Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:
______________________________ ______________
Matthew Cavallo April 21st, 2022

Examining the Correlation and Lead Time Between Reported COVID-19 Cases and SARS-CoV-2 RNA Concentrations in Wastewater Influent Catchment Areas in Fulton County, Georgia

By

Matthew Cavallo

Degree to be awarded: MPH

Epidemiology

Dr. Christine Moe

Committee Chair

Examining the Correlation and Lead Time Between Reported COVID-19 Cases and SARS-CoV-2 RNA Concentrations in Wastewater Influent Catchment Areas in Fulton County, Georgia

By

Matthew Cavallo

Bachelor of Science, Microbiology

Ohio State University

2019

Thesis Committee Chair: Dr. Christine Moe

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2022

Abstract

Examining the Correlation and Lead Time Between Reported COVID-19 Cases and SARS-CoV-2 RNA Concentrations in Wastewater Influent Catchment Areas in Fulton County, Georgia

By Matthew Cavallo

INTRODUCTION: Wastewater based surveillance quickly gained popularity during the COVID-19 pandemic as an effective and non-invasive surveillance system aimed at understanding incidence within a wastewater facility catchment area. The ability to detect SARS-CoV-2 RNA in cases that aren't reported, or asymptomatic cases, provides a unique advantage over traditional clinical testing. OBJECTIVE: The objective of this analysis was to understand how SARS-CoV-2 RNA concentrations in wastewater correlate to reported COVID-19 cases. The time with which SARS-CoV-2 RNA concentrations precede reported COVID-19 cases was also examined. METHODS: Wastewater samples were collected from influent lines servicing a known catchment area. Samples would undergo viral concentration and extraction, followed by RT-PCR and quantification. Reported COVID-19 case data was collected from the Georgia Department of Public Health, then geocoded by patient address, to determine weekly number of reported COVID-19 cases for each influent catchment area basin. Linear regression analysis as well as correlation analysis was conducted to examine the relationship between these variables. RESULTS: SARS-CoV-2 RNA in wastewater and reported COVID-19 cases showed a positive correlation for each of the six influent line catchment areas. The line of best fit for each linear regression was weak, potentially due to a variety of biological and environmental factors. The correlation analysis indicated that SARS-CoV-2 RNA concentrations preceded reported COVID-19 cases by about one week for five influent line catchment areas, and about two weeks for one influent line catchment area. DISCUSSION: Future research will need to be conducted that takes into account environmental factors such as wastewater facility flow rate and precipitation levels, as well as biological factors such as variable viral shedding. A one-week lead time is another advantage of wastewater surveillance that allows officials to mobilize resources and prepare healthcare for potential case spikes. As the pandemic dies down and incidence decreases, wastewater surveillance will be an important tool for understanding community burden and mitigating potential outbreaks.

Examining the Correlation and Lead Time Between Reported COVID-19 Cases and SARS-CoV-2 RNA Concentrations in Wastewater Influent Catchment Areas in Fulton County, Georgia

By

Matthew Cavallo

Bachelor of Science, Microbiology

Ohio State University

2019

Thesis Committee Chair: Dr. Christine Moe

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2022

Table of Contents

Literature Review

Methods

Results

Discussion

Literature Review

Introduction

The COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has had a significant global impact that is still being felt years after its emergence in 2019. About two years after the first case was identified in Wuhan, China, the United States has been one of the hardest hit countries as it approaches one million deaths (CDC 2022). Introduction of a safe and effective vaccine has reduced the rate of hospitalizations and deaths; however, vaccine hesitancy has left hospitals in some parts of the country overloaded. Current vaccines have had limited effectiveness on reducing COVID-19 transmission, and there has been mixed results regarding the length of immunity following vaccination or natural infection. Harmful variants stemming from vaccine inequity and other environmental or social factors such as Delta and Omicron, in combination with pandemic fatigue, have made it difficult for public health officials to maintain effective health and safety protocols leading to unprecedented spread across the United States. Recent periods of low incidence have presented public health officials with an opportunity to introduce new interventions aimed at mitigating community spread. Throughout the pandemic, testing has been the most important tool towards measuring its course as well as informing government and healthcare officials on what regions need the most resources allocated (Xiao et al., 2022). However, limitations such as cases choosing not to get tested or asymptomatic infections impact these case counts, and therefore may not provide the full picture (Lau et al., 2021). The cost of a robust testing system is also prohibitively high for many cities across the US, and residents of more rural parts of the country may find it difficult to access testing sites.

Wastewater surveillance is a tool that has been used for many decades to detect a variety of community health patterns such as disease outbreaks, environmental exposures, and degree of illicit drug use (Larsen et al., 2021). Wastewater surveillance involves collecting wastewater samples from a known catchment area and processing through appropriate lab methods to isolate a desired target. This is usually followed by RNA quantification and analysis to transform the data into more meaningful results. When conducting analysis, it is important to consider variables that could dilute the true concentration of RNA in the sample such as treatment facility flow rate and recent rainfall. Treatment facilities servicing different population sizes will experience different rates of water flowing through them, measured in millions of gallons per day. Estimates of flow rate can be found through publicly available data or through working with local partners. Like flow rate, rainfall has the potential to dilute the true RNA concentration. This variable can be difficult to measure and incorporate into analysis. In order to most accurately measure RNA concentrations, it is best to avoid sampling during or directly after significant rainfall. The COVID-19 pandemic renewed enthusiasm for wastewater-based surveillance and saw its implementation in national, state, and local governments and organizations (Farkas et al., 2020). It has also proven to be a non-invasive and cheaper way to collect information about potential or current outbreaks since you don't have to require participation from those that are infected.

Wastewater surveillance however does have its own limitations. Limitations to the number of samples a team can collect, lost samples in the field, and vandalism of equipment can all influence the precision of wastewater sampling efforts and affect estimations of the true COVID-19 burden in these regions. Additionally, effective surveillance requires a multidisciplinary team from a variety of backgrounds and perspectives. A lack of resources and manpower could limit the effectiveness of these surveillance systems in rural parts of the country.

Wastewater Surveillance for Viral Pathogens

Extensive research shows that enteric diseases as well as enveloped viruses such as SARS-CoV-2 can be detected in wastewater from community wastewater treatment facilities (Farkas et al., 2020). The potential for wastewater surveillance also goes beyond the typical enteric virus surveillance, which would be the logical group to target for wastewater-based surveillance. Viruses such as Zika, dengue, influenza, and coronaviruses have all been detected in community wastewater (Bisseux et al., 2018). Detection of these diseases in wastewater also correlates to disease prevalence in the surrounding community (Bisseux et al., 2018). These pathogens are able to be detected in wastewater due to viral shedding in feces from infected subjects. Individuals with COVID-19 will begin to shed virus through their feces prior to exhibiting symptoms as well as throughout the recovery process, with the peak occurring shortly after symptom onset (Miura et al., 2021). Although the process of viral shedding has been well documented, the duration and proportion of individuals actually shed the virus as well as the concentration of virus they are shedding can vary based on a variety of biological and environmental factors. Modelling in combination with observed of hospitalized cases estimate the duration of viral shedding to be about 26 days (95% CI 21.7 – 34.9) with an estimated RNA concentration of 3.4 log copies per gram of feces (95% CI 0.24 – 6.5) over the whole shedding period (Miura et al., 2021). The wide confidence intervals for both estimates highlight the variability of these calculations, which could introduce bias when estimating incidence from these concentrations. Concentration values can also be affected by the time of day at which samples are taken, length of time between sample collection and processing, and extraction efficiency.

SARS-CoV-2, like other respiratory viruses such as the flu, exhibits some seasonality in its transmission patterns (Liu et al., 2021). Typically, there is increased incidence in colder months during the winter as more people stay inside, and less incidence during warmer months in the summer when more people are outside, and transmission is less likely to occur. This pattern was not observed during the beginning of the pandemic due to lockdowns and lack of immunity in the population, however the introduction of vaccines as well as more people gaining natural immunity has led to a more seasonal pattern as the worst of the pandemic winds down (Su et al., 2020). Just as this pattern can be observed in clinical case counts, this pattern should also be observed in RNA concentrations from wastewater. Lower concentrations are observed during the summer as there are less cases and viral shedding, and vice versa during the winter. However, wastewater sampling is more impacted by this seasonality than traditional clinical testing. Increased rainfall during spring months has the potential to dilute the true RNA concentrations in wastewater relative to months with less rainfall such as the fall or summer. In similar studies, microplastic concentration studies showed a statistically significant difference between the wet and dry months, indicating runoff dilution effects (Kittipongvises et al., 2019). The same study also showed negative correlations between microplastic concentrations and both wastewater flow and precipitation. This relationship indicates that increased precipitation can mask true concentrations and limit the effectiveness of wastewater surveillance systems in areas with high annual rainfall. School sampling becomes limited in the summer as students are released for their break. This means that thousands of susceptible individuals are scattered throughout the population, without an effective and convenient method for mass testing such as wastewater surveillance. The winter months likely exhibit the closest relationship between wastewater RNA concentrations and clinical testing as people spend more time inside. This is due to the period of very high community transmission, which results in increased testing and potentially a larger proportion of asymptomatic individuals. Overall, the seasonality and climate of a region should be taken into account when developing wastewater surveillance systems.

Wastewater Surveillance Applications

By monitoring changes in temporal SARS-CoV-2 RNA concentration in wastewater, it is possible to draw some conclusion about the true COVID-19 burden in regions serviced by the wastewater treatment facilities where the wastewater samples are collected. Understanding the changes in case counts at the community level is important as it provides a better understanding about the trajectory of the pandemic and enables more strategic allocation of resources. (Epps et al., 2021). Along with diagnostic testing, wastewater surveillance can also be applied to critical institutions such as schools or nursing homes, as a way to understand COVID-19 transmission and possibly prevent large outbreaks. Being able to identify the presence of even one unknown case in a critical setting such as a school or retirement home, could enable a response and be the difference between an isolated case and a full outbreak (Betancourt et al., 2020). Wastewater surveillance at an individual building level needs to have a relatively quick turnaround time along with accurate results, in order to provide actionable information. If the wastewater results indicate that there is a COVID-19 case somewhere in the building, officials can take the necessary precautions to limit spread such as mass testing or implementing masks if a mandate is not already in place.

Wastewater- based surveillance is also advantageous compared to clinical diagnostic testing as viral concentrations in wastewater include SARS-CoV-2 excreted by cases without clinical symptoms. Asymptomatic cases can be a significant driver of COVID-19 transmission and identification of these cases may only occur through routine or mandatory testing, or through contact tracing. At the community level, burden of asymptomatic cases can be identified from wastewater surveillance by comparing case count estimations based on RNA concentrations in wastewater to reported clinical case counts. However, estimates of COVID cases from SARS-CoV-2 concentrations in wastewater have considerable uncertainly due to a variety of environmental and biological factors that have been discussed could result in concentrations lower than the true value. There is also under-ascertainment of symptomatic COVID-19 who do not get tested or whose test results are not reported by using things such as home testing kits.

One important factor to consider when doing wastewater surveillance is the limit of detection for the pathogen you are trying to detect, which can vary by method. The definition for limit of detection can vary by lab but is generally regarded as the lowest concentration that can be detected in a sample with 95% certainty. This can be defined in terms of either CT values or quantified RNA concentrations. For CT values, the limit of detection is typically 40, and CT values higher than 36 are considered unquantifiable. Limits of detection for quantified SARS-CoV-2 RNA in wastewater can be as low as 4000 gene copies per 50mL of sample (Fung et al., 2020). The limit of detection is determined by running a dilution series of a standard. Starting with a known concentration of SARS-CoV-2 RNA, the sample is diluted by a factor of 10 until you have a range of concentrations. These concentrations are then run through PCR, with the lowest concentration that still produces a CT value being considered the limit of detection. Highly sensitive RT-qPCR assays have the potential to detect as low as 10 gene copies per reaction (Ahmed et al., 2022). It is important to consider that unquantifiable samples, or samples that test above the assays limit of detection, aren’t necessarily free of SARS-CoV-2 RNA.

Continuing to improve this limit of detection will be important as it will allow scientists to identify smaller numbers of cases in larger populations. Through wastewater testing, any group is unlikely be able to detect a singular case in a building or catchment area regardless of how sensitive their methods are as the virus shed from a single individual is unlikely to cross PCR limit of detection. It is difficult to determine what the lowest number of cases is that can be identified from wastewater surveillance due to the variability of shedding in infected individuals. Further tests would need to be done involving testing of individual buildings containing known numbers of infected individuals to explore this limit of detection.

Understanding the relationship between SARS-CoV-2 RNA concentrations in wastewater and clinical cases in the population is a crucial factor in using the results to make informed policy decisions. Several studies have already shown the correlation between SARS-CoV-2 concentrations and clinical case counts in the population (Bar-Or et al., 2020; Medema et al., 2020). This validates the general idea that wastewater can be used as an effective surveillance tool. Taking this correlation a step further, the lead time from the wastewater surveillance data can be estimated. Lead time is defined as the time with which RNA concentrations precede observed clinical case counts. By lagging clinical results in relation to wastewater results, estimates can be made about how much lead time finding concentrations of SARS-CoV-2 RNA gives us before we see it translate into clinical results. Understanding the lead time, which can vary widely depending on the robustness of clinical testing, sampling frequency, and population size, is vital for public health officials. Estimations of lead time are difficult to generalize to other cities, as wastewater flow rates and the population being served by each plant differs between each city as well. A variety of other health and biological factors such as the delay between symptom onset and viral shedding in stool, as well as what proportion of people shed virus and an individual's level of ability to seek treatment can all impact RNA concentrations in wastewater (Olesen et al., 2021). It is also important to consider the benefit of this lead time in the context of testing availability. In a population where there is abundant testing, the lead time for wastewater data would not be as effective since those experiencing symptoms or have been exposed are free to get tested as they please. In a population with limited testing, this lead time can be vitally important as it allows officials to enact mitigation measures before individuals begin to show symptoms, as well as allocate testing resources to these regions.

While simple presence absence of SARS-CoV-2 from wastewater surveillance is effective for individual buildings to guide mitigation efforts, results from the community often need to be quantified to allow officials to tailor strategies based on the severity of the outbreak (McMahan et al., 2021). In order to transform these quantified values into more meaningful case counts, SARS-CoV-2 RNA gene copies per day can be utilized in an SEIR model, which some studies have utilized to predict COVID-19 transmission (Carcione et al., 2020). Outside of fitting these concentrations into a model, plotting case counts on the y axis alongside quantified RNA values on the x axis, estimations can be made about what concentrations correspond to in clinical case counts using a linear regression.

Understanding the prevalence of COVID-19 in certain regions is especially important in determining the overall trajectory of the pandemic as well as the stress being placed on hospitals in the area. Being able to estimate case counts from RNA concentrations is also important for communicating findings to community partners. When communicating results to partners without science backgrounds such as school or government officials, it is important to not focus too heavily on the data and analysis that goes into creating those results. Officials want to be able to understand a value of how many cases they are dealing with and what their next options are to contain it. Understanding both the lead time and the estimate case counts associated with wastewater results is vital to transforming wastewater surveillance into an important clinical tool that can be used to make informed policy and mitigation decisions.

Fulton County Demographics

Effective COVID-19 surveillance is especially important in large metropolitan cities like Atlanta, with a diverse population and neighborhoods accommodating people from a variety of socioeconomic backgrounds. Fulton County Georgia, one of the three counties that make up metro Atlanta, is the largest in the state and the only county that has over one million residents. Sampling efforts from this project focused primarily on the residents located in the southern part of Fulton County due to the high minority population in this area (42% African American), as well as limited testing capability and vaccine rates (55% fully vaccinated), based on data available from the Fulton County government. Additionally, this part of the county is close to the Hartsfield-Jackson International Airport meaning new and potentially infected individuals are constantly entering and leaving the population. Working closely with officials throughout the city of Atlanta, any positive results from the wastewater data were quickly communicated to allow for a quick and appropriate response. Another important factor to consider is the large minority population in Atlanta, considering these populations has been underserved and hit the hardest by the pandemic (Reyes 2020). Black Americans have carried the largest burden of cases, hospitalizations, and deaths throughout the pandemic, followed closely by Latino Americans. This could be due to a variety of factors such as inability to work from home, limited access to testing or testing sites, and reduced vaccine uptake due to vaccine hesitancy (Reyes 2020). Although our understanding of the social determinants of health associated with COVID-19 is limited, wastewater surveillance can be a tool used to reduce mortality disparities and build trust with local communities. An obvious intervention is to improve clinical data collection efficiency and accuracy. This data collection should also focus on collecting variables such as socio-economic status, race, and place of residence. Wastewater surveillance is unable to provide any information of these variables, however it does provide consistent results that are more accurate to the true burden in a region compared to looking at clinical data alone. Wastewater surveillance, in combination with demographic information can still be an effective tool in understanding racial health disparities and is especially important in communities with limited access to testing.

Research Question

Most COVID-19 wastewater surveillance efforts in the US have focused on data from samples collected at municipal wastewater treatment plants and examined the relationship to city-wide COVID-19 case reports and hospitalizations. In Atlanta, wastewater samples were collected from six influent lines at two wastewater treatment facilities and quantified concentrations of SARS-CoV-2 RNA in these samples. We also collected geocoded data from the GA Department of Public Health on reported COVID-19 cases in the six geographic catchment areas represented by these influent line samples. Using these two data sources, this study examined:

1) temporal trends in SARS-CoV-2 RNA concentrations and temporal trends in reported COVID-19 cases for each catchment area to determine:

A) if the wastewater results provided advance warning of subsequent rises in COVID-19 prevalence in the catchment area, and

B) what was the length of lead time observed in the wastewater results.

2) if there was a consistent relationship between the SARS-CoV-2 RNA concentrations in the influent line samples and the reported COVID-19 cases in the geographic catchment area served by the influent line.

This analysis will take into account local flow rates and population characteristics. Based on the results, recommendations will be made for how these results can be used in the community and how they can help underserved populations. Results from this thesis will be important to public health officials around Atlanta as provides information from local wastewater data, and it takes into account local factors. This analysis can also be expanded further in the future to include more variables, and potentially to examine spatial-temporal changes in SARS-CoV-2 RNA concentrations and clinical case counts.

Methods

Wastewater Sampling

Wastewater grab samples were collected from six influent lines entering two wastewater treatment facilities (three at the Utoy Creek Water Reclamation Center and three at the South River Water Reclamation Center). Each of these influent lines serve geographically distinct areas of the southern part of Fulton County, Georgia. Characteristics of these sites as well as their location and catchment area are shown in Table 1 and Figure 1. Population estimates were calculated using 2020 census block data and finding the total population from each census block that intersects with each wastewater facility catchment area. Population estimates were not calculated for Old Winn Dixie due to its small size and overlap with other catchment areas.

The samples were typically collected early Monday mornings to coincide with residents getting ready for work or school. After lifting the manhole cover, collection was done using a painter's pole with a seat for a one-liter bottle attached at the end. A sterile polypropylene bottle was then lowered into the appropriate wastewater stream and held there until full. The samples were then transported from the field to the lab on ice and kept at 4°C until processing, typically conducted within one hour of arrival. Extensive work was done with the department of watershed management in order to identify manhole catchment areas and new sampling sites. Moore swab samples were also set at various community sites around the county; however, these are set out over a 24-hour period to serve as a presence/absence indicator and therefore can't be quantified and will not be important to this analysis. COVID-19 case data was collected using publicly available information from the Georgia Department of Health. Reported COVID-19 cases were then geocoded by patient address, rather than testing location, to determine weekly number of reported COVID-19 cases for each influent catchment area basin. COVID-19 case report data was only used from late March of 2021 through late February of 2022 as this was the period when the SARS-CoV-2 RNA was reliably quantified in the wastewater samples and when the COVID-19 case counts were available.

Virus Concentration and RNA Extraction

Samples were processed within 24 hours of collection using a Nanotrap Kingfisher concentration/extraction and MagMax extraction kit following the protocol from Sabon et al., 2022. Bovine Respiratory Syncytial Virus (BRSV) was used as a positive extraction control for every sample processed. Following concentration and extraction using this method, an average of 50µl purified viral RNA was produced. This RNA was then analyzed by real-time quantitative reverse transcription polymerase chain reaction (RT-qPCR) including a SARS-CoV-2 dilution series standard curve consisting of known RNA concentration values for relative quantification. Primers and probes targeting the N1 gene of SARS-CoV-2 as well as the B gene of BRSV were used in the RT-qPCR reaction. BRSV concentrations were not quantified and were treated as a simple presence absence to serve as an extraction control. Extrapolation from the standard curve allowed conversion of CT values into estimated number of gene copies in the reaction tube. All samples were tested in duplicate. Quantifiable positive samples were defined as those with both duplicate tubes having a CT value of 36 or less and the CT values of the duplicate tubes were within two units of each other. Some positive wastewater results were excluded from the analysis because their CT value was greater than 36 or there was greater than two CT difference between duplicate tubes. Estimated SARS-CoV-2 RNA concentrations from every influent line were measured weekly to examine temporal trends in concentrations and examine the relationship to reported COVID-19 case counts in each catchment area.

Data Analysis

All data analyses were conducted using R Studio (RStudio PBC, Boston Massachusetts). In order to estimate case counts from the wastewater data, the SARS-CoV-2 RNA copies per 100mL were log transformed and reported clinical case count data by each wastewater basin was utilized. Weekly SARS-CoV-2 RNA concentrations were matched to weekly reported COVID-19 case counts by influent line catchment area. A linear regression was used to assess the association between these two variables. The linear regression for each catchment basin generates a simple formula that can be used to estimate reported COVID-19 cases in the population of that geographic area.

$$\text{Reported Case Count} = \text{y-intercept} + \text{slope} \times \log_{10} \text{SARS-CoV-2 RNA concentration}$$

The linear regression also generates an R-squared value to determine goodness of fit and how much variance in the log viral concentrations truly impacts changes in observed case counts. This regression does not include other variables that could impact viral concentrations such as flow rate or typical rainfall for the month. Previous analysis shows that out of viral load, flow rate, and daily rainfall, viral load was the only significant predictor variable for estimating case counts (Vallejo et al., 2020). Removing flow rate and daily rainfall from the analysis also yielded a better line of fit compared to a multivariate analysis with the variables included (Vallejo et el., 2020).

Estimating the lead time between increasing SARS-CoV-2 RNA concentrations in wastewater and reported number of COVID-19 cases was performed by examining the levels of correlation between the two variables at different time periods. These values were also plotted in order to observe trends and simple correlation between the variables. The trend lines were smoothed using smoothing techniques from the Tidyverse package in R in order to observe patterns in the data and minimize overplotting. This smoothing makes the relationship between the variables easier to compare and visualize.

Results

Reported COVID-19 cases and SARS-CoV-2 genome copies trends

Throughout the study period from February 2021 through April 2022, a total of 257 wastewater samples were tested, and 40,294 COVID-19 cases were reported (Table 3). The average number of SARS-CoV-2 genome copies was also reported, with the lowest and highest detected values also reported. For each influent line catchment area, there were large temporal variations in reported COVID-19 cases and in SARS-COV-2 RNA concentration in the samples from the influent lines (Table 2). Reported COVID-19 cases and SARS-CoV-2 RNA concentrations peaked during the winter months which was in line with trends across the country. Figure 2 illustrates the temporal variability in both the reported COVID-19 cases and the SARS-CoV-2 RNA concentrations for each influent basin. For some catchment areas, there were times when the SARS-CoV-2 RNA concentrations in the wastewater were higher than expected relative to the number of reported COVID-19 cases such as Phillip Lee in the month of October or Jonesboro in the month of November. This is likely due to gaps in clinical testing resulting from asymptomatic cases or cases not getting tested. For other catchment areas, the SARS-CoV-2 RNA concentrations were lower than expected given the number of reported COVID-19 cases in the area. For example, the Phillip Lee influent line consistently had a relatively low average concentration of SARS-CoV-2 genome copies compared to the number of reported COVID-19 cases in that area. This could be due to inhibition in quantifying the RNA from this site or resulting from some lab methods not being able to detect the true concentrations for this area. For this analysis, samples that were considered negative or unquantifiable were assigned a value of 66.67 genomic copies per 100mL. This value is used in the analysis to signify that negative samples likely were not completely free of SARS-CoV-2 RNA, and this value is the lowest number of copies able to be detected in wastewater. Therefore, results from table 2 with values of 66.67 genomic copies per 100mL represent months when no positive wastewater samples were detected, or the positive samples that were detected were considered unquantifiable. Table 3 is a summary of the results from table 2 and shows how many samples were collected over the time period.

Estimating COVID-19 population prevalence from measured SARS-CoV-2 RNA concentration in wastewater

Figures 3a-3f represents the results from the linear regression analysis of reported COVID-19 cases and quantified SARS-CoV-2 RNA values in wastewater for each influent basin. There was a positive relationship between these two variables in all six catchment basins as well as a formula to estimate case counts from the graphs produced. The regression for the Jonesboro influent line showed the best line of fit, while the regression for South Fulton showed the weakest line of best fit. The slope for each influent line showed a moderately positive relationship. The flint influent line regression shows a log concentration of 10.81gc/100mL with an estimated case count for the week January 3rd, 2022, of 1861 cases. Similar concentrations in the analysis estimate that this log concentration would be associated with between 200 and 300 weekly cases. This relationship is seen at every influent site for the first week in January. Despite each regression graph showing at least one outlier, removal of the outlier actually resulted is a less fitting line and these results were therefore included in the analysis. Standard error bars in each graph indicate the variance in these estimations. The graphs also show how COVID-19 case estimations can vary by treatment plant. Phillip Lee, an influent line that often had positive wastewater results and relatively high reported numbers of COVID-19 cases, showed a positive relationship between the two variables with a relatively strong line of best fit ($r^2 = 0.243$). In contrast, other sites like Flint that also frequently had positive wastewater results and high case counts exhibited a less strong line of best fit ($r^2 = 0.167$). Other sites like the South Fulton influent line, often had negative wastewater results despite having high numbers of reported COVID-19 cases per week. This lack of association between the variables for this site is also reflected in the poor line of best fit ($r^2 = 0.103$).

Potential factors contributing to the variation between the influent line catchment areas as well as factors that could impact the reported COVID-19 cases and SARS-CoV-2 RNA quantification in the wastewater samples will be discussed later in the paper. Overall, the linear regression analysis indicates that the relationship between reported number of COVID-19 cases and SARS-CoV-2 RNA concentration show decent predictive ability and can be used to predict the number of infected individuals in a given region based on viral load concentrations from wastewater data. Overall, the relationship between SARS-CoV-2 RNA concentrations in wastewater and reported COVID-19 case counts is influenced by a variety of factors that future research can address which is discussed later in this analysis. Due to this, it is difficult to conclude that there is good predictive ability for COVID-19 incidence from SARS-CoV-2 RNA concentrations in wastewater.

Lead Time Estimation

The ability to estimate the lead time between the rise of SARS-CoV-2 concentrations in wastewater and the rise in the reported number of COVID-19 cases varies by influent catchment area. Table 4 compares the correlation between reported COVID-19 cases and SARS-CoV-2 RNA concentrations in wastewater with different lag times to examine the temporal relationships between these variables. Overall, a one-week lag between SARS-CoV-2 RNA concentrations was most strongly correlated to the number of reported COVID-19 cases. Only in the Jonesboro catchment basin was there a stronger correlation with a two-week lag time. For each site the true lead time is likely somewhere between one and two weeks. No lag time between the two variables had the lowest correlation, suggesting that the SARS-CoV-2 RNA concentrations in wastewater were a better indication of the number of cases that would be reported in the week following the wastewater sample rather than the same week that the wastewater samples were collected. Although the correlation between the wastewater results and reported COVID-19 cases was examined for up to twelve weeks of lag time (results not shown), only the one and two-week lag times had statistically significant correlation values. Results from the correlation analysis showed that there were no significant correlations between the reported COVID-19 cases and the SARS-CoV-2 RNA concentrations in wastewater samples collected in the following weeks. These results are supported by the temporal trend lines shown in Figure 2 where the trend line for SARS-CoV-2 RNA concentration almost always rises before the trend line for the number of reported COVID-19 cases.

The red line shows the trends in case counts for each influent line, which follow a similar pattern to trends observed in other parts of the country with peaks in the winter and lower incidence in the summer (Kittipongvises et al., 2022). Black dots represent the weekly RNA concentrations collected from these sites during this time period. The black trend line for these values represents the best fitting line for these values in order to visualize the changes in concentrations over time. Observing the patterns of the temporal trends highlights the correlation between these two variables and illustrates the lead time the wastewater surveillance can provide.

Discussion

The objectives of this analysis were to determine the relationship between SARS-CoV-2 RNA concentrations in wastewater and reported COVID-19 concentrations, as well as determine whether these concentrations had any temporal lead time on observed cases. Results showed a weak to moderate relationship between these variables, and about a one-week lead time for SARS-CoV-2 RNA concentrations. The weak relationship between the variables was likely due to the exclusion of important variables that impact wastewater concentrations such as flow rate and precipitation levels. Limitations to this analysis, as well as future recommendations to improve it are discussed.

Applications

Following extensive analysis of wastewater samples for SARS-CoV-2 RNA as well as reported COVID-19 cases in Fulton County, Georgia, it is important to interpret these findings, identify potential limitations, and make recommendations for future research and implementation. Although there is a clearly demonstrated relationship between SARS-CoV-2 RNA wastewater concentrations and reported COVID-19 cases, it is important to discuss the significance of this relationship and how public health officials can take advantage of it. As expected, and demonstrated in Table 2 and Figure 2, dips in reported case counts also resulted in lower wastewater RNA concentrations. Reported case count spikes in August and December 2021 also resulted in much higher concentrations relative to other months with less cases. By identifying an approach that public health officials can use to estimate COVID-19 case counts from SARS-CoV-2 RNA concentrations in wastewater, officials can not only prepare for upcoming spikes from these results, but they may also able to estimate just how big the upcoming burden may be. This is especially useful when working with organizations involved with diagnostic testing or vaccine distribution. By estimating how many additional COVID-19 cases are expected, officials can send additional tests or vaccines to these areas without sending too much or too few supplies. Operating at a higher efficiency in this way will allow for more robust clinical surveillance, as well as more targeted diagnostic testing and vaccine distribution, with the long-term goal of reducing COVID-19 incidence in geographic areas with high SARS-CoV-2 RNA concentrations in wastewater. The estimated lead time of one to two weeks also plays a factor in the efficiency of implementing diagnostic testing and vaccination scale efforts. The lead time creates a deadline for emergency outbreak response initiatives and gives hospitals advanced warning to prepare for surge capacity. However, it is difficult to initiate these interventions based on wastewater data alone due to the logistics of implementing these interventions over a short period of time as well as the reliance on clinical data to inform policy decisions.

Anticipating that a large spike in cases is coming can be useful for increasing testing in certain areas, but more cases doesn’t necessarily lead to an increase in hospitalizations. Since you are unable to gather demographic information from wastewater sampling, it is nearly impossible to determine whether the future spike is primarily occurring in a vulnerable population. These spikes could also be occurring in populations that are already vaccinated and boosted, therefore reducing the need for increased vaccine availability in these areas. Limitations to the demographic data that can be collected is a significant limitation to wastewater-based surveillance as it is difficult to determine whether spikes in cases are of public health importance without the proper context regarding underlying health conditions, race, socioeconomic status, and vaccination rates in the geographic areas with the increased SARS-CoV-2 RNA concentrations. Further analysis can be done to explore the relationship between SARS-CoV-2 RNA concentrations and hospitalizations. It is expected that these variables would follow similar trends as reported COVID-19 cases and hospitalizations are closely correlated depending on the SARS-CoV-2 dominant strain at the time.

Existing Literature

Relationship Between SARS-CoV-2 RNA Concentrations and Reported COVID-19 Cases

Results from this analysis support previously conducted studies assessing the correlation between wastewater SARS-CoV-2 RNA concentrations and reported COVID-19 case counts both in the United States around the world. Similar studies in Canada (Pileggi et al., 2022) and Spain (Sanjuan et al., 2021) successfully demonstrate this correlation between SARS-CoV-2 RNA concentrations from samples from wastewater treatment facilities servicing large cities encompassing over one million people. Both studies reached similar conclusions that wastewater surveillance is an effective tool for predicting COVID-19 outbreaks, especially during low incidence periods, and is unbiased by the presence of large numbers of asymptomatic cases. Most previous studies of wastewater-based surveillance for COVID-19 focus sample collection at wastewater treatment facilities servicing thousands of people in the catchment area as opposed to examining the relationship between SARS-CoV-2 RNA concentrations in wastewater and COVID-19 cases reported at a small community level or even an individual building level.

Wastewater surveillance of hospitals has been used by researchers as a way to monitor the presence of antibiotic resistance genes and other harmful pathogens in the community (Timraz et al., 2016). A recent study conducted in tertiary care hospitals around Calgary utilized wastewater sampling directly from hospitals with high COVID-19 incidence as a way to understand the epidemiology of SARS-CoV-2 specifically in a healthcare setting (Acosta et al., 2021). Understanding how COVID-19 spreads in a healthcare setting is important since perception about the risk of contracting COVID-19 in these settings has led to hospital avoidance, which could lead to negative outcomes associated with other health conditions. The study found that not only did spikes in hospital wastewater SARS-CoV-2 RNA concentrations correlate to case spikes in the community, but also correlated to increases in hospital acquired COVID-19 infections. The study was even able to predict outbreaks at the unit level of each hospital based on the SARS-CoV-2 RNA concentrations in wastewater, however the lead time between the rise in SARS-CoV-2 RNA concentrations in wastewater and the occurrence of an outbreak was not reported by researchers. It is also important to note that hospital SARS-CoV-2 RNA concentrations in wastewater are not impacted by flow rate and precipitation like for municipal wastewater treatment facilities. However, chlorine and other chemical disinfectants were observed which could impact virus concentrations in wastewater. This study is important as it shows that the correlation between SARS-CoV-2 RNA wastewater concentrations and reported case counts applies even at an individual building level and does not require thousands of cases or large catchment areas to observe trends.

Other studies at an individual building level highlight the importance of wastewater surveillance when dealing with potentially asymptomatic cases. Wastewater surveillance was conducted on an apartment complex after an initial round of diagnostic testing, where positive cases and close contacts were quarantined appropriately (Wong et al., 2022). Despite no additional confirmed cases, researchers detected increasing SARS-CoV-2 RNA concentrations in the wastewater, suggesting new cases in the building. A second round of testing found one additional case, which the increasing SARS-CoV-2 concentrations in wastewater were attributed to. After removal of this individual, SARS-CoV-2 concentrations in the wastewater persisted for a few days which was attributed to recovered individuals returning to the complex. This study reveals two key findings. First, the researchers were able to detect a single case in the building from wastewater surveillance alone. This is something that cannot be done at a population level or in individual buildings with high incidence such as hospitals, which highlights this tools effectiveness for potentially monitoring high risk buildings such as nursing homes. Second, researchers still detected SARS-CoV-2 RNA in the wastewater after confirmed cases recovered from their symptoms and returned home. This indicates that viral shedding can occur both before and after symptoms are present. Applying this to the population level, recovering individuals could contribute to observed SARS-CoV-2 concentrations in wastewater without being considered an active case. This would result in higher-than-expected SARS-CoV-2 RNA concentrations in wastewater which could affect the correlation to reported COVID-19 case counts. Further research will need to be conducted to understand COVID-19 viral shedding as discussed later in this section.

Lead Time Estimation

Studying the lead time wastewater surveillance provides is difficult to do at an individual building or community level as it is difficult to generalize these results to a larger population, and inconsistent cases make it difficult to conduct a long-term study. Several studies across the globe have attempted to study lead time at a population level with mixed results, and these were examined through a meta-analysis (Kumar et al., 2022). The meta-analysis concluded that the lead time wastewater surveillance provides varies widely depending on population dynamics, climate conditions, and connectedness of the sewer network. Local temperature and precipitation can have a significant impact on SARS-CoV-2 concentrations, which can vary widely based on geographic location. This supports considerations taken specifically for this analysis, which found different results based on influent line, and indicates that lead time observed in one region can not be generalized to others without taking into account local factors. Public health officials hoping to rely on wastewater surveillance as an early detection tool will need to be aware of studies conducted in similar geographic locations and climates and will still need to exercise caution when making assumptions about lead time without conducting studies specific to the population being monitored.

Limitations

Although this analysis successfully demonstrates the relationship between SARS-CoV-2 RNA concentrations in wastewater and reported COVID-19 case counts, it does present several weaknesses. First, despite the positive linear relationship log concentrations and weekly case totals, the line of best fit for each regression doesn't fit as well as similar analysis in existing literature which typically have an r squared value ranging from 0.6 to 0.8 (Vallejo et al., 2020). This could be due to several factors including data quality, differences in RNA quantification, and other unknown confounders that could influence RNA concentrations in wastewater or impact COVID-19 case reporting. The most likely factor is that the analysis did not include the variable flow rates for each wastewater treatment facility and also doesn't take into account precipitation levels, both of which could impact SARS-CoV-2 RNA concentrations. Flow rates are also much higher in colder months or during lockdowns as people spent more time inside compared to the warmer months (Farkas et al., 2020). The variability of flow rate for individual wastewater treatment facilities is also something that should be taken into account in future analysis.

Another potential limitation to this analysis is that it specifically focuses on Fulton County, despite the metro Atlanta area also consisting of parts of DeKalb and Gwinnett Counties. Many people reside in these counties and commute to Fulton County for work or other recreation. Without understanding the COVID-19 burden in these areas, it is difficult to enact meaningful health and safety protocols for the city. People travelling to Fulton County for work could contribute to the wastewater concentrations however would not contribute to the clinical case counts since this is based on residential address. Case counts in each of the three counties, although high relative to other counties around Georgia, are still quite different from each other. Cumulatively through April 7th, 2021, according to the Georgia Department of Public Health (GA DPH), Fulton County has had 186,812 confirmed cases since the start of the pandemic. This is compared to 131,158 cases in Dekalb County and 175,268 cases in Gwinnett County. This also only includes those that have been tested and doesn’t include symptomatic cases that didn’t get tested or asymptomatic cases, both of which would still contribute to the observed wastewater SARS-CoV-2 RNA concentrations. High case counts in surrounding counties that compose a large metro city makes it difficult to enact policy decisions focusing on just one county and could influence both the case counts and RNA concentrations used in this analysis. Finally, the analysis did not take into account the rapidly evolving nature of COVID-19. The original strain that caused the initial lockdowns later evolved into the Delta variant followed by the Omicron variant. Limited information on the amount of virus infected individuals shed in their stool, especially related to the different variants, leaves a large gap in information regarding how these wastewater concentrations can change over time. The Omicron variant, more infectious and less severe, led to rapid spikes in clinical cases across the country. However, it is unknown whether the variants decreased in severity also leads to less viral shedding. A shorter duration of illness as well as less cases of "long Covid" can also impact the amount of virus being shed. Analysis differentiating by variant may have resulted in different correlations and could yield better insight into how case counts, and wastewater RNA concentrations have changed over time.

Recommendations

Wastewater based surveillance has many potential applications and avenues for future research that can further develop our understanding of community disease incidence. In terms of application, it is important to develop robust wastewater surveillance systems across the country, including more rural parts that may be difficult. This work is currently in the early stages at the national level with government entities like the CDC and private companies such as Biobot. Conducting this surveillance at the national level is important for understanding the overall health of the country and could give insight into where and what conditions the government should prioritize with their public health policies. In order to establish this system at a more regional level, more efficient and cost-effective laboratory and sampling methods will need to be developed. Current methods often rely on expensive processing equipment as well as specialized training to gather and analyze the resulting data. Long commutes and inaccessible manholes provide significant challenges for rural sampling efforts. Cooperation between research laboratories, government entities, and private industries would make the process more efficient by combining resources and allowing for a larger sampling area.

Another way to improve on the application of this system is to expand into surveillance of other diseases besides COVID-19. Wastewater-based surveillance became more popular during the pandemic as an effective way estimate COVID-19 incidence while facing a clinical testing shortage. Now that there is an abundance of clinical testing, wastewater-based surveillance can be expanded into understanding the incidence of diseases with less robust surveillance systems such as norovirus or influenza. Certain communities could also benefit from using wastewater surveillance to understand the extent of illicit drug use and change their policies or resources accordingly. This can be achieved without an increase to the number of samples collected by organizations. However, this will result in different laboratory process and procedures, increasing the time for training and overall costs associated for surveillance. Testing for a variety of different diseases is likely achievable at the national level, however public health departments and other wastewater surveillance organizations at the regional level will need to target a few specific diseases that they consider high priority in order to reduce costs. This could result in disjointed surveillance efforts at the state level, and results would only be applicable to the region being tested without being able to draw conclusions about what is happening at the state level.

In terms of future research, there is still a lot to be done to understood about COVID-19 viral shedding rates, environmental impacts on RNA concentrations in wastewater, and what other diseases wastewater-based surveillance could be effective for. As stated previously, more research needs to be done on whether there are differences in viral shedding rates by variant. If there were differences, this would greatly impact the effectiveness of these surveillance systems as researchers would constantly need to update how they analyze these concentrations based on whatever the most prevalent variant is at the time. Additionally, more research needs to be done on what biological or environmental factors affect rates of viral shedding. The rate is likely to change based on patient sex, lifestyle, BMI, and other genetic factors. Much is unknown about how much each of these factors impacts the rates and makes it difficult to determine whether wastewater RNA concentrations are truly representative of the whole population. To start, research can be done on whether there are differences in viral shedding rates between symptomatic and asymptomatic cases. This will be vital as one of the biggest benefits to wastewater surveillance is the ability to pick up on asymptomatic cases. If asymptomatic cases were to shed the virus significantly less compared to symptomatic cases, the system would be less robust as it would still be missing a significant number of cases similar to clinical testing.

In addition to variable flow rates and precipitation levels that have the potential to dilute RNA concentrations in wastewater, there are other environmental pollutants and chemicals added to wastewater either during processing or as run off that could impact concentrations. The reality is that it is often difficult to find pretreated wastewater without unwanted environmental pollutants for many parts of the country. These pollutants often have an impact on the pH of the wastewater, which in turn could degrade the RNA found in it. First, it will be important to understand how varying pH can impact RNA extraction efficiency. This could be impacted either by RNA degradation in the water itself or could result in incompatibility with certain laboratory methods for wastewater processing. It is likely impossible to stop RNA degradation due to environmental factors prior to sample collection, which would result in lower-than-expected concentrations. However, if the pH is only impacting laboratory processes, a pH adjustment can be made. Future research will also need to be done to develop more sensitive laboratory methods for processing wastewater samples, including methods to remove environmental pollutants in samples. This could result in additional costs and longer processing times but would greatly improve the efficiency of RNA extraction from wastewater.

Finally, it is important to explore what other diseases or conditions can be monitored through wastewater surveillance. Although there are currently a large number of diseases that can be detected, there are still many conditions with a severe impact on the US healthcare systems that could have ineffective clinical surveillance systems. One such condition is diabetes, which is a large problem in the US due to obesity and inability to access healthcare (Xu et al., 2018). Diabetes is likely underreported both as a clinical condition and as a cause of death (CDC 2008). Diabetes is a largely lifestyle related condition that requires prompt treatment and changes in order to avoid negative outcomes. Using wastewater surveillance to determine the true burden the US healthcare system experiences due to diabetes could significantly increase awareness and resources to certain parts of the country with large previously unknown burdens. However, conditions such as diabetes could also present ethical concerns as it does have a genetic component (Xu et al., 2018). Since surveillance for this condition wouldn't be as simple as extracting virus RNA, there would likely need to be a sequencing component to the surveillance possibly involving human mitochondrial DNA. The sequencing component of this approach would also raise ethical concerns that require more resources and lengthy sample processing times. Overall wastewater-based surveillance is a relatively unexplored tool, with a variety of potential future applications.

Conclusions

Wastewater based surveillance has been a vital tool in curbing the effects of the COVID-19 pandemic. Strong correlations between SARS-CoV-2 RNA concentrations in wastewater and clinical case counts highlight this systems accuracy and effectiveness. Public health and government officials are able to rely on results from these surveillance systems in order to enact policy changes and allocate resources accordingly. Although this tool is not new, the pandemic should highlight the strength of these surveillance systems and should be implemented more widely in the future. Further research will need to be conducted in order to answer some of the unknowns both about wastewater surveillance and COVID-19. As the pandemic comes closer to a conclusion, wastewater surveillance will be an important tool in stopping potential outbreaks and for understanding disease incidence as mask mandates and testing requirements are lifted. As more states and departments embrace wastewater surveillance for its benefits, the system will become more efficient and innovative, creating an efficient and robust surveillance tool to better understand the health status of the country.

Appendix

Figure 1. City of Atlanta sewerage map indicating six influent line catchment areas included in this study

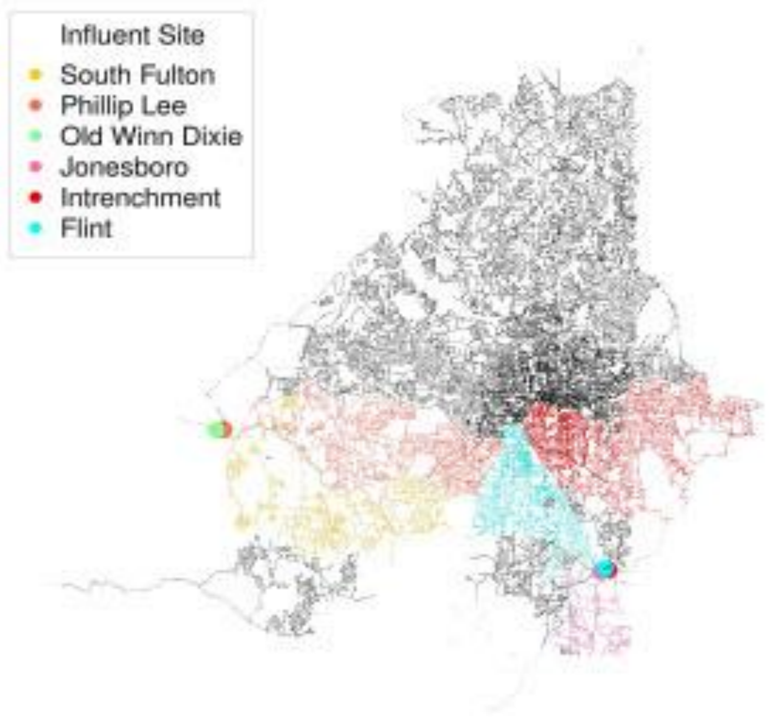


Image created by Dr. Yuke Wang, Senior Biostatistician, Emory University

Table 1. Wastewater influent line characteristics

Influent Line	Wastewater Treatment Facility	Catchment Size (Number of Manholes)	Average Volume (Million Gallons per Day)	Estimated Population in Catchment Area
South Fulton	Utoy Creek	2489	4	39,377
Phillip Lee	Utoy Creek	3646	11	37,552
Old Winn Dixie	Utoy Creek	13	1	N/A
Jonesboro	South River	770	1	5,427
Intrenchment	South River	7108	13	79,502
Flint	South River	2981	3.2	29,519

Table 2. Average reported COVID-19 cases per month (average SARS-CoV-2 genome copies/100mL per month) by influent line catchment area, April 2021 – February 2022

Influent Site	**Apr-21**	**May-21**	**Jun-21**	**Jul-21**	**Aug-21**	**Sep-21**	**Oct-21**	**Nov-21**	**Dec-21**	**Jan-22**	**Feb-22**
South Fulton	93 (139.60)	54 (66.67)	289 (264.24)	181 (N/A)	39 (66.67)	58 (66.67)	76 (174.83)	63 (66.67)	927 (8092.83)	1030 (978.21)	303 (66.67)
Phillip Lee	626 (1771.10)	284 (573.11)	307 (575.90)	471 (N/A)	926 (8666.89)	746 (6027.63)	889 (8159.13)	801 (5056.46)	1356 (17330.68)	1980 (10278.91)	678 (1330.96)
Old Winn Dixie	478 (899.29)	163 (181.20)	57 (104.89)	556 (1121.39)	1126 (17154.95)	569 (3689.36)	864 (4982.13)	633 (2683.66)	1792 (21278.54)	1451 (9381.58)	349 (380.37)
Jonesboro	95 (161.95)	301 (272.60)	24 (66.67)	588 (1612.30)	1345 (2494.38)	171 (475.09)	211 (395.76)	477 (985.22)	1504 (12456.69)	1051 (7184.19)	210 (201.11)
Intrenchment	310 (699.78)	600 (1991.12)	89 (166.98)	389 (N/A)	1299 (2706.11)	216 (2104.07)	679 (2923.12)	729 (4333.40)	1399 (11634.37)	1124 (8110.72)	150 (66.67)
Flint	317 (659.01)	92 (173.32)	44 (66.67)	871 (4441.99)	1417 (20606.90)	417 (4444.40)	470 (5724.00)	1009 (10628.34)	621 (10542.52)	2110 (13072.22)	791 (2233.75)

Table 2a quantifies the trend lines shown in Figure 2. Average case counts by each influent line are shown with average number of genomic copies over the same time period. "N/A" in the month of July indicates no quantifiable results since samples from South Fulton, Phillip Lee, and Intrenchment were not collected. Samples that were considered negative or unquantifiable were assigned a value of 66.67 genomic copies per 100mL which is the lowest limit of detection for the wastewater assay.

Table 3. Summary of wastewater samples tested and reported COVID-19 cases per influent line catchment area and average SARS-COV-2 genome copies per 100mL of wastewater, April 2021 – February 2022

Influent Site	Total number of Samples Tested	Total Number of Cases	Average Genomic Copies per 100mL (min - max)
South Fulton	42	2,932	992.31 (66.67 – 8092.83)
Phillip Lee	43	8,593	4417.08 (573.11 – 17330.68)
Old Winn Dixie	45	8,038	5623.40 (104.89 – 21278.54)
Jonesboro	44	5,977	2391.45 (66.67 – 12456.69)
Intrenchment	40	6,595	3467.63 (66.67 – 11634.37)
Flint	43	8,159	6398.47 (66.67 – 20606.90)
Total	257	40,294	6,119.10 (66.67 – 21278.54)

Summary of results from table 2a. Samples from each site were collected weekly and quantified over the time period of April 2021 to February 2022. The total genomic copies represent the average across all influent lines for the study period, with the greatest maximum and lowest minimum.

Table 4. Correlation between reported COVID-19 cases by influent catchment area and SARS-CoV-2 genomic copies per 100mL wastewater in the influent line samples

Influent Site	No Lag (r)	One Week Lag (r)	Two Week Lag (r)
Flint	0.373	0.415	0.329
Intrenchment	0.483	0.804	0.582
Old Winn Dixie	0.273	0.607	0.472
Phillip Lee	0.349	0.693	0.530
Jonesboro	0.525	0.700	0.829
South Fulton	0.137	0.628	0.503

Table 3 shows the correlation between clinical case counts and genomic copies per 100mL. The lag is the correlation between these variables with wastewater concentrations delayed by one and two weeks.

Figure 2. Trend lines of clinical case counts and SARS-CoV-2 gene copies per 100mL

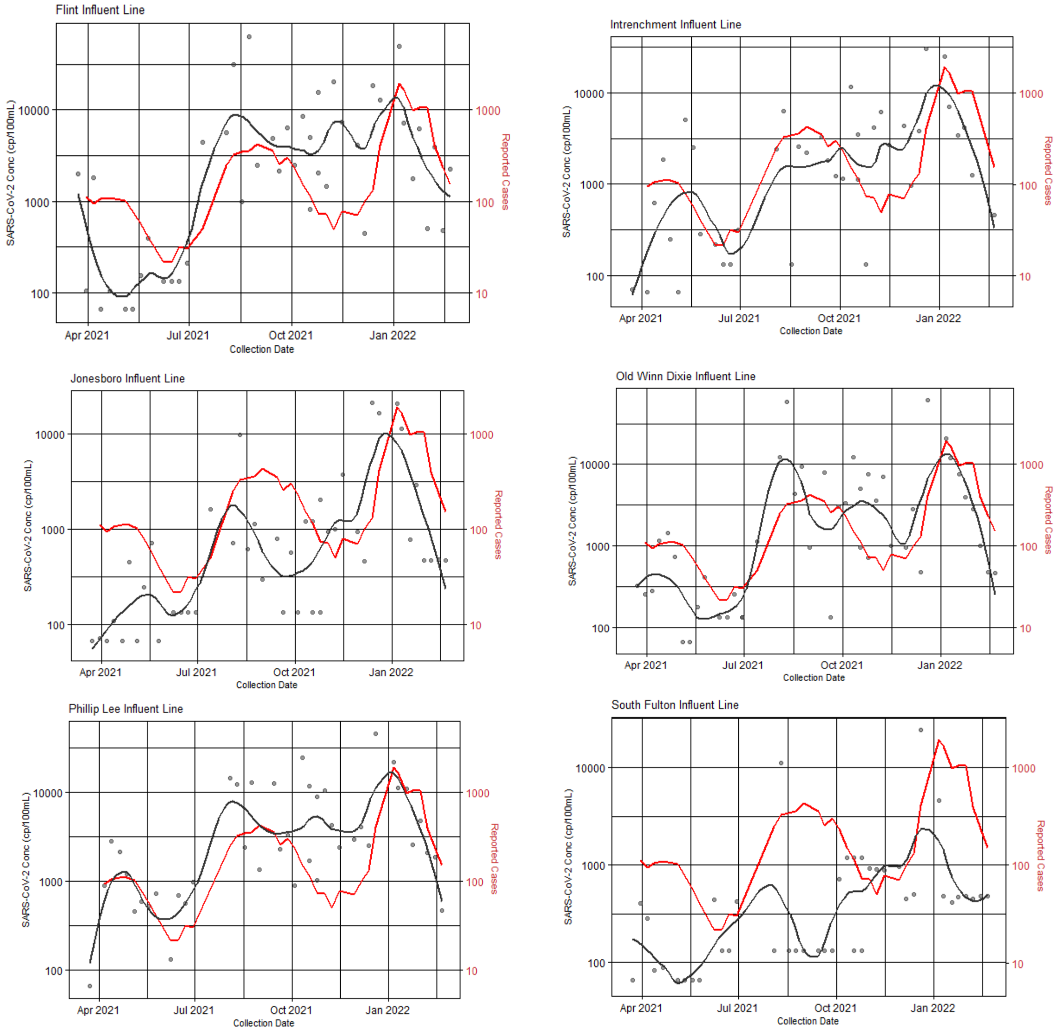


Figure 2 shows trend lines of both clinical case counts and genomic copies per 100mL of SARS-CoV-2 RNA. The black dots represent these concentrations, with a black trend line of best fit to illustrate the trends in these values. The red trend line represents changes in cases over time from April 2021 to February 2022.

Figure 3a-f. Linear regression of log10 viral concentrations (genomic copies/100mL) and clinical case counts by influent catchment area. Each figure represents RNA concentrations and case counts specific to each site. A unique formula for each site is used to estimate case counts from concentrations.

Figure 3a.

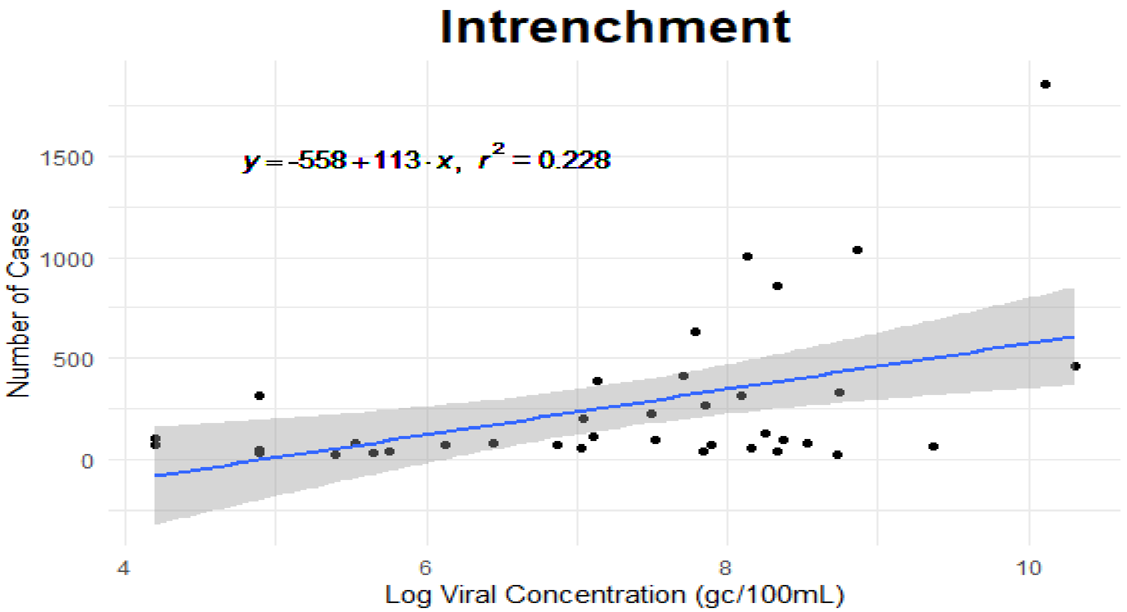


Figure 3b.

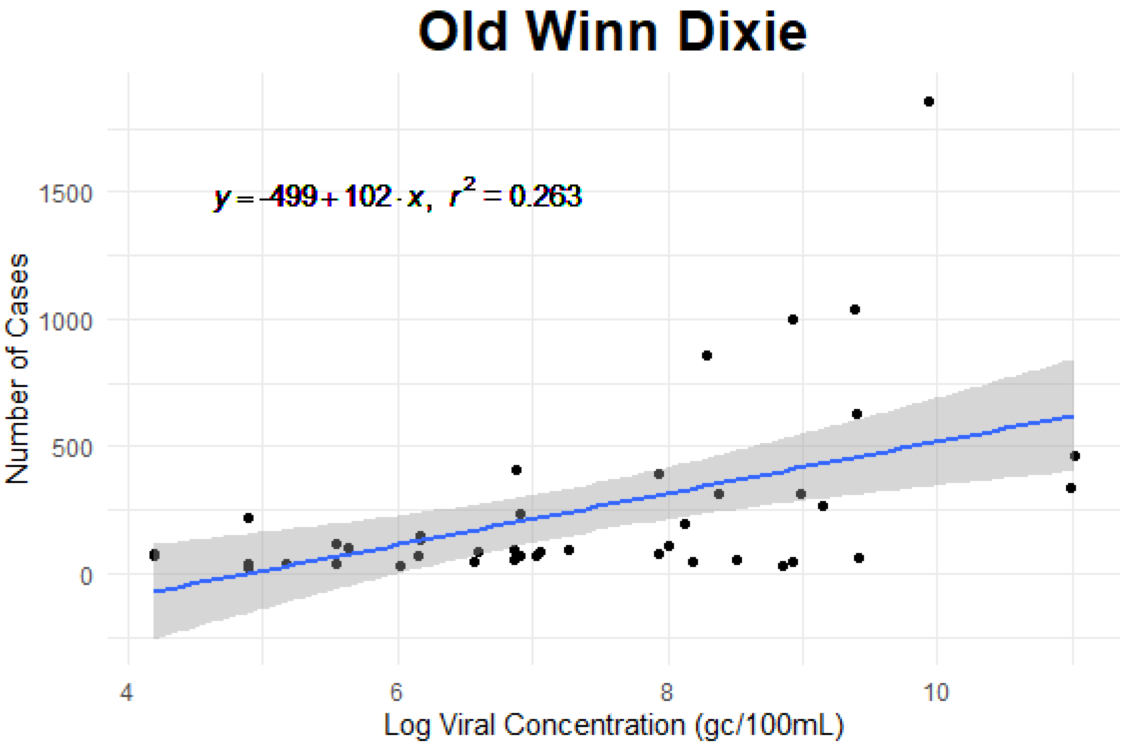


Figure 3c.

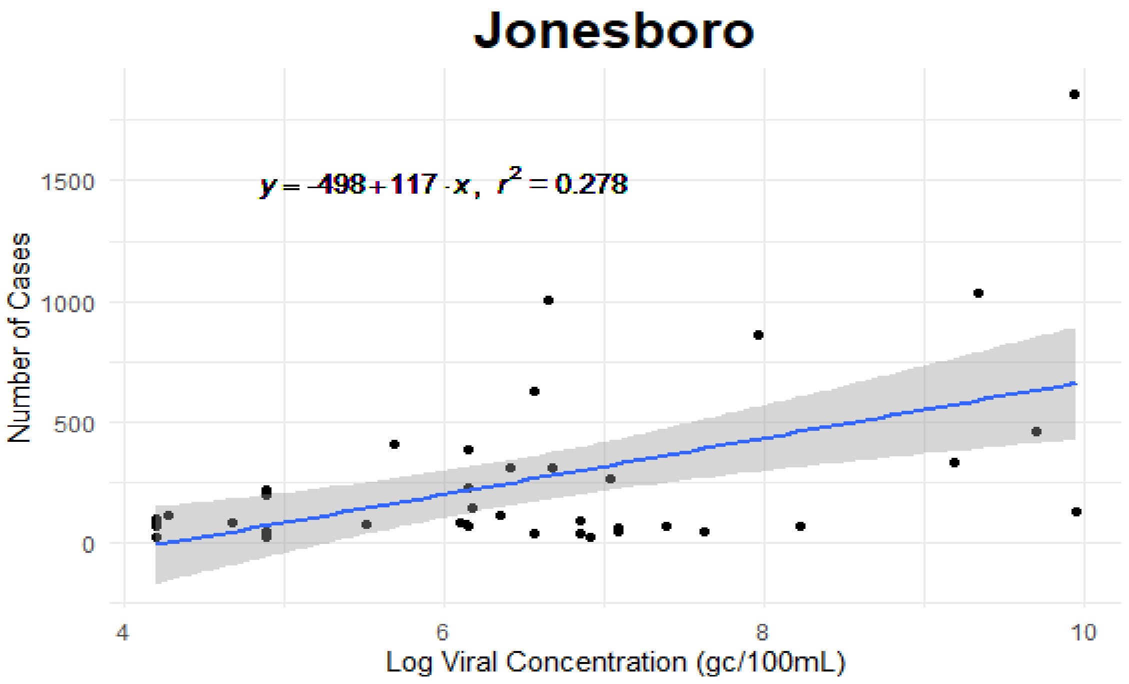

Jonesboro

Figure 3d.

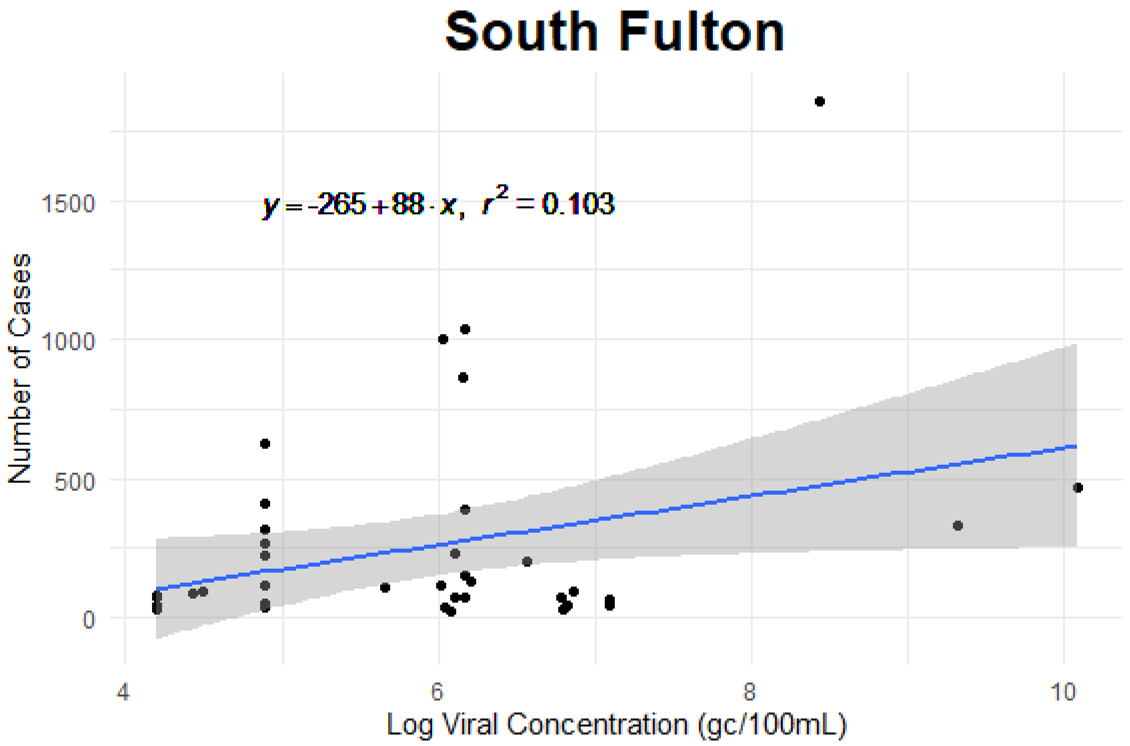

South Fulton

Figure 3e.

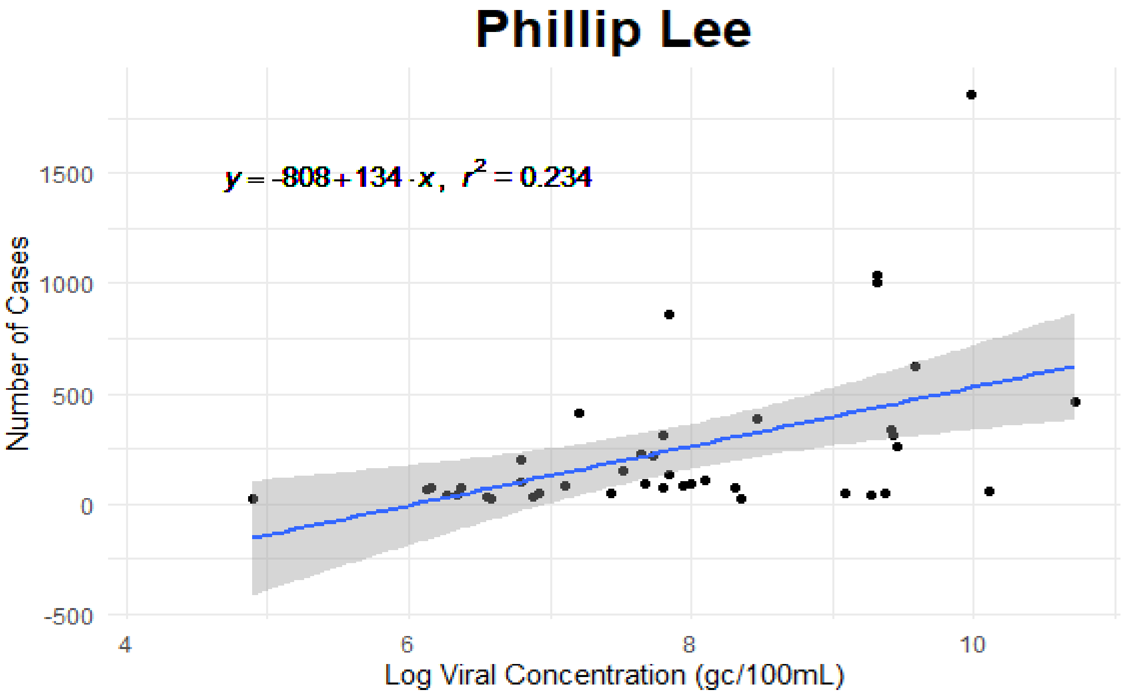
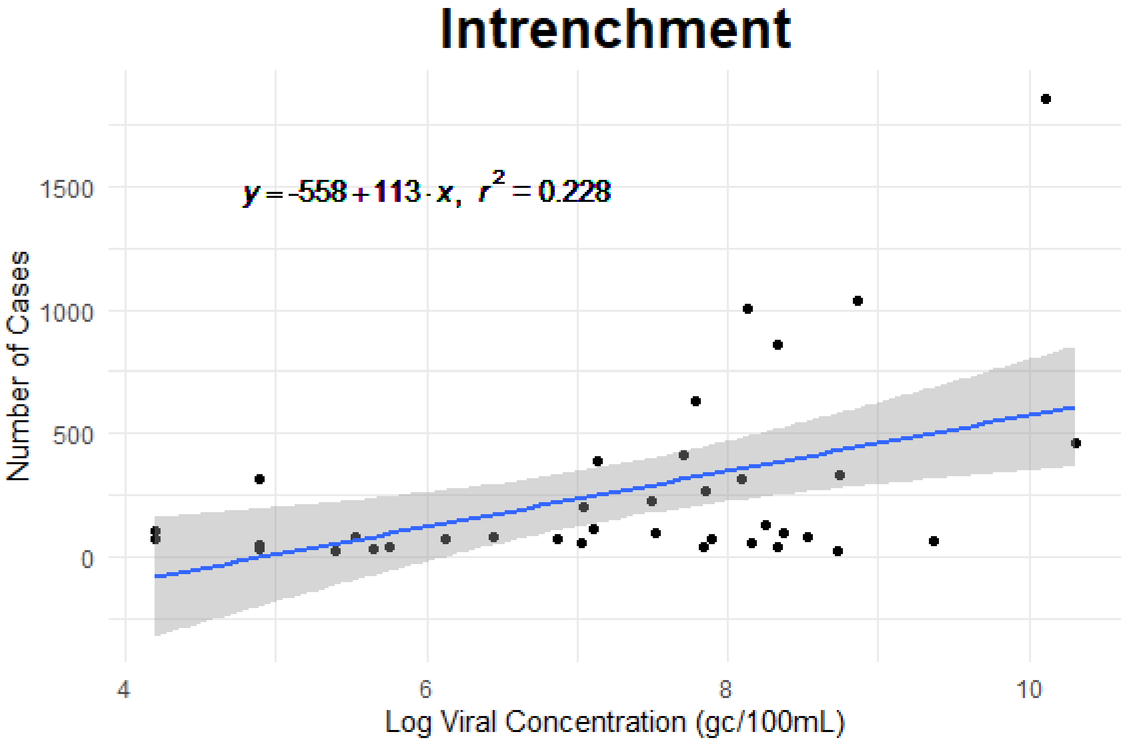


Figure 3f.



References

Analytics, B. (2020, September 12). Mining wastewater data to refine COVID-19 case estimation. *Medium*. https://biobotanalytics.medium.com/mining-wastewater-data-to-refine-covid-19-case-estimation-8dc58d67f706

Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis | PNAS. (n.d.). Retrieved April 20, 2022, from https://www.pnas.org/doi/10.1073/pnas.2109229118

Betancourt, W. W., Schmitz, B. W., Innes, G. K., Brown, K. M. P., Prasek, S. M., Stark, E. R., Foster, A. R., Sprissler, R. S., Harris, D. T., Sherchan, S. P., Gerba, C. P., & Pepper, I. L. (2020). *Wastewater-based Epidemiology for Averting COVID-19 Outbreaks on The University of Arizona Campus* (p. 2020.11.13.20231340). medRxiv. https://doi.org/10.1101/2020.11.13.20231340

Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A.-M., Abravanel, F., Izopet, J., Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.-L., & Henquell, C. (2018). Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015. *Eurosurveillance*, *23*(7), 17. https://doi.org/10.2807/1560-7917.ES.2018.23.7.17-00237

CDC COVID Data Tracker: Home. (n.d.). Retrieved April 20, 2022, from https://covid.cdc.gov/covid-data-tracker/#datatracker-home

COVID-19. (n.d.). Retrieved April 21, 2022, from https://www.fultoncountyga.gov/covid-19

Farkas, K., Hillary, L. S., Malham, S. K., McDonald, J. E., & Jones, D. L. (2020). Wastewater and public health: The potential of wastewater surveillance for monitoring COVID-19. *Current Opinion in Environmental Science & Health*, *17*, 14–20. https://doi.org/10.1016/j.coesh.2020.06.001

Highly predictive regression model of active cases of COVID-19 in a population by screening wastewater viral load | medRxiv. (n.d.). Retrieved April 20, 2022, from https://www.medrxiv.org/content/10.1101/2020.07.02.20144865v2.full

Iii, O. S., VanTassell, J., Raymond, J., marlene.wolfe, Liu, P., & Moe, C. (2022, January 21). *Nanotrap® KingFisher^{TM} Concentration/Extraction & MagMAX KingFisher^{TM} Extraction*. Protocols.Io. https://www.protocols.io/view/nanotrap-kingfisher-concentration-extraction-amp-m-b2nkqdcw

Kittipongvises, S., Phetrak, A., Hongprasith, N., & Lohwacharin, J. (2022). Unravelling capability of municipal wastewater treatment plant in Thailand for microplastics: Effects of seasonality on detection, fate and transport. *Journal of Environmental Management*, *302*, 113990. https://doi.org/10.1016/j.jenvman.2021.113990

Kumar, M., Jiang, G., Kumar Thakur, A., Chatterjee, S., Bhattacharya, T., Mohapatra, S., Chaminda, T., Kumar Tyagi, V., Vithanage, M., Bhattacharya, P., Nghiem, L. D., Sarkar, D., Sonne, C., & Mahlknecht, J. (2022). Lead time of early warning by wastewater surveillance for

COVID-19: Geographical variations and impacting factors. *Chemical Engineering Journal*, *441*, 135936. https://doi.org/10.1016/j.cej.2022.135936

Larsen, D. A., Green, H., Collins, M. B., & Kmush, B. L. (2021). Wastewater monitoring, surveillance and epidemiology: A review of terminology for a common understanding. *FEMS Microbes*, *2*, xtab011. https://doi.org/10.1093/femsmc/xtab011

Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J., & Khosrawipour, V. (2021). Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*, *27*(2), 110–115. https://doi.org/10.1016/j.pulmoe.2020.05.015

Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., & Yang, K. (2021). The role of seasonality in the spread of COVID-19 pandemic. *Environmental Research*, *195*, 110874. https://doi.org/10.1016/j.envres.2021.110874

Making waves: Defining the lead time of wastewater-based epidemiology for COVID-19—ScienceDirect. (n.d.). Retrieved April 20, 2022, from https://www.sciencedirect.com/science/article/pii/S004313542100631X?via%3Dihub

McCall, C., Wu, H., Miyani, B., & Xagoraraki, I. (2020). Identification of multiple potential viral diseases in a large urban center using wastewater surveillance. *Water Research*, *184*, 116160. https://doi.org/10.1016/j.watres.2020.116160

McMahan, C. S., Self, S., Rennert, L., Kalbaugh, C., Kriebel, D., Graves, D., Colby, C., Deaver, J. A., Popat, S. C., Karanfil, T., & Freedman, D. L. (2021). COVID-19 wastewater epidemiology: A model to estimate infected populations. *The Lancet Planetary Health*, *5*(12), e874–e881. https://doi.org/10.1016/S2542-5196(21)00230-8

Miura, F., Kitajima, M., & Omori, R. (2021). Duration of SARS-CoV-2 viral shedding in faeces as a parameter for wastewater-based epidemiology: Re-analysis of patient data using a shedding dynamics model. *The Science of the Total Environment*, *769*, 144549. https://doi.org/10.1016/j.scitotenv.2020.144549

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

Profile of Metro Atlanta: Metro Atlanta Demographics Overview. (n.d.). Retrieved April 20, 2022, from https://www.metroatlantachamber.com/resources/reports-and-information/executive-profile

Street, R., Mathee, A., Mangwana, N., Dias, S., Sharma, J. R., Ramharack, P., Louw, J., Reddy, T., Brocker, L., Surujlal-Naicker, S., Berkowitz, N., Malema, M. S., Nkambule, S., Webster, C., Mahlangeni, N., Gelderblom, H., Mdhluli, M., Gray, G., Muller, C., & Johnson, R. (2021). Spatial and Temporal Trends of SARS-CoV-2 RNA from Wastewater Treatment Plants over 6 Weeks in Cape Town, South Africa. *International Journal of Environmental Research and Public Health*, *18*(22), 12085. https://doi.org/10.3390/ijerph182212085

Su, M., Peng, S., Chen, L., Wang, B., Wang, Y., Fan, X., & Dong, Z. (2020). A Warm Summer is Unlikely to Stop Transmission of COVID-19 Naturally. *GeoHealth*, *4*(12), e2020GH000292. https://doi.org/10.1029/2020GH000292

Vallejo, J. A., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B. K., Nasser, M., Cid, Á., Veiga, M. C., Acevedo, A., Lamora, C., Bou, G., Cao, R., & Poza, M. (2020). *Predicting the number of people infected with SARS-COV-2 in a population using statistical models based on wastewater viral load* (p. 2020.07.02.20144865). medRxiv. https://doi.org/10.1101/2020.07.02.20144865

Vasquez Reyes, M. (2020). The Disproportional Impact of COVID-19 on African Americans. *Health and Human Rights*, *22*(2), 299–307.

Wang, M. L., Behrman, P., Dulin, A., Baskin, M. L., Buscemi, J., Alcaraz, K. I., Goldstein, C. M., Carson, T. L., Shen, M., & Fitzgibbon, M. (2020). Addressing inequities in COVID-19 morbidity and mortality: Research and policy recommendations. *Translational Behavioral Medicine*, *10*(3), 516–519. https://doi.org/10.1093/tbm/ibaa055

Wong, J. C. C., Tan, J., Lim, Y. X., Arivalan, S., Hapuarachchi, H. C., Mailepessov, D., Griffiths, J., Jayarajah, P., Setoh, Y. X., Tien, W. P., Low, S. L., Koo, C., Yenamandra, S. P., Kong, M., Lee, V. J. M., & Ng, L. C. (2021). Non-intrusive wastewater surveillance for monitoring of a residential building for COVID-19 cases. *Science of The Total Environment*, *786*, 147419. https://doi.org/10.1016/j.scitotenv.2021.147419

Xiao, A., Wu, F., Bushman, M., Zhang, J., Imakaev, M., Chai, P. R., Duvallet, C., Endo, N., Erickson, T. B., Armas, F., Arnold, B., Chen, H., Chandra, F., Ghaeli, N., Gu, X., Hanage, W. P., Lee, W. L., Matus, M., McElroy, K. A., … Alm, E. J. (2022). Metrics to relate COVID-19 wastewater data to clinical testing dynamics. *Water Research*, *212*, 118070. https://doi.org/10.1016/j.watres.2022.118070

Xu, G., Liu, B., Sun, Y., Du, Y., Snetselaar, L. G., Hu, F. B., & Bao, W. (2018). Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: Population based study. *The BMJ*, *362*, k1497. https://doi.org/10.1136/bmj.k1497