in a Mini-Beadbeater 24 (3,500 rpm; 112011, BioSpec) for 1 minute. Samples were then transferred to ice for 2 minutes followed by another round of bead beating for 1 minute. After bead beating, 40 µL of Proteinase K was added to each sample tube and briefly vortexed to mix. Sample tubes were transferred to a water bath and incubated at 56 C for 10 minutes before centrifugation for 2 minutes at 17,000 g. Following centrifugation, 350 μL of supernatant was transferred to the first well of a Maxwell cartridge along with 300 μL of Lysis Buffer. The remainder of the nucleic acid extraction was performed on the Maxwell RSC 48 automated platform and eluted into 50 µL of elution buffer. Eluates were stored at -20 C for no more than 24 hours until quantification.

The Rice laboratory switched extraction to the chemagic 360 automated platform using the Viral DNA/RNA 300 Kit H96 (CMG-1433, PerkinElmer), following the manufacturer’s protocol. After bead beating and centrifugation, 300 µL of supernatant was transferred to a deep well plate where each well contained 300 µL of lysis buffer, 4 µL of Poly(A) RNA and 10 µL of Proteinase K. The remainder of the extraction was performed on the chemagic 360 resulting in 50 µL of eluate for each sample.

Quantification. RT-ddPCR was performed on a QX200 AutoDG Droplet Digital PCR System (Bio-Rad) and a C1000 Thermal Cycler (Bio-Rad) in 96-well optical plates. SARS-CoV-2 N1 and N2 gene targets were quantified in wastewater samples using a one-step RT-ddPCR assay. Briefly, a 22 µl reaction mix containing 10 µl of nucleic acid was reverse transcribed using the One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad) following the manufacturer’s protocol. Reaction mix compositions and thermal cycling conditions are detailed in Tables (S2-S5). Each 96-well plate was processed with no-template controls. Droplets were read on a QX200 Droplet Reader (Bio-Rad) and analyzed using QuantaSoft v1.7.4 software. Droplets for SARS-CoV-2 N1/N2 were manually thresholded per channel and data were exported to an Excel file for further analysis using custom R scripts.

A limit of detection (LOD) 3 positive droplets and an acceptable total generated droplet count of at least 10,000 was established for all sample wells as recommended by the manufacturer. A per-plate LOD for N1/N2 was also calculated by assigning the copy number corresponding to wells with 3 droplets as the initial LOD concentration for that plate. If there was more than one sample with 3 droplets, the average concentration for all samples having three droplets was considered. If there were no 3-droplet samples on the plate, a copy number of 0.7 was assumed. This copy number was derived from the number assigned to 3 droplets given 10,000 total droplets. This is the most conservative estimate. The initial LOD concentration was then added to the limit of blank (LOB) to obtain the final LOD concentration for the entire plate. The LOB was determined by adding the average concentration of all negative control samples on the plate to 1.6 times the standard deviation of the negative controls.

**Table S2.** Primers and probes used for quantification of SARS-CoV-2 with RT-ddPCR.

|  |  |  |  |
| --- | --- | --- | --- |
| Target | Primer/Probe | Sequence (5’-3’) | Ref. |
| N1 | Forward Primer  Reverse Primer  Probe | GACCCCAAAATCAGCGAAAT  TCTGGTTACTGCCAGTTGAATCTG  HEX - ACCCCGCATTACGTTTGGTGGACC - BHQ-1 | CDC |
| N2 | Forward Primer  Reverse Primer  Probe | TTACAAACATTGGCCGCAAA  GCGCGACATTCCGAAGAA  FAM - ACAATTTGCCCCCAGCGCTTCAG - Zen, Iowa Black FQ | CDC |

**Table S3.** Final concentrations of N1/N2 Prime-Probe mix

|  |  |  |  |
| --- | --- | --- | --- |
| Target | Reagent | Final conc. (µM) | 20x conc.  (µM) |
| N1 | Probe | 0.25 | 5 |
| Primer F | 0.9 | 18 |
| Primer R | 0.9 | 18 |
| N2 | Probe | 0.25 | 5 |
| Primer F | 0.9 | 18 |
| Primer R | 0.9 | 18 |

**Table S4.** Reaction composition for SARS-CoV-2 N1/N2 RT-ddPCR assay.

|  |  |  |
| --- | --- | --- |
| Component | Volume (µL) | Concentration |
| One-Step RT-ddPCR Supermix | 5.5 | 1x |
| Reverse Transcriptase (10x) | 2.2 |  |
| 300 mM DTT | 1.1 |  |
| N1/N2 Primer-Probe mix\* | 1.1 | 1x |
| RNase/DNase free water | 2.1 |  |
| RNA template | 10 |  |

\*Volume represents 1x of the 20x concentration.

**Table S5.** Thermal cycling conditions for SARS-CoV-2 one-step RT-ddPCR assay.

|  |  |  |  |
| --- | --- | --- | --- |
| Cycling Step | Temperature  °C | Time | Number of Cycles |
| Reverse Transcription | 50 | 60 min |  |
| Enzyme activation | 95 | 10 min |  |
| Denaturation | 95 | 30 sec | 40 |
| Annealing/Extension | 60 | 60 sec |
| Enzyme Deactivation | 98 | 10 min |  |
| Hold (optional) | 4 | Inf |  |

*HHD Methods.*

Concentration & Extraction. Concentration and extraction were performed identically to the Rice method using the chemagic 360 extraction platform with the exception of the volume of lysate supernatant transferred into the RNA extraction plate (600 µL instead of 300 µL) and the final eluate volume (XX µL instead of 50 µL).

Quantification. Quantification of SARS-CoV-2 was performed using the IDEXX Water SARS-COV-2 RT-PCR test kit on the Applied Biosystem QuantStudio Dx. The IDEXX test kit contains the N1 and N2 primer and probe sequences which are described by the U.S. CDC design1. The concentration of the primers and probe is provided by the manufacturer. The SARS-CoV-2 Mix includes primers and probes for the detection of SARS-CoV-2 RNA when amplified with the included RNA Master Mix (RNA MMx). Each reaction is prepared by adding 8µL of SARS-CoV-2 mix, 8µL of RNA Master Mix and 4µL of nucleic acid (Table S6). SARS-CoV-2 RNA targets (N1 and N2) were both detected on the FAM channel. Absolute quantification of SARS-CoV-2 in wastewater was performed using a standard curve produced ATCC- SARS-CoV-2 RNA reference material. In addition, the Water SARS-CoV-2 RT-PCR Test utilizes the Positive Control (PC) and PCR Grade Water (Negative Control). The standard curve was created by diluting the reference material in a series of known concentrations and then measuring a SARS-CoV-2 Ct value by RT-qPCR (FAM channel) for each concentration. The calculated relationship is then used to convert the CT value measured for an unknown wastewater sample into a concentration of RNA expressed as genome copy number per Liter of tested wastewater.

**Table S6.** Reaction composition for RT-qPCR

|  |  |  |  |
| --- | --- | --- | --- |
| Master Mix Reagent | Stock Vol. | Working Concentration | Req. vol/Rx (µL) |
| SARS-CoV-2 mix prime/probe | 1 mL | 1X | 8 |
| RNA Master Mix | 1 mL | 1X | 8 |
| Sample RNA |  |  | 4 |
| Total |  |  | 20 |

**Table S7**. Thermocycling cycling conditions for RT-qPCR

|  |  |  |  |
| --- | --- | --- | --- |
| Cycling Step | Temperature | Time | Cycles |
| RT Step | 50 | 5 min |  |
| Enzyme Activation | 95 | 20 sec |  |
| Denaturation | 95 | 3 sec |  |
| Annealing, Extension | 60 | 30 sec | 40 |

**Table S8.** Acceptable Range for Standard Curve

|  |  |
| --- | --- |
|  | Range |
| Slope | 3.3 (range from 3.1-3.7) |
| Efficiency | 90 - 110 |
| R2 | >0.980 |

*Observations below the LOD*: Measurements below half the LOD were replaced with a random number from a uniform distribution with endpoints of ½ the LOD and the LOD.

*Aggregating data across labs.* Prior to calibrating HHD results with the HHD regression model, the results from each lab are adjusted for difference in reported gene target quantification of a sample. Rice measures and reports N1 copies per liter for a sample separately from N2 copies per liter for the sample, while HHD measures and reports the combined N1 and N2 copies per liter for a sample. As such, the N1 copies per liter and N2 copies per liter for a Rice sample result are added and divided by 2. The HHD sample result is divided by 2. This adjustment was also applied to the data used in calculating the HHD regression model.

*Adjustments for method changes and across labs using regressions.* The measurements from Rice using the Maxwell platform and the measurements from HHD were adjusted to the levels of the measurement from Rice using the chemagic 360 platform. In each case, regression models are carefully considered, and the best model was selected. Also, in each case, non-detects were removed from the data used to build the regression models.

Rice Maxwell Adjustments

For n = 24 samples, we concentrated 4 replicates and extracted 2 replicates using the Maxwell platform and 2 replicates using the chemagic 360 platform for a head-to-head comparison. We performed a linear regression to determine the adjustment to apply to the Maxwell values.

Linear model:

Text, letter

Description automatically generated

Note: Ccpl = Chemagic and Mcpl = Maxwell

HHD to Rice chemagic Adjustments

To aggregate data across labs, we performed a regression analysis using samples collected between April 12, 2021 to August 16, 2021 (n = 387). Samples were split between the Rice and HHD labs, and 2 replicates were processed in each lab. We compared a cubic-polynomial to a linear regression model for the HHD to Rice chemagic adjustment. We chose the cubic-polynomial regression model because it had a higher adjusted R2 as compared to the linear model.

Cubic-polynomial model:

Text

Description automatically generated

*Cross-correlation analysis.* We accounted for autocorrelation when assessing whether the time-step cross-correlations between the wastewater viral load and the other series (positivity rate, ED visit rate, general bed use rate, and ICU bed use rate) were significantly different from no lead/lag2. Briefly, we computed the autocorrelation for each series individually using SAS Enterprise Guide (version 8.1). Next, for each cross-correlation, we took the product of the autocorrelation of the wastewater viral load and the autocorrelation of the other series. We summed this product over 0 to 35 days, and then multiplied it by 2 and subtracted 1 (to account for the autocorrelation at lag 0). We calculated the variance by dividing this value by the number of days in the given wave (this was repeated for each wave). For each cross-correlation and each wave, the standard error was computed by taking the square root of the variance to determine a lower bound 90% and 95% confidence interval. If the Pearson correlation coefficient for a given lag value was greater than the lower confidence interval, it was deemed significant.

**Results**

*Wastewater data.* The number of measurements performed during the study period and across both labs totaled 18,642. Out of those, 927 (5%) were below the LOD and 606 (3.2%) were below half the LOD and imputed with a random number from a uniform distribution with endpoints of ½ the LOD and the LOD. The data is available for download from the Rice University Kinder Institute’s Urban Data Platform3.

*Time-step correlation analysis results.*

In Table S9 below, the half-width or margin-of-error (denoted margin) of the cross-correlation 95% and 90% confidence intervals are shown. The margin-of-error is adjusted for the presence of autocorrelation in each series. The cross-correlations are statistically significantly greater than zero if the lower confidence bound given by (point estimate – margin) is greater than zero. The range of days lagged (from -7 to +21) that were significantly greater than zero are also indicated. For example, in the 1st wave for the wastewater spline and general bed rate, the Pearson correlation coefficients for days lagged between +4 and +21 were above the calculated 95% margin-of-error (0.9334). For the 2nd and 3rd wave, the margin-of-error is above one in all cases indicating the cross-correlations are not statistically significantly different from zero. This finding is due to the small number of measurements that were used to estimate the cross-correlations in the 2nd and 3rd waves.

**Table S9**. Margin-of-error (95% and 90%) for each cross-correlation performed between wastewater and other indicator metric for each wave and over the entire time series.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1st WAVE** | | | | |
|  | 95% margin | Lag Figure Range days > 95% | 90% margin | Lag Figure Range days > 90% |
| WW x Positivity | 0.8908 | -4 to +16 | 0.7327 | -7 to +21 |
| WW x ED Visits | 0.9385 | -5 to +16 | 0.7719 | -7 to +21 |
| WW x General Bed | 0.9334 | +4 to +21 | 0.7677 | -7 to +21 |
| WW x ICU Bed | 0.96218 | +6 to +14 | 0.7914 | -7 to +21 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **September 1, 2020 to April 25, 2022** | | | | |
|  | 95% margin | Lag Figure Range days > 95% | 90% margin | Lag Figure Range days > 90% |
| WW x Positivity | 0.4326 | -7 to +21 | 0.3558 | -7 to +21 |
| WW x ED Visits | 0.4558 | -7 to +21 | 0.3749 | -7 to +21 |
| WW x General Bed | 0.4533 | -7 to +21 | 0.3729 | -7 to +21 |
| WW x ICU Bed | 0.4673 | 0 to +21 | 0.3844 | -3 to +21 |

**References**

(1) Borchardt, M. A.; Boehm, A. B.; Salit, M.; Spencer, S. K.; Wigginton, K. R.; Noble, R. T. The Environmental Microbiology Minimum Information (EMMI) Guidelines: QPCR and DPCR Quality and Reporting for Environmental Microbiology. *Environ. Sci. Technol.* **2021**, *55* (15), 10210–10223. https://doi.org/10.1021/acs.est.1c01767.

(2) Shumway, R. H.; Stoffer, D. S.; Stoffer, D. S. *Time Series Analysis and Its Applications*; Springer, 2000; Vol. 3.

(3) Ensor, K. B.; Schneider, R.; Stadler, L. B.; Hopkins, L. City of Houston SARS-CoV-2 Wastewater Monitoring Data (Version 1) [Data Set]. *Rice Univ.-Kinder Inst. UDP* **2022**. https://doi.org/10.25612/837.K4JMWR0ZPZOK.