Changes in Sewage Sludge Chemical Signatures During a COVID-19 Community Lockdown, Part 2: Nontargeted Analysis of Sludge and Evaluation with COVID-19 Metrics

Sara L. Nason,a,* Elizabeth Lin,b Krystal J. Godri Pollitt,b and Jordan Pecciac

aConnecticut Agricultural Experiment Station, New Haven, Connecticut, USA
bDepartment of Environmental Health, Yale School of Public Health, New Haven, Connecticut, USA
cDepartment of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut, USA

Abstract: Sewage sludge and wastewater include urine and feces from an entire community, and it is highly likely that this mixture contains chemicals whose presence is dependent on levels of SARS-CoV-2 in the community. We analyzed primary sewage sludge samples collected in New Haven, Connecticut, USA, during the initial wave of the COVID-19 pandemic using liquid chromatography coupled with high-resolution mass spectrometry and performed an exploratory investigation of correlations between chemical features and COVID-19 metrics including concentrations of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) RNA in the sludge and local COVID-19 case numbers and hospital admissions. Inclusion of all chemical features in this analysis is key for discovering potential indicator compounds for COVID-19, whose structures may not be known. We found correlations with COVID-19 metrics for several identified chemicals as well as many unidentified features in the data, including three potential indicator molecules that are recommended for prioritization in future studies on COVID-19 in wastewater and sludge. These features have molecular weights of 108.0935, 318.1214, and 331.1374. While it is not possible to achieve prediction of COVID-19 epidemiological metrics from the one data set used in the present study, advances in this research area are important to share as scientists worldwide work on discovering efficient methods for tracking SARS-CoV-2 in wastewater and the environment. Environ Toxicol Chem 2022;41:1193–1201. © 2021 SETAC

Keywords: Wastewater; COVID-19; High-resolution mass spectrometry; Nontargeted analysis

INTRODUCTION

Approximately 34 billion gallons of raw wastewater are treated daily in US treatment facilities (US Environmental Protection Agency, 2019). Untreated wastewater includes urine and feces from entire communities as well as all of the chemicals that are disposed of down the drain. Sewage sludge and wastewater are high-dimensional sample matrices which contain valuable community-scale epidemiological information that can be used as population health indicators. In past wastewater-based epidemiology studies, various compounds/substances have been assessed for monitoring the usage of illicit drugs (i.e., cocaine, heroin, methamphetamine; Khan & Nicell, 2011; Zuccato et al., 2008) and licit drugs (i.e., caffeine, nicotine, alcohol; Senta et al., 2015), pharmaceuticals (e.g., benzodiazepines, opioids, antidepressants, antibiotics, and antimicrobials; Castiglioni et al., 2006; Langford & Thomas, 2011), and consumer product ingredients (e.g., ultraviolet filters, synthetic musks, pesticides, plasticizers, and flame retardants; Snyder et al., 2003). In addition, human pathogens, stress, and diet markers have been investigated in wastewater/sludge to assist with human biomonitoring and assessing, monitoring, and managing population health (Chen et al., 2014; Choi et al., 2018; Senta et al., 2015, 2020; Yang et al., 2015).

Recently, wastewater and sludge have been used extensively for surveillance of COVID-19 outbreaks, where concentrations of RNA from the novel severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) correlate with COVID-19 cases and hospital admissions (Mandal et al., 2020; Medema et al., 2020; Peccia et al., 2020; Wurtzer et al., 2020). Wastewater and sludge samples used in these (and other) surveillance studies have also been leveraged for evaluating the scale and the direct and
indirect impacts of outbreaks and associated lockdowns by examining changes in drug consumption, chemical usage, and chemical biomarkers (Choi et al., 2018; Lu et al., 2020; Reinstadler et al., 2021; Senta et al., 2020; Varghese et al., 2021; Z. Z. Wang et al., 2021). In Part 1 of this series, we used both targeted and suspect screening approaches with liquid chromatography—high resolution mass spectrometry (LC-HRMS) to identify chemicals of interest in primary sewage sludge samples collected over the first 15 weeks of the COVID-19 pandemic in New Haven, Connecticut, USA, and investigated trends over time during the initial lockdown and reopening phases (Nason et al., 2022). However, this initial study used only a fraction of the data collected with LC-HRMS analysis and did not evaluate the relationship of detected chemicals in sludge with levels of SARS-CoV-2 RNA that were measured in the same environmental samples or other COVID-19 metrics (Peccia et al., 2020).

In Part 2 we present an exploratory analysis featuring results from nontargeted screening of these same primary sewage sludge samples and evaluation of the relationships among detected chemical features and parallel measurements of COVID-19 metrics including levels of SARS-CoV-2 RNA in the sludge, COVID-19 case numbers, and hospital admissions. Sewage sludge includes urine and feces from the entire community, and we hypothesize that this mixture contains biologically produced chemicals whose presence is dependent on levels of SARS-CoV-2 in the community. For infected patients, COVID-19 causes changes in the metabolome (Blasco et al., 2020), and therefore is likely to cause changes in the community-wide exposome as well. Previous wastewater epidemiology studies have identified biomarkers for oxidative stress and inflammation that can be used for analysis of community health, and it is likely that markers for specific diseases are present as well (Blasco et al., 2020; Choi et al., 2018; Sims & Kasprzyk-Hordern, 2020; Vitale et al., 2021). Inclusion of all data features in this analysis allows for discovery of potential biomarker compounds for COVID-19, whose structures may not be known. While it is not possible to achieve prediction of COVID-19 epidemiological metrics from one data set, the present study works toward that broader goal by presenting findings based on samples that were used directly for both biological and chemical analyses. Our data and statistical methods will be of use to other researchers who are conducting similar research investigating sludge as an environmental matrix for evaluation with health-related parameters.

**METHODS**

**Sample collection and preparation and instrumental analysis**

All sample collection and instrumental analysis methods have been described previously (Nason et al., 2022; Peccia et al., 2020), and all method details and quality control information are provided in the Supporting Information. Briefly, primary sewage sludge samples were collected at the East Shore Water Pollution Abatement Facility, New Haven, Connecticut, USA, daily from March 19 to June 30, 2020. This wastewater management facility treats sewage from four nearby communities, including New Haven, Hamden, East Haven, and Woodbridge, which are primarily residential (93.7% of plant customers) but also have significant land zoned for commercial and industrial use (6.5 and 0.2% of plant customers, respectively). Part of the catchment area contains combined sewers, and travel time from wastewater sources to the treatment plant is less than 4 h for all areas served. The study time period encompassed the first wave of COVID-19 infections in New Haven County, the initial statewide stay-at-home order, Phase 1 reopening, and the beginning of Phase 2 reopening. Two types of samples were analyzed over different time periods (Figure 1): daily sludge samples were analyzed for 4 weeks from March 19 through April 15 and weekly composite sludge samples were analyzed for 15 weeks from March 19 through June 30. To form weekly composites, sludge solids and liquids from seven samples were pooled separately after collection and prior to extraction. Solid and liquid portions (all sample types) were separated via centrifugation. The solid fraction was homogenized prior to subsampling. The solid portion was extracted with acetonitrile, then equal portions of the extract and the liquid portion were combined and

![Figure 1](https://example.com/figure1.png)

**Figure 1:** Time series showing the number of new COVID-19 cases in the study area by date of first positive specimen collection (date format month/day/year). The gray line shows daily numbers, and the solid line shows a 7-day rolling average. The red (solid) and blue (dashed) lines below the plot show the daily and weekly sample analysis time periods, respectively.
filtered. After extraction, three 5-week composite samples were generated using the weekly samples. Daily and weekly samples were extracted and analyzed in separate batches.

Samples were analyzed using an Ultimate 3000 liquid chromatograph equipped with a Hypersil Gold C-18 column and coupled with a Q-Exactive mass spectrometer (Thermo Scientific) with positive electrospray ionization. Daily and weekly samples were analyzed using a full MS and all ion fragmentation (AIF) instrument method. Five-week composite samples were included in the weekly sample full MS-AIF instrument run and analyzed separately using a data-dependent MS2 (ddMS2) instrument method. We used an iterative inclusion method to ensure collection of as many ddMS2 spectra as possible (similar advantages to other iterative methods; Koelmel et al., 2017, 2020; Nason et al., 2022). The 5-week composite samples were used for compound identification analysis and overall sample characterization (i.e., in principal component analysis [PCA]) but were not used in the main quantitative analysis. The ddMS2 spectra were used in subsequent data analysis for filtering peaks and providing structural annotations.

**Compound identification and data list formation**

Data processing methods for compound identification are fully described in Part 1 of this series and in the Supporting Information. The full list of identified compounds is included in Part 1 of this series but is also used in the present analysis, so the methods are described briefly. We used both TraceFinder 4.1 and Compound Discoverer 3.1 software (Thermo Scientific) to identify and (semi-)quantify compounds in the samples. The TraceFinder analysis method used full MS and AIF scan data to identify compounds based on an internal database of compounds that includes many pesticides, veterinary drugs, toxins/ poisons, as well as other chemicals. The Compound Discoverer method used a peak picking algorithm that was independent of any databases or libraries and then used ddMS2 data for analyzed peaks to identify compounds using spectral match scores with compounds in the mzCloud database. Both methods used a 5 ppm mass window for mass-to-charge ratio values, and each identified compound was assigned a confidence level that corresponds to the software used and the certainty of the match (1 = confirmed with standard, 2 = probable identification, 3 = tentative identification; Nason et al., 2022; Schymanski et al., 2014). For further data analysis, we formed a list of all identified compounds and their concentration (targeted compounds) or peak area (suspect screening results) in each of the analyzed samples (Supporting Information, Tables S8 and S9). This list is subsequently referred to as the “targeted list” and overlaps with the data included in Part 1.

We used the full peak list from Compound Discoverer, including both identified and unidentified compounds, for further analyses. Background features (ratio of average sample peak area to blank peak area of <5) were removed, and only peaks with ddMS2 spectra were included. We filtered the list by presence of ddMS2 spectra because this eliminated many of the poorly integrated, low-intensity peaks in the data set and so that we would have evidence toward determining structures for interesting features. Because of the iterative inclusion data collection strategy, we collected ddMS2 spectra for the majority of features in the data set (6204 features with ddMS2 spectra). All analyses using this list was based on peak area data, which were reported for all daily and weekly composite samples. This list is provided in Supporting Information, Tables S10 and S11, and is referred to as the “nontargeted list.”

The SARS-CoV-2 RNA concentrations in our primary sludge samples have previously been published, and a methods summary is provided in the Supporting Information (Peccia et al., 2020). The number of reported COVID-19 positive tests (cases) and hospital admission data were originally obtained from the Connecticut Department of Public Health and the Yale New Haven Hospital. The case number data were aligned with date of specimen collection and represent Connecticut’s reported count for the cities/towns served by the wastewater-treatment plant from which sludge samples were collected: New Haven, Hamden, Woodbridge, and East Haven. The hospital admission numbers represent the daily number of residents from New Haven, Hamden, Woodbridge, and East Haven who were diagnosed with COVID-19 and admitted to Yale New Haven Hospital with COVID-19. All data were provided on a per-day basis.

**Structural determination**

Compiled structural information for unidentified features of interest is provided based on predicted composition and ChemSpider search data from Compound Discoverer. In addition, we used the fragment ion search (FISh) scoring tool in Compound Discoverer to calculate MS2 spectral match scores for proposed compound identifications based on the ChemSpider database. The FISh scoring tool creates simulated MS2 spectra for known chemical structures and compares the fragment ion masses with those found in sample data. We used a mass tolerance of 5 ppm and a signal-to-noise ratio of 3. Information on the FISh algorithm is provided in the Supporting Information.

**Statistical methods**

The PCA was completed in Compound Discoverer using all detected features. Because of the broad range of variability in our data, we mean-centered the data and scaled it by dividing by standard deviation. Points were exported and plotted using Excel 365 (Microsoft). Based on PCA results, daily and weekly sample data were analyzed separately for all other tests.

Both daily and weekly data (targeted and nontargeted lists) showed a nonnormal distribution. Spearman’s rank-order correlation analysis (α = 0.05) was performed to explore the relationship between feature peak areas/concentrations and SARS-CoV-2 RNA copies, reported COVID-19 cases, and
hospital admissions. Spearman’s rank-order correlation analysis assesses monotonic relationships between variables, which may or may not be linear and can be used with nonnormal data. It has previously been used to investigate trends between chemical levels and COVID-19 metrics (S. Wang et al., 2020). For correlation analysis using daily chemical results (4 weeks, March 19–April 15), daily SARS-CoV-2 RNA copies, number of hospital admissions, and number of COVID-19 cases were used. For correlations using weekly chemical results (15 weeks, March 19–June 30), 7-day average SARS-CoV-2 RNA concentrations were used. All RNA results corresponding to the weeks 1 to 11, because of missing hospital administration information from weeks 12 to 15. Spearman’s rank-order correlation analysis was conducted in RStudio 4.0.3 (R Foundation for Statistical Computing, 2020). All data used in the correlations are provided in Supporting Information, Tables S8–S11.

RESULTS

PCA

The PCA was conducted for the purpose of determining whether daily and weekly samples could be analyzed together as one data set and to determine if there were changes in overall sample composition over time. When daily and weekly samples were included in the same PCA, they clustered separately (Supporting Information, Figure S2A). The standard and quality control samples analyzed during each run indicate that instrument sensitivity was higher during the daily sample LC-HRMS run, though quantified values were similar in the rerun samples; all subsequent analyses were performed separately on weekly and daily sample data. We did not see clear clusters or evidence of overall change over time in the daily sample only PCA (Supporting Information, Figure S2B); we saw progressive change over 15 weeks of weekly sample analysis (Supporting Information, Figure S2C), indicating that we are likely to see changes over time for many detected chemical features.

Targeted chemical correlations with COVID-19 metrics

Tables 1 and 2 show all statistically significant correlations (Spearman’s analysis, p < 0.05) for daily and weekly samples, among compound levels identified through targeted analysis and suspect screening and SARS-CoV-2 RNA concentrations, number of COVID-19 case numbers, and number of COVID-19 hospitalizations. All Spearman’s correlation results are provided in Supporting Information, Table S12. Overall, our findings relating sludge chemicals and SARS-CoV-2 RNA levels are similar to the trends found for changes in chemical levels in sludge samples over time that are presented in Part 1 (Nason et al., 2022). This makes sense because, for the daily sampling analysis (March 19–April 15), SARS-CoV-2 RNA levels increased nearly linearly, as did several of the identified chemicals. Some of these increases are likely related to the pandemic (e.g., triclocarban, a disinfectant, and sertraline, an antidepressant), whereas others are expected seasonal changes (e.g., imazalil, a pesticide, and diphenhydramine, an allergy medication). Two of the negatively correlated compounds, levorphanol and edaravone, are drugs which are not related to COVID-19 treatment and are primarily administered in clinical settings. Decreased levels may be due to clinic closures, canceled elective procedures, and many patients staying home. All correlation coefficients for daily assessment of sludge chemicals and COVID-19 metrics were <0.6.

TABLE 1: Daily sample chemical correlations with COVID-19 metrics

| Compound          | Confidence level | RNA copies | COVID-19 cases | No. hospitalizations |
|-------------------|------------------|------------|----------------|---------------------|
| Benzotriazole     | CD-1             | −0.52      | −0.44          | −0.37               |
| Betanechol        | TF-3             | −0.41      | −0.56          | −0.46               |
| Codeine           | TF-1a            | —          | —              | −0.42               |
| Caffeine          | TF-2a            | —          | −0.55          | −0.39               |
| Diphenhydramine   | CD-1             | 0.42       | —              | —                   |
| Edaravone         | CD-2             | −0.56      | —              | —                   |
| Gabapentin        | CD-2a            | —          | −0.37          | —                   |
| Imazalil          | TF-2             | 0.47       | —              | —                   |
| Ipronidazole      | TF-3             | −0.50      | —              | −0.38               |
| Levorphanol       | TF-1             | −0.58      | −0.43          | −0.46               |
| Metformin         | TF-3             | —          | −0.37          | —                   |
| Oxybenzone        | CD-1             | —          | −0.57          | —                   |
| Piperonyl butoxide| TF-2             | −0.41      | −0.44          | —                   |
| Sertraline        | CD-1             | 0.45       | —              | —                   |
| THCA              | TF-3             | 0.36       | —              | —                   |
| Triclocarban      | CD-1             | 0.47       | 0.38           | —                   |

Levels 1–3 represent confirmed, probable, and tentative identifications, respectively.

*Only statistically significant values (p < 0.05) are included. Dashes represent nonsignificant values.
CD = compound discoverer; TF = TraceFinder; THCA = tetrahydrocannabinolic acid.
In both daily and weekly results, chemical levels had more statistically significant correlations with RNA copies than with case numbers or hospitalizations, likely because the chemicals and RNA were both measured in the same sludge samples. Although testing during the early pandemic was conducted at very low rates, COVID-19 cases were correlated with more chemicals measured in sludge samples compared to COVID-19-related hospitalizations (Peccia et al., 2020). This observation likely resulted from the lag measurements of positive case testing and reporting and hospital admissions (Peccia et al., 2020); in the study by Peccia et al., SARS-CoV-2 viral loads were 0–2 days ahead of SARS-CoV-2-positive test results by date of specimen collection and 1–4 days ahead of local hospital admissions, using the same wastewater samples.

### Nontargeted feature correlations with COVID-19 metrics

We found correlations (Spearman’s rank-order correlation, \( p < 0.05 \)) with SARS-CoV-2 RNA levels for 1398 features in the daily sample data (23% of all features) and 1751 features in the weekly sample data (28% of all features; Supporting Information, Table S12). We hypothesize that this high degree of correlation is due to the presence of factors (e.g., surface runoff, storm water) that have large effects on the overall composition of the sludge both biologically and chemically. Our results also may contain false positives because of the high number of multiple comparisons in the analysis. The purpose of analyzing the broader list of peaks was to find chemical features that correlate with the virus concentrations and may be used as a means for exposure assessment in epidemiological studies evaluating health outcomes. Therefore, the bulk of our analysis focused only on strong positive correlations with SARS-CoV-2 RNA levels. Table 3 shows the results from weekly sample data and SARS-CoV-2 RNA levels with Spearman’s correlation coefficients \( \geq 0.9 \). Relative signal intensities are shown in Figure 2. Results are manually curated, and features with poor quantitation due to overlapping peaks, high amounts of noise, or low signal intensity have been removed.

In addition, we investigated correlations between the weekly sample nontargeted peak areas and COVID-19 cases and hospitalizations. Similar to the SARS-CoV-2 RNA correlations, we focused on positive correlations with Spearman’s \( \rho \geq 0.9 \) that were manually curated for integration quality. We found 20 features that strongly correlated with the number of COVID-19 cases and two features that strongly correlated with the hospitalization data (Supporting Information, Tables S13 and S14). The three lists of strongly correlated features do not contain any overlap, though there is substantial overlap in the broader lists of statistically significant correlations. Two features from Table 3 had a statistically significant correlation (\( p < 0.05 \)) with both number of COVID-19 cases and hospitalizations, and an additional two had a statistically significant correlation only with case numbers (Table 3). We propose that the features that correlate with all three virus-related metrics be prioritized for further investigation at other locations and time points. The daily sample nontargeted features did not have any positive

### Table 2: Weekly sample chemical correlations with COVID-19 metrics

| Compound          | Confidence level | RNA copies | COVID-19 cases | No. hospitalizations |
|-------------------|------------------|------------|----------------|---------------------|
| Anhydroycegonine  | CD-2a            | 0.58       | 0.63           | —                   |
| Avobenzene        | CD-2             | -0.79      | —              | —                   |
| Azithromycin      | TF-1             | 0.69       | 0.66           | —                   |
| Benzoctazolide    | CD-1             | -0.80      | —              | -0.71               |
| Benzylocegonine   | TF-2a            | -0.66      | —              | —                   |
| Berberine         | TF-1             | —          | —              | 0.61                |
| Caffeine          | TF-2a            | -0.86      | —              | -0.65               |
| Cinchophenol      | CD-2             | -0.74      | —              | —                   |
| Ecoregonine methyl| TF-2a            | -0.78      | —              | —                   |
| Eserine           | CD-2             | 0.57       | —              | —                   |
| Gabapentin        | CD-2a            | -0.66      | —              | —                   |
| Hydromorphone     | TF-1a            | -0.75      | —              | —                   |
| Imazalil          | TF-2             | 0.65       | 0.73           | —                   |
| Levorphanol       | TF-1             | 0.73       | 0.70           | —                   |
| Losartan          | TF-1             | 0.64       | 0.64           | —                   |
| Methadone         | TF-1             | -0.80      | —              | —                   |
| Nithazine         | TF-3             | 0.80       | 0.76           | —                   |
| Octocrylene       | CD-2             | -0.76      | —              | —                   |
| Oxybenzone        | CD-1             | -0.88      | —              | —                   |
| Piperonyl butoxide| TF-2             | —          | —              | -0.71               |
| Pramocaine        | CD-2             | -0.63      | —              | —                   |
| TFMPP             | TF-3             | 0.66       | 0.67           | —                   |
| Tolycaene         | CD-2             | 0.72       | 0.69           | —                   |
| Zalcitabine       | TF-3             | 0.55       | 0.55           | —                   |

Levels 1–3 represent confirmed, probable, and tentative identifications, respectively.

*Only statistically significant values (\( p < 0.05 \)) are included. Dashes represent nonsignificant values.

CD = compound discoverer; TF = TraceFinder; TFMPP = trifluoromethylphenylpiperazine.

Correlation coefficients for the weekly sample (March 19–June 30) RNA correlations were higher but also strongly related to the trends over time. The SARS-CoV-2 RNA levels increased rapidly over the first 4 weeks of the study period, then decreased for the remaining 11 weeks. Therefore, the correlation results include chemicals that decreased over the duration of the study period (or increased for negative correlations). None of the positively correlated compounds showed the characteristic hump during the first 4 weeks. Several of the negatively correlated compounds (benzotriazole, caffeine, and oxybenzone) decreased in the initial weeks before increasing for the remainder of the sampling period (Nason et al., 2022). These trends are likely due to pandemic-related lifestyle changes in driving, sleeping, and cosmetic use. Berberine concentration in wastewater was positively correlated only with number of hospitalizations (Spearman’s \( \rho = 0.61 \); Table 2; Warowicka et al., 2020). Berberine is a substance that shows a broad range of biochemical and pharmacological activities (i.e., anticancer, anti-inflammatory, antiviral, and antidepressant) and has been proposed for COVID treatment in multiple studies (Varghese et al., 2021; Z. Z. Wang et al., 2021; Warowicka et al., 2020). We did not find any significant correlation for hydroxychloroquine, a highly publicized potential treatment for COVID-19.
TABLE 3: Weekly sample nontargeted feature correlations with COVID-19 metrics

| Molecular weight | RT (min) | No. adducts | SARS-CoV2 RNA copies | COVID-19 cases | No. hospitalizations |
|------------------|----------|-------------|----------------------|----------------|---------------------|
| 993.7166         | 44.53    | 5           | 0.94                 | —              | —                   |
| 281.0758         | 4.14     | 1           | 0.93                 | 0.57           | —                   |
| 126.1157         | 2.52     | 2           | 0.93                 | —              | —                   |
| 139.0993         | 2.84     | 1           | 0.92                 | 0.53           | —                   |
| 331.1374         | 3.13     | 2           | 0.92                 | —              | —                   |
| 400.9319         | 2.45     | 2           | 0.91                 | —              | —                   |
| 108.0935         | 38.34    | 3           | 0.90                 | 0.78           | 0.70                |
| 318.1214         | 12.57    | 3           | 0.90                 | 0.76           | 0.63                |
| 409.9154         | 2.44     | 2           | 0.90                 | —              | —                   |
| 334.1440         | 40.23    | 1           | 0.90                 | —              | —                   |

*Only features with Spearman’s ρ ≥ 0.9 for RNA copies are listed. All statistically significant Spearman’s ρ (p < 0.05) values are provided for COVID-19 cases and hospitalizations for listed features. All are positive. Dashes represent nonsignificant values.

RT = retention time; SARS-CoV2 = severe acute respiratory syndrome–coronavirus 2.

The top 10 nontargeted features that correlate with levels of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) RNA in primary sewage sludge samples (March 19–June 30, 2020, weekly composite samples). The RNA levels are shown as solid black triangles (right y-axis). Relative peak areas for nontargeted features are shown as colored open shapes (left y-axis). Relative peak areas for each feature were calculated by dividing all peak areas for each feature by the week 1 peak area. Feature 331 (yellow circles) most closely matches the peaks in SARS-CoV-2 RNA in the first half of the sampling period.

Structural determinations

Determination of chemical structures for nontargeted compounds was not a primary goal of the present study. However, we selected the three most likely biomarker candidates from our data and compiled the information available from feature characterization in Compound Discoverer. The features with molecular weights of 108.0935 and 318.1214 (features 108 and 318, respectively) were found to have Spearman’s correlation coefficients >0.9 with the SARS-CoV-2 RNA levels as well as statistically significant correlations with number of COVID-19 cases and COVID-19 hospitalizations. The predicted composition for feature 108 was C6H12, and the ChemSpider search identified multiple potential isomers which we are not able to differentiate between. The FISH match score for each of the first five options was 18.18%, with two of nine fragments matched: the parent mass ([M + H]+) and 67.0550 (C6H7 [M-e]+). This relatively low score indicates that all of the listed options are unlikely to be a true match for the compound identification. The retention time for this feature was quite late, so a hydrocarbon structure that is relatively hydrophobic makes sense. It is also possible that this feature represents an in-source fragment of a larger parent compound that is present in the samples. The MS2 spectra are provided in Supporting Information, Figure S4. Feature 318 has a predicted composition of C16H10N2O5. We calculated FISH scores for the top 20 ChemSpider match results (all had the same molecular formula). Indolebutyroyl aspartic acid (Chemical Abstracts Service [CAS] no. 101289-65-0) and N-[(1,3-dioxo-1,3-dihydro-2H-isidindol-2-yl)acetyl]leucine (CAS no. 6707-71-7) were the top matches, with scores of 75% with nine of 12 fragment ions matched. Both of these structures are potential identifications for these features, though other related structures are also possible. There is little information available about the sources and uses of these chemicals, so their potential relationship with SARS-CoV-2 is unclear. Annotated MS2 spectra are provided in Supporting Information, Figure S5.

While it does not have the highest correlation coefficient, the feature with molecular weight 331.1374 (feature 331) best matches the peaks in the SARS-CoV-2 RNA levels in primary sludge (Figure 2). In addition, it has the strongest linear correlation with SARS-CoV-2 concentrations (Supporting Information, Figure S3). Feature 331 has a predicted formula of C19H19F2NO2. The top ChemSpider database result, flazalone, has a FISH score of 26.67% with four of 15 fragment ions detected. The annotated MS2 spectra and flazalone structure are shown in Supporting Information, Figure S6. All the matched fragments contain the tertiary amine group present in the flazalone structure, which may indicate that a similar amine is present in the true structure. It is highly unlikely that flazalone is the true structural match for feature 331, both because it has a low FISH score and because it is quite hydrophobic, whereas feature 331 had a very early retention time on our reverse-phase LC column.
DISCUSSION

Although other groups have also analyzed wastewater samples that were collected during the pandemic era using LC-MS, our data show unique trends. We provide information on individual chemicals in primary sewage sludge samples that correlate with concentrations of SARS-CoV-2 RNA and other COVID-19 metrics and identify multiple molecular features as potential COVID-19 biomarkers. Previously, a study conducted by S. Wang et al. in upstate New York, USA, explored the relationship between population substance use patterns, COVID-19 positivity, and SARS-CoV-2 RNA detection frequency and found that consumption rates of six substance groups—antidepressants, antiepileptics, antihistamines, antihypertensives, synthetic opioids, and central nervous system stimulants—were positively correlated with SARS-CoV-2 RNA detection frequency and COVID-19 test positivity (S. Wang et al., 2020). While our analysis focused on individual chemicals, we identified several of the same compounds and can compare the trends. We found positive correlations for serotonin and diphenhydramine with SARS-CoV-2 RNA levels in daily sludge samples. This finding corresponds with the results reported by Wang et al. on antidepressants and antihistamines. For synthetic opioids, we found a positive correlation between levorphanol and SARS-CoV-2 levels in weekly samples but negative correlations for methadone in weekly samples and levorphanol in daily samples. Our findings differed from those of S. Wang et al. for antiepileptic drugs; we found a negative correlation between gabapentin and COVID-19 metrics in both daily and weekly samples and no trends for the other antiepileptics we identified. Similarly conflicting, we found negative correlations with SARS-CoV-2 levels for two out of three cocaine metabolites in weekly samples. We found no correlations for antihypertensive drugs. In addition, S. Wang et al. employed nontargeted analysis and tentatively identified piperine as a covariate with SARS-CoV-2 RNA copies in wastewater samples collected in central New York treatment plants; though we also tentatively identified piperine through nontargeted screening in the present study, the compound was not correlated with SARS-CoV-2 RNA copies, positive cases, or hospital admissions. Despite the relative geographic proximity and study design similarity between our analysis and that of S. Wang et al., our results are substantially different. This may be due to differences in sample collection, preparation, and analysis methods or to regional differences in pandemic behavior changes. Overall, it appears that the classes of compounds identified in both studies are not necessarily related to SARS-CoV-2 RNA levels in sludge and wastewater. Closer analysis of nontargeted data features, for which chemical identifications are not available, across data sets may enable additional comparisons to be evaluated.

In another wastewater study conducted during the pandemic, Bowers and Subedi (2021) identified and validated four biomarkers (8-iso-prostaglandin F2α, 2,3-dinor-iso-prostaglandin F2α−III, prostaglandin E2, and 5-iso-prostaglandin F2α−VI) for oxidative stress monitoring in wastewater using LC-MS. None of the identified oxidative stress biomarkers were detected in our study results, which is unsurprising because of differences in sample preparation methods. Moreover, alcohol indicator compounds were not found in our data. Previously, ethyl sulfate was used as a biomarker for alcohol use during the pandemic (Bade et al., 2021), but we did not have a feature at this mass. In addition, SARS-CoV-2 RNA levels and antibiotic resistance have been observed to coincide in India, implying increased use of antimicrobial drugs with increased COVID-19 (Kumar et al., 2021). We did not observe this trend, except for a correlation between the disinfectant triclocarban and SARS-CoV-2 RNA levels in the daily sample data.

More broadly, wastewater analysis during past influenza outbreaks has shown a strong correlation between disease cases and levels of antiviral drugs (Azuma et al., 2012, 2013, 2015). We expected to see more correlation between potential COVID-19 treatments and COVID-19 metrics; several antiviral drugs were included in our quantitative, targeted analysis for this purpose. However, hydroxychloroquine was the only antiviral detected in our samples, and its levels appear to be driven more by media announcements than COVID-19 levels (Nason et al., 2022). We hypothesize that analysis of samples from later in the pandemic, when treatments for severe COVID-19 were more developed, would show more correlation between SARS-CoV-2 RNA levels and chemicals used in the clinical treatment of COVID-19-positive patients.

CONCLUSIONS

The present study applied targeted, suspect, and nontargeted screening methods using LC-HRMS, to explore analyte features in primary sewages sludge samples that were strongly correlated with SARS-CoV-2 RNA copies, COVID-19 cases, and hospital admissions. Unlike previous studies on chemicals in wastewater-related matrices during the COVID-19 pandemic, our analysis includes both identified and unidentified chemical features. We found multiple features that are potential COVID-19 biomarkers and have not previously been studied in relation to wastewater and SARS-CoV-2; structural annotation information was provided for the top three. All three showed strong correlations with SARS-CoV-2 RNA levels, and two correlated with additional COVID-19 metrics (SARS-CoV-2 RNA copies, COVID-19 cases, and hospital admissions). For future screening in wastewater and sludge samples, these three chemical features should be considered as top analyte candidates for surveillance. In addition, we provide information on pharmaceuticals and other anthropogenic compounds that correlate with COVID-19 metrics. Major limitations in our study include that it only examines one wastewater-treatment plant and relies on peak area data for quantitative analysis. Despite these drawbacks, our data provide valuable information that helps to move the field forward as the scientific community comes together to assess the direct and indirect effects of the pandemic on communities.

Supporting Information—The Supporting Information are available on the Wiley Online Library at https://doi.org/10.1002/etc.5226. The Supporting Information includes detailed
analytical methods information, all data used for Spearman’s correlation analysis, linear correlation analysis for nontargeted features, detailed principal component analysis results, tables of nontargeted features that correlate with COVID-19 cases and hospital admissions, and MS2 spectra for structural determinations.

Acknowledgments—We thank the Greater New Haven Water Pollution Control Authority for providing access to samples and assisting with daily sampling, D. Brackney for initiating contact between S. L. Nason and J. Peccia, and J. Koelmel for helpful input. The present study was supported by US Department of Agriculture’s National Institute of Food and Agriculture Hatch funds (CONH00789).

Conflict of Interests—The authors declare that there are no conflict of interests.

Author Contributions Statement—S. L. Nason led sample analysis, data analysis in Compound Discoverer, and manuscript writing. E. Lin led data analysis in R, performed all Spearman’s correlations, and assisted with sample analysis, other data analysis, and manuscript writing. K. J. Godri Pollitt assisted with statistical analyses. J. Peccia led sample collection and provided access to the samples. All contributed to manuscript edits and revisions. S. L. Nason and E. Lin contributed equally to the manuscript.

Data Availability Statement—The data lists used for analysis are provided in the Supporting Information. The RAW instrumental data files used in the present study are available as a data set on MassIVE (ftp://MSV000086676@massive.ucsd.edu) along with the full peak list produced in our Compound Discoverer analysis. Data, associated metadata, and calculation tools are also available from the corresponding author (sara.nason@ct.gov).

REFERENCES
Azuma, T., Nakada, N., Yamashita, N., & Tanaka, H. (2012). Synchronous dynamics of observed and predicted values of anti-influenza drugs in environmental waters during a seasonal influenza outbreak. Environmental Science & Technology, 46(23), 12873–12881.
Azuma, T., Nakada, N., Yamashita, N., & Tanaka, H. (2013). Mass balance of anti-influenza drugs discharged into the Yodo River system, Japan, under an influenza outbreak. Chemosphere, 93(9), 1672–1677.
Azuma, T., Nakada, N., Yamashita, N., & Tanaka, H. (2015). Prediction, risk and control of anti-influenza drugs in the Yodo River basin, Japan during seasonal and pandemic influenza using the transmission model for infectious disease. Science of the Total Environment, 521–522, 68–74.
Bade, R., Simpson, B. S., Ghetia, M., Nguyen, L., White, J. M., & Gerber, C. (2021). Changes in alcohol consumption associated with social distancing and self-isolation policies triggered by COVID-19 in South Australia: A wastewater analysis study. Addiction, 116(6), 1600–1605.
Blasco, H., Bessy, C., Plantier, L., Lefevre, A., Piver, E., Bernard, L., Marlet, J., Stefic, K., Benz-de Bretagne, I., Cannet, P., Lumbi, H., Morel, T., Bouard, P., Andres, C. R., Vourch, P., Herault, O., Guillou, A., & Emond, P. (2020). The specific metabolite profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytokine metabolism. Scientific Reports, 10(1), Article 16824.
Bowers, I., & Subedi, B. (2021). Isoprostanes in wastewater as biomarkers of oxidative stress during COVID-19 pandemic. Chemosphere, 271, Article 129489.
Castiglioni, S., Bagni, R., Fanelli, R., Pomati, F., Calamari, D., & Zuccato, E. (2006). Removal of pharmaceuticals in sewage treatment plants in Italy. Environmental Science & Technology, 40(1), 357–363.
Chen, C., Kostakis, C., Gerber, J. P., Tschirke, B. J., Irvine, R. J., & White, J. M. (2014). Towards finding a population biomarker for wastewater epidemiology studies. Science of the Total Environment, 487, 621–628.
Choi, P. M., Tschirke, B. J., Donner, E., O’Brien, J. W., Grant, S. C., Kaserzon, S. L., Mackie, R., O’Malley, E., Crosbie, N. D., Thomas, K. V., & Mueller, J. F. (2018). Wastewater-based epidemiology biomarkers: Past, present and future. Trends in Analytical Chemistry, 105, 453–469.
Khan, U., & Nicell, J. A. (2019). Refined sewer epidemiology mass balances and their application to heroin, cocaine and ecstasy. Environmental International, 37(7), 1236–1252.
Koelmel, J. P., Kroeger, N. M., Gill, E. L., Ulmer, C. Z., Bowden, J. A., Patterson, R. E., Yost, R. A., & Garrett, T. J. (2017). Expanding lipidomic coverage using LC-MS/MS data-dependent acquisition with automated list exclusion generation. Journal of the American Society for Mass Spectrometry, 28(5), 908–917.
Koelmel, J. P., Li, X., Stow, S. M., Sartain, M. J., Murali, A., Kemperman, R., Tsugawa, H., Takahashi, M., Vasiliiou, V., Bowden, J. A., Yost, R. A., Garrett, T. J., & Kitagawa, N. (2020). Lipid Annotator: Towards accurate annotation in nontargeted liquid chromatography high-resolution tandem mass spectrometry (LC-HRMS/MS) lipidomics using a rapid and user-friendly software. Metabolites, 10(3), Article 101.
Kumar, M., Sarma, D. K., Shubbham, S., Kumawat, M., Verma, V., Nina, P. B., Jp, D., Kumar, S., Singh, B., & Tiwari, R. R. (2021). Futuristic non-antibiotic therapies to combat antibiotic resistance: A review. Frontiers in Microbiology, 12, Article 609459.
Langford, K., & Thomas, K. V. (2011). Input of selected human pharmacological metabolites into the Norwegian aquatic environment. Journal of Environmental Monitoring, 13(2), 416–421.
Lu, D., Huang, Z., Luo, J., Zhang, X., & Sha, S. (2020). Primary concentration—The critical step in implementing the wastewater based epidemiology for the COVID-19 pandemic: A mini-review. Science of the Total Environment, 747, Article 141245.
Mandai, P., Gupta, A. K., & Dubey, B. K. (2020). A review on presence, survival, disinfection/removal methods of coronavirus in wastewater and progress of wastewater-based epidemiology. Journal of Environmental Chemical Engineering, 8(5), Article 104317.
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.
Nason, S., Lin, E. Z., Eitzer, B., Koelmel, J. P., & Peccia, J. (2022). Changes in sewage sludge chemical signatures during a COVID-19 community lockdown, part 1: Traffic, drugs, mental health, and disinfectants. Environmental Toxicology and Chemistry, 41(5), 1179–1192. https://doi.org/10.1002/etc.5217
Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wehler, W. A., Warren, J. J., Weinberger, D. M., Arnold, W., & Omer, S. B. (2020). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nature Biotechnology, 38(10), 1164–1167.
R Foundation for Statistical Computing. (2020). R: A language and environment for statistical computing. https://www.R-project.org/
Reinstadler, V., Ausweger, V., Grabher, A. L., Kreidl, M., Huber, S., Grander, J., Haslacher, S., Singer, K., Schlapp-Hackl, M., Sorg, M., Biber, H., & Oberacher, G. (2020). Monitoring drug consumption in Innsbruck during coronavirus disease 2019 (COVID-19) lockdown by wastewater analysis. Science of the Total Environment, 757, Article 144006.
Schymanski, E. L., Jeon, J., Gulde, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Environmental Science & Technology, 48(4), 2097–2098.
Senta, I., Gracia-Lor, E., Borsotti, A., Zuccato, E., & Castiglioni, S. (2015). Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. *Water Research, 74*, 23–33.

Senta, I., Rodríguez-Mozaz, S., Corominas, L., & Petrovic, M. (2020). Wastewater-based epidemiology to assess human exposure to personal care and household products—A review of biomarkers, analytical methods, and applications. *Trends in Environmental Analytical Chemistry, 28*, Article e00103.

Sims, N., & Kasprzyk-Hordern, B. (2020). Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environment International, 139*, Article 105689.

Snyder, S. A., Westerhoff, P., Yoon, Y., & Sedlak, D. L. (2003). Pharmaceuticals, personal care products, and endocrine disruptors in water: Implications for the water industry. *Environmental Engineering Science, 20*(5), 449–469.

US Environmental Protection Agency. (2019, February 4). The sources and solutions: Wastewater. Nutrient Pollution. Retrieved May 28, 2021, from https://www.epa.gov/nutrientpollution/sources-and-solutions-wastewater#text=Wastewater%20treatment%20facilities%20in%20the%20gallons%20of%20wastewater%20every%20day

Varghese, F. S., van Woudenbergh, E., Overheul, G. J., Eleveld, M. J., Kunver, L., van Heerbeek, N., van Laarhoven, A., Miesen, P., den Hartog, G., de Jonge, M. I., & van Rij, R. P. (2021). Berberine and obatoclax inhibit SARS-CoV-2 replication in primary human nasal epithelial cells in vitro. *Viruses, 13*(2), Article 282.

Vitale, D., Suárez-Varela, M. M., & Picó, Y. (2021). Wastewater-based epidemiology, a tool to bridge biomarkers of exposure, contaminants, and human health. *Current Opinion in Environmental Science & Health, 20*, Article 100229.

Wang, S., Green, H. C., Wilder, M. L., Du, Q., Kmush, B. L., Collins, M. B., Larsen, D. A., & Zeng, T. (2020). High-throughput wastewater analysis for substance use assessment in central New York during the COVID-19 pandemic. *Environmental Science: Processes & Impacts, 22*(11), 2147–2161.

Wang, Z. Z., Li, K., Maskey, A. R., Huang, W., Toutov, A. A., Yang, N., Sivastava, K., Geliebter, J., Tiwari, R., Miao, M., & Li, X. M. (2021). A small molecule compound berberine as an orally active therapeutic candidate against COVID-19 and SARS: A computational and mechanistic study. *FASEB Journal, 35*(4), Article e21360.

Warowicka, A., Nawrot, R., & Gozdzicka-Jozefiak, A. (2020). Antiviral activity of berberine. *Archives of Virology, 165*(9), 1935–1945.

Wurtzer, S., Marechal, V., Mouchel, J. M., Maday, Y., Teyssou, R., Richard, E., Almayrac, J. L., & Moulin, L. (2020). Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in wastewater, Greater Paris, France, 5 March to 23 April 2020. *Euro Surveillence, 25*(50), Article 2000776.

Yang, Z., Angles d’Auriac, M., Goggins, S., Kasprzyk-Hordern, B., Thomas, K. V., Frost, C. G., & Estrela, P. (2015). A novel DNA biosensor using a ferrocenyl intercalator applied to the potential detection of human population biomarkers in wastewater. *Environmental Science & Technology, 49*(9), 5609–5617.

Zuccato, E., Chiabrando, C., Castiglioni, S., Bagnati, R., & Fanelli, R. (2008). Estimating community drug abuse by wastewater analysis. *Environmental Health Perspectives, 116*(8), 1027–1032.