FORECASTING VIRAL AND BACTERIAL OUTBREAKS THROUGH ENVIRONMENTAL SURVEILLANCE

By

Liang Zhao

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Environmental Engineering – Doctor of Philosophy

2024

ABSTRACT

In the recent decades we have witnessed numerous outbreaks worldwide, resulting in millions of infections and deaths. Examples include the 1918 H1N1 virus, the 1968 H3N2 virus, the 2003 SARS coronavirus, the 2012 MERS-CoV, and the 2019 SARS-CoV-2. Factors including rapid population growth, escalating climate change crisis, recurring natural disasters, booming immigration and globalization, and concomitant sanitation and wastewater management challenges are anticipated to exacerbate the frequencies of disease outbreaks in the years to come. The traditional disease detection system primarily relies on the diagnostic analysis of specimens collected from infected individuals in clinical settings. This approach has significant limitations in predicting and providing early warnings for impending disease outbreaks. Infected individuals are often tested only after the development of symptoms, and health authorities are usually notified following the inception of a disease surge. Consequently, health authorities respond reactively instead of taking proactive measures during a pandemic. Additionally, clinical data collected by traditional disease surveillance systems often fail to accurately reflect actual infections in communities when asymptomatic infections are dominating, clinical testing is incapable to capture comprehensive infections, limitations in testing supplies and accessibility, and patients' testing behaviors. Environmental surveillance, especially wastewater surveillance or wastewater-based epidemiology, allows analyses of environmental community composite samples. Municipal wastewater samples are composite biological samples of an entire community that represent a snapshot of the disease burden of the population covered by the corresponding sewer-shed. Collecting and analyzing untreated wastewater samples from centralized wastewater treatment plants and neighborhood manholes for specific viral and bacterial targets at a regular cadence can reveal the trends of pathogen concentrations in wastewater. These trends represent the viral and

bacterial loads shed by infected individuals, whether they are symptomatic or asymptomatic. Based on measured wastewater concentrations of disease pathogens and other available datasets such as clinical and demographic datasets, researchers can establish models to predict disease incidences before clinical reporting and develop tools to provide early warnings of upcoming surges of diseases. This crucial information can help public health officials in making informed decisions regarding the implementation of preparedness measures and the allocation of resources. The primary objective of this dissertation is to develop comprehensive laboratorial, technological, and translational methodologies for forecasting viral and bacterial outbreaks through wastewater-based epidemiology.

I would like to dedicate this dissertation to my parents and my wife.
Their love and belief in me have been the cornerstone of my life.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Irene Xagoraraki, for her constant support, care, help, and guidance throughout my studies, research, and career development. Dr. Xagoraraki introduced me to the world of science and research, opened the door of possibilities for me to explore, fostered my growth as a junior researcher, inspired me to overcome obstacles, and played a crucial and indispensable role in my academic journey. I am forever indebted and grateful for her guidance and contributions to my development. I would like to thank my committee members, Dr. Alison Cupples, Dr. Syed Hashsham, Dr. Rene Hinojosa, for their kind support, guidance, availability, and feedback. I also would like to thank the Detroit project partnership, Dr. Russell Faust, Dr. John Norton, Dr. Randy David, Dr. Anna Mehrotra, Dr. Johnathon Sheets, for their kind support and invaluable advice. Additionally, many thanks to my lab mates Dr. Yangyang Zou, Dr. Yabing Li, Dr. Heidy Peidro Guzman, Brijen Miyani, Zachary Gentry, Maddie Spooner, Sydney Jacobi, Reegan Kelly, Lake Van Natter, for their kind support, friendship, and collaborations. I also want to extend my sincere gratitude to the faculty and staff in the Department of Civil and Environmental Engineering for their help whenever it was needed. Finally, I would like to thank my family for their love, support, and encouragement, which have always been fundamental and inspirational at each step in my journey. To my parents, thank you for your unwavering love, for encouraging me to pursue my dreams, for trusting in my abilities, and for empowering me to strive for the best. To my wife, Yingying, thank you for encouraging me to embrace challenges, inspiring me to learn and grow during difficult times, and being the source of unlimited strength and motivation all time.

TABLE OF CONTENTS

INTRODUCTION

Viral and bacterial outbreaks resulted in devastating and uncontrollable negative impacts on human lives and development globally over the past decades. The recent COVID-19 pandemic is a striking example. At the same time, the past few decades have also experienced unprecedented societal, demographic, and climatic changes. These include escalating movements of populations via flights, railways, automobiles, and maritime transportations; an increasing influx of populations into already densely populated urban areas; as well as drastically changing climatic conditions. These changes collectively contribute to the increasing risk of infectious disease outbreaks on a global scale.

Traditional disease surveillance system operated by health authorities depends on clinical examinations of symptomatic individuals who seek healthcare services. Infected individuals are typically tested after the onset of symptoms, and health authorities are notified subsequently. This process results in relays between exposure to the disease pathogen and symptoms onset, as well as between symptoms onset and confirmation by a diagnostic test. These delays can further result in the disease becoming widespread in communities before any intervention measures could be implemented. A prolonged incubation time of some diseases can also exacerbate widespread transmissions. Therefore, the traditional disease surveillance system based on clinical data collection often fail to detect infected individuals before their symptoms onset, and the system is incapable of providing early warnings of disease outbreaks (Xagoraraki & O'Brien, 2020).

The traditional disease surveillance system also often relies on voluntary testing and willingness of testing by infected individuals, leading to biased dataset of disease prevalence. Infected individuals with mild or no symptoms are less likely to undergo a diagnostic test in clinical settings. The capacity of clinical testing can also be inadequate, especially during a surge of infections or in areas with limited testing resources, resulting in a significant portion of

underreported infections (Wang et al., 2022). Lastly, conducting population-wide screening by clinical testing can impose a heavy financial burden on communities (Wang et al., 2022). Overall, factors including biased testing populations, underreported asymptomatic populations, limited testing capability, undetermined testing behaviors, collectively have an enormous impact on the accuracy and inclusiveness of clinical datasets (Safford et al., 2022).

Environmental surveillance, especially wastewater surveillance or wastewater-based epidemiology (WBE), can provide an inclusive snapshot of disease prevalence in communities and exhibit numerous advantages comparing to the traditional disease surveillance system. First, WBE can help health authorities predict and provide early warnings of surging cases. For instance, Chapters 1 and 2 in this dissertation demonstrate both advanced statistical models to predict COVID-19 cases and simple statistical tools to provide early warnings of impending surges of COVID-19 cases, respectively. Researchers also developed models to predict COVID-19 cases based on WBE datasets, such as artificial neural network models (Galani et al., 2022), vector autoregression models (Cao & Francis, 2021), and automatic regression integrated moving average models (Matheri et al., 2022). Particularly, Bibby et al., indicated that WBE's early warning of COVID-19 cases were prominent in locations with limited clinical testing capacity, or significant delays in clinical results reporting, or where prevalent asymptomatic infections exhibit (Bibby et al., 2021). Second, WBE can provide disease infections data for communities with limited healthcare services. In socioeconomically disadvantaged communities, healthcare service access is often limited, therefore, leading to inaccuracy of clinical datasets collected in these areas. Additionally, in most rural areas, healthcare services are more limited than in urban areas, and public health authorities in rural areas have less reliable disease dataset for decision makings comparing to their urban counterparts (Cohen et al., 2024). Therefore, potential public health benefits of WBE in these areas are substantial. Third, WBE can complement clinical data while

asymptomatic infections are dominating, or clinical testing is delayed, or clinical testing is insufficient to capture comprehensive infections (Mac Mahon et al., 2022). For instance, during the initial stage of the COVID-19 pandemic, researchers witnessed extremely high numbers of undetected and underreported cases and indicated that the actual infections were much higher than reported cases (Lau et al., 2021). Other researchers reported the case-to-report ratio for COVID-19 as 26 to 32 in India, and the value was even higher (82 to 130) at the early stage when testing was less available (Murhekar et al., 2021). During these periods, WBE is particularly effective and beneficial as a disease surveillance tool since clinical testing is time and resource restricted, and inevitably delayed for identifying and tracking underreported cases. A recent study also indicated that WBE can be beneficial when severe cases are rare but asymptomatic infections persist during the final phase of disease eradication (Daughton, 2020). Finally, WBE can also help health authorities save financial resources to circumvent massive clinical testing in populated regions. Researchers indicated that strategically replacing some clinical testing with WBE could save financial resources without compromising surveillance accuracy (Safford et al., 2022; Wang et al., 2022).

WBE was originally implemented to monitor illicit drugs and later was recognized worldwide as an epidemiological tool to gauge infections of viral and bacterial diseases in communities. For human viruses, they do not replicate outside of a host and can remain stable in the environment for significant periods. Thus, WBE can provide a comprehensive overview of the viral disease burden in communities. For human bacteria, the WBE workflow used for viruses may not be effective. Specific methodologies and workflow need to be designed, developed, and implemented for targeting bacterial species in wastewater. For instance, Chapter 6 in this dissertation describes one of the first bacterial wastewater surveillance studies on monitoring *Chlamydia trachomatis* and *Treponema pallidum* in Detroit's wastewater and demonstrates

advantages of WBE in monitoring bacterial species such as back-estimating infections and complementing clinical data.

In December 2019, Coronavirus Disease 2019 (COVID-19) was first identified in Wuhan, China. The disease is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a positive-sense single-stranded RNA virus that is highly contagious in humans. COVID-19 rapidly spread worldwide, resulting in a pandemic, which was declared by the World Health Organization (WHO) on March 11th, 2020. In response, WBE was quickly implemented to monitor concentrations of SARS-CoV-2 in wastewater globally, including in countries such as the United States, the Netherlands, Japan, Australia, among others (Ahmed et al., 2020; Haramoto et al., 2020; Miyani et al., 2020; Sherchan et al., 2020). This dissertation comprises six chapters that introduce significant advancements in WBE with a focus on the Tri-County Detroit Area (TCDA) in Michigan, U.S., and explore the development and implementation of laboratorial, technical, and translational methodologies for monitoring and forecasting viral and bacterial disease outbreaks, including SARS-CoV-2 and sexually transmitted infections. The first five chapters primarily focus on advancements of WBE methodologies on monitoring SARS-CoV-2 and the sixth chapter focuses on expanding WBE applications on monitoring sexually transmitted infections caused by bacterial pathogens, including Chlamydia and Syphilis. Briefly, chapter 1 introduces advanced statistical models based on wastewater surveillance datasets that can identify and predict COVID-19 peaks in clinical cases in the TCDA five weeks prior to peaks of reported clinical data. Disease characteristics such as incubation time, shedding duration, and shedding dynamics were incorporated in a mechanistic model to estimate the time lag between measured viral concentrations in wastewater and reported clinical data. Chapter 2 introduces simple statistical methodologies to determine early warnings of COVID-19 surges that are used for informed decision making by public health officials. This chapter also introduces statistical methodologies

to determine peaks of COVID-19 cases and approaches to evaluate the proposed early warning methods. Chapter 3 compares time series data of wastewater measurements and clinical cases through statistical tools to optimize early warning potential of three commonly used wastewater concentration methods. This chapter identifies the optimal wastewater concentration method to provide early warnings of viral diseases. Chapter 4 demonstrates time lags between SARS-CoV-2 concentrations in wastewater and diverse clinical metrics through time lagged cross correlation methods. This chapter identifies dynamically changing time lags and various parameters affecting time lags. Chapter 5 introduces experimental and statistical methodologies to compare three commonly applied U.S. Centers for Disease Control and Prevention (CDC) assays targeting SARS-CoV-2 in wastewater, including N1, N2 and SC2 assays. Through comparative analyses, this chapter identifies the optimal assay for testing SARS-CoV-2 in wastewater to predict COVID-19 cases in the TCDA. Chapter 6 explores bacterial wastewater surveillance to monitor widespread sexually transmitted infections, including Chlamydia and Syphilis, through developing a wastewater surveillance workflow and optimizing laboratory molecular biology methods. This chapter also demonstrates approaches to back-estimate infections based on WBE datasets.

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of the Total Environment*, *728*, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Bibby, K., Bivins, A., Wu, Z., & North, D. (2021). Making waves: Plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research*, *202*(July), 117438. https://doi.org/10.1016/j.watres.2021.117438

Cao, Y., & Francis, R. (2021). On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Science of The Total Environment*, *786*, 147451. https://doi.org/10.1016/j.scitotenv.2021.147451

Cohen, A., Vikesland, P., Pruden, A., Krometis, L.-A., Lee, L. M., Darling, A., Yancey, M., Helmick, M., Singh, R., Gonzalez, R., Meit, M., Degen, M., & Taniuchi, M. (2024). Making waves: The benefits and challenges of responsibly implementing wastewater-based surveillance for rural communities. *Water Research*, *250*, 121095. https://doi.org/10.1016/j.watres.2023.121095

Daughton, C. G. (2020). Wastewater surveillance for population-wide Covid-19: The present and future. *Science of The Total Environment*, *736*, 139631. https://doi.org/10.1016/j.scitotenv.2020.139631

Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Adamopoulos, P. G., Peccia, J., Thompson, D. C., Kontou, A., Karagiannidis, A., Lianidou, E. S., Avgeris, M., Paraskevis, D., Tsiodras, S., Scorilas, A., Vasiliou, V., Dimopoulos, M.-A., & Thomaidis, N. S. (2022). SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Science of The Total Environment*, *804*, 150151. https://doi.org/10.1016/j.scitotenv.2021.150151

Haramoto, E., Malla, B., Thakali, O., & Kitajima, M. (2020). First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan. *Science of The Total Environment*, *737*, 140405. https://doi.org/10.1016/j.scitotenv.2020.140405

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

Mac Mahon, J., Criado Monleon, A. J., Gill, L. W., O'Sullivan, J. J., & Meijer, W. G. (2022). Wastewater-based epidemiology (WBE) for SARS-CoV-2 – A review focussing on the significance of the sewer network using a Dublin city catchment case study. *Water Science and Technology*, *86*(6), 1402–1425. https://doi.org/10.2166/wst.2022.278

Matheri, A. N., Belaid, M., Njenga, C. K., & Ngila, J. C. (2022). Water and wastewater digital

surveillance for monitoring and early detection of the COVID-19 hotspot: industry 4.0. *International Journal of Environmental Science and Technology*. https://doi.org/10.1007/s13762-022-03982-7

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Murhekar, M. v, Bhatnagar, T., Selvaraju, S., Saravanakumar, V., Thangaraj, J. W. V., Shah, N., Kumar, M. S., Rade, K., Sabarinathan, R., Asthana, S., Balachandar, R., Bangar, S. D., Bansal, A. K., Bhat, J., Chopra, V., Das, D., Deb, A. K., Devi, K. R., Dwivedi, G. R., … Zaman, K. (2021). SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey. *The Lancet Global Health*, *9*(3), e257–e266. https://doi.org/10.1016/S2214-109X(20)30544-1

Safford, H. R. , Karen, S., & Heather, B. N. (2022). Wastewater analysis can be a powerful public health tool—if it's done sensibly. *Proceedings of the National Academy of Sciences*, *119*(6), e2119600119. https://doi.org/10.1073/pnas.2119600119

Sherchan, S. P., Shahin, S., Ward, L. M., Tandukar, S., Aw, T. G., Schmitz, B., Ahmed, W., & Kitajima, M. (2020). First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA. *Science of The Total Environment*, *743*, 140621. https://doi.org/10.1016/j.scitotenv.2020.140621

Wang, Y., Liu, P., Zhang, H., Ibaraki, M., VanTassell, J., Geith, K., Cavallo, M., Kann, R., Saber, L., Kraft, C. S., Lane, M., Shartar, S., & Moe, C. (2022). Early warning of a COVID-19 surge on a university campus based on wastewater surveillance for SARS-CoV-2 at residence halls. *Science of The Total Environment*, *821*, 153291. https://doi.org/10.1016/j.scitotenv.2022.153291

Xagoraraki, I., & O'Brien, E. (2020). Wastewater-Based Epidemiology for Early Detection of Viral Outbreaks. In D. J. O'Bannon (Ed.), *Women in Water Quality: Investigations by Prominent Female Engineers* (pp. 75–97). Springer International Publishing. https://doi.org/10.1007/978-3-030-17819-2_5

CHAPTER 1: FIVE-WEEK WARNING OF COVID-19 PEAKS PRIOR TO THE OMICRON SURGE IN DETROIT, MICHIGAN USING WASTEWATER SURVEILLANCE

Published in *Science of the Total Environment*:

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, *844*, 157040.

Abstract

Wastewater-based epidemiology (WBE) is useful in predicting temporal fluctuations of COVID-19 incidence in communities and providing early warnings of pending outbreaks. To investigate the relationship between SARS-CoV-2 concentrations in wastewater and COVID-19 incidence in communities, a 12-month study between September 1, 2020, and August 31, 2021, prior to the Omicron surge, was conducted. 407 untreated wastewater samples were collected from the Great Lakes Water Authority (GLWA) in southeastern Michigan. N1 and N2 genes of SARS-CoV-2 were quantified using RT-ddPCR. Daily confirmed COVID-19 cases for the City of Detroit, and Wayne, Macomb, Oakland counties between September 1, 2020, and October 4, 2021, were collected from a public data source. The total concentrations of N1 and N2 genes ranged from 714.85 to 7145.98 gc/L and 820.47 to 6219.05 gc/L, respectively, which were strongly correlated with the 7-day moving average of total daily COVID-19 cases in the associated areas, after 5 weeks of the viral measurement. The results indicate a potential 5-week lag time of wastewater surveillance preceding COVID-19 incidence for the Detroit metropolitan area. Four

statistical models were established to analyze the relationship between SARS-CoV-2 concentrations in wastewater and COVID-19 incidence in the study areas. Under a 5-week lag time scenario with both N1 and N2 genes, the autoregression model with seasonal patterns and vector autoregression model were more effective in predicting COVID-19 cases during the study period. To investigate the impact of flow parameters on the correlation, the original N1 and N2 gene concentrations were normalized by wastewater flow parameters. The statistical results indicated the optimum models were consistent for both normalized and non-normalized data. In addition, we discussed parameters that explain the observed lag time. Furthermore, we evaluated the impact of the omicron surge that followed, and the impact of different sampling methods on the estimation of lag time.

1. Introduction

Wastewater-based epidemiology (WBE) for the prediction of viral outbreaks was proposed in 2019 and 2020 (Xagoraraki and O'Brien, 2019; O'Brien & Xagoraraki, 2019; Xagoraraki, 2020), and has been applied for the early detection of COVID-19 (Ahmed et al., 2020a, 2021a; Miyani et al., 2020). One of the critical utilities of WBE is the possibility to forecast upcoming fluctuations of disease with a lag time. The lag time in our study is defined as the lag between peaks in measured concentrations of SARS-CoV-2 in wastewater and peaks in reported COVID-19 cases based on clinical testing. Lag times observed in published studies since the inception of COVID-19 pandemic (summarized in Table 1. 1.) vary widely between 2 days and 28 days. The observed lag times may depend on multiple parameters. These parameters include disease characteristics of SARS-CoV-2 such as incubation time and shedding duration summarized in Tables 1S. 1 and 1S. 2, respectively, that may change with novel variants (Yaniv et al., 2021). The parameters that affect observed lag times may also involve contributing populations and their

demographic characteristics, including age (Omori et al., 2021), sex (Syangtan et al., 2021), racial ancestry (Allan-Blitz et al., 2021; Feehan et al., 2021), and traveling history of infected populations (Xiao et al., 2020). Furthermore, the hydraulic influence in the sewage network, including dilution events (Foladori et al., 2020), and sorption and desorption of the virus in wastewater (Yin et al., 2018) as well as the methods of wastewater sampling (that may be include viruses sorbed on particles or supernatant viruses) are critically affecting the observed lag times. Moreover, manners of reporting the clinical data, including the accessibility to the testing (Wiens et al., 2021), and traveling time to the testing sites (Rader et al., 2020) are important.

Some of the factors contributing to the lag time between wastewater-based data peaks and clinical data peaks are visualized in the timeline shown in Figure 1. 1. The majority of published studies show empirical evidence that the range of the incubation time of SARS-CoV-2 prior to the Omicron surge was 0 to 14 days (shown in Table 1S. 1. and summarized in Figure 1. 1.). Clinical testing is often delayed from the manifestation of clinical symptoms by several days, due to limited availability of testing supplies, limited ability to reach testing sites, or resistance of people to seek testing (Rader, 2020; Torres et al., 2021; Wiens et al., 2021). In some cases, the mean delay in the reporting of confirmed COVID-19 cases is 5 days with 15% of cases reported after day 10 (Harris, 2020). Some clinical studies have demonstrated a delay of approximately 7 days from illness onset to clinical testing (Huang et al., 2020). Henceforth, the delay in gathering clinical data could vary significantly, particularly in demographically and socioeconomically varied populations, like that of Detroit, Michigan. Generally, between day 19 and 25, clinical data will become publicly available, after an estimated delay of 3 to 9 days of clinical data collection and processing time (Garg et al., 2020; Harris, 2020; Rader, 2020). Additionally, Figure 1. 1. demonstrates the temporal progress of data collection of viral loadings in wastewater. The detention time of wastewater in the collection network is estimated as 12 to 24 hours (Table 1S. 4.). In Figure 1. 1. it is assumed

that in most cases, wastewater laboratory tests are completed within a day upon sample collection. A compilation of all the above-mentioned timelines indicates that the lag time may be estimated to be between 3 and 4 weeks (Figure 1. 1.). This is expected to vary with different variants and different sampling methods.

Here we present a twelve-month consecutive study, prior to the omicron surge, using N1 and N2 gene RT-ddPCR to monitor SARS-CoV-2 concentrations in untreated wastewater samples collected from the WRRF of GLWA that serves the city of Detroit, as well as Wayne, Macomb, and Oakland counties in Michigan. We applied a sampling method that captured suspended viruses in the supernatant of wastewater to circumvent the potential input of "old" viruses via desorption of settled viruses during high flows. We investigated lag time through statistical analyses and established four models to predict COVID-19 clinical cases using normalized and non-normalized SARS-CoV-2 concentrations in wastewater. The performance of each model is evaluated using the Root Mean Square Error (RSME) and Pearson correlation between the predicted case number and actual clinical case number. Future incidences of COVID-19 were predicted based on SARS-CoV-2 concentrations in wastewater during the study period, using statistical models. In addition, we examined the influence of the omicron surge to the early waring lag time. We also performed another widely applied sampling method for comparison purposes to demonstrate the benefits of our current method in terms of providing early warning for COVID-19 incidences in Detroit.

Table 1. 1. Lag time in published studies prior to Omicron surge

Location	Sample type	Lag time	Test method	References
Milan & Rome, Italy	wastewater	within a few days	RT-qPCR	(La Rosa et al., 2020)
Ottawa, Canada	wastewater	2 days	RT-qPCR	(D'Aoust et al., 2021)
Montana, USA	wastewater	2-4 days	RT-qPCR	(Nemudryi et al., 2020)
Wisconsin, USA	wastewater	0-6 days	RT-qPCR	(Feng et al., 2021)
New York, USA	wastewater	3 days	RT-qPCR	(Larsen et al., 2021)
New Haven, Connecticut, USA	sludge	0-2, 1-4, 6-8 days under given scenarios	RT-qPCR	(Peccia et al., 2020)
Charlotte, North Carolina	wastewater	5-12 days	RT-qPCR RT-ddPCR	(Barua et al., 2022)
Gandhinagar, Gujarat, India	wastewater	7-14 days	RT-PCR	(Kumar et al., 2021)
Paris, France	wastewater	8 days	RT-qPCR	(Wurtzer et al., 2020)
Minnesota, USA	wastewater	statewide: 15-17 days, regional level: 4-20 days	RT-qPCR	(Melvin, 2021)
Massachusetts, USA	wastewater	4-10 days	RT-qPCR	(Wu et al., 2022)
Netherlands	wastewater	4 days	RT-qPCR	(Lodder & de Roda Husman, 2020)
Netherlands	sewage samples	6 days	RT-qPCR	(Medema et al., 2020)
Utah, USA	wastewater	7 days	RT-qPCR	(Weidhaas et al., 2021)
Milan Metropolitan Area, Italy,	raw and treated wastewater	8 days	RT-qPCR	(Rimoldi et al., 2020)
Spain	wastewater	12-16 days	RT-qPCR	(Randazzo et al., 2020)
Australia	wastewater	21 days	RT-qPCR	(Ahmed et al., 2021a)
Gothenburg, Sweden	wastewater	19-21 days	RT-qPCR	(Saguti et al., 2021)
Australia	wastewater	28 days	RT-qPCR	(Ahmed et al., 2020a)

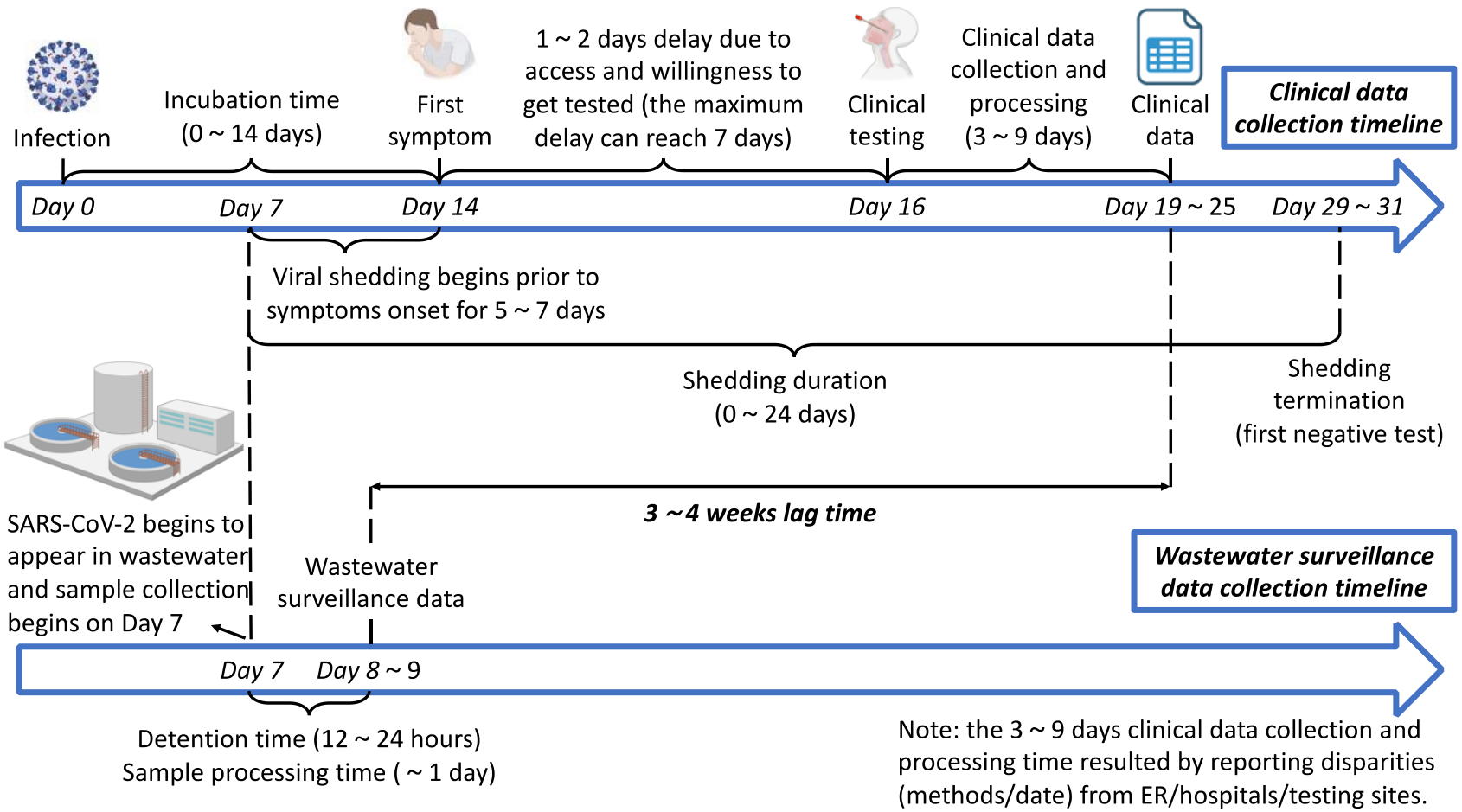


Figure 1. 1. Time scale of clinical data collection and wastewater surveillance
(incubation time and shedding duration are summarized in Tables 1S. 1. and 1S. 2., respectively)

2. Materials and Methods

2.1 Wastewater treatment plant and sample collection

Untreated wastewater samples were collected from the WRRF of GLWA located in southeastern Michigan between September 1, 2020, and August 31, 2021, prior to the Omicron surge. The WRRF in Detroit is the largest single-site wastewater treatment plant in the U.S. with a primary treatment capacity of 1,700 million gallons per day (MGD) and a secondary treatment capacity of 930 MGD. GLWA's WRRF has a semi-combined sewer-shed system, which collects and treats stormwater along with residential, industrial, and commercial waste, depending on service areas. The WRRF serves the three most populous Michigan counties: Wayne, Oakland, and Macomb. Figure 1. 2. shows all ZIP codes captured by the three main interceptors that discharge into the WRRF. The WRRF receives wastewater via three main interceptors including the Detroit River Interceptor (DRI), the North Interceptor-East Arm (NIEA), and Oakwood-Northwest-Wayne County Interceptor (ONWI) from its service areas which are shown in Figure 1. 3. The samples were collected from all three interceptors at the point of discharge into the WRRF.

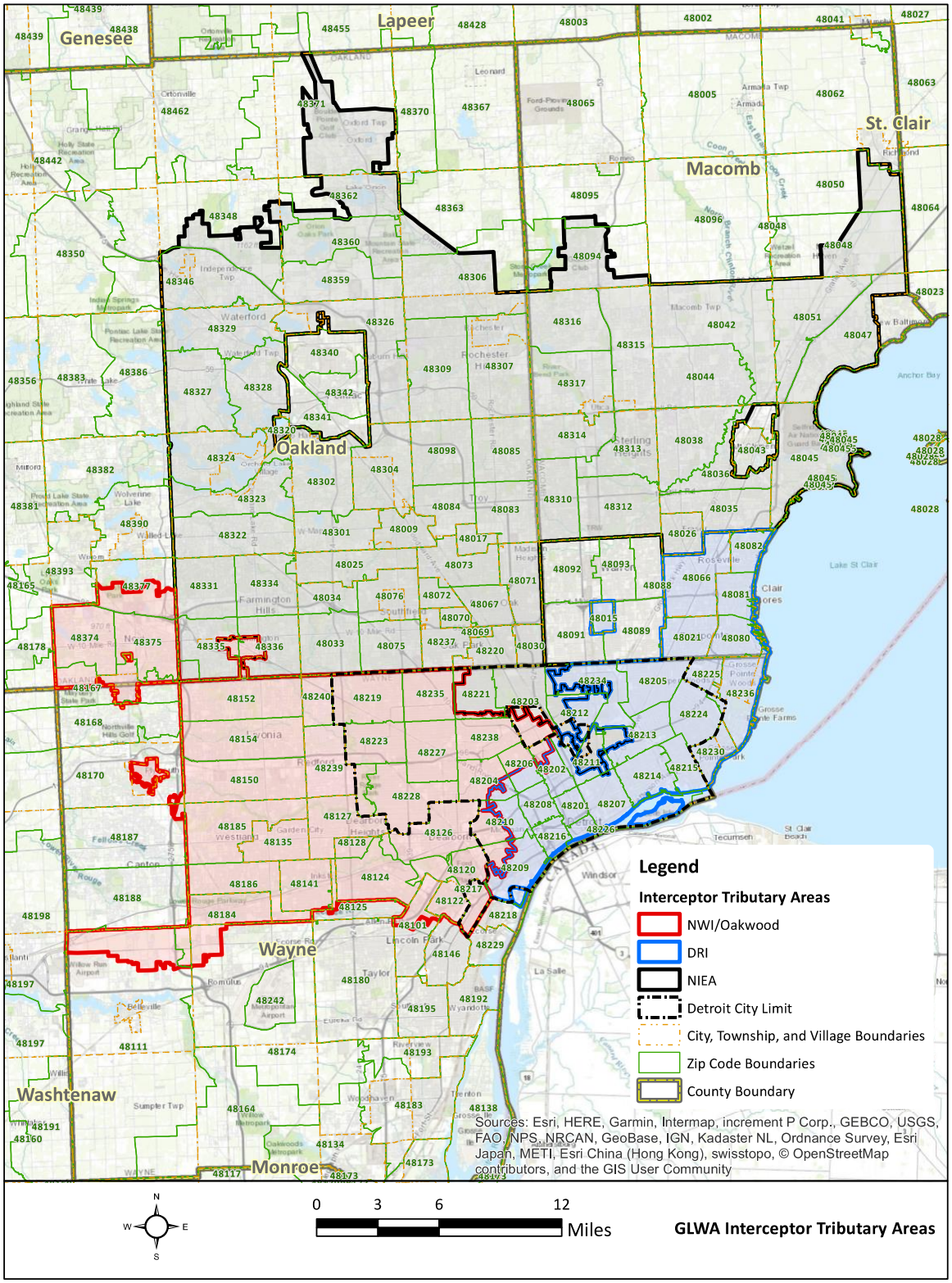


Figure 1. 2. GLWA WRRF tributary areas

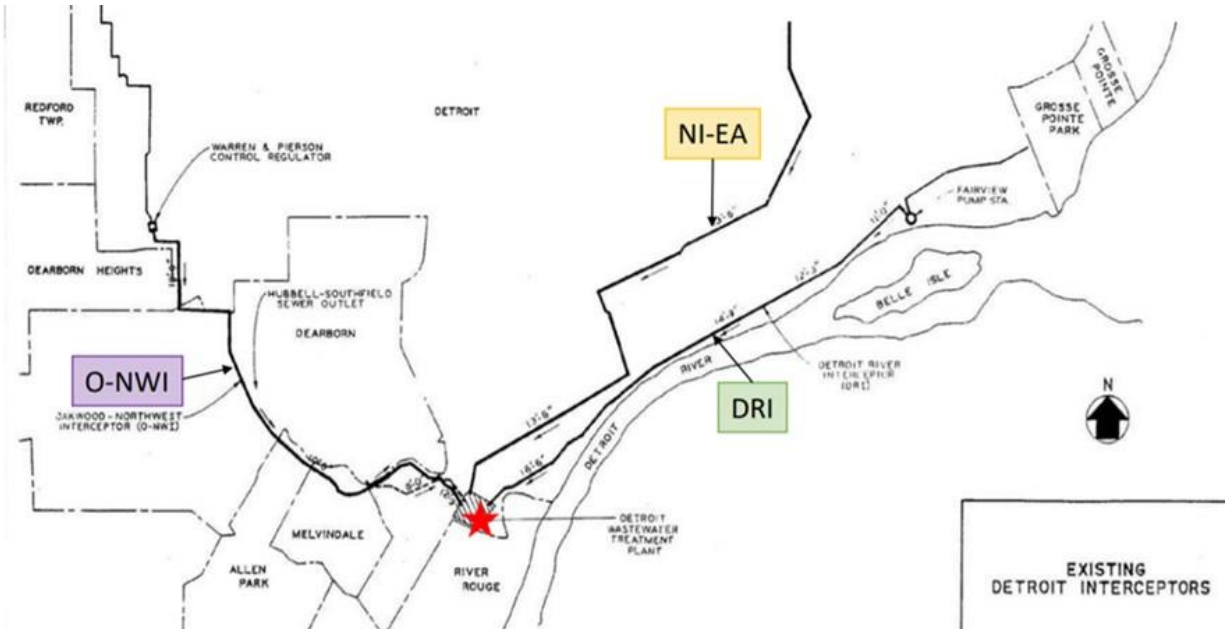


Figure 1. 3. Locations of the GLWA WRRF three interceptors

Estimated populations served by each interceptor, daily flows, and other characteristics of the three interceptors, between September 2020 and August 2021, are shown in Table 1S. 26. Sampling occurred weekly between September 1, 2020, and August 30, 2021. A total of 407 untreated wastewater samples were collected at the influent of the WRRF, including 146, 117, and 144 samples from ONWI, NIEA, and DRI, respectively. In addition, to evaluate the effect of omicron surge in the estimated lag time, we performed testing between August 1, 2021, and February 28, 2022, using the same methods, with a total of 249 untreated wastewater samples from the same sites.

2.2 Sampling methods

Viruses were collected and isolated from wastewater using electropositive NanoCeram column filters (Argonide, Sanford, FL) based on the EPA Virus Adsorption-Elution (VIRADEL) method (Xagoraraki et al., 2014; Miyani et al., 2021b). Depending on the suspended solids of wastewater, approximately 10 to 50 L of raw wastewater passed through NanoCeram

electropositive cartridge filters at a rate not more than 11 L/min using a previously described method (USEPA 2001; USEPA 2014; Miyani et al., 2021b). Flow meter readings were recorded at the inception and termination of each sampling event. After sampling, all NanoCeram column filters were placed in sealed plastic bags, on ice, and transported to the laboratory within 24 hours for downstream analysis.

In addition to the VIRADEL method, for method comparison purposes, 1L of 24-hr composite samples were collected weekly between August 1, 2021, and February 28, 2022, to conduct polyethylene glycol precipitation (PEG) for the virus concentration (Ahmed et al., 2020b, 2020c; D'Aoust et al., 2021; Kaya et al., 2022).

2.3 Virus elution, RNA extraction, RT-ddPCR, and variants testing

Viruses were eluted within 24 hours after sampling based on a previously described method (Miyani et al., 2021b). Bacteriophage Phi6 was used as a proxy virus to estimate losses during virus elution and concentration (Kantor et al., 2021; Ye et al., 2016). The recoveries obtained were from 10.37% to 58.96%, with a mean recovery of 24.91% (±22.89%). Viral RNA was extracted using Viral RNA QIAGEN kit (QIAGEN, Germantown, Maryland), following the manufacturer's protocol with the modification described previously (Miyani et al., 2021b).

RT-ddPCR was performed on a QX200 AutoDG Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA), using the One-step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA). Per the CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel for SARS-CoV-2 detection (www.cdc.gov), the primers and probe targeting N1 and N2 of SARS-CoV-2 were shown in Table 1S. 5., which were proven to be the most sensitive assays to identify SARS-CoV-2 (Ahmed et al., 2022; Bivins et al., 2021) and were chosen in the current study. N1 N2 gene Duplex Assay Reaction Mixture is shown in Table 1S. 6. Samples were then

run on a C1000 Touch Thermal Cycler (Bio-Rad, Hercules, CA, USA) using the conditions shown in Table 1S. 7., following a measurement of fluorescence on a QX200 Droplet Reader (Bio-Rad, Hercules, CA, USA).

For each RT-ddPCR run, three positive controls (PTCs) and three negative controls (NTCs), and process negative controls (including virus elution and RNA extraction process controls) were included. 102 gc/µL Twist Bioscience Twist Synthetic SARS-CoV-2 RNA Control 2 (MN908947.3) was used for PTCs. Nuclease-free water was used for NTCs. Nanopure water was used as a substitute for 1.5% beef extract in virus elution, as process negative controls. Sterile nuclease-free water was used as a substitute for 140 µL of sample for RNA extraction, as process negative controls. All samples were run in triplicate.

Determination of Limit of Blank (LOB) and Limit of Detection (LOD) were based on the methods described in the manufacturer's (Bio-Rad) guidelines for evaluating analytical sensitivity and validation of RT-ddPCR (Bio-Rad, Hercules, CA, USA). The Limit of Blank (LOB) was determined by testing three types of samples using RT-ddPCR, across four consecutive days including the prior-to COVID-19 pandemic samples collected from the same interceptors, nuclease-free water, and negative process control samples from elution and extraction processes. The purpose of testing the LOB across four separate days was to include the unnoticeable impacts when the tests are performed on different days. The prior-to COVID-19 pandemic interceptor samples were collected on February 18, 2018, from the ONWI, NIEA, and DRI interceptors. LOB for N1 gene ddPCR was determined to be 0.09 gc/µL, and the LOB for N2 gene ddPCR was determined to be 0.08 gc/µL. The Limit of Detection (LOD) was determined for each sample using 10^(-4) and 10^2 gc/µL of positive SARS-CoV-2 RNA across nine consecutive days using a nonparametric method, as-described in the manufacturer's (Bio-rad) guidelines aforementioned.

An LOD of 0.1 gc/µL with 72.92% confidence for the N1 gene ddPCR and 0.1 gc/µL with 81.25% confidence for the N2 gene ddPCR were determined.

To elucidate the impact of SARS-CoV-2 variants on the lag time, the mutations of dominant SARS-CoV-2 variants including Alpha, Beta, Gamma, Delta, and Omicron variants were tested in the wastewater samples using the GT Molecular kits (Pulicharla et al., 2021), during the time when N1 and N2 measurements reached three peaks, which were defined in the current study as Peak I (10/6/20 – 10/28/20) and Peak II (2/17/21 – 3/8/21). For comparison purposes data were collected during the omicron surge Peak III (12/15/21 – 1/12/22) and are presented in the supplementary information. Table 1S. 8. shows the time of report of the dominant SARS-CoV-2 variants which are corresponded to the peaks' periods that were chosen for variant tests.

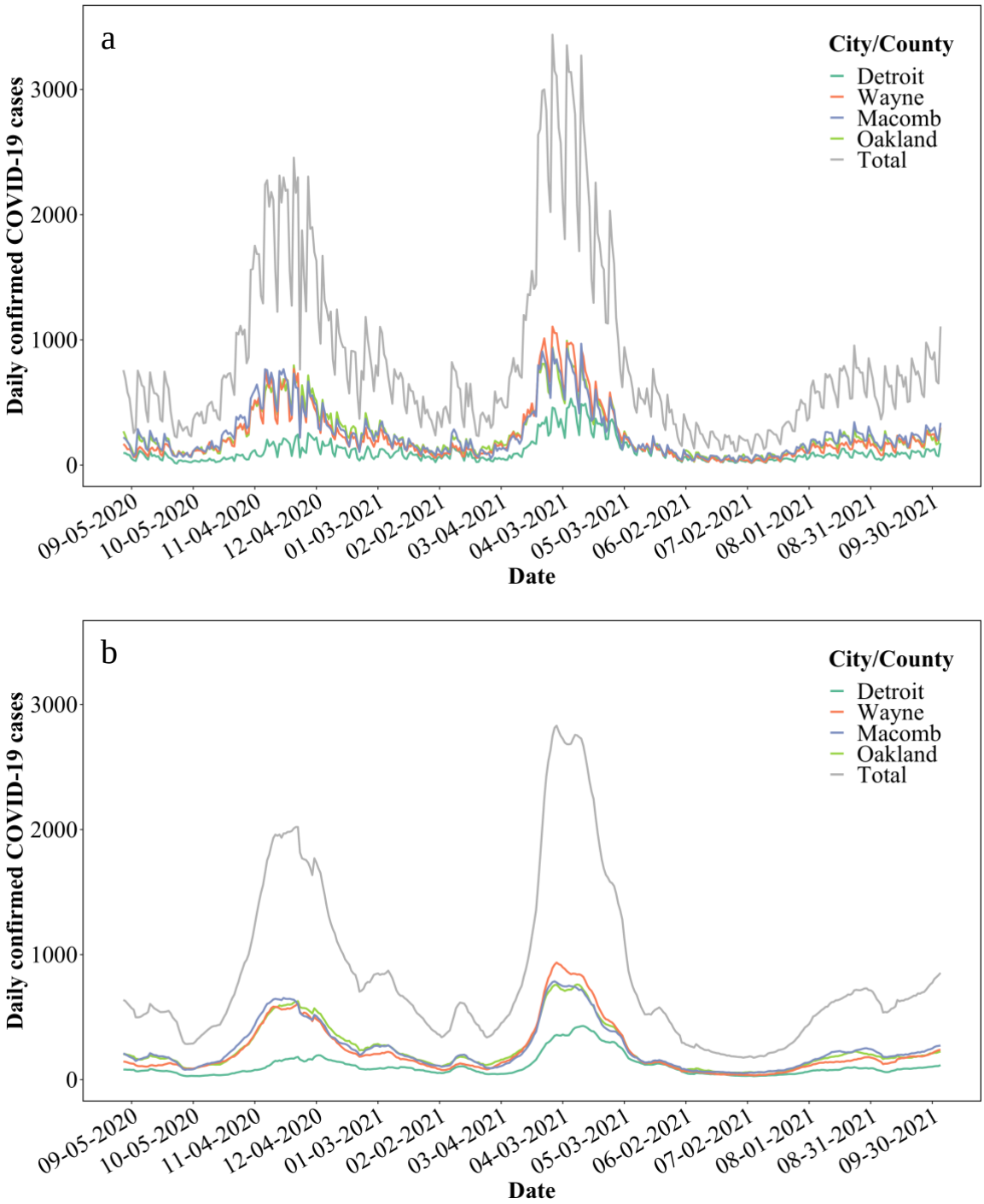


Figure 1. 4. a. Total confirmed COVID-19 cases between September 1, 2020, and October 4, 2021, in the city of Detroit, and Wayne, Macomb, and Oakland counties; b. Total confirmed COVID-19 cases between September 1, 2020, and October 4, 2021, with 7-day moving averages in the city of Detroit, and Wayne, Macomb, and Oakland counties.

2.4 Clinical data of COVID-19

Publicly available data of confirmed COVID-19 incidence in the city of Detroit, as well as Wayne, Macomb, and Oakland counties were used for this study (michigan.gov/coronavirus/). The supplying database was accessed on January 10, 2022 (Figure 1. 4. a.). The range of clinical data was between September 1, 2020, and October 4, 2021. The database was accessed again on April 1, 2022, to supply data for the comparative analysis during the omicron surge. Data were reported as follows: "county" is based on the county of residence (or city in the case of Detroit); "cases" are aggregated by the date of onset of COVID-19 symptoms, if known, otherwise by laboratory specimen date, if known, otherwise by case referral date; "confirmed cases" only include individuals who have had a positive diagnostic laboratory test for COVID-19. Clinical data considering a 7-day moving average was used for further statistical analysis (Figure 1. 4. b.). COVID-19 data were only available per city/country for the Detroit metropolitan area. Moreover, each interceptor received wastewater from portions of each city/county, thus, only the total SARS-CoV-2 concentrations could be correlated to the total COVID-19 cases of city/counties.

2.5 Contributing populations, flow rates, and normalization

The contributing population of each interceptor was estimated from 2020 calculations, provided by the Southeast Michigan Council of Governments by traffic analysis zone (TAZ). Geographic information systems (GIS) analysis was adopted to intersect the TAZ boundaries with ZIP Code boundaries and proportionally allocate population from each TAZ to the intersected areas. ZIP Code boundaries were also intersected with the interceptor service areas to allow for a calculation of population by interceptor area.

Daily flow rates for the three interceptors were estimated from the daily influent flow to the WRRF, calculated from GLWA-reported primary influent flow minus WRRF recycle flows,

and a calibrated hydrologic and hydraulic model developed for the GLWA collection system. The collection system model was developed with the U.S. Environmental Protection Agency's (EPA's) Stormwater Management Model (SWMM) 5 as part of the GLWA Wastewater Master Plan. The SWMM model represents sanitary wastewater and infiltration/inflow, hydraulics in all physical assets of the collection system and at the WRRF entrances, and stages and flows in the Rouge and Detroit rivers.

To account for the changing flow and dilution of the wastewater where those parameters are highly variable day to day, two approaches of normalizing the N1 and N2 gene concentrations of SARS-CoV-2 in gc/L were adopted using Eq. (1). And Eq. (2). The normalized and non-normalized N1 and N2 gene concentrations were used for statistical analysis and modeling.

(1) $C_{(1)} = C_{N1 \text{ or } N2 \text{ gene}} \times V \times f$

(2) $C_{(2)} = C_{N1 \text{ or } N2 \text{ gene}} / S$

$C_{(1)}$: Normalized concentration of SARS-CoV-2 (gc/d)

$C_{N1 \text{ or } N2 \text{ gene}}$: N1 or N2 gene concentrations of SARS-CoV-2 (gc/L)

V: Volume of wastewater flowing into WWTP interceptors during sampling events

f: 3.8 × 10^6, conversion factor between liter and million gallons

$C_{(2)}$: Normalized concentration of SARS-CoV-2 (gc/L of sanitary flow)

S: Sanitary flow percentage of wastewater flowing into WWTP interceptors during sampling events (%)

2.6 Data analyses and visualization

Data were tracked and organized with Microsoft Excel. MATLAB of a 2019b edition was applied to perform the model regression analyses. All figures were generated using R Statistical Computing Software and MATLAB (2019b), depending essentially on ggplot2 package for

visualization, and Forecast and VAR package for autoregression model. The inputs of the models are 7-day moving average of the clinical cases (Barua et al., 2022) and total N1 or N2 gene concentrations for the three interceptors.

Model 1: Linear regression

$Y = bx + a$ (3)

where *Y* is reported clinical cases, *x* is SARS-CoV-2 concentration from the wastewater samples, *b* is the slope of the linear regression line, and *a* is the intercept from the line. The method uses least squares regression.

Model 2: Linear regression model with autoregressive model (ARIMA model)

Autoregressive Integrated Moving Average (ARIMA) is one of the most widely used forecasting methods for time series data. It applies to time series data which have a trend. In this study, ARIMA model is used to fit the residuals from the linear regression between clinical cases and SARS-CoV-2 concentration. The ARIMA model is characterized by autoregressive (AR) and moving average (MA), and number of differencing required to make the time series stationary. For example, for AR model, depends only on its own lags Y_{t-1}, Y_{t-2}, …, Y_{t-p}.

$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \zeta$ (4)

where, if Y_t is clinical case of week t, Y_{t-1} is the previous one-week (t-1) value of clinical case. β_1 is the coefficient of lag 1-week that the model estimates and α is the intercept term. The criterion to choose the best ARIMA model in this study are: 1) lower Akaike information criterion (AIC) value, a lower AIC score indicates a more predictive model 2) white noise after adjusting residuals 3) low standard error (SE) value.

Model 3: Regression model with Autoregressive model has a seasonal pattern (SARIMA model)

Seasonal ARIMA or SARIMA model is an extension of ARIMA model that applies to time series

data with a seasonal component. In this study, ARIMA has a limitation fitting clinical cases and SARS-CoV-2 concentration time series data because it does not apply to these two time series data which have a seasonal pattern.

Model 4: Vector autoregressive model (VAR)

VAR model is a forecasting model that can be used when two or more time series influence each other. In this study, clinical cases and SARS-CoV-2 concentration time series data are considered as two time series data, and there is a relationship between these two time series data.

$$Y_{1,t} = \alpha_1 + \beta_{11,1}Y_{1,t-1} + \beta_{12,1}Y_{2,t-1} + \ldots + \zeta_{1,t} \quad (5)$$

$$Y_{2,t} = \alpha_2 + \beta_{21,1}Y_{1,t-1} + \beta_{22,1}Y_{2,t-1} + \ldots + \zeta_{2,t} \quad (6)$$

where, $Y_{1,t}$ is the clinical cases at week t and $Y_{2,t}$ is the SARS-CoV-2 concentration at the week t. $Y_{1,t-1}$ is the lag of one-week values of clinical cases, $Y_{2,t-1}$ is the lag of one-week values of SARS-CoV-2 concentration.

3. Results and Discussion

3.1 Observed SARS-CoV-2 RNA in wastewater samples and observed COVID-19 cases

During the 12-month study, N1 and N2 genes of SARS-CoV-2 were detected and quantified in all 407 wastewater samples using RT-ddPCR shown in Table 1. 2. The overall observed trends of the total N1 and N2 gene concentrations increased steeply from mid-September 2020 and reached a peak in mid-October 2020 (Figure 1. 5.), which heralded the first peak of COVID-19 infections, in mid-November 2020. Both N1 and N2 gene concentrations stayed comparatively steady between November 2020 and the end of January 2021, following a sharp increase in February 2021 and reached a peak by the end of the month. This brought about the second peak of COVID-19 infections, from approximately late March to April 2021. Subsequently, the total SARS-CoV-2 concentrations measured by N1 and N2 gene RT-ddPCR in

wastewater samples reached comparatively lower peaks in June 2021, which preceded the increase of COVID-19 incidences towards the end of July and August 2021 (Figure 1. 5.). The following decrease of N1 and N2 gene concentrations after each peak was largely due to the termination of shedding events which last a few weeks as demonstrated in Figure 1. 1. and Table 1S. 2.

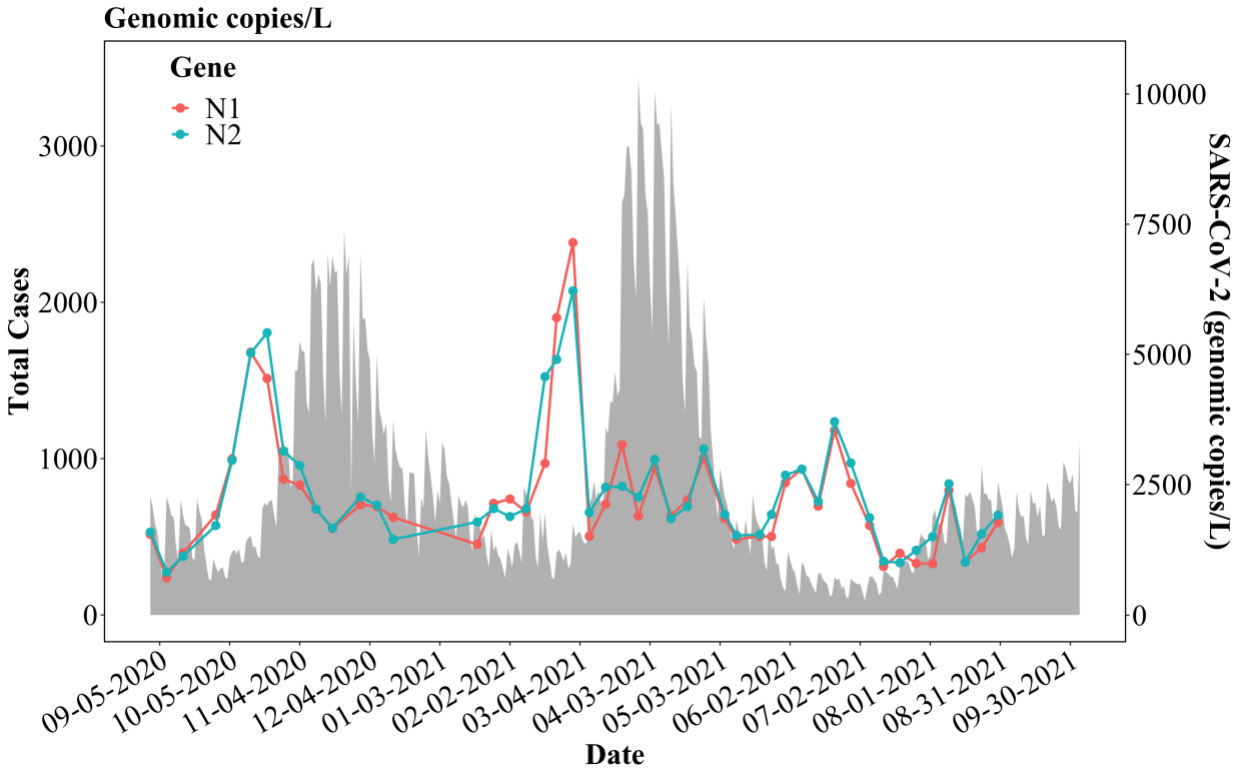


Figure 1. 5. Total N1 and N2 gene concentrations in gc/L for the three interceptors and total confirmed COVID-19 cases in the city of Detroit, as well as Wayne, Macomb, and Oakland counties

3.2 Correlation between SARS-CoV-2 and confirmed COVID-19 cases

A series of statistical analysis with normalizations using different parameters, including BOD, TSS, wastewater flow volume and sanitary percentage of wastewater were conducted. Wastewater flow volume and sanitary percentage of wastewater were chosen for subsequent analysis in terms of stronger Pearson's correlation. Table 1. 2. presents the N1 and N2 gene concentrations normalized by total flow volume and sanitary percentage, respectively. The fluctuations of the normalized results in gc/d and gc/L of sanitary flow stay relatively comparable

to the original results in gc/L of the N1 and N2 genes, which predicted the peaks of COVID-19 incidence by 5 weeks.

Table 1. 2. N1 and N2 gene concentrations measured by RT-ddPCR in wastewater samples collected from GLWA WRRF (SF stands for "Sanitary Flow")

Unit	Gene		Interceptor		
			ONWI	NIEA	DRI
gc/l	N1	Maximum	5.77E+03	1.52E+03	1.41E+03
		Minimum	2.56E+02	2.05E+02	1.85E+02
		Mean	1.07E+03	6.59E+02	5.50E+02
	N2	Maximum	4.83E+03	2.58E+03	1.93E+03
		Minimum	2.81E+02	2.03E+02	2.10E+02
		Mean	1.05E+03	7.29E+02	6.01E+02
gc/d	N1	Maximum	5.23E+12	1.29E+12	1.87E+12
		Minimum	1.66E+11	1.57E+11	1.61E+11
		Mean	7.68E+11	4.36E+11	4.41E+11
	N2	Maximum	4.38E+12	1.75E+12	1.87E+12
		Minimum	1.56E+11	1.61E+11	1.82E+11
		Mean	7.58E+11	4.81E+11	4.79E+11
gc/l of SF	N1	Maximum	2.54E+04	3.87E+03	9.99E+03
		Minimum	8.05E+02	4.71E+02	8.57E+02
		Mean	3.73E+03	1.31E+03	2.35E+03
	N2	Maximum	2.13E+04	5.23E+03	9.99E+03
		Minimum	7.58E+02	4.82E+02	9.71E+02
		Mean	3.68E+03	1.44E+03	2.55E+03

SARS-CoV-2 gene concentrations in wastewater were influenced by absolute clinical case numbers of COVID-19 on single consecutive days. Therefore, 7-day moving averages of clinical cases were performed to smooth the data and reduce outliers. Confirmed COVID-19 clinical data were then aligned with SARS-CoV-2 concentrations (N1/N2) on the exact sampling dates. Missing data from samples were filled using linear interpolation (Lepot et al., 2017). After aligning these two datasets, each SARS-CoV-2 concentrations (N1/N2) had a corresponding COVID-19 case data to be analyzed for pairwise correlation and statistical modeling in the next sections.

We proposed the hypothesis that the fluctuations of SARS-CoV-2 concentrations in wastewater correlate to confirmed COVID-19 cases with a prescribed range of lag times. Hence, Pearson's correlations were first conducted to investigate the strength of a linear correlation

between the total N1 and N2 gene concentrations in gc/L, gc/d, and gc/L of sanitary flow and total confirmed COVID-19 cases in the study areas with various lag times (Table 1S. 27.). To estimate the approximate lag time, week-shift of the clinical cases was adopted. Notably, SARS-CoV-2 concentrations in wastewater strongly correlated with the 5-week shifting forward of a 7-day moving average of COVID-19 incidences (Pearson's r = 0.62 for N1 gene in gc/L, and Pearson's r = 0.64 for N2 gene in gc/L).

Our suggested lag time of 5 weeks between peaks in SARS-CoV-2 genes in wastewater and peaks in reported COVID-19 clinical tests is relatively similar to the longest lag times reported in the literature (Table 1. 1.). For example, a lag of 21 days was reported in Australia (Ahmed et al., 2021) and Sweden (Saguti et al., 2021), and 28 days in Australia (Ahmed et al., 2020a). The increased lag time may be in partly due to our sampling method (VIRADEL) that focuses on viruses in the supernatant of untreated wastewater rather than in the precipitates and solids. These samples represent the near real-time increase in clinical cases in the community. This is important for samples collected in large interceptors where precipitation and resuspension of solids occurs to a large extent. Factors affecting our observed lag time are discussed in section 3.4.

Our analysis also demonstrated that the normalization of clinical data using wastewater flow rate and percentage of sanitary flow did not significantly improve the correlation between the SARS-CoV-2 concentrations in wastewater and COVID-19 clinical cases in communities during the study period. Similar observations were demonstrated in another recent study (Ai et al., 2021).

3.3 Statistical models on the relationship between wastewater surveillance of SARS-CoV-2 and COVID-19 incidence

To better estimate the relationship between the measured SARS-CoV-2 concentrations in wastewater and reported COVID-19 incidence in the communities, four statistical models were

established naming: linear model (L), autoregression model (A), autoregression with time effect model (AT), and vector autoregression model (VA). Confirmed COVID-19 cases with lag times of 3, 4, and 5 weeks were chosen to correlate with N1 and N2 gene concentrations in gc/L, gc/d, and gc/L of sanitary flow using the aforementioned models. The Root Mean Square Error (RMSE) and Pearson's coefficient between actual cases and predicted cases were calculated to estimate the performance of each model shown in Supplementary Tables S1. 10. to S1. 15. From the previous result in section 3.2, a lag time of 5 weeks exhibits a stronger correlation, in agreement with the models in Table 1. 3. and Tables 1S. 29. and 1S. 30., based on Pearson's r. For 5 weeks, in Table 1S. 29., linear regression only considers SARS-CoV-2 concentration as the predictor in the model, the correlation between actual case number and predicted case number is range from 0.4-0.62. Seasonal patterns from the residual of linear regression were also observed. Thus, we consider that there is an autocorrelation effect from the case number. The models were therefore improved with autoregressive errors using ARIMA. The correlation then consequently increased to 0.4-0.67. It is important to note that there is a limitation of ARIMA, it does not apply to seasonal data. After a seasonal component was included in the ARIMA model, the correlation further improved to 0.94-0.95. Thus, the analyses suggest that there is a seasonal pattern present in clinical COVID-19 cases. VA is also a better model (r ranges from 0.95 to 0.96) because it considers both SARS-CoV-2 concentration and case number as the predictors in the model and use their past values to predict current case number. Similarly, the modeling results of N2 gene concentrations show agreement with modeling results of N1 gene concentrations.

Table 1. 3. Statistical modeling results between N1 and N2 gene concentrations and total COVID-19 cases during the 5-week lag time study period in city of Detroit, as well as Wayne, Macomb, and Oakland counties (* is shown in Figure 1. 6.)

RMSE Values		N1-based results			N2-based results		
Unit of N1/N2 gene		gc/L	gc/d	gc/L of SF	gc/L	gc/d	gc/L of SF
3-week lag time	Linear	7.22	135.76	11.78	12.40	926.30	5.62
	Autoregression	135.65	780.00	250.52	341.27	901.23	700.34
	Autoregression+ time effect	10.18	10.18	10.97	11.34	12.34	15.33
	Vector Autoregression	8.32	7.85	8.97	8.89	9.90	9.90
4-week lag time	Linear	7.26	123.56	9.18	16.37	104.45	8.33
	Autoregression	182.92	234.90	635.69	132.35	730.74	500.62
	Autoregression+ time effect	7.50	7.47	7.20	9.75	7.39	7.33
	Vector Autoregression	8.00	7.99	8.62	6.88	8.31	7.62
5-week lag time	Linear	1.83	48.97	2.62	13.95	36.19	2.36
	Autoregression	105.81	417.57	642.83	548.14	570.56	100.95
	Autoregression+ time effect*	1.47*	1.60	1.60	3.21*	1.60	1.42
	Vector Autoregression*	0.35*	0.53	4.44	7.57*	4.37	1.03

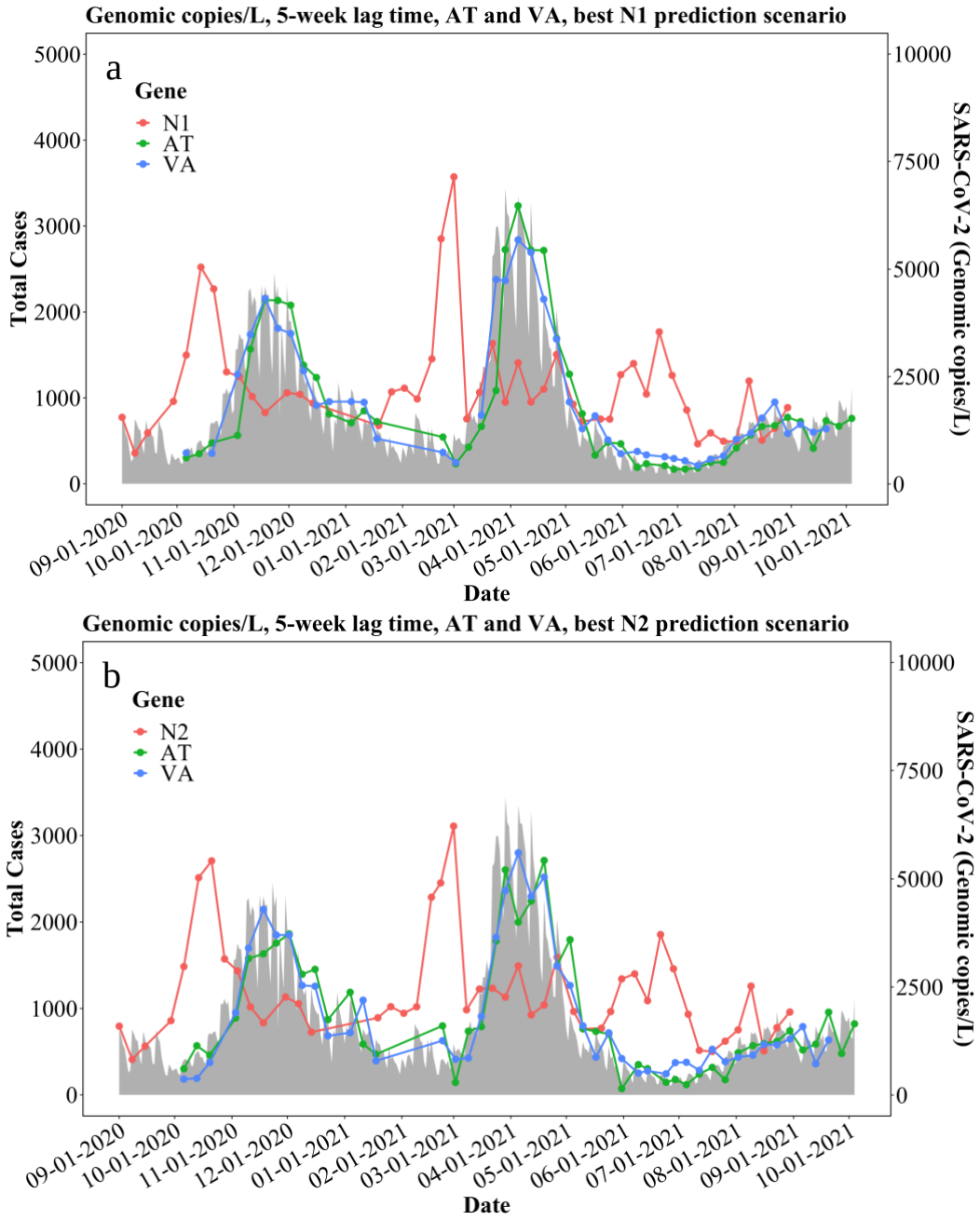


Figure 1. 6. Best prediction models based on (a) N1 gene concentrations (gc/L) and (b) N2 gene concentrations (gc/L) with a 5-week lag time

Comparing the non-normalized data with normalized data in Tables 1S. 28 and Table 1S. 29., for the L and A models, using normalized data does not improve the correlation. For AT and VA models, using normalized data shows similar results comparing to using non-normalized data. It may indicate that these two models reduce the effect of normalizing data. It also indicates that normalizations of the original N1 and N2 gene concentrations did not significantly improve the performance of the modeling and vice versa. All other detailed results are shown in Tables S10 – S15. Same models and analysis were also applied to both the measurements by VIRADEL method (results were shown in Tables 1S. 16. to 1S. 21.) and PEG sampling method results (shown in Tables 1S. 22. and 1S. 23., Figure 1S. 2.) of the study period between August 2021 and February 2022 for comparison purposes as explained in section 3.4.

3.4 Factors affecting lag time

3.4.1 Variants

All the discussions above are based on the study period between September 2020 and August 2021 prior to the Omicron surge. Variations of lag times could not be explicitly elucidated without addressing the changing epidemiological characteristics of emerging SARS-CoV-2 variants, involving incubation time, shedding durations, shedding dynamics, and so forth. Given the shortened incubation time (median of 3 days, (Baker et al., 2022; Brandal et al., 2021; Jansen et al., 2021)) and shedding duration (less than 10 days, (Lamers et al., 2022)) during the Omicron surge, we identified a 2-week lag time, between August 2021 and February 2022, with the same VIRADEL method and same modeling methods (Figure 1S. 2.). The incubation time (Baker et al., 2022) and shedding duration (Lamers et al., 2022) were apparently shorter comparing to the parental and previous variants, inevitably leading to a shorter lag time. Besides, the shedding dynamics changed amid Omicron surge which inevitably affected the lag time (Table 1S. 3.). The

variant test results shown in Table 1S. 9. demonstrate the different mutations of dominant variants identified in the samples which in addition correspond to the reported emerging variants shown in Table 1S. 8. Changing epidemiological characteristics of SARS-CoV-2 variants play a crucial role in affecting the lag time.

3.4.2 Sampling method

An additional factor that may influence lag time is the sampling method. The VIRADEL electropositive filtration method focuses on supernatant virus in wastewater and avoids the inclusion of large wastewater organic solids where viruses may adsorb onto the surfaces. This method has been recommended by the EPA (USEPA 2001, USEPA 2014) and has been extensively applied in the field (Miyani et al., 2020, 2021a; McCall et al., 2020, 2021). On the other hand, grab or composite sampling followed by PEG precipitation incorporates viruses attached onto larger solid particles, which tend to settle in large interceptors and generally re-suspend during periods of high flow. If these larger particles are included in a grab or composite sample, they may include a portion of the viruses that have been settled for a while, having been excreted earlier into that sewer-shed, thus interfering with the desired prediction which is the objective of this work. This becomes a critical factor in interceptors of large urban centers. For small catchment areas with short hydraulic detention times in neighborhood sewer lines the importance of this factor is expected to decrease.

To compare the two methods, we simultaneously collected samples from the same locations. This comparison took place during the omicron surge and the data are shown in the supplementary information. Between August 2021 and February 2022, 24-hour composite samples followed by PEG were collected at the day as the VIRADEL electropositive filtration was performed. This allowed us to investigate on the impact of sampling methods on lag time and

demonstrate the potential for providing early warning signals of the VIRADEL method through comparison.

The PEG measurements (Figure 1S. 1.) and its Pearson's correlation (Table 1S. 25.) demonstrated that the N1 and N2 concentrations did not correlate with COVID-19 cases with a lag time. Results from the same models presented above are shown in Tables 1S. 22. and 1S. 23. for PEG measurements. Results demonstrate that N1 and N2 concentrations based on PEG method did not provide an early warning of COVID-19 cases or significant correlations with the COVID-19 cases in the study area in terms of Pearson's r and RMSE.

3.4.3 Clinical data uncertainties

The report time for associated clinical cases tends to be uncertain due to potential reluctance to be tested (Feng et al., 2021), disparities in reporting clinical data and availability of different testing methods (Ai et al., 2021), limited testing supplies and limited testing sites under rapid testing demand and so forth (Hasan & Nasution, 2021). Since the inception of the COVID-19 pandemic, communities in the U.S. as well as many countries across the world had troubles and limited access to COVID-19 testing and real-time reporting, resulting in delay of real-time tracking and monitoring the clinical cases (Bibby et al., 2021; Larsen et al., 2021).

The lag time could not be explicitly elucidated without addressing all the challenges facing clinical testing, especially amid the early stages of the unprecedented pandemic. In the early stage of the pandemic, the health departments were struggling to provide prompt testing across the country, such as in Detroit and its surrounding counties, rendering a notable delay in clinical data collection and reporting, inevitably leading to a longer lag time (Rader, 2020; Torres et al., 2021; Wiens et al., 2021). Throughout the past two years' efforts amid COVID-19 pandemic, governmental agencies and LHDs have been adapting to increasing COVID-19 cases as well as

testing demands by gradually building the testing capabilities, improving clinical data processing and organizing, pushing clinical data releasing and so forth (Alexander et al., 2022; Powell et al., 2021). The significant improvement and adaptation of LHD to the rapid changing transmissions and clinical incidences of COVID-19 enable a quicker and prompt clinical data reporting and releasing, inevitably leading to a shorter lag time. Despite these caveats, WBE is widely accepted as an effective tool for forewarning community fluctuations in COVID-19 infections (Bibby et al., 2021). Notwithstanding the uncertainties discussed above, we put forth that WBE is an important tool in predicting future fluctuations of COVID-19 infections.

Overall, this study demonstrates the effectiveness of applying wastewater-based epidemiology (WBE) as an early warning tool for the prediction of fluctuations of COVID-19 cases in communities in the Detroit metropolitan area. To our knowledge, this is one of the first studies to systematically evaluate the lag time between peaks in measured concentrations of SARS-CoV-2 in wastewater and peaks in reported COVID-19 cases based on clinical testing. Also, this is, to our knowledge, one of the first studies to propose the use of an autoregression with seasonal pattern model and a vector autoregression model in predicting clinical COVID-19 incidences based on the N1 and N2 gene measurements in wastewater. Though WBE demonstrates great promise, potential limitations and challenges remain. More research is warranted to establish a standard framework for modeling the latency between early detection of COVID-19 and presentation of clinical cases. Future studies should include establishing predictive models to optimize wastewater surveillance for early warning of clinical manifestation.

4. Conclusions

- During the 12-month study, prior to the omicron surge, 407 wastewater samples were collected and analyzed for SARS-CoV-2 genes using RT-ddPCR. Measured

concentrations of SARS-CoV-2 ranged from 714.85 to 7145.98 gc/L by total N1 gene RT-ddPCR and 820.47 to 6219.05 gc/L by total N2 gene RT-ddPCR.

- Lag time, the latency from surge in viral concentration in wastewater and peak in clinical cases was estimated as 5 weeks prior to the Omicron surge.
- As compared to linear regression and autoregression (ARIMA) models, the autoregression model with seasonal patterns and vector autoregression model were more effective in predicting COVID-19 cases during the study period for the 5-week lag scenario.
- Original N1 and N2 gene concentrations were normalized by total flow volumes and sanitary percentage. The statistical results indicated the optimum 5-week prediction models were consistent for both normalized and non-normalized data.
- Surveillance through wastewater sampling and analysis can be employed for predicting infections and monitoring health conditions in large metropolitan areas such as Detroit.

5. Acknowledgements

We thank the Michigan Department of Health and Human Services (MDHHS), Michigan Department of Environment, Great Lakes and Energy (MI-EGLE), and the Great Lakes Water Authority (GLWA) for funding this work. We furthermore thank Michigan State University AgBioresearch and Michigan State University Institute for Global Health for supporting this work.

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. v., & Mueller, J. F. (2020a). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of the Total Environment*, *728*, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Ahmed, W., Bertsch, P. M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B. R., Mueller, J. F., Simpson, S. L., Smith, W. J. M., Symonds, E. M., Thomas, K. v., Verhagen, R., & Kitajima, M. (2020b). Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total Environment*, *739*, 139960. https://doi.org/10.1016/j.scitotenv.2020.139960

Ahmed, W., Bivins, A., Bertsch, P. M., Bibby, K., Choi, P. M., Farkas, K., Gyawali, P., Hamilton, K. A., Haramoto, E., Kitajima, M., Simpson, S. L., Tandukar, S., Thomas, K. V., & Mueller, J. F. (2020c). Surveillance of SARS-CoV-2 RNA in wastewater: Methods optimization and quality control are crucial for generating reliable public health information. *Current Opinion in Environmental Science & Health*, *17*, 82–93. https://doi.org/10.1016/j.coesh.2020.09.003

Ahmed, W., Bivins, A., Metcalfe, S., Smith, W. J. M., Verbyla, M. E., Symonds, E. M., & Simpson, S. L. (2022). Evaluation of process limit of detection and quantification variation of SARS-CoV-2 RT-qPCR and RT-dPCR assays for wastewater surveillance. *Water Research*, *213*, 118132. https://doi.org/10.1016/j.watres.2022.118132

Ahmed, W., Tscharke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T. M. H., O'Brien, J. W., Simpson, S. L., Sherman, P., Thomas, K. v., Verhagen, R., Zaugg, J., & Mueller, J. F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Science of the Total Environment*, *761*, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ru, P., Pan, X., Bohrerova, Z., & Lee, J. (2021b). Wastewater SARS-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States. *Science of the Total Environment*, *801*, 149757. https://doi.org/10.1016/j.scitotenv.2021.149757

Alexander, M., Unruh, L., Koval, A., & Belanger, W. (2022). United States response to the COVID-19 pandemic, January–November 2020. *Health Economics, Policy and Law*, *17*(1), 62–75. https://doi.org/DOI: 10.1017/S1744133121000116

Allan-Blitz, L. T., Hertlein, F., & Klausner, J. D. (2021). Heterogeneity in SARS-CoV-2 Positivity by Ethnicity in Los Angeles. *Journal of Racial and Ethnic Health Disparities*, 2–5. https://doi.org/10.1007/s40615-021-01062-z

Agarwal, V., Venkatakrishnan, A. J., Puranik, A., Kirkup, C., Lopez-Marquez, A., Challener, D. W., Theel, E. S., O'Horo, J. C., Binnicker, M. J., Kremers, W. K., Faubion, W. A., Badley, A. D., Williams, A. W., Gores, G. J., Halamka, J. D., Morice, W. G., & Soundararajan, V. (2020). Long-term SARS-CoV-2 RNA shedding and its temporal association to IgG seropositivity. *Cell Death Discovery*, *6*(1). https://doi.org/10.1038/s41420-020-00375-y

Auwaerter, P. Coronavirus COVID-19 (SARS-CoV-2). The Johns Hopkins University; 2021.

Baker, M. A., Rhee, C., Tucker, R., Badwaik, A., Coughlin, C., Holtzman, M. A., Hsieh, C., Maguire, A., Mermel Blaeser, E., Seetharaman, S., Solem, O., Vaidya, V., & Klompas, M. (2022). Rapid control of hospital-based SARS-CoV-2 Omicron clusters through daily testing and universal use of N95 respirators. *Clinical Infectious Diseases*, ciac113. https://doi.org/10.1093/cid/ciac113

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., … Feng, T. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, *20*(8), 911–919. https://doi.org/10.1016/S1473-3099(20)30287-5

Bibby, K., Bivins, A., Wu, Z., & North, D. (2021). Making waves: Plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research*, *202*(July), 117438. https://doi.org/10.1016/j.watres.2021.117438

Bivins, A., Kaya, D., Bibby, K., Simpson, S. L., Bustin, S. A., Shanks, O. C., & Ahmed, W. (2021). Variability in RT-qPCR assay parameters indicates unreliable SARS-CoV-2 RNA quantification for wastewater surveillance. *Water Research*, *203*, 117516. https://doi.org/10.1016/j.watres.2021.117516

Brandal, L. T., MacDonald, E., Veneti, L., Ravlo, T., Lange, H., Naseer, U., Feruglio, S., Bragstad, K., Hungnes, O., Ødeskaug, L. E., Hagen, F., Hanch-Hansen, K. E., Lind, A., Watle, S. V., Taxt, A. M., Johansen, M., Vold, L., Aavitsland, P., Nygård, K., & Madslien, E. H. (2021). Outbreak caused by the SARS-CoV-2 Omicron variant in Norway, November to December 2021. *Eurosurveillance*, *26*(50). https://doi.org/https://doi.org/10.2807/1560-7917.ES.2021.26.50.2101147

Bui, L. V., Nguyen, H. T., Levine, H., Nguyen, H. N., Nguyen, T. A., Nguyen, T. P., Nguyen, T. T., Do, T. T. T., Pham, N. T., & Bui, M. H. (2020). Estimation of the incubation period of COVID-19 in Vietnam. *PLoS ONE*, *15*(12 December), 1–9. https://doi.org/10.1371/journal.pone.0243889

Cevik, M., Tate, M., Lloyd, O., Maraolo, A. E., Schafers, J., & Ho, A. (2021). SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, *2*(1), e13–e22. https://doi.org/10.1016/S2666-5247(20)30172-5

Chan, V. W. S., Chiu, P. K. F., Yee, C. H., Yuan, Y., Ng, C. F., & Teoh, J. Y. C. (2020). A systematic review on COVID-19: urological manifestations, viral RNA detection and special considerations in urological conditions. *World journal of urology*, 1-12.

Cheng, H.-Y., Jian, S.-W., Liu, D.-P., Ng, T.-C., Huang, W.-T., Lin, H.-H., & Team, T. C.-19 O. I. (2020). Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. *JAMA Internal Medicine*, *180*(9), 1156–1163. https://doi.org/10.1001/jamainternmed.2020.2020

Dai, J., Yang, L., & Zhao, J. (2020). Probable longer incubation period for elderly covid-19 cases: Analysis of 180 contact tracing data in hubei province, china. *Risk Management and Healthcare Policy*, *13*, 1111–1117. https://doi.org/10.2147/RMHP.S257907

D'Aoust, P. M., Graber, T. E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Servos, M. R., Srikanthan, N., MacKenzie, M., Figeys, D., Manuel, D., Jüni, P., MacKenzie, A. E., & Delatolla, R. (2021). Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Science of the Total Environment*, *770*. https://doi.org/10.1016/j.scitotenv.2021.145319

Dhouib, W., Maatoug, J., Ayouni, I., Zammit, N., Ghammem, R., Fredj, S. Ben, & Ghannem, H. (2021). The incubation period during the pandemic of COVID-19: a systematic review and meta-analysis. *Systematic Reviews*, *10*(1), 1–14. https://doi.org/10.1186/s13643-021-01648-y

Ejima, K., Kim, K. S., Ludema, C., Bento, A. I., Iwanami, S., Fujita, Y., Ohashi, H., Koizumi, Y., Watashi, K., Aihara, K., Nishiura, H., & Iwami, S. (2021). Estimation of the incubation period of COVID-19 using viral load data. *Epidemics*, *35*. https://doi.org/10.1016/j.epidem.2021.100454

Feehan, A. K., Velasco, C., Fort, D., Burton, J. H., Price-Haywood, E. G., Katzmarzyk, P. T., Garcia-Diaz, J., & Seoane, L. (2021). Racial and workplace disparities in seroprevalence of SARS-CoV-2, Baton Rouge, Louisiana, USA. *Emerging Infectious Diseases*, *27*(1), 314–317. https://doi.org/10.3201/eid2701.203808

Feng, S., Roguet, A., Mcclary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G., Mclellan, S. L., & Mclellan, S. (2021). Evaluation of sampling frequency and normalization of SARS-CoV-2 wastewater concentrations for capturing COVID-19 burdens in the community. *MedRxiv*, *1*(414), 2021.02.17.21251867.

Foladori, P., Cutrupi, F., Segata, N., Manara, S., Pinto, F., Malpei, F., Bruni, L., & La Rosa, G. (2020). SARS-CoV-2 from faeces to wastewater treatment: What do we know? A review. *Science of the Total Environment*, *743*, 140444. https://doi.org/10.1016/j.scitotenv.2020.140444

Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., Prill, M., Chai, S., Kirley, P., Alden, N., & Kawasaki, B. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with. *Morbidity and Mortality Weekly Report, US Department of Health and Human Services/Centers for Disease Control and Prevention*, *69*(15), 458–464.

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., … Zhong, N. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, *382*(18), 1708–1720. https://doi.org/10.1056/nejmoa2002032

Harris, J. E. (2020). Overcoming reporting delays is critical to timely epidemic monitoring: the case of COVID-19 in New York City. *MedRxiv*, 2020.08.02.20159418.

Hasan, A., & Nasution, Y. (2021). A compartmental epidemic model incorporating probable cases to model COVID-19 outbreak in regions with limited testing capacity. *ISA Transactions*, *xxxx*. https://doi.org/10.1016/j.isatra.2021.01.029

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., … Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, *26*(5), 672–675. https://doi.org/10.1038/s41591-020-0869-5

Hua, C. Z., Miao, Z. P., Zheng, J. S., Huang, Q., Sun, Q. F., Lu, H. P., Su, F. F., Wang, W. H., Huang, L. P., Chen, D. Q., Xu, Z. W., Ji, L. D., Zhang, H. P., Yang, X. W., Li, M. H., Mao, Y. Y., Ying, M. Z., Ye, S., Shu, Q., … Fu, J. F. (2020). Epidemiological features and viral shedding in children with SARS-CoV-2 infection. *Journal of Medical Virology*, *92*(11), 2804–2812. https://doi.org/10.1002/jmv.26180

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Jansen, L., Tegomoh, B., Lange, K., Showalter, K., Figliomeni, J., Abdalhamid, B., Iwen, P. C., Fauver, J., Buss, B., & Donahue, M. (2021). Investigation of a SARS-CoV-2 B.1.1.529 (Omicron) Variant Cluster - Nebraska, November-December 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*(5152), 1782–1784. https://doi.org/10.15585/mmwr.mm705152e3

Jiehao, C., Jin, X., Daojiong, L., Zhi, Y., Lei, X., Zhenghai, Q., Yuehua, Z., Hua, Z., Ran, J., Pengcheng, L., Xiangshi, W., Yanling, G., Aimei, X., He, T., Hailing, C., Chuning, W., Jingjing, L., Jianshe, W., & Mei, Z. (2020). A case series of children with 2019 novel coronavirus infection: Clinical and epidemiological features. *Clinical Infectious Diseases*, *71*(6), 1547–1551. https://doi.org/10.1093/cid/ciaa198

Kantor, R. S., Nelson, K. L., Greenwald, H. D., & Kennedy, L. C. (2021). Challenges in Measuring the Recovery of SARS-CoV-2 from Wastewater. *Environmental Science & Technology*, *55*(6), 3514–3519. https://doi.org/10.1021/acs.est.0c08210

Kaya, D., Niemeier, D., Ahmed, W., & Kjellerup, B. v. (2022). Evaluation of multiple analytical methods for SARS-CoV-2 surveillance in wastewater samples. *Science of The Total Environment*, *808*, 152033. https://doi.org/10.1016/j.scitotenv.2021.152033

Kumar, M., Joshi, M., Patel, A. K., & Joshi, C. G. (2021). Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environmental Research*, *196*(December 2020), 110946. https://doi.org/10.1016/j.envres.2021.110946

La Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L., & Suffredini, E. (2020). First detection of SARS-CoV-2 in untreated wastewaters in Italy. *Science of the Total Environment*, *736*, 139652. https://doi.org/10.1016/j.scitotenv.2020.139652

Lamers, M. M., Mykytyn, A. Z., Breugem, T. I., Groen, N., Knoops, K., Schipper, D., van Acker, R., van den Doel, P. B., Bestebroer, T., Koopman, C. D., Reusken, C., Muraro, M. J., GeurtsvanKessel, C. H., van Royen, M. E., Peters, P. J., Zhang, J., & Haagmans, B. L. (2022). SARS-CoV-2 Omicron efficiently infects human airway, but not alveolar epithelium. *BioRxiv*, 2022.01.19.476898. https://doi.org/10.1101/2022.01.19.476898

Larsen, D. A., Collins, M. B., Du, Q., Hill, D., Insaf, T. Z., Kilaru, P., Kmush, B. L., Middleton, F., Stamm, A., Wilder, M. L., Zeng, T., & Green, H. (2021). Coupling freedom from disease principles and early warning from wastewater surveillance to improve health security. *MedRxiv*, 2021.06.11.21258797.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, *172*(9), 577–582. https://doi.org/10.7326/M20-0504

Lepot, M., Aubin, J. B., & Clemens, F. H. L. R. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. In *Water (Switzerland)* (Vol. 9, Issue 10). https://doi.org/10.3390/w9100796

Li, N., Wang, X., & Lv, T. (2020). Prolonged SARS-CoV-2 RNA shedding: Not a rare phenomenon. *Journal of Medical Virology*, *92*(11), 2286–2287. https://doi.org/10.1002/jmv.25952

Ling, Y., Xu, S. B., Lin, Y. X., Tian, D., Zhu, Z. Q., Dai, F. H., Wu, F., Song, Z. G., Huang, W., Chen, J., Hu, B. J., Wang, S., Mao, E. Q., Zhu, L., Zhang, W. H., & Lu, H. Z. (2020). Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. *Chinese Medical Journal*, *133*(9), 1039–1043. https://doi.org/10.1097/CM9.0000000000000774

Lodder, W., & de Roda Husman, A. M. (2020). SARS-CoV-2 in wastewater: potential health risk, but also data source. *The Lancet Gastroenterology and Hepatology*, *5*(6), 533–534. https://doi.org/10.1016/S2468-1253(20)30087-X

Long, B., Carius, B. M., Chavez, S., Liang, S. Y., Brady, W. J., Koyfman, A., & Gottlieb, M. (2022). Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *The American Journal of Emergency Medicine*, *54*, 46–57. https://doi.org/https://doi.org/10.1016/j.ajem.2022.01.028

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., Wall, P., Green, M., O'Grady, L., & More, S. J. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, *10*(8), 1–9. https://doi.org/10.1136/bmjopen-2020-039652

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/https://doi.org/10.1016/j.watres.2020.116160

McCall, C., Wu, H., O'Brien, E., & Xagoraraki, I. (2021). Assessment of enteric viruses during a hepatitis outbreak in Detroit MI using wastewater surveillance and metagenomic analysis. *Journal of Applied Microbiology*, *131*(3), 1539–1554. https://doi.org/https://doi.org/10.1111/jam.15027

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 and Technology Letters*, *7*(7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Melvin, R. (2021). Predictive power of SARS-CoV-2 wastewater surveillance for diverse populations across a large geographical range. *MedRxiv*, *3*(2), 58–66.

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Miyani, B., McCall, C., & Xagoraraki, I. (2021a). High abundance of human herpesvirus 8 in wastewater from a large urban area. *Journal of Applied Microbiology*, *130*(5), 1402–1411. https://doi.org/https://doi.org/10.1111/jam.14895

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., Norton, J., & Xagoraraki, I. (2021b). Early Warnings of COVID-19 Second Wave in Detroit. *Journal of Environmental Engineering*, *147*(8), 06021004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001907

Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K. K., Wilkinson, R., & Wiedenheft, B. (2020). Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Reports Medicine*, *1*(6), 100098. https://doi.org/10.1016/j.xcrm.2020.100098

O'Brien, E., & Xagoraraki, I. (2019). A water-focused one-health approach for early detection and prevention of viral outbreaks. *One Health*, *7*, 100094. https://doi.org/10.1016/J.ONEHLT.2019.100094

Omori, R., Miura, F., & Kitajima, M. (2021). Age-dependent association between SARS-CoV-2 cases reported by passive surveillance and viral load in wastewater. *Science of the Total Environment*, *792*, 148442. https://doi.org/10.1016/j.scitotenv.2021.148442

Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., 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. https://doi.org/10.1038/s41587-020-0684-z

Powell, M. A., Erwin, P. C., & Bermejo, P. M. (2021). Comparing the COVID-19 Responses in Cuba and the United States. *American Journal of Public Health*, *111*(12), 2186–2193. https://doi.org/10.2105/AJPH.2021.306526

Qi, L., Yang, Y., Jiang, D., Tu, C., Wan, L., Chen, X., & Li, Z. (2020). Factors associated with the duration of viral shedding in adults with COVID-19 outside of Wuhan, China: a retrospective cohort study. *International Journal of Infectious Diseases*, *96*, 531–537. https://doi.org/10.1016/j.ijid.2020.05.045

Qian, G. Q., Chen, X. Q., Lv, D. F., Ma, A. H. Y., Wang, L. P., Yang, N. Bin, & Chen, X. M. (2020). Duration of SARS-CoV-2 viral shedding during COVID-19 infection. *Infectious Diseases*, *52*(7), 511–512. https://doi.org/10.1080/23744235.2020.1748705

Quesada, J. A., López-Pineda, A., Gil-Guillén, V. F., Arriero-Marín, J. M., Gutiérrez, F., & Carratala-Munuera, C. (2021). Incubation period of COVID-19: A systematic review and meta-analysis. *Revista Clínica Española (English Edition)*, *221*(2), 109–117. https://doi.org/10.1016/j.rceng.2020.08.002

Rader, B. (2020). Increased travel times to United States SARS-CoV-2 testing sites: a spatial modeling study. *MedRxiv*.

Rader, B., Astley, C. M., Sy, K. T. L., Sewalk, K., Hswen, Y., Brownstein, J. S., & Kraemer, M. U. G. (2020). Increased travel times to United States SARS-CoV-2 testing sites: a spatial modeling study. *MedRxiv*, 2020.04.25.20074419. https://doi.org/10.1101/2020.04.25.20074419

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*, *181*. https://doi.org/10.1016/j.watres.2020.115942

Rimoldi, S. G., Stefani, F., Gigantiello, A., Polesello, S., Comandatore, F., Mileto, D., Maresca, M., Longobardi, C., Mancon, A., Romeri, F., Pagani, C., Cappelli, F., Roscioli, C., Moja, L., Gismondo, M. R., & Salerno, F. (2020). Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. *Science of the Total Environment*, *744*, 140911. https://doi.org/10.1016/j.scitotenv.2020.140911

Saguti, F., Magnil, E., Enache, L., Churqui, M. P., Johansson, A., Lumley, D., Davidsson, F., Dotevall, L., Mattsson, A., Trybala, E., Lagging, M., Lindh, M., Gisslén, M., Brezicka, T.,

Nyström, K., & Norder, H. (2021a). Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Research, 189*. https://doi.org/10.1016/j.watres.2020.116620

Saguti, F., Magnil, E., Enache, L., Churqui, M. P., Johansson, A., Lumley, D., Davidsson, F., Dotevall, L., Mattsson, A., Trybala, E., Lagging, M., Lindh, M., Gisslén, M., Brezicka, T., Nyström, K., & Norder, H. (2021b). Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Research, 189*. https://doi.org/10.1016/j.watres.2020.116620

Silverman, J. D., Hupert, N., & Washburne, A. D. (2020). Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. *Science Translational Medicine*, *12*(554). https://doi.org/10.1126/SCITRANSLMED.ABC1126

Syangtan, G., Bista, S., Dawadi, P., Rayamajhee, B., Shrestha, L. B., Tuladhar, R., & Joshi, D. R. (2021). Asymptomatic SARS-CoV-2 Carriers: A Systematic Review and Meta-Analysis. *Frontiers in Public Health*, *8*(January), 1–10. https://doi.org/10.3389/fpubh.2020.587374

Tan, W. Y. T., Wong, L. Y., Leo, Y. S., & Toh, M. P. H. S. (2020). Does incubation period of COVID-19 vary with age? A study of epidemiologically linked cases in Singapore. *Epidemiology and Infection*. https://doi.org/10.1017/S0950268820001995

Torres, T. S., Luz, P. M., Coelho, L. E., Jalil, C., Falco, G. G., Sousa, L. P., Jalil, E., Bezerra, D. R. B., Cardoso, S. W., Struchiner, C. J., Veloso, V. G., & Grinsztejn, B. (2021). SARS-CoV-2 testing disparities across geographical regions from a large metropolitan area in Brazil: Results from a web-based survey among individuals interested in clinical trials for COVID-19 vaccines. *The Brazilian Journal of Infectious Diseases*, *25*(4), 101600. https://doi.org/10.1016/j.bjid.2021.101600

Tu, Y., Wei, Y., Zhang, D., Chen, C., Hu, X., & Fei, G. (2020). Analysis of factors affected the SARS-CoV-2 viral shedding time of COVID-19 patients in Anhui, China: a retrospective study. https://doi.org/10.21203/rs.3.rs-20954/v1

USEPA. (2001). Concentration and processing of waterborne viruses by positive charge 1MDS cartridge filters and organic flocculation. Chap. 14. USEPA Manual of Methods of Virology, USEPA, Washington, DC.

USEPA. (2014). Method 1615 Measurement of Enterovirus and Norovirus Occurrence in Water by Culture and RT-PCR.

Wang, C., Xu, M., & Zhang, Z. (2021). A case of COVID-19 with long duration of viral shedding. *Journal of Microbiology, Immunology and Infection*, *54*(1), 152–153. https://doi.org/10.1016/j.jmii.2020.05.008

Wang, K., Zhang, X., Sun, J., Ye, J., Wang, F., Hua, J., Zhang, H., Shi, T., Li, Q., & Wu, X. (2020). Differences of Severe Acute Respiratory Syndrome Coronavirus 2 Shedding Duration in Sputum and Nasopharyngeal Swab Specimens Among Adult Inpatients With Coronavirus

Disease 2019. *Chest*, *158*(5), 1876–1884. https://doi.org/10.1016/j.chest.2020.06.015

Warabi, Y., Tobisawa, S., Kawazoe, T., Murayama, A., Norioka, R., Morishima, R., Inoue, T., Shimizu, T., & Takahashi, K. (2020). Effects of oral care on prolonged viral shedding in coronavirus disease 2019 (COVID-19). *Special Care in Dentistry*, *40*(5), 470–474. https://doi.org/10.1111/scd.12498

Weidhaas, J., Aanderud, Z. T., Roper, D. K., VanDerslice, J., Gaddis, E. B., Ostermiller, J., Hoffman, K., Jamal, R., Heck, P., Zhang, Y., Torgersen, K., Laan, J. Vander, & LaCross, N. (2021). Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of the Total Environment*, *775*, 145790. https://doi.org/10.1016/j.scitotenv.2021.145790

Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., & Prescott, H. C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, *324*(8), 782–793. https://doi.org/10.1001/jama.2020.12839

Wiens, K. E., Mawien, P. N., Rumunu, J., Slater, D., Jones, F. K., Moheed, S., Caflish, A., Bior, B. K., Jacob, I. A., Lako, R. L. L., Guyo, A. G., Olu, O. O., Maleghemi, S., Baguma, A., Hassen, J. J., Baya, S. K., Deng, L., Lessler, J., Demby, M. N., … Wamala, J. F. (2021). Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Juba, South Sudan: a population-based study. *MedRxiv : The Preprint Server for Health Sciences*, 1–28. https://doi.org/10.1101/2021.03.08.21253009

Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M. A., Bushman, M., Chai, P. R., Duvallet, C., Erickson, T. B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W. P., Huang, K. H., Lee, W. L., Matus, M., … Alm, E. J. (2022). SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Science of The Total Environment*, *805*, 150121. https://doi.org/10.1016/j.scitotenv.2021.150121

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G., & Huang, X. (2020). Correspondence Prolonged presence of SARS-CoV-2 viral RNA in. In *The Lancet Gastroenterology & Hepatology* (Vol. 5, Issue 5, pp. 434–435).

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 waste water, Greater Paris, France, 5 March to 23 April 2020. *Eurosurveillance*, *25*(50). https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000776

Xagoraraki, I., Yin, Z., & Svambayev, Z. (2014). Fate of viruses in water systems. *Journal of Environmental Engineering*, *140*(7), 04014020.

Xagoraraki, I. (2020). *Can We Predict Viral Outbreaks Using Wastewater Surveillance?* (p. J. Environ. Eng., 2020, 146(11): 01820003). Journal of Environmental Engineering.

Xagoraraki, I., & O'Brien, E. (2020). Wastewater-based epidemiology for early detection of

viral outbreaks. *Women in water quality: Investigations by prominent female engineers*, 75-97.

Xiao, Z., Xie, X., Guo, W., Luo, Z., Liao, J., Wen, F., Zhou, Q., Han, L., & Zheng, T. (2020). Examining the incubation period distributions of COVID-19 on Chinese patients with different travel histories. *Journal of Infection in Developing Countries*, *14*(4), 323–327. https://doi.org/10.3855/JIDC.12718

Xing, Y. H., Ni, W., Wu, Q., Li, W. J., Li, G. J., Wang, W. Di, Tong, J. N., Song, X. F., Wing-Kin Wong, G., & Xing, Q. S. (2020). Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. *Journal of Microbiology, Immunology and Infection*, *53*(3), 473–480. https://doi.org/10.1016/j.jmii.2020.03.021

Xu, K., Chen, Y., Yuan, J., Yi, P., Ding, C., Wu, W., Li, Y., Ni, Q., Zou, R., Li, X., Xu, M., Zhang, Y., Zhao, H., Zhang, X., Yu, L., Su, J., Lang, G., Liu, J., Wu, X., … Li, L. (2020a). Factors associated with prolonged viral RNA shedding in patients with coronavirus disease 2019 (COVID-19). *Clinical Infectious Diseases*, *71*(15), 799–806. https://doi.org/10.1093/cid/ciaa351

Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., Zhang, H., Liu, H., Xia, H., Tang, J., Zhang, K., & Gong, S. (2020b). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature Medicine*, *26*(4), 502–505. https://doi.org/10.1038/s41591-020-0817-4

Yan, D., Liu, X. Y., Zhu, Y. N., Huang, L., Dan, B. T., Zhang, G. J., & Gao, Y. H. (2020). Factors associated with prolonged viral shedding and impact of lopinavir/ritonavir treatment in hospitalised non-critically ill patients with SARS-CoV-2 infection. *European Respiratory Journal*, *56*(1).

Yang, R., Gui, X., & Xiong, Y. (2020). Comparison of Clinical Characteristics of Patients with Asymptomatic vs Symptomatic Coronavirus Disease 2019 in Wuhan, China. *JAMA Network Open*, *3*(5), 2019–2022. https://doi.org/10.1001/jamanetworkopen.2020.10182

Yaniv, K., Ozer, E., & Kushmaro, A. (2021). SARS-CoV-2 variants of concern, Gamma (P.1) and Delta (B.1.617), sensitive detection and quantification in wastewater employing direct RT-qPCR. *MedRxiv*, 2021.07.14.21260495.

Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science & Technology*, *50*(10), 5077–5085. https://doi.org/10.1021/acs.est.6b00876

Yin, Z., Voice, T. C. , Tarabara, V. V. , & Xagoraraki, Irene. (2018). Sorption of Human Adenovirus to Wastewater Solids. *Journal of Environmental Engineering*, *144*(11). https://doi.org/10.1061/(ASCE)EE.1943-7870.0001463

Yu, X., Sun, X., Cui, P., Pan, H., Lin, S., Han, R., Jiang, C., Fang, Q., Kong, D., Zhu, Y., Zheng, Y., Gong, X., Xiao, W., Mao, S., Jin, B., Wu, H., & Fu, C. (2020). Epidemiological and clinical characteristics of 333 confirmed cases with coronavirus disease 2019 in Shanghai, China. *Transboundary and Emerging Diseases*, *67*(4), 1697–1707. https://doi.org/10.1111/tbed.13604

Zaki, N., & Mohamed, E. A. (2021). The estimations of the COVID-19 incubation period: A scoping reviews of the literature. *Journal of Infection and Public Health*, *14*(5), 638–646. https://doi.org/10.1016/j.jiph.2021.01.019

Zhang, N., Gong, Y., Meng, F., Shi, Y., Wang, J., Mao, P., Chuai, X., Bi, Y., Yang, P., & Wang, F. (2021). Comparative study on virus shedding patterns in nasopharyngeal and fecal specimens of COVID-19 patients. *Science China Life Sciences*, *64*(3), 486–488. https://doi.org/10.1007/s11427-020-1783-9

Zhang, W. Y., Yu, L. Q., Huang, J. A., & Zeng, D. X. (2020). Prolonged Viral RNA Shedding Duration in COVID-19. *American Journal of Therapeutics*, *2*, 1–2. https://doi.org/10.1097/MJT.0000000000001200

Zhou, B., She, J., Wang, Y., & Ma, X. (2020a). Duration of Viral Shedding of Discharged Patients with Severe COVID-19. *Clinical Infectious Diseases*, *71*(16), 2240–2242. https://doi.org/10.1093/cid/ciaa451

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020b). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, *395*(10229), 1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3

Zhou, R., Li, F., Chen, F., Liu, H., Zheng, J., Lei, C., & Wu, X. (2020c). Viral dynamics in asymptomatic patients with COVID-19. *International Journal of Infectious Diseases*, *96*, 288–290. https://doi.org/10.1016/j.ijid.2020.05.030

APPENDIX

Table 1S. 1. Incubation time of SARS-CoV-2 prior to Omicron surge in literature

Incubation time (days)	Age	Gender	Sample population	Location or agency	References
0 to 14	-	-	-	World Health Organization	(Zaki & Mohamed, 2021)
2 to 14	-	-	-	European Centre for Disease Prevention and Control	(Zaki & Mohamed, 2021)
Median of 3	-	-	-	-	(Yang et al., 2020)
Median of 4	Median age of 47	41.9% female	1099 COVID-19 patients	522 hospitals in mainland China	(Guan et al., 2020)
Median of 4.8	Median age of 45	187 males, 204 females	391 cases	Shenzhen, China	(Bi et al., 2020)
Median of 5	Median age of 42	-	-	Singapore	(Tan et al., 2020)
Median of 5	-	-	-	United States	(Silverman et al., 2020)
Median of 5.1	-	-	-	50 provinces and regions outside Wuhan, China	(Lauer et al., 2020)
Median of 5, mean of 7.8	-	-	-	-	(Zaki & Mohamed, 2021)
5.2 to 6.65	-	no age, gender, or ethnicity restrictions		42 studies done primarily in China	(Dhouib et al., 2021)
5.6 to 6.7	-	no age, gender, or ethnicity restrictions		Analysis of literature from January to March 2020	(Quesada et al., 2021)
Median of 5.8, mean of 5.1	-	-	-	24 published studies for meta-analysis	(McAloon et al., 2020)
Median of 5.8	Mean age of 46.1	97 males, 82 females	180 cases	Shiyan, China	(Dai et al., 2020)
Median of 5.85	-	-	-	Japan	(Ejima et al., 2021)
Median of 5.6, mean of 6.4	Mean age of 36, median age of 31	58.2% female	265 cases	Vietnam	(Bui et al., 2020)
Median of 6.5	Mean age of 6.17	4 males, 6 females	10 pediatric cases	Wuhan, China	(Jiehao et al., 2020)
Median of 9.1	Mean of 8.2	-	43 pediatric cases	Zhejiang, China	(Hua et al., 2020)
Median of 7.2 to 9	Median age of 50 years	172 (51.7%) males, 161 (48.3%) females	333 cases	Shanghai, China	(Yu et al., 2020)

Table 1S. 2. Shedding duration of SARS-CoV-2 prior to Omicron surge in literature

Shedding duration (days)	Age	Gender	Sample population	Location or agency	References
7 to 8	-	-	31 cases	China	(Zhou et al., 2020c)
Median of 9.5	Median age of 44	38 males, 28 females	66 cases	China	(Ling et al., 2020)
10 to 22	Median age of 48	12 males, 11 females	23 people	Beijing, China	(Zhang et al., 2021)
7 to 16	-	-	-	Anhui, China	(Tu et al., 2020)
Mean of 31.6	Mean age of 60	Male : female is 2:6	-	Tokyo Metropolitan Neurological Hospital, Japan	(Warabi et al., 2020)
Median of 12	-	-	-	China	(Qian et al., 2020)
Median of 17	-	-	133 cases	Wuhan, China	(Xu et al., 2020a)
14 to 30	-	-	-	Analysis from previous published work	(Ahmed et al., 2021; Jiehao et al., 2020; Wu et al., 2020; Xu et al., 2020b).
> 30	-	-	378 cases	Published governmental documents of China	(Li et al., 2020)
Median of 17	Mean age of 42	67 males, 80 females	147 cases	Changsha, China	(Qi et al., 2020)
Median of 17.2	-	-	586 individuals	13 studies	(Cevik et al., 2021)
17.3 to 22.7	-	-	851 cases	United States	(Agarwal et al., 2020)
Median of 20	Mean age of 56	119 males, 72 females	191 cases	Jinyintan Hospital, Wuhan Pulmonary Hospital, China	(Zhou et al., 2020b)
Median of 23	Median age of 52	52 male, 78 female	120 cases	Hubei, China	(Yan et al., 2020)
45	34	Male	-	Wuhan, China	(Zhang et al., 2020)
Median of 31	Median age of 58	22 males, 19 females	41 cases	-	(Zhou et al., 2020a)
Median of 34	-	-	68 cases	Taikang Tongji Hospital, Huoshenshan Hospital, China	(Wang et al., 2020)
49	59	Female	-	Wuhan, China	(Wang et al., 2021)
2 to 49	-	-	3714 cases	Review of 21 studies	(Chan et al., 2020)
2 to 20	Under 10	-	3 cases	Qingdao, China	(Xing et al., 2020)
More than 70	Median age of 8.2	-	-	Zhejiang, China	(Hua et al., 2020)

Table 1S. 3. Time of shedding prior to symptoms onset

Study period	Time of shedding prior to symptoms onset	References
September 2020 to August 2021	5 to 7 days before symptoms onset	(Cheng et al., 2020; He et al., 2020)
August 2021 to February 2022	2 days before symptoms onset	(Auwaerter, 2021)
	3 days before symptoms onset	(Long et al., 2022; Wiersinga et al., 2020)

Table 1S. 4. Travel time (hours) to WRRF for each interceptor

Interceptor	Weighted Average	Minimum	Maximum
DRI	12.3	0.2	41.8
NIEA	22.5	0.7	51.2
NWI	8.6	0.1	25.9

Table 1S. 5. Sequences of the primers and probe used to detect SARS-CoV-2

Name	Description	Oligonucleotide Sequence (5'>3')
2019-nCoV_N1-F	2019-nCoV_N1 Forward Primer	GAC CCC AAA ATC AGC GAA AT
2019-nCoV_N1-R	2019-nCoV_N1 Reverse Primer	TCT GGT TAC TGC CAG TTG AAT CTG
2019-nCoV_N1-P	2019-nCoV_N1 Probe	FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1
2019-nCoV_N2-F	2019-nCoV_N2 Forward Primer	TTA CAA ACA TTG GCC GCA AA
2019-nCoV_N2-R	2019-nCoV_N2 Reverse Primer	GCG CGA CAT TCC GAA GAA
2019-nCoV_N2-P	2019-nCoV_N2 Probe	FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1

Table 1S. 6. N1 N2 gene duplex assay reaction mixture

Name	Volume	Final concentration
One-Step RT-Supermix (20x)	5.5 µL	1x
Reverse Transcriptase (RT)	2.2 µL	20 units/µL
300 mM DTT	1.1 µL	15 mM
N1 primer probe mix	3.3 µL	900 nM/250nM*
N2 primer probe mix	3.3 µL	900 nM/250nM*
PCR-grade water	1.1 µL	-
Sample RNA	5.5 µL	-

Note: *Forward and reverse primers are at a final concentration of 900 nM and each probe is at a final concentration of 250 nM.

Table 1S. 7. Thermocycling conditions for N1 and N2 duplex reaction

Temperature	Time	Note
25°C	3 min	-
50°C	60 min	-
95°C	10 min	-
95°C	30 sec	40 cycles, ramp speed of 2°C/second
55°C	1 min	40 cycles, ramp speed of 2°C/second
98°C	10 min	-
4°C	∞	-

Table 1S. 8. Timeline of SARS-CoV-2 variants

Name of dominate Variants	Lineage	First reported	Country of Origin	References
Alpha	B.1.1.7	September 2020	United Kingdom	cdc.gov
Beta	B.1.351	October 2020	South Africa	
Gamma	P.1	January 2021	Japan	
Delta	B.1.617.2	February 2021	India	
Omicron	B.1.1.529	November 2021	South Africa	

Table 1S. 9. Test of variants for the time during the three highest peaks of N1 and N2 gene measurements using GT Molecular Kits

Study period for peaks of N1/N2 measurements	Alpha (B.1.1.7)		Beta (B.1.351)	Gamma (P.1)	Delta (B.1.617.2)		Omicron (B.1.1.529)	
Mutations	N501Y	Del69-70	K417N	K417T	T478K	L452R	N679K	Q954H
Peal I (10/6/20 - 10/28/20)	+	+	+	-	-	-	-	-
Peak II (2/17/21 – 3/8/21)	+	+	+	+	+	+	+	-
Peak III (12/15/21 – 1/12/22)	-	+	+	+	+	+	+	+

Table 1S. 10. Modeling results of N1 gene in gc/L for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/L)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 0.015\ x_{t(3)} + 534.96$	7.22	0.26
	Autoregression	$y_{t(3)}-0.2928y_{t(3)-1} -0.3741\ y_{t(3)-2}= 0.015(x_{t(3)}-0.2928x_{t(3)-1} – 0.3741x_{t(3)-2}) + 534.96$	135.65	0.07
	Autoregression+ time effect	$y_{t(3)} = 1052 – 1.7923t(3) -0.044dx_{t(3)}$ $y_{t(3)}*= y_{t(3)} + 0.5857y_{t(3)-1}$	10.18	0.76
	Vector Autoregression	$y_{t(3)} = 1.30y_{t(3)-1} +0.12\ y_{t(3)-2} -0.49\ x_{t(3)-1} – 0.01x_{t(3)-2} -73.46$	8.32	0.72
4 weeks	Linear	$y_{t(4)} = 0.29\ x_{t(4)} + 227.04$	7.26	0.51
	Autoregression	$y_{t(4)}-0.1706y_{t(4)-1} -0.2799\ y_{t(4)-2}= 0.29(x_{t(4)}-0.2354x_{t(4)-1} – 0.2799x_{t(4)-2}) + 227.04$	182.92	0.50
	Autoregression+ time effect	$y_{t(4)} = 1730 + 6.78t(4) + 0.06dx_{t(4)}$ $y_{t(4)}*= y_{t(4)} + 0.59y_{t(4)-1}$	7.50	0.92
	Vector Autoregression	$y_{t(4)} = 1.42y_{t(4)-1} +0.20\ y_{t(4)-2} –0.61\ x_{t(4)-1} + 0.004x_{t(4)-2} + 109.89$	8.00	0.86
5 weeks	Linear	$y_{t(5)} = 0.35\ x_{t(5)} + 93.13$	1.83	0.62
	Autoregression	$y_{t(5)}+0.2362y_{t(5)-1} -0.0785\ y_{t(5)-2}= 0.35(0.2362x_{t(5)-1} – 0.0785x_{t(5)-2}) + 93.13$	105.81	0.67
	Autoregression+ time effect	$y_{t(5)} = 1337 + 20.20t(5)-0.011dx_{t(5)}$ $y_{t(5)}*= y_{t(5)} + 0.64y_{t(5)-1}$	1.47	0.95
	Vector Autoregression	$y_{t(5)} = 1.54y_{t(5)-1} -0.02\ y_{t(5)-2} -0.68\ x_{t(5)-1} -0.04\ x_{t(5)-2} – 191.18$	0.35	0.96

Notes: (1) In the table, X represents the measured SARS-CoV-2 concentrations in wastewater, while Y represents COVID-19 incidences. Y* is the new estimated values based on ARIMA (SARIMA) model. Pearson r is between actual case and predicted case. Pearson's correlation is between the actual clinical cases and predicted clinical cases. (2) All tables from the Table 1S. 10. To the Table 1S. 21. are based on the VIRADEL method. Tables 1S. 22. To the 1S. 23. are based on the PEG method.

Table 1S. 11. Modeling results of N1 gene in gc/d for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/d)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 6.07e\text{-}11\ x_{t(3)} + 783.22$	135.76	0.16
	Autoregression	$y_{t(3)}\text{-}0.4325y_{t(3)-1}\ \text{-}0.1264\ y_{t(3)-2} = 6.07e\text{-}11(x_{t(3)}\ \text{-}\ 0.4325x_{t(3)-1} - 0.1264x_{t(3)-2}) + 783.22$	780.00	0.07
	Autoregression+ time effect	$y_{t(3)} = 1245\ \text{-}1.507\,t(3) - 7.23e\text{-}11dx_{t(3)}$ $y_{t(3)}{*} = y_{t(3)} + 0.5794y_{t(3)-1}$	10.18	0.75
	Vector Autoregression	$y_{t(3)} = 1.5461y_{t(3)-1} + 0.0425\ y_{t(3)-2}\ \text{-}0.7104\ x_{t(3)-1}\ \text{-}\ 0.0006x_{t(3)-2} + 43.8610$	7.85	0.70
4 weeks	Linear	$y_{t(4)} = 1.62e\text{-}10x_{t(4)} + 628.35$	123.56	0.51
	Autoregression	$y_{t(4)}\text{-}0.3145y_{t(4)-1}\ \text{-}0.1923\ y_{t(4)-2} = 1.62e\text{-}10(x_{t(4)}\text{-}\ 0.3245x_{t(4)}\text{-}1 - 0.1923x_{t(4)-2}) + 628.35$	234.90	0.51
	Autoregression+ time effect	$y_{t(4)} = 1292\ \text{-}16.65\ t(4)\ \text{-}1.75dx_{t(4)}$ $y_{t(4)}{*} = y_{t(4)} + 0.5857y_{t(4)-1}$	7.47	0.90
	Vector Autoregression	$y_{t(4)} = 1.3541y_{t(4)-1} + 0.0235y_{t(4)-2} + 0.006x_{t(4)-1} + 1.4532e\text{-}2x_{t(4)-2} + 580.33$	7.99	0.85
5 weeks	Linear	$y_{t(5)} = 1.93e\text{-}10x_{t(5)} + 590.75$	48.97	0.55
	Autoregression	$Y_{t(5)} + 0.0559y_{t(5)-1} + 0.0436\ y_{t(5)-2} = 1.93e\text{-}10(x_{t(5)}\ \text{-}\ 0.0559x_{t(5)-1} - 0.0436x_{t(5)\ -2)} + 590.75$	417.57	0.54
	Autoregression+ time effect	$y_{t(5)} = 1337\text{-}\ \text{-}17.84\,t(5) + 2.08e\text{-}10dx_{t(5)}$ $y_{t(5)}{*} = y_{t(5)} + 0.6408y_{t(5)-1}$	1.60	0.94
	Vector Autoregression	$y_{t(5)} = 1.3529y_{t(5)-1} + 0.0007\ y_{t(5)-2} + 0.0008\ x_{t(5)-1}\ \text{-}\ 0.0006x_{t(5)-2} + 407.89$	0.53	0.95

Table 1S. 12. Modeling results of N1 gene in gc/L of sanitary flow for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/L of sanitary flow)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 0.010x_{t(3)} + 795.32$	11.78	0.09
	Autoregression	$Y_{t(3)}+0.4477y_{t(3)-1} +0.1154\ y_{t(3)-2}= 0.010(x_{t(3)}- 0.4477x_{t(3)-1} + 0.1154x_{t(3)-2}) + 795.32$	250.52	0.10
	Autoregression+ time effect	$y_{t(3)} = 1244 – 15.06t(3) + 0.0143dx_{t(3)}$ $y_{t(3)}*= y_{t(3)} + 0.5783y_{t(3)-1}$	10.97	0.75
	Vector Autoregression	$y_{t(3)} = 1.4395y_{t(3)-1} +0.0170\ y_{t(3)-2} -0.6046\ x_{t(3)-1} - 0.005x_{t(3)-2} -68.6976$	8.97	0.72
4 weeks	Linear	$y_{t(4)} = 0.032x_{t(4)}+ 642.02$	9.18	0.33
	Autoregression	$y_{t(4)}-0.3304y_{t(4)-1} -0.1952\ y_{t(4)-2}= 0.032(x_{t(4)}- 0.3304x_{t(4)-1} – 0.1952x_{t(4)-2}) + 642.02$	635.69	0.33
	Autoregression+ time effect	$y_{t(4)} = 1291- 16.53t(4) +0.03\ dx_{t(4)}$ $y_{t(4)}*= y_{t(4)} + 0.5892y_{t(4)-1}$	7.20	0.90
	Vector Autoregression	$y_{t(4)} = 1.5254y_{t(4)-1} -0.003\ y_{t(4)-2} -0.6839\ x_{t(4)-1} + 0.0008x_{t(4)-2} + 180.32$	8.62	0.85
5 weeks	Linear	$y_{t(5)} = 0.04x_{t(5)} + 607.55$	2.62	0.40
	Autoregression	$y_{t(5)}+0.0572y_{t(5)-1} -0.0444\ y_{t(5)-2}= 0.04(x_{t(5)}- 0.0572x_{t(5)-1} – 0.0444x_{t(5)-2}) + 607.55$	642.83	0.40
	Autoregression+ time effect	$y_{t(5)} = 1337-17.83t(5)-0.0329dx_{t(5)}$ $y_{t(5)}*= y_{t(5)} + 0.6424y_{t(5)-1}$	1.60	0.95
	Vector Autoregression	$y_{t(5)} = 1.5570y_{t(5)-1} -1.0251e^{-5}\ y_{t(5)-2} -0.710\ x_{t(5)-1} - 5.4944\ _{t(5)-2} +182.207$	4.44	0.95

Table 1S. 13. Modeling results of N2 gene in gc/L for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/L)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 0.168x_{t(3)} + 483.35$	12.40	0.28
	Autoregression	$y_{t(3)}$-0.3067$y_{t(3)-1}$ -0.3836 $y_{t(3)-2}$= 0.168($x_{t(3)}$-0.3067$x_{t(3)-1}$ – 0.3836$x_{t(3)-2}$) + 483.35	341.27	0.20
	Autoregression+ time effect	$y_{t(3)} = 1566 + 0.22t(3) - 0.008dx_{t(3)}$ $y_{t(3)}$*= $y_{t(3)} + 0.5825y_{t(3)-1}$	11.34	0.75
	Vector Autoregression	$y_{t(3)} = 1.31y_{t(3)-1} + 0.12\ y_{t(3)-2} - 0.50\ x_{t(3)-1} + 0.02x_{t(3)-2} - 69.302$	8.89	0.73
4 weeks	Linear	$y_{t(4)} = 0.307x_{t(4)} + 163.10$	16.37	0.51
	Autoregression	$y_{t(4)}$-0.2354$y_{t(4)-1}$ -0.3320 $y_{t(4)-2}$= 0.307($x_{t(4)}$-0.2354$x_{t(4)-1}$ – 0.3320$x_{t(4)-2}$) + 163.70	132.35	0.29
	Autoregression+ time effect	$y_{t(4)} = 1506 + 3.21t(4) - 0.004\ dx_{t(4)}$ $y_{t(4)}$*= $y_{t(4)} + 0.3213y_{t(4)-1}$	9.75	0.90
	Vector Autoregression	$y_{t(4)} = 2.07y_{t(4)-1} + 0.70\ y_{t(4)-2} - 2.07\ x_{t(4)-1} + 0.17x_{t(4)-2} + 304.844$	6.88	0.86
5 weeks	Linear	$y_{t(5)} = 0.152x_{t(5)} + 534.03$	13.95	0.64
	Autoregression	$y_{t(5)}$-0.2928$y_{t(5)-1}$ -0.3741$y_{t(5)-2}$= 0.152($x_{t(5)}$-0.2928$x_{t(5)-1}$ – 0.3741$x_{t(5)-2}$) + 534.03	548.14	0.65
	Autoregression+ time effect	$y_{t(5)} = 1342 - 0.060t(5) - 0.008dx_{t(5)}$ $y_{t(5)}$*= $y_{t(5)} + 0.5825y_{t(5)-1}$	3.21	0.92
	Vector Autoregression	$y_{t(5)} = 1.30y_{t(5)-1} + 0.12\ y_{t(5)-2} - 0.49\ x_{t(5)-1} - 0.01\ x_{t(5)-2}$ – 73.46	7.57	0.95

Table 1S. 14. Modeling results of N2 gene in gc/d for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/d)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 6.040e\text{-}11x_{t(3)} + 779.15$	926.30	0.10
	Autoregression	$y_{t(3)}+0.4402y_{t(3)-1} +0.1109\ y_{t(3)-2}= 6.04e\text{-}11(x_{t(3)}+0.4402x_{t(3)-1} + 0.1109x_{t(3)-2}) + 690.03$	901.23	0.08
	Autoregression+ time effect	$y_{t(3)} = 1245 -15.08t(3) +7.61e\text{-}11dx_{t(3)}$ $y_{t(3)}*= y_{t(3)} + 0.5776y_{t(3)-1}$	12.34	0.75
	Vector Autoregression	$y_{t(3)} = 1.448y_{t(3)-1} +6.8e\text{-}11\ y_{t(3)-2} -0.6142\ x_{t(3)-1} -2.45e^{-11}x_{t(3)-2} + 80.6042$	9.90	0.72
4 weeks	Linear	$y_{t(4)} = 1.48e\text{-}10x_{t(4)} + 639.54$	104.45	0.26
	Autoregression	$y_{t(4)}+ 0.3572y_{t(4)-1} + 0.1745\ y_{t(4)-2}= 1.48e\text{-}10(x+0.3572x_{t(4)-1} + 0.1745x_{t(4)-2}) + 639.54$	730.74	0.25
	Autoregression+ time effect	$y_{t(4)} = 1292 – 1.65t(4) + 1.669dt_{t(4)}$ $y_{t(4)}*= y_{t(4)} + 0.5806y_{t(4)-1}$	7.39	0.90
	Vector Autoregression	$y_{t(4)} = 1.5112y_{t(4)-1} -1.004e^{-11}y_{t(4)-2} -6.72\ x_{t(4)-1} – 2.60e^{-12}x_{t(4)-2} + 1734$	8.31	0.90
5 weeks	Linear	$y_{t(5)} = 1.78e\text{-}10x_{t(5)} + 602.38$	36.19	0.30
	Autoregression	$y_{t(5)}+0.1487y_{t(5)-1} -0.1046\ y_{t(5)-2}= 1.78e\text{-}10(x_{t(5)}-0.1487xt\text{-}1 – 0.1046x_{t(5)-2}) + 602.38$	570.56	0.31
	Autoregression+ time effect	$y_{t(5)} = 1337\text{-}1.784t(5) + 1.985e^{-10}dx_{t(5)}$ $y_{t(5)}*= y_{t(5)} + 0.6359y_{t(5)-1}$	1.60	0.95
	Vector Autoregression	$y_{t(5)} = 1.5559y_{t(5)-1} -6.1112e^{-12}\ y_{t(5)-2} -7.22e^{-1}\ x_{t(5)-1} -0.218e^{-11}\ x_{t(5)-2} +193.243$	4.37	0.95

Table 1S. 15. Modeling results of N2 gene in gc/L of sanitary flow for September 2020 to August 2021 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/L of sanitary flow)	RMSE	Pearson r
3 weeks	Linear	$y_{t(3)} = 0.010x_{t(3)} + 802.33$	5.62	0.10
	Autoregression	$y_{t(3)}-0.11230y_{t(3)-1} -0.1392\ y_{t(3)-2}= 0.010(x_{t(3)}- 0.1123x_{t(3)-1} – 0.1392x_{t(3)-2}) + 802.33$	700.34	0.10
	Autoregression+ time effect	$y_{t(3)} = 2231 + 28.33t(3) -0.001dx_{t(3)}$ $y_{t(3)}*= y_{t(3)} + 0.5922y_{t(3)-1}$	15.33	0.74
	Vector Autoregression	$y_{t(3)} = 1.0334y_{t(3)-1} +1.3420\ y_{t(3)-2} -0.003\ x_{t(3)-1} + 0.001x_{t(3)-2} +83.22$	9.90	0.73
4 weeks	Linear	$y_{t(4)} = 0.052x_{t(4)} + 600.30$	8.33	0.31
	Autoregression	$y_{t(4)}-0.1363y_{t(4)-1} + 0.2311\ y_{t(4)-2}= 0.052(x_{t(4)}- 0.1363x_{t(4)-1} + 0.2311x_{t(4)-2}) + 600.30$	500.62	0.31
	Autoregression+ time effect	$y_{t(4)} = 1345 + 5.22t(4) -0.005\ dx_{t(4)}$ $y_{t(4)}*= y_{t(4)} + 0.5277y_{t(4)-1}$	7.33	0.91
	Vector Autoregression	$y_{t(4)} = 1.324y_{t(4)-1} +0.0021\ y_{t(4)-2} +0.1942x_{t(4)-1} - 0.1230xt\text{-}2 + 730.52$	7.62	0.87
5 weeks	Linear	$y_{t(5)} = 0.065x_{t(5)} + 469.92$	2.36	0.40
	Autoregression	$y_{t(5)}+0.1044y_{t(5)-1} -0.2935\ y_{t(5)-2}= 0.065(0.1044x_{t(5)-1} – 0.2935x_{t(5)-2}) + 469.92$	100.95	0.40
	Autoregression+ time effect	$y_{t(5)} = 1553 + 0.345t(5) + 3.12dx_{t(5)}$ $y_{t(5)}*= y_{t(5)} + 0.6391y_{t(5)-1}$	1.42	0.95
	Vector Autoregression	$y_{t(5)} = 1.2039y_{t(5)-1} -0.3341\ y_{t(5)-2} -1.3423\ x_{t(5)-1}+0.0007\ x_{t(5)-2} + 300.49$	1.03	0.96

Table 1S. 16. Modeling results of N1 gene in gc/L for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/L)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 1.26x_{t(0)} - 206.25$	0.082	0.53
	Autoregression	$y_{t(0)}-0.0713y_{t(0)-1} = 1.26(x_{t(0)}-0.0713x_{t(0)-1}) - 206.25$	272.53	0.47
	Autoregression+ time effect	$y_{t(0)} = 813.90 + 48.33t(0) + 1.15dx_{t(0)}$ $y_{t(0)}*= y_{t(0)} + 0.5267y_{t(0)-1}$	1.990	0.90
	Vector Autoregression	$y_{t(0)} = 1.1919y_{t(0)-1} +0.4965\ y_{t(0)-2} -0.4097\ x_{t(0)-1} + 0.0237x_{t(0)-2} -430.5077$	0.023	0.94
1 weeks	Linear	$y_{t(1)} = 1.48x_{t(1)} - 717.29$	11.827	0.60
	Autoregression	$y_{t(1)}-0.4613y_{t(1)-1} -0.5982\ y_{t(1)-2}= 1.48(x_{t(1)}-0.4613x_{t(1)-1} - 0.5982x_{t(1)-2}) - 717.29$	182.92	0.58
	Autoregression+ time effect	$y_{t(1)} = 924.27 + 41.98t(1) + 1.47\ dx_{t(1)}$ $y_{t(1)}*= y_{t(1)} + 0.3047y_{t(1)-1}$	1.193	0.90
	Vector Autoregression	$y_{t(1)} = 1.2909y_{t(1)-1} +0.0946\ y_{t(1)-2} -0.3850\ x_{t(1)-1} - 0.1435x_{t(1)-2} + 262.92$	0.024	0.92
2 weeks	Linear	$y_{t(2)} = 1.35x_{t(2)} - 411.50$	4.0017	0.62
	Autoregression	$y_{t(2)} - 0.3419y_{t(2)-1} -0.5811\ y_{t(2)-2}= 1.35(0.3419x_{t(2)-1} - 0.4811x_{t(2)-2}) - 411.50$	309.72	0.60
	Autoregression+ time effect	$y_{t(2)} = 1062 + 41.92t(2) + 1.375dx_{t(2)}$ $y_{t(2)}*= y_{t(2)} + 0.3444y_{t(2)-1}$	1.468	0.93
	Vector Autoregression	$y_{t(2)}= 1.2019y_{t(2)-1} -0.23\ y_{t(2)-2} -0.181\ x_{t(2)-1} -0.222\ x_{t(2)-2} - 752.19$	0.25	0.94

Table 1S. 17. Modeling results of N1 gene in gc/d for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/d)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 8.02e\text{-}10\ x_{t(0)} + 820.33$	1772	0.32
	Autoregression	$y_{t(0)}\text{-}0.1424y_{t(0)-1}\ \text{-}0.1559\ y_{t(0)-2} = 8.02e\text{-}10(x_{t(0)}\text{-}0.1424x_{t(0)-1} - 0.1559x_{t(0)-2}) + 820.33$	>1000	0.32
	Autoregression+ time effect	$y_{t(0)} = 813\ \text{-}4.83t(0) - 7.17e\text{-}10dx_{t(0)}$ $y_{t(0)}^{*} = y_{t(0)} + 0.5205y_{t(0)-1}$	123.22	0.86
	Vector Autoregression	$y_{t(0)} = 1.1821y_{t(0)-1} + 2.8964e^{-10}\ y_{t(0)-2}\ \text{-}3.6328\ x_{t(0)-1}\ \text{-}2.3945e^{-10}x_{t(0)-2}\ \text{-}386.1032$	19.30	0.89
1 weeks	Linear	$y_{t(1)} = 1.24e\text{-}9x_{t(1)} + 121.02$	123.56	0.57
	Autoregression	$y_{t(1)}\text{-}0.332y_{t(1)-1}\ \text{-}0.0629\ y_{t(1)-2} = 1.62e\text{-}10(x_{t(1)}\text{-}0.3245x_{t(1)-1} - 0.0629x_{t(1)-2}) + 628.35$	>1000	0.57
	Autoregression+ time effect	$y_{t(1)} = 950\ \text{-}4.03t(1)\ \text{-}1.77dx_{t(1)}$ $y_{t(1)}^{*} = y_{t(1)} + 0.5856y_{t(1)-1}$	325.33	0.90
	Vector Autoregression	$y_{t(1)} = 1.2767y_{t(1)-1} + 1.3001e^{-10}y_{t(1)-2} - 4.0889x_{t(1)-1} + 2.3333e^{-10}\ x_{t(1)-2}\ \text{-}2.9652$	9.27	0.92
2 weeks	Linear	$y_{t(2)} = 1.37e\text{-}9x_{t(2)}\ \text{+}2.14$	135.22	0.60
	Autoregression	$y_{t(2)}\text{+}0.2944y_{t(2)-1} + 0.0950\ y_{t(2)-2} = 1.37e^{-9}\ (x_{t(2)}\text{-}0.2944x_{t(2)-1} - 0.0950x_{t(2)-2}) + 2.14$	>1000	0.61
	Autoregression+ time effect	$y_{t(2)} = 1062 + 41.93_{t(2)} + 1.31e^{-9}\ dx_{t(2)}$ $y_{t(2)}^{*} = y_{t(2)} + 0.3450y_{t(2)-1}$	15.30	0.85
	Vector Autoregression	$y_{t(2)} = 1.3666y_{t(2)-1}\ \text{-}2.07193\text{-}10\ y_{t(2)-2}\ \text{-}4.5682\ x_{t(2)-1}\ \text{-}1.3492e^{-10}x_{t(2)-2}\ \text{+}283.62$	14.32	0.90

Table 1S. 18. Modeling results of N1 gene in gc/L of sanitary flow for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N1, gc/L of sanitary flow)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 0.17x_{t(0)} + 858.95$	49.83	0.27
	Autoregression	$y_{t(0)}+0.1611y_{t(0)-1} +0.1654\ y_{t(0)-2}= 0.17(x_{t(0)}- 0.1611x_{t(0)-1} + 0.1654x_{t(0)-2}) + 858.95$	1023	0.25
	Autoregression+ time effect	$y_{t(0)} = 813 + 48.33t(0) + 0.153dx_{t(0)}$ $y_{t(0)}*= y_{t(0)} + 0.5165y_{t(0)-1}$	13.93	0.90
	Vector Autoregression	$y_{t(0)} = 1.2922y_{t(0)-1} +0.0566\ y_{t(0)-2} -0.4510\ x_{t(0)-1} - 0.0495x_{t(0)-2} -337.9133$	12.47	0.89
1 weeks	Linear	$y_{t(1)} = 0.27x_{t(1)} + 157.49$	56.25	0.49
	Autoregression	$y_{t(1)} +0.2493y_{t(1)-1} -0.1654\ y_{t(1)-2}= 0.27(x_{t(1)}- 0.2493x_{t(1)}-1 – 0.1654x_{t(1)-2}) + 157.49$	227.38	0.50
	Autoregression+ time effect	$y_{t(1)} = 945- 40.61t(1) +0.26\ dx_{t(1)}$ $y_{t(1)}*= y_{t(1)} + 0.5262y_{t(1)-1}$	10.90	0.90
	Vector Autoregression	$y_{t(1)} = 1.0422y_{t(1)-1} -0.3233\ y_{t(1)-2} -0.3926\ x_{t(1)-1} + 0.1224x_{t(1)-2} + 208.33$	10.73	0.90
2 weeks	Linear	$y_{t(2)} = 0.33x_{t(2)} – 101.06$	74.69	0.56
	Autoregression	$y_{t(2)}+0.2557y_{t(2)-1} -0.1652y_{t(2)-2}= 0.04(x_{t(2)}- 0.0572x_{t(2)-1} – 0.0444x_{t(2)-2}) + 607.55$	245.97	0.57
	Autoregression+ time effect	$y_{t(2)} = 1177-34.58t(2)-0.3147dx_{t(2)}$ $y_{t(2)}*= y_{t(2)} + 0.5233y_{t(2)}-1$	13.48	0.89
	Vector Autoregression	$y_{t(2)} = 1.2521y_{t(2)-1} -0.0033y_{t(2)-2} -0.4322\ x_{t(2)-1} - 0.0009\ x_{t(2)-2} +283.02$	12.07	0.90

Table 1S. 19. Modeling results of N2 gene in gc/L for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/L)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 1.40x_{t(0)} - 447.13$	10.65	0.48
	Autoregression	$y_{t(0)} + 0.0356y_{t(0)-1} = 1.40(x_{t(0)} - 0.0356x_{t(0)-1}) - 447.13$	712.33	0.37
	Autoregression+ time effect	$y_{t(0)} = 813.83 + 48.30t(0) + 1.26dx_{t(0)}$ $y_{t(0)}^* = y_{t(0)} + 0.5139y_{t(0)-1}$	0.953	0.92
	Vector Autoregression	$y_{t(0)} = 1.1756y_{t(0)-1} + 0.5199\, y_{t(0)-2} - 0.3786\, x_{t(0)-1} + 0.0882x_{t(0)-2} - 640.73$	0.022	0.92
1 weeks	Linear	$y_{t(1)} = 1.72x_{t(1)} - 1106.03$	14.70	0.60
	Autoregression	$y_{t(1)} - 0.4055y_{t(1)-1} - 0.5911\, y_{t(1)-2} = 1.72(x_{t(1)} - 0.4055x_{t(1)-1} - 0.5911x_{t(1)-2}) - 1106.03$	173.7	0.59
	Autoregression+ time effect	$y_{t(1)} = 924.25 + 41.95t(1) + 1.66\, dx_{t(1)}$ $y_{t(1)}^* = y_{t(1)} + 0.3049y_{t(1)-1}$	1.005	0.92
	Vector Autoregression	$y_{t(1)} = 1.3016y_{t(1)-1} + 0.1017\, y_{t(1)-2} - 0.3786\, x_{t(1)-1} - 0.2298x_{t(1)-2} + 367.63$	0.008	0.94
2 weeks	Linear	$y_{t(2)} = 1.55x_{t(2)} - 730.15$	2.68	0.62
	Autoregression	$y_{t(2)} - 0.3212y_{t(2)-1} - 0.5774\, y_{t(2)-2} = 1.35(0.3212x_{t(2)-1} - 0.5774x_{t(2)-2}) - 730.15$	205.03	0.63
	Autoregression+ time effect	$y_{t(2)} = 1062 + 41.93t(2) + 1.506dx_{t(2)}$ $y_{t(2)}^* = y_{t(2)} + 0.3463y_{t(2)-1}$	27.96	0.93
	Vector Autoregression	$y_{t(2)} = 1.1479y_{t(2)-1} - 0.3116\, y_{t(2)-2} - 0.1036\, x_{t(2)-1} - 0.3904\, x_{t(2)-2} + 1120.17$	0.05	0.94

Table 1S. 20. Modeling results of N2 gene in gc/d for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/d)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 6.10e\text{-}10x_{t(0)} + 1089$	157	0.21
	Autoregression	$y_{t(0)}\text{-}0.3354y_{t(0)-1}\ \text{-}0.1977\ y_{t(0)-2} = 6.10e^{-10}(x_{t(0)}\text{-}0.3354x_{t(0)-1} - 0.1977x_{t(0)-2}) + 1089$	>1000	0.24
	Autoregression+ time effect	$y_{t(0)} = 832\ \text{-}9.83t(0) - 5.50e^{-10}\ dx_{t(0)}$ $y_{t(0)}\text{*} = y_{t(0)} + 0.5312y_{t(0)-1}$	123.22	0.89
	Vector Autoregression	$y_{t(0)} = 1.2474y_{t(0)-1} + 2.0000e^{-10}\ y_{t(0)-2}\ \text{-}4.0097\ x_{t(0)-1}\ \text{-}2.4903e^{-10}\ x_{t(0)-2}\ \text{-}366.122$	60.22	0.89
1 weeks	Linear	$y_{t(1)} = 1.16e\text{-}9x_{t(1)} + 247.99$	222.30	0.47
	Autoregression	$y_{t(1)}\text{-}0.2592y_{t(1)-1}\ \text{-}0.0232\ y_{t(1)-2} = 1.16e^{-9}(x_{t(1)}\text{-}0.2592x_{t(1)-1} - 0.0232x_{t(1)-2}) + 247.99$	>1000	0.49
	Autoregression+ time effect	$y_{t(1)} = 924\text{-}4.19t(1)\ \text{-}1.12e^{-9}dx_{t(1)}$ $y_{t(1)}\text{*} = y_{t(1)} + 0.5856y_{t(1)-1}$	14.29	0.90
	Vector Autoregression	$y_{t(1)} = 1.294y_{t(1)-1}\ \text{-}\ 2.7231e^{-13}\ y_{t(1)-2} - 0.4789x_{t(1)-1} + 1.8463e^{-10}\ x_{t(1)-2} + 30.074$	20.07	0.92
2 weeks	Linear	$y_{t(2)} = 1.31e\text{-}9x_{t(2)} + 107$	181.22	0.51
	Autoregression	$y_{t(2)}\text{+}0.2642y_{t(2)-1} + 0.0592\ y_{t(2)-2} = 1.31e^{-9}(x_{t(2)}\text{-}0.2642x_{t(2)-1} - 0.0592x_{t(2)-2}) + 107.23$	>1000	0.53
	Autoregression+ time effect	$y_{t(2)} = 106 + 40.33t(2) + 1.26e^{-9}dx_{t(2)}$ $y_{t(2)}\text{*} = y_{t(2)} + 0.3402y_{t(2)-1}$	15.62	0.89
	Vector Autoregression	$y_{t(2)} = 1.3426y_{t(2)-1}\ \text{-}2.2162\text{-}10\ y_{t(2)-2}\ \text{-}0.4196\ x_{t(2)-1}\ \text{-}5.4200e^{-11}x_{t(2)-2} + 389.185$	8.86	0.90

Table 1S. 21. Modeling results of N2 gene in gc/L of sanitary flow for August 2021 to February 2022 (measurements are based on the VIRADEL method)

Lag time	Model	Equation (N2, gc/L of sanitary flow)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 0.11x_{t(0)} + 1265$	11.33	0.15
	Autoregression	$y_{t(0)}+0.4132y_{t(0)-1} +0.1114\, y_{t(0)-2}= 0.11(x_{t(0)}- 0.4132x_{t(0)-1} + 0.1114x_{t(0)-2}) + 1265$	>1000	0.19
	Autoregression+ time effect	$y_{t(0)} = 810 + 48.21t(0) + 0.0973dx_{t(0)}$ $y_{t(0)}*= y_{t(0)} + 0.5273y_{t(0)-1}$	13.89	0.87
	Vector Autoregression	$y_{t(0)} = 1.3329y_{t(0)-1} +0.0397\, y_{t(0)-2} -0.4705\, x_{t(0)-1} - 0.0547x_{t(0)-2} -301.0673$	10.44	0.85
1 weeks	Linear	$y_{t(1)} = 0.23x_{t(1)} + 391$	85.43	0.38
	Autoregression	$y_{t(1)} +0.1417y_{t(1)-1} -0.0218\, y_{t(1)-2}= 0.23(x_{t(1)}- 0.1417x_{t(1)-1} – 0.0218x_{t(1)-2}) + 391$	830.22	0.38
	Autoregression+ time effect	$y_{t(1)} = 924- 41.98t(1) +0.237\, dx_{t(1)}$ $y_{t(1)}*= y_{t(1)} + 0.3673y_{t(1)-1}$	12.35	0.89
	Vector Autoregression	$y_{t(1)} = 1.2811y_{t(1)-1} -0.0005\, y_{t(1)-2} -0.4350\, x_{t(1)-1} + 0.0563x_{t(1)-2} – 82.8092$	9.81	0.91
2 weeks	Linear	$y_{t(2)} = 0.30x_{t(2)} +120.30$	49.73	0.46
	Autoregression	$y_{t(2)}+0.2021y_{t(2)-1} -0.0243y_{t(2)-2}= 0.30(x_{t(2)}- 0.2021x_{t(2)-1} – 0.0243x_{t(2)-2}) + 120.30$	28.36	0.48
	Autoregression+ time effect	$y_{t(2)} = 1062-41.95t(2)-0.288dx_{t(2)}$ $y_{t(2)}*= y_{t(2)} + 0.3499y_{t(2)-1}$	13.52	0.89
	Vector Autoregression	$y_{t(2)} = 1.3403y_{t(2)-1} -0.0352y_{t(2)-2} -0.4442\, x_{t(2)-1} - 0.0211\, x_{t(2)-2} +292.4333$	3.34	0.89

Table 1S. 22. Modeling results of N1 gene in gc/L by PEG for August 2021 to February 2022 (measurements are based on the PEG method)

Lag time	Model	Equation (N1, gc/L)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 0.005x_{t(0)} + 1226$	62.147	0.32
	Autoregression	$y_{t(0)}-0.1222y_{t(0)-1} -0.5774\ y_{t(0)-2}= 0.005(x_{t(0)}- 0.1222x_{t(0)-1} -0.5774x_{t(0)-2}) + 1226$	1633	0.31
	Autoregression+ time effect	$y_{t(0)} = 245.20 + 9.83t(0) + 0.0035dx_{t(0)}$ $y_{t(0)}*= y_{t(0)}+ 0.7592y_{t(0)-1}$	170.83	0.64
	Vector Autoregression	$y_{t(0)} = 1.5283y_{t(0)-1} -6.6103\ y_{t(0)-2} -6.8042\ x_{t(0)-1} – 7.222e^{-4}x_{t(0)-2} -430.5077$	2.94	0.84
1 weeks	Linear	$y_{t(1)} = 0.001x_{t(1)} + 1895$	24.74	0.08
	Autoregression	$y_{t(1)}-0.4123y_{t(1)-1} -0.5921\ y_{t(1)-2}= 0.001(x_{t(1)}- 0.4613x_{t(1)-1} – 0.5921x_{t(1)-2}) + 1895$	1924	0.06
	Autoregression+ time effect	$y_{t(1)} = 409.2 + 92.77t(1) – 9.01e^{-4}dx_{t(1)}$ $y_{t(1)}*= y_{t(1)} + 0.4575y_{t(1)-1}$	719.68	0.70
	Vector Autoregression	$y_{t(1)} = 1.3014y_{t(1)-1}-0.0006\ y_{t(1)-2} –0.4453\ x_{t(1)-1} – 0.0011x_{t(1)-2} + 587.16$	30.33	0.86
-1 weeks	Linear	$y_{t(-1)} = 0.007x_{t(-1)} + 882.57$	18.998	0.47
	Autoregression	$y_{t(-1)} – 0.2893y_{t(-1)-1} -0.5712\ y_{t(-1)-2}= 0.007(0.3419x_{t(-1)-1} – 0.5712x_{t(-1)-2}) +882.57$	983.95	0.45
	Autoregression+ time effect	$y_{t(-1)} = -40.495 + 109.64t(-1) + 0.0054dx_{t(-1)}$ $y_{t(-1)}*= y_{t(-1)} + 0.6023y_{t(-1)-1}$	0.934	0.90
	Vector Autoregression	$y_{t(-1)} = 1.4950y_{t(-1)-1} -2.228e^{-4}\ y_{t(-1)-2} -6.608\ x_{t(-1)-1} - 5.257e^{-4}\ x_{t(-1)-2} – 385.00$	8.472	0.93

Table 1S. 23. Modeling results of N2 gene in gc/L by PEG for August 2021 to February 2022 (measurements are based on the PEG method)

Lag time	Model	Equation (N2, gc/L)	RMSE	Pearson r
0 weeks	Linear	$y_{t(0)} = 0.005x_{t(0)} + 1325$	63.69	0.42
	Autoregression	$y_{t(0)}-0.2232y_{t(0)-1} -0.5566\ y_{t(0)-2}= 0.005(x_{t(0)}- 0.2232x_{t(0)-1} -0.5566x_{t(0)-2}) + 1325$	266.91	0.31
	Autoregression+ time effect	$y_{t(0)} = 245.50 + 9.85t(0) - 0.037dx_{t(0)}$ $y_{t(0)}*= y_{t(0)} + 0.6603y_{t(0)-1}$	47.66	0.62
	Vector Autoregression	$y_{t(0)} = 1.5283y_{t(0)-1} -6.6103\ y_{t(0)-2} -6.8042\ x_{t(0)-1} – 7.222e^{-4}x_{t(0)-2} -430.5077$	7.33	0.90
1 weeks	Linear	$y_{t(1)} = 0.0007x_{t(1)} + 1969$	5.17	0.04
	Autoregression	$y_{t(1)}-0.4430y_{t(1)-1} -0.5466\ y_{t(1)-2}= 0.0007(x_{t(1)}- 0.4430x_{t(1)-1} – 0.5466x_{t(1)-2}) + 1969$	1924	0.13
	Autoregression+ time effect	$y_{t(1)} = 409.2 + 92.71t(1) – 0.0010dx_{t(1)}$ $y_{t(1)}*= y_{t(1)} + 0.4573y_{t(1)-1}$	7.33	0.68
	Vector Autoregression	$y_{t(1)} = 1.3014y_{t(1)-1} -0.0006\ y_{t(1)-2} –0.4453\ x_{t(1)-1} – 0.0011x_{t(1)-2} + 587.16$	3.1	0.89
-1 weeks	Linear	$y_{t(-1)} = 0.007x_{t(-1)} + 925.83$	44.79	0.42
	Autoregression	$y_{t(-1)} – 0.2498y_{t(-1)-1} -0.5628y_{t(-1)-2}= 0.007(0.2498x_{t(-1)-1} – 0.5628x_{t(-1)-2}) +925.83$	261.17	0.39
	Autoregression+ time effect	$y_{t(-1)}= -40.494 + 109.64t(-1) + 0.0058dx_{t(-1)}$ $y_{t(-1)}*= y_{t(-1)} + 0.6091y_{t(-1)-1}$	24.07	0.91
	Vector Autoregression	$y_{t(-1)} = 1.4950y_{t(-1)-1} -2.228e^{-4}\ y_{t(-1)-2} -6.608\ x_{t(-1)-1} - 5.257e^{-4}\ x_{t(-1)-2} – 385.00$	1.187	0.92

Table 1S. 24. Pearson's r between 7-day moving average of total cases (with lag times by days) and total N1 and N2 concentrations (VIRADEL method) between 8/1/21 and 2/28/22 amid the Omicron surge

Lag time (by day)	N1	N2
0	0.48	0.43
1	0.52	0.47
2	0.55	0.50
3	0.57	0.52
4	0.58	0.55
5	0.61	0.57
6	0.64	0.60
7	0.69	0.65
8	0.72	0.68
9	0.73	0.70
10	0.74	0.70
11	0.75	0.72
12	0.76	0.73
13	0.74	0.71
14	0.71	0.68
15	0.67	0.64

Table 1S. 25. Pearson's r between 7-day moving average of total cases (with lag times by days) and total N1 and N2 concentrations (PEG method) between 8/1/21 and 2/28/22 amid the Omicron surge

Lag time (by day)	N1	N2
0	0.47	0.42
1	0.44	0.40
2	0.42	0.38
3	0.38	0.34
4	0.35	0.31
5	0.32	0.28
6	0.29	0.25
7	0.24	0.21
8	0.20	0.17
9	0.16	0.13
10	0.12	0.09
11	0.08	0.06
12	0.04	0.01
13	0.00	-0.02
14	-0.04	-0.06
15	-0.07	-0.08

Table 1S. 26. Interceptor population, flow, and water quality parameters between September 1, 2020, and August 30, 2021

Parameter	ONWI	NIEA	DRI
Population Served	840,600	1,482,000	492,000
Flow (MGD)	179 ± 72	165 ± 72	200 ± 55
Estimated Fraction that is Sanitary (%)	33 ± 7	53 ± 12	27 ± 5
BOD (mg/L)	113 ± 39	179 ± 69	70 ± 31
TSS (mg/L)	100 ± 48	191 ± 79	100 ± 47

Notes: (1) ML/d = 10^6 liters per day; MGD = million gallons per day; cBOD = carbonaceous biochemical oxygen demand; TSS = total suspended solids; TP = total phosphorus. (2) Values for flow, BOD, TSS and TP are shown as average ± one standard deviation

Table 1S. 27. Pearson's correlation between N1 and N2 gene concentrations in various units and total COVID-19 cases in city of Detroit, as well as Wayne, Macomb, and Oakland counties with lag times

Lag time	Unit	N1 vs. total cases	N2 vs. total cases
3 weeks	gc/l	0.27	0.28
	gc/d	0.11	0.11
	gc/l of sanitary flow	0.10	0.09
4 weeks	gc/l	0.51	0.52
	gc/d	0.29	0.26
	gc/l of sanitary flow	0.28	0.25
5 weeks	gc/l	0.62	0.64
	gc/d	0.34	0.31
	gc/l of sanitary flow	0.33	0.30

Table 1S. 28. Statistical modeling results between N1 gene concentrations and total COVID-19 cases in city of Detroit, as well as Wayne, Macomb, and Oakland counties

Lag time	Model	RMSE			Pearson r		
		N1, gc/L	N1, gc/d	N1, gc/L of sanitary flow	N1, gc/L	N1, gc/d	N1, gc/L of sanitary flow
3-week	Linear	7.22	135.76	11.78	0.26	0.16	0.09
	Autoregression	135.65	780.00	250.52	0.07	0.07	0.10
	Autoregression+ time effect	10.18	10.18	10.97	0.76	0.75	0.75
	Vector Autoregression	8.32	7.85	8.97	0.72	0.70	0.72
4-week	Linear	7.26	123.56	9.18	0.51	0.51	0.33
	Autoregression	182.92	234.90	635.69	0.50	0.51	0.33
	Autoregression+ time effect	7.50	7.47	7.20	0.92	0.90	0.90
	Vector Autoregression	8.00	7.99	8.62	0.86	0.85	0.85
5-week	Linear	1.83	48.97	2.62	0.62	0.55	0.40
	Autoregression	105.81	417.57	642.83	0.67	0.54	0.40
	Autoregression+ time effect	1.47	1.60	1.60	0.95	0.94	0.95
	Vector Autoregression	0.35	0.53	4.44	0.96	0.95	0.95

Table 1S. 29. Statistical modeling results between N2 gene concentrations and total COVID-19 cases in city of Detroit, as well as Wayne, Macomb, and Oakland counties

Lag time	Model	RMSE			Pearson r		
		N2, gc/L	N2, gc/d	N2, gc/L of sanitary flow	N2, gc/L	N2, gc/d	N2, gc/L of sanitary flow
3-week	Linear	12.40	926.30	5.62	0.28	0.10	0.10
	Autoregression	341.27	901.23	700.34	0.20	0.08	0.10
	Autoregression+ time effect	11.34	12.34	15.33	0.75	0.75	0.74
	Vector Autoregression	8.89	9.90	9.90	0.73	0.72	0.73
4-week	Linear	16.37	104.45	8.33	0.51	0.26	0.31
	Autoregression	132.35	730.74	500.62	0.29	0.25	0.31
	Autoregression+ time effect	9.75	7.39	7.33	0.90	0.90	0.91
	Vector Autoregression	6.88	8.31	7.62	0.86	0.90	0.87
5-week	Linear	13.95	36.19	2.36	0.64	0.30	0.40
	Autoregression	548.14	570.56	100.95	0.65	0.31	0.40
	Autoregression+ time effect	3.21	1.60	1.42	0.94	0.95	0.95
	Vector Autoregression	7.57	4.37	1.03	0.95	0.95	0.96

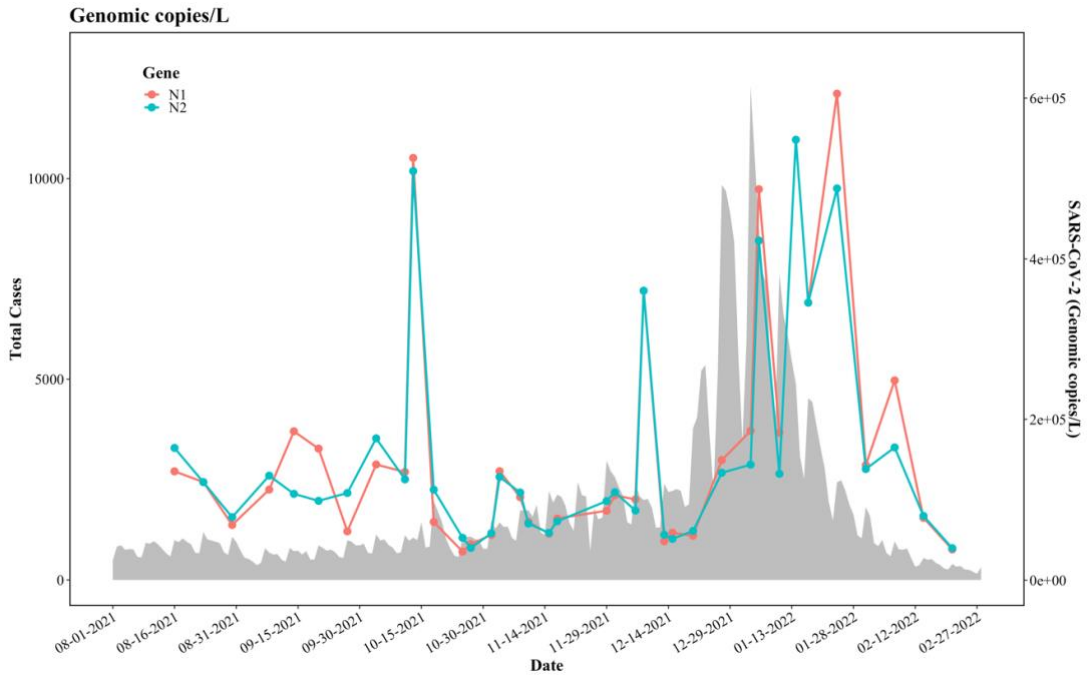


Figure 1S. 1. Total N1 and N2 gene concentrations in gc/L measured by PEG method for the three interceptors and total confirmed COVID-19 cases in the city of Detroit, as well as Wayne, Macomb, and Oakland counties.

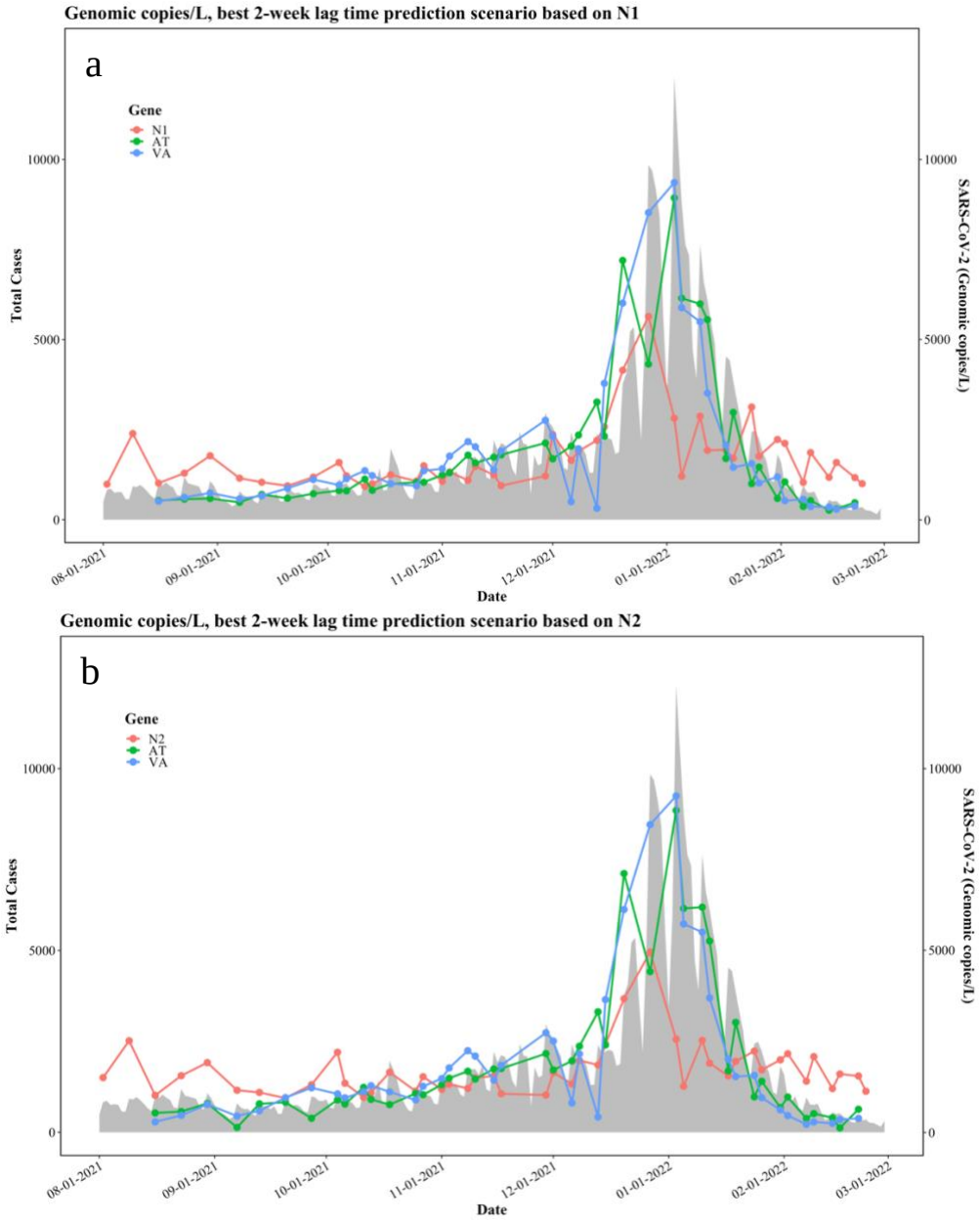


Figure 1S. 2. Modeling results of linear model (L), autoregression model (A), autoregression model with time effect (AT), and vector autoregression model (VA) based on N1 and N2 measurements, between August 1, 2021, and February 28, 2022, amid the Omicron surge, with 2-week lag time. a. based on N1; b. based on N2.

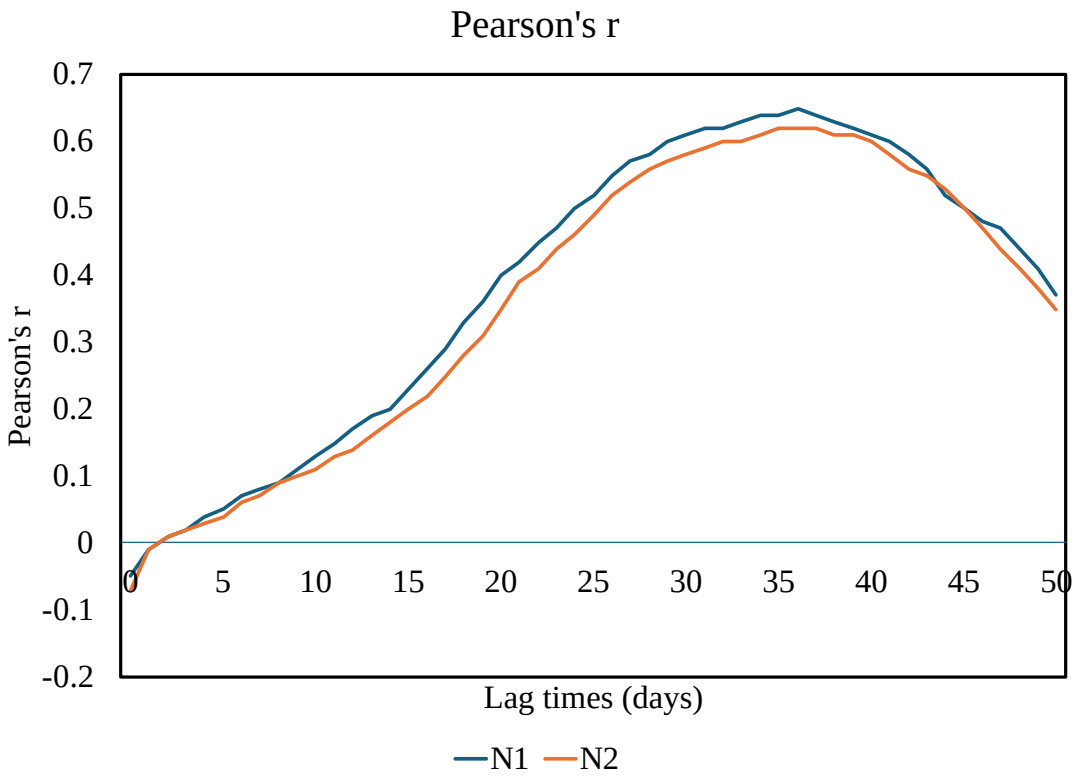


Figure 1S. 3. Pearson's r between 7-day moving average of total cases (with lag times by days) and total N1 and N2 concentrations (VIRADEL method) prior to the Omicron surge, between 9/1/20 and 8/31/21

CHAPTER 2: SIMPLE METHODS FOR EARLY WARNINGS OF COVID-19 SURGES: LESSONS LEARNED FROM 21 MONTHS OF WASTEWATER AND CLINICAL DATA COLLECTION IN DETROIT, MICHIGAN, UNITED STATES

Published in *Science of the Total Environment*:

Zhao, L., Zou, Y., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Norton, J., & Xagoraraki, I. (2023). Simple methods for early warnings of COVID-19 surges: Lessons learned from 21 months of wastewater and clinical data collection in Detroit, Michigan, United States. *Science of The Total Environment, 864*, 161152.

Abstract

Wastewater-based epidemiology (WBE) has drawn great attention since the Coronavirus disease 2019 (COVID-19) pandemic, not only due to its capability to circumvent the limitations of traditional clinical surveillance, but also due to its potential to forewarn fluctuations of disease incidences in communities. One critical application of WBE is to provide "early warnings" for upcoming fluctuations of disease incidences in communities which traditional clinical testing is incapable to achieve. While intricate models have been developed to determine early warnings based on wastewater surveillance data, there is an exigent need for straightforward, rapid, broadly applicable methods for health departments and partner agencies to implement. Our purpose in this study is to develop and evaluate such early-warning methods and clinical-case peak-detection methods based on WBE data to mount an informed public health response. Throughout an extended wastewater surveillance period across Detroit, MI metropolitan area (the entire study period is from September 2020 to May 2022) we designed eight early-warning methods (three real-time and five post-factum). Additionally, we designed three peak-detection methods based on clinical epidemiological data. We demonstrated the utility of these methods for providing early

warnings for COVID-19 incidences, with their counterpart accuracies evaluated by hit rates. “Hit rates” were defined as the number of early warning dates (using wastewater surveillance data) that captured defined peaks (using clinical epidemiological data) divided by the total number of early warning dates. Hit rates demonstrated that the accuracy of both real-time and post-factum methods could reach 100%. Furthermore, the results indicate that the accuracy was influenced by approaches to defining peaks of disease incidence. The proposed methods herein can assist health departments capitalizing on WBE data to assess trends and implement quick public health responses to future epidemics. Besides, this study elucidated critical factors affecting early warnings based on WBE amid the COVID-19 pandemic.

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been spreading worldwide since its first identification in Wuhan, China, in December 2019. Since SARS-CoV-2 persists in human bodily fluids and excretions, including saliva, sputum, urine, and feces, numerous studies have applied wastewater-based epidemiology (WBE), also known as wastewater surveillance, to monitor COVID-19 infections in various global settings (Ahmed et al., 2020a; Ahmed et al., 2021, 2022; Barua et al., 2022; Bivins et al., 2021; Corchis-Scott et al., 2021; Li et al., 2022; Miyani et al., 2020, 2021; Sherchan et al., 2020; Xiao et al., 2022; Zhao et al., 2022; Zhu et al., 2022). WBE is a comparatively inexpensive and less laborious tool than clinical surveillance for tracking disease incidence and/or prevalence within a large-scale community (Safford et al., 2022; Xagoraraki 2020a; Xagoraraki et al., 2020b). WBE provides a comprehensive and anonymous surveillance of both symptomatic and asymptomatic viral disease, making it an ideal complimentary approach to traditional clinical surveillance testing (Bibby et al., 2021; Safford et al., 2022). A recent study

reported that WBE can conserve financial resources without altering surveillance accuracy by replacing some of clinical surveillance programs with WBE-based surveillance (Safford et al., 2022). Perhaps most critically, WBE has the potential to provide early warnings of impending disease outbreaks or surges, if translated effectively to a public health setting (Zhao et al., 2022). "Early warnings" in this context refers to the early detection of relevant pathogen fluctuations within a community, providing a critical window to mount a public health response, prior to lagging parallel trends in clinical cases (Bibby et al., 2021; Olesen et al., 2021). Recent studies have proposed early warning algorithms predicated on intricate statistical models, such as autoregressive time series models (Zhao et al., 2022), artificial neural network models (Zhu et al., 2022), and other advanced statistical and machine learning models (Table 2S. 1.). These sophisticated models are resource- and time-intensive for health departments to calculate and interpret, particularly in the context of a possible emerging threat. Currently, there are scant studies that have investigated or tested early warning methods for COVID-19, using straightforward, reliable, and rapid approaches for health departments. Some attempts to determine early warnings for COVID-19 clinical cases include: using thresholds for wastewater viral RNA concentrations (Zhu et al., 2021b), calculating Epidemic Volatility Index (EVI) (Kostoulas et al., 2021), implementing statistical thresholds such as mean plus two standard deviations (Bowman et al., 2016; Prabdial-Sing et al., 2021), mean and variance (O'Brien et al., 2021), kurtosis and skewness (Harris et al., 2020), assessing the ratio between wastewater viral concentrations and clinical cases (w/c ratio) (D'Aoust et al., 2022; Xiao et al., 2022), estimating the percentage change of wastewater viral concentrations and their relationships to clinical cases (Kumar et al., 2021), etc.

Until now, resources have been spent to generate WBE data that are not fully understood or applied. Wastewater surveillance for pathogens is only beneficial if public health practitioners

and partner agencies can apply the results to inform policy decisions and guide actions. Henceforth, we propose three clinical case peak-defining methods (Table 2. 1.) and eight simple-to-calculate early warning methods (Table 2. 2.) that can be smoothly implemented by public health departments and partner agencies to provide prompt warnings of impending disease incidences or surges. One of the advantages of these methods is that they encompass both real-time and post-factum analyses. Moreover, these methods combine WBE data with clinical data, though a w/c ratio (D'Aoust et al., 2022; Xiao et al., 2022) using the post-factum methods. Lastly, accuracy of these methods was evaluated via "hit rate", which is subsequently defined. Results indicate that hit rates for all real-time methods and four post-factum methods could reach 100% under different circumstances, demonstrating successful discernment of clinical case peaks. Thus, these methods can equip local public health officials with a toolset that integrates wastewater surveillance with traditional clinical surveillance data, to provide early warnings for disease outbreaks or surges, and alert officials and the public when action is needed based on warnings identified by the methods. Besides, we also elucidated the impact of policy changes due to COVID-19, and social events in the Detroit metropolitan area on the wastewater viral concentrations and clinical cases over the past two years. In addition, factors that affect applying WBE for disease surveillance and accuracy of early warning methods are discussed.

2. Materials and Methods

2.1 Sample collection and laboratory analysis

Untreated wastewater samples were collected twice weekly from the Water Resource Recovery Facility (WRRF) of the Great Lakes Water Authority (GLWA) located in Detroit, Michigan, USA, between September 1, 2020, and May 31, 2022. The WRRF receives wastewater via three main interceptors including the Detroit River Interceptor (DRI), the North Interceptor-

East Arm (NIEA), and the Oakwood-Northwest-Wayne County Interceptor (ONWI) from its service area that encompasses greater Detroit. Samples were collected from the three interceptors at the point of discharge into the WRRF. Depending on the suspended solids of wastewater, approximately 10 to 50 L of untreated wastewater passed through NanoCeram electropositive cartridge filters at a rate not more than 11 L/min using a previously described method (Miyani et al., 2021; Zhao et al., 2022). Viruses were eluted within 24 h after sampling, based on a previously described method (Supplementary Materials: Sampling and Virus Elution) (Miyani et al., 2021; Zhao et al., 2022). Bacteriophage Phi6 was used as a proxy virus to evaluate recovery during virus elution and concentration (Kantor et al., 2021; Ye et al., 2016; Zhao et al., 2022). Recoveries obtained ranged from 10.37 % to 58.96 %, with a mean recovery of 24.91 % (±22.89 %) (Zhao et al., 2022). Viral RNA was extracted using Viral RNA QIAGEN kit (QIAGEN, Germantown, MD, USA), following the manufacturer's protocol with the method described previously (Supplementary Materials: RNA Extraction) (Miyani et al., 2021; Zhao et al., 2022). RT-ddPCR was performed on a QX200 AutoDG Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA), using the One-step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA) in a previously described method (Supplementary Materials: RT-ddPCR) (Zhao et al., 2022). The Limit of Blank (LOB) was determined by examining three types of samples using RT-ddPCR, across four consecutive days, including interceptor samples collected before COVID-19 pandemic, nuclease-free water, and negative process control samples from elution and extraction. The samples before COVID-19 pandemic were collected on February 18, 2018, from the ONWI, NIEA, and DRI interceptors at GLWA using the same methods. Limit of Blank (LOB) for N1 gene ddPCR was determined to be 0.09 gc/μL, and the LOB for N2 gene ddPCR was determined to be 0.08 gc/μL (Zhao et al., 2022). Limit of Detection (LOD) of 0.1 gc/μL with 72.92 %

confidence for the N1 gene ddPCR and 0.1 gc/μL with 81.25 % confidence for the N2 gene ddPCR were determined (Zhao et al., 2022).

2.2 WBE and clinical data of COVID-19

Throughout our 21-month surveillance of wastewater in the Detroit metropolitan area, the wastewater surveillance data (September 2020 to May 2022) together with clinical data were implemented with eight proposed early-warning methods and three proposed clinical case peak-defining methods. Publicly available clinical data were accessed on August 30, 2022, for the period between September 25, 2020, and May 31, 2022, for the city of Detroit, as well as Wayne, Macomb, and Oakland counties (michigan.gov) shown in Figure 2. 1. a. Clinical data presented as a 7-day moving average (Menkir et al., 2021) was used for further statistical analysis (Figure 2. 1. b.). COVID-19 data was only available per city or county within study area. Lastly, each interceptor received wastewater from portions of each city or county, thus, only the total SARS-CoV-2 concentrations could be correlated to the total COVID-19 cases in each jurisdiction.

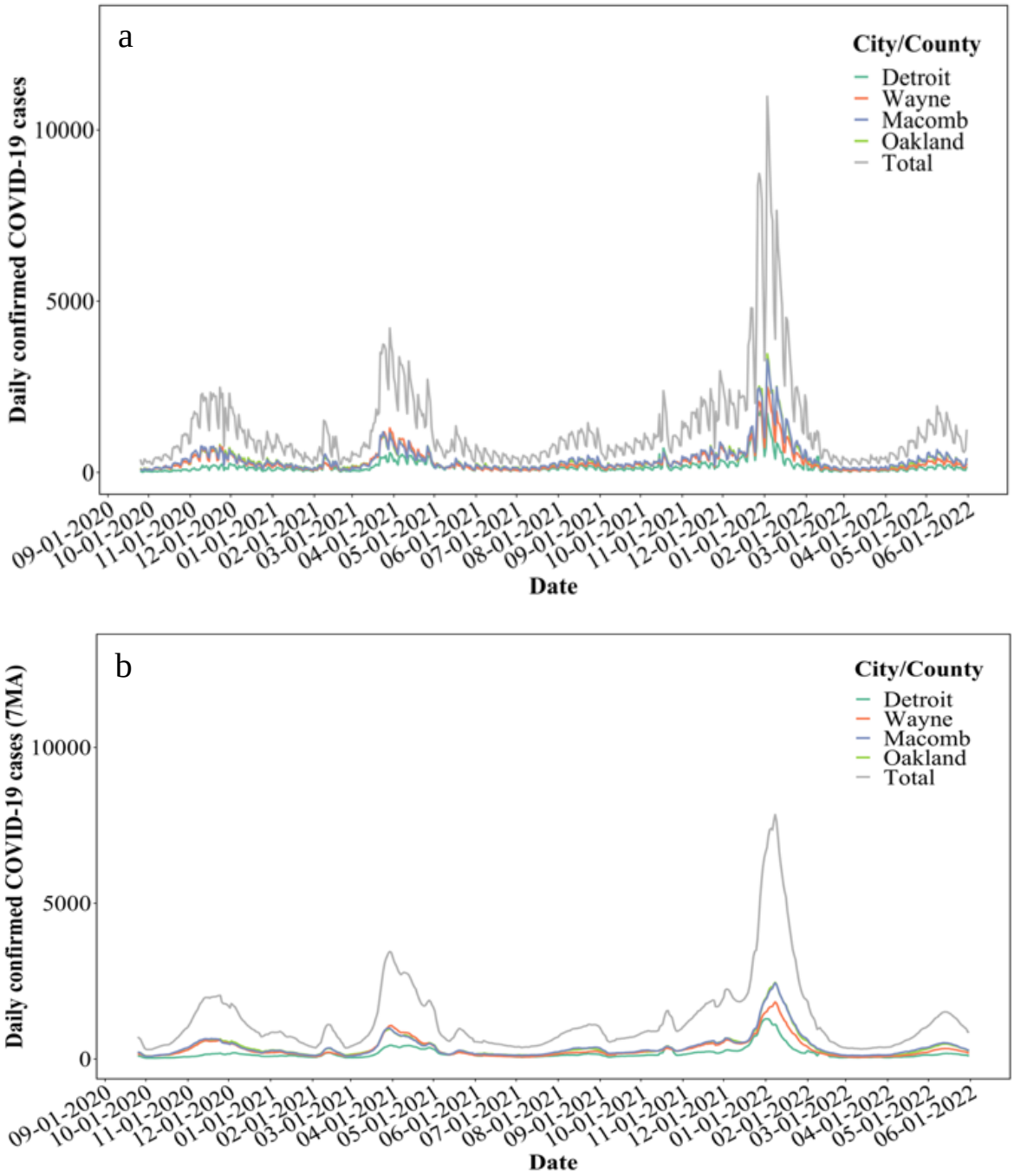


Figure 2. 1. a. COVID-19 cases in the city of Detroit, as well as Wayne, Macomb, and Oakland counties; b. 7-day moving average of COVID-19 cases

2.3 Data analysis and visualization

Data were tracked and organized with Microsoft Excel (version 16.66.1, Microsoft co. ltd). MATLAB of a 2019b edition (MatLab, 2018) and R version 4.1.3 (Team, 2022) were utilized to perform the early warning analyses, depending primarily on the ggplot2 package for visualization, and the DescTools package for standard deviation, mean, variance, skewness, kurtosis, and quantile for calculation. Eq. (1). depicts the ratio between the wastewater viral gene concentrations and clinical cases (w/c ratio) (D'Aoust et al., 2022; Xiao et al., 2022), which was first proposed as an indicator of disease incidence based on wastewater surveillance in a recent study (Xiao et al., 2022).

$C_{N1 \text{ or } N2 \text{ gene}}$: N1 or N2 gene concentrations (genomic copies/L, gc/L)

Clinical case: daily confirmed COVID-19 cases (7-day moving average)

$$\text{w/c ratio} = C_{N1 \text{ or } N2 \text{ gene}} / \text{Clinical case} \quad (1)$$

2.3.1 Methods of defining peaks of COVID-19 cases

Few recent studies have discussed on approaches to defining peaks of clinical cases of COVID-19. O'Brien et al. defined the peak-range of COVID-19 clinical cases to be when the first derivative of cases remains positive for seven consecutive data points (O'Brien et al., 2021). Similarly, in this study, we define the peak-range to be when the 7-day moving average of clinical cases continues increasing for over 14 consecutive days (method I, shown in Table 2. 1. and Figure 2. 2. a.). The uptick begins at day 0, peaks at the maximum, and ends symmetrically. Method II defines the peak where the intersection values are greater than an established variance/mean threshold, shown in Figure 2. 2. b. (Eq. (2)). Similarly, method III uses mean-0.5standard deviation threshold to define the case-peak where the intersection values are greater than the threshold, shown in Figure 2. 2. c. (Eq. (3)). Both methods II and III measure the distribution of the clinical

cases in the Detroit metropolitan area. The specific “peak ranges” or “surges” determined by these three methods are summarized in Table 2. 1.

variance/mean threshold = V_c/M_c (2)

V_c represents the variance of clinical cases

M_c represents the mean of clinical cases

mean – 0.5standard deviation threshold = $M_c - 0.5 \times S_c$ (3)

S_c represents the standard deviation of clinical cases

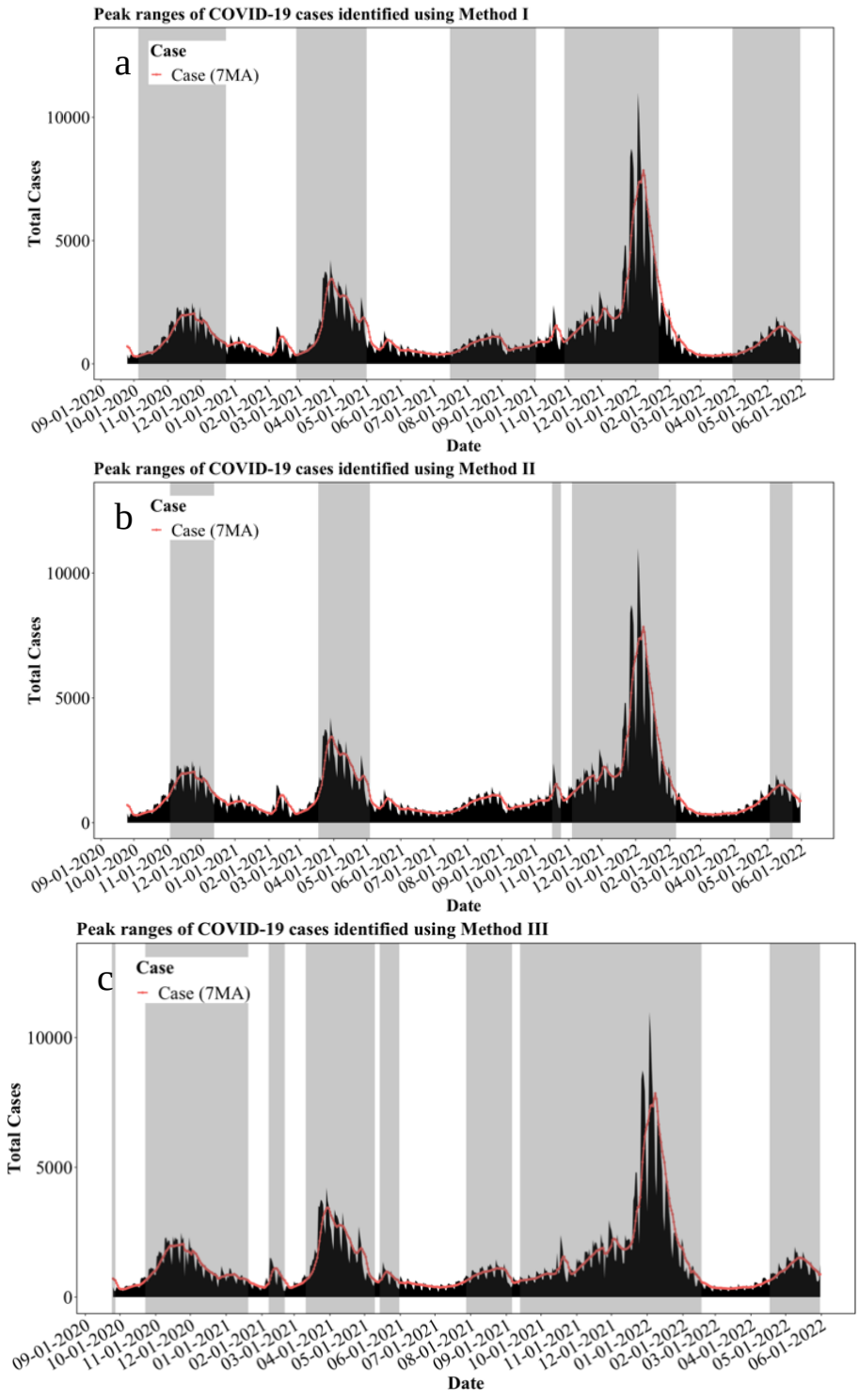


Figure 2. 2. Methods of defining peaks for total COVID-19 cases: a. Method I defined peaks (gray shaded area) of total COVID-19 cases; b. Method II defined peaks (gray shaded area) of total COVID-19 cases; c. Method III defined peaks (gray shaded area) of total COVID-19 cases

2.3.2 Real-time early warning methods

As aforementioned, the peak ranges of clinical cases are defined using three methods, namely, methods I, II, and III (Table 2. 1. and Figure 2. 2.). Subsequently, eight methods of early warnings determination (Table 2. 2.) were proposed based on literature studies that were elucidated in section 1. Among them, OBMN1N2, PPCN1N2, and PPCS200 are applied to real-time analysis. OBMN1N2 is applied to real-time N1 and N2 gene concentrations (gc/L) in wastewater, with first early warning dates determined on the final of three consecutively increasing measurements. This method (OBMN1N2) reduces the possibility a "false warning" due to high possible variations of the measured data since OBM requires three consecutive increasing data points to issue a warning. It is also a "non-quantification" method, consistently applicable regardless of the degree of increasing values. The PPCN1N2 method identifies early warnings when the positive percentage change of N1 and N2 gene concentrations (gc/L) are greater than 40% (Kumar et al., 2021), depicted as Eq. (4):

$$PPC_{N1N2} = (C_{N1 \text{ or } N2 \text{ gene (n)}} - C_{N1 \text{ or } N2 \text{ gene (n-1)}}) / C_{N1 \text{ or } N2 \text{ gene (n-1)}} \times 100\% \quad (4)$$

n indicates the current measurement

n-1 indicates the previous measurement

Based on the characteristics of our WBE dataset for the Detroit area, PPCS200 requires the positive percentage change of slope of N1 and N2 gene concentrations (gc/L) to be greater than 200% to issue early warnings which is depicted in Eq. (6). Percentage change of the slope (Sk) measures the degree of increase between values and can identify values that increased significantly. We have assigned the threshold to be 200% for the PPCS method, to capture the most meaningful warnings from our WBE data:

$$\text{Slope } S_k = (C_{N1 \text{ or } N2 \text{ gene (n)}} - C_{N1 \text{ or } N2 \text{ gene (n-1)}}) / (\text{Date}_{(n)} - \text{Date}_{(n-1)}) \quad (5)$$

$PPCS_{200} = (S_k - S_{(k-1)}) / S_{(k-1)} * 100\%$ (6)

n indicates the current measurement or date; n-1 indicates the previous measurement or date; k indicates the current slope; k-1 indicates the previous slope.

2.3.3 Post-factum early warning methods

Post-factum methods identify early warnings when wastewater surveillance data exceed the thresholds proposed in this study. These methods are designed for post-factum implementations, where both wastewater gene concentration data and clinical data have been reported. An early warning is triggered when the threshold criteria is exceeded. The threshold is computed for an investigation period of interest, after the surges of disease have occurred. For instance, researchers have proposed two standard deviations as an early warning threshold and the time of early warning was determined by the first time the signal exceeded the threshold (Drake et al., 2010). Five statistical thresholds including mean plus two standard deviations (MSD), variance divided by mean (VAM), skewness (SKE), kurtosis (KUR), and 90th-percentile (PER90), were calculated using N1 and N2 gene concentrations (gc/L) and w/c ratio (gc/L/case), to determine early warnings. MSD targets the upper bound limit generated by the two standard deviations away from the mean (Gao et al., 2021; Wang et al., 2017), which is equivalent to a 95% confidence interval. While PER90 targets the top 10% of data from the distribution, other studies have applied 70th percentiles, or 80th percentiles to inform early warnings for hand, foot, and mouth disease in China (Gao et al., 2021). VAM is a similar method to MSD, which identifies the variability of the data away from the mean. SKE measures asymmetry of distribution about its mean, and KUR measures the combined weight of a distribution's tails to its center (i.e., whether the plotted shape of the distribution is too sharply "peaked") (Harris et al., 2020).

Table 2. 1. Methods of defining peak-ranges of confirmed COVID-19 cases

Method	Method description	Early warning level / Cutoff level	Peak range
Method I	Data sequence numerical increase or decrease	Continual increase for $\geqslant$ 14 consecutive data points. Peak range begins at day 0. Peaks at the maximum, and ends symmetrically	10/6/20 – 12/24/20, 2/27/21 – 5/1/21, 7/17/21 – 10/2/21, 10/29/21 – 1/22/22, 3/31/22 – 5/31/22
Method II	Variance / mean	Intersection values > variance/mean threshold	11/4/20-12/13/20, 3/19/21-5/4/21, 10/18/21-10/25/21, 11/5/21-2/7/22, 5/4/22-5/24/22
Method III	Mean – 0.5standard deviation (SD)	Intersection values > mean-0.5SD threshold	9/25/20-9/27/20, 10/24/20-1/20/21, 2/8/21-2/21/21, 3/12/21-5/10/21, 5/15/21-5/31/21, 7/29/21-9/6/21, 9/14/21-2/17/22, 4/18/22-5/31/22

Table 2. 2. Early warning methods

Type of analysis	Early warning parameter	Early warning level / cutoff level	Abbreviation	Data type
Real-time analysis	Data sequence numerical increase or decrease	Keep increasing for 3 consecutive data points	OBM_{N1N2}*	N1 N2 gene concentrations
	Positive percentage change	> 40%	PPC_{N1N2}	N1 N2 gene concentrations
	Positive percentage change of slope	> 200%	$PPCS_{200}$*	N1 N2 gene concentrations
Post-factum analysis	Mean + 2 standard deviation	Intersection values higher than the corresponding threshold	MSD (B1) *	N1 N2 gene concentrations and w/c ratio
	Variance / mean		VAM (B2)	
	Skewness		SKE (B3)	
	Kurtosis		KUR (B4)	
	90 percentile		PER_{90} (B5) *	

Note: (1) B1, B2, B3, B4, B5 are short representation of each statistical method for visualization purposes. (2) * marked methods were demonstrated in figures in the main text.

2.3.4 Hit rate

Hit rate is introduced to appraise the accuracy of each method – it is defined as the ratio of the number of early warning dates "m" capturing the defined peaks to the total number of early warning dates "n" (see Eq. (7)). The hit rate was calculated in this manner for all early warning methods. As reported by our recent study (Zhao et al., 2022), wastewater signals of N1 and N2 genes preceded the reported clinical cases by up to 5 weeks in the Detroit metropolitan area. Thus, for this study, the number of early warning dates "m" that are said to capture defined peaks must satisfy two criteria: (1) identified early warning dates are located inside the defined peak regions (shaded gray areas in Figure 2. 2.); (2) identified early warning dates are located within a five-week window preceding the defined peak regions.

Hit rate = (number of early warning dates "m" capturing defined peaks) / (total number of identified early warning dates "n") × 100% (7)

m: number of early warning dates identified by eight early warning methods, capturing defined

peaks, identified by three peak-defining methods

n: total number of early warning dates identified by eight early warning methods

3. Results and Discussion

3.1 Wastewater viral concentrations precede disease incidence and can relate to public health policy or community social events

Analysis of our 21-month wastewater surveillance data reveals that the trend of total N1 and N2 gene concentrations preceded and forewarned the trend of total COVID-19 clinical cases (Figure 2. 3. a.). Both wastewater viral concentrations and clinical data were compared with calendar dates of major statewide, citywide, and countywide public health policies (Figure 2. 3. a.). For instance, both N1 and N2 gene concentrations began to increase shortly after the State of Michigan allowed the reopening of gyms, pools, and permitted organized sports on September 3, 2020. This is suggestive of populations shedding the virus into wastewater after being infected by COVID-19 likely due to unregulated social gatherings (Figure 2. 3. a.). Subsequently, both N1 and N2 gene concentrations began to gradually decrease, potentially due to the reduction in SARS-CoV-2 shedding, which persisted up to 24 days (Zhao et al., 2022), as well as due to the implementation of new COVID-19 public health orders and guidelines for Detroit on October 9 and October 14, 2020, respectively (Figure 2. 3. a.). Decreasing trends of both wastewater viral concentrations and clinical cases were observed after city of Detroit extended the emergency epidemic order on January 1, 2022 (Figure 2. 3. a.). Xiao et al. reported similar trends of wastewater viral concentrations, as affected by state public health policy in Massachusetts, USA (Xiao et al., 2022). Major public health orders and guidelines can affect wastewater viral concentrations as well as subsequent COVID-19 clinical cases by regulating everyday social gatherings and mitigation efforts (Xiao et al., 2022). Wastewater viral concentrations and clinical

data were also compared with public holidays and known large-scale social events in the Detroit metropolitan area (Figure 2. 3. b.). Public holidays and social events celebrated in Detroit were seen to be reflected in both the wastewater viral concentration and the clinical data (Figure 2. 3. b.). It was observed that both N1 and N2 gene concentrations increased after Labor Day (September 7th, 2020) (Figure 2. 3. b.), likely resulting from social gatherings during the holiday, as well as the policy of easing COVID-19 restrictions (September 3rd, 2020) in Michigan (Figure 2. 3. a.), leading to potentially high transmissions of COVID-19. Similarly, both N1 and N2 gene concentrations began to increase after Martin Luther King Jr. Day (January 18th, 2021) and peaked shortly after Presidents Day (February 15th, 2021). This increase in gene concentration preceded an increase in COVID-19 cases by 4 to 5 weeks (Zhao et al., 2022). Similar observations can be identified with a steeper increase of wastewater viral concentrations as well as the clinical cases after Veterans Day (November 11th, 2021). Clinical cases surged again in early January of 2022 after the Thanksgiving, Christmas, and New Year's day holidays, when social gatherings might be expected during holiday celebrations (Figure 2. 3. b.). Notably, the increase of both N1 and N2 gene concentrations in early September of 2020 and early February of 2021, preceding the increase of clinical cases for 4 to 5 weeks (Zhao et al., 2022), can be related to opening of schools, colleges, and universities in fall and spring semesters (Xiao et al., 2022).

Wastewater surveillance has the ability to monitor virtually most members of a community (with an integrated sewage system), regardless of the presence of disease symptoms or inequity in testing accessibility (Bibby et al., 2021). To quantitively compare wastewater data and clinical cases, a ratio (w/c ratio) between the wastewater viral concentrations and 7-day moving average of daily confirmed COVID-19 clinical cases is adopted from recent studies for this purpose (D'Aoust et al., 2022; Xiao et al., 2022). The w/c ratio could reflect potential undercounting or

overcounting of actual clinical cases (D'Aoust et al., 2022; Xiao et al., 2022). Undercounting occurs when people do not seek clinical testing, or when access to testing is restricted due to resource limitations, or when there is an elevated rate of asymptomatic infections. This scenario was evident in the summer of 2021 from June to August where the w/c ratio of both N1 and N2 genes was high, but the confirmed cases were low (Figure 2. 3. c.). Furthermore, during the summer, cases were likely to be undercounted due to lack of testing and potentially increasing spreading during summer social activities. Some studies have reported similar issues that can result in undercounting, when the actual number of cases is 12 times larger than reported cases (Lau et al., 2021), and the case-to-report ratio could reach 26 to 32 at the early stage of the pandemic when testing sources were limited (Murhekar et al., 2021). Conversely, overcounting can occur when testing resources are abundant and infected populations get tested multiple times during the entire period of infection. These individuals are counted and reported repeatedly as individual clinical cases, since the shedding duration of SARS-CoV-2 can persist up to 24 days prior to the Omicron surge (Zhao et al., 2022) and throat/nasal swab PCR tests can also remain positive up to nearly 20 days (Xiao et al., 2022). Note that the introduction of w/c ratio also eliminates the noises of N1 and N2 gene concentrations among peaks and accentuate the prominent peaks of N1 and N2 gene w/c ratio preceding major peaks of clinical cases (Figure 2. 3. c.). From the inception of the pandemic, the increasing trend of w/c ratio in September and October 2020 provided early warnings of upcoming peaks of clinical cases in late October and November 2020 (Figure 2. 3. c.). Similarly, the peaks of w/c ratio in late February 2021 forewarned the impending surge of clinical cases in March 2021. The w/c ratio stayed relatively low and stable before the peak of clinical cases in late December 2021 and early January 2022, which indicated that testing sources were sufficient. Moreover, the State of Michigan (michigan.gov/coronavirus) reported that testing

capacity statewide increased to approximately 50,000 test results per day in December 2021, corroborating the accuracy of the low w/c ratio in this period. Because as testing capacity increased throughout November and December of 2021, delayed clinical cases were likely reported simultaneously with newly reported cases but the delayed clinical cases did not contribute to the current wastewater viral concentrations, perhaps resulting from reduced viral shedding, leading to lower w/c ratio. Moreover, the surge of the Omicron variant in November and December 2021 in the Detroit metropolitan area could have contributed to shifting transmission dynamics, leading to a substantial increase in reported clinical cases, and thus, contributing to the relatively low w/c ratio (Auwaerter et al., 2022; Long et al., 2022; Wiersinga et al., 2020; Zhao et al., 2022).

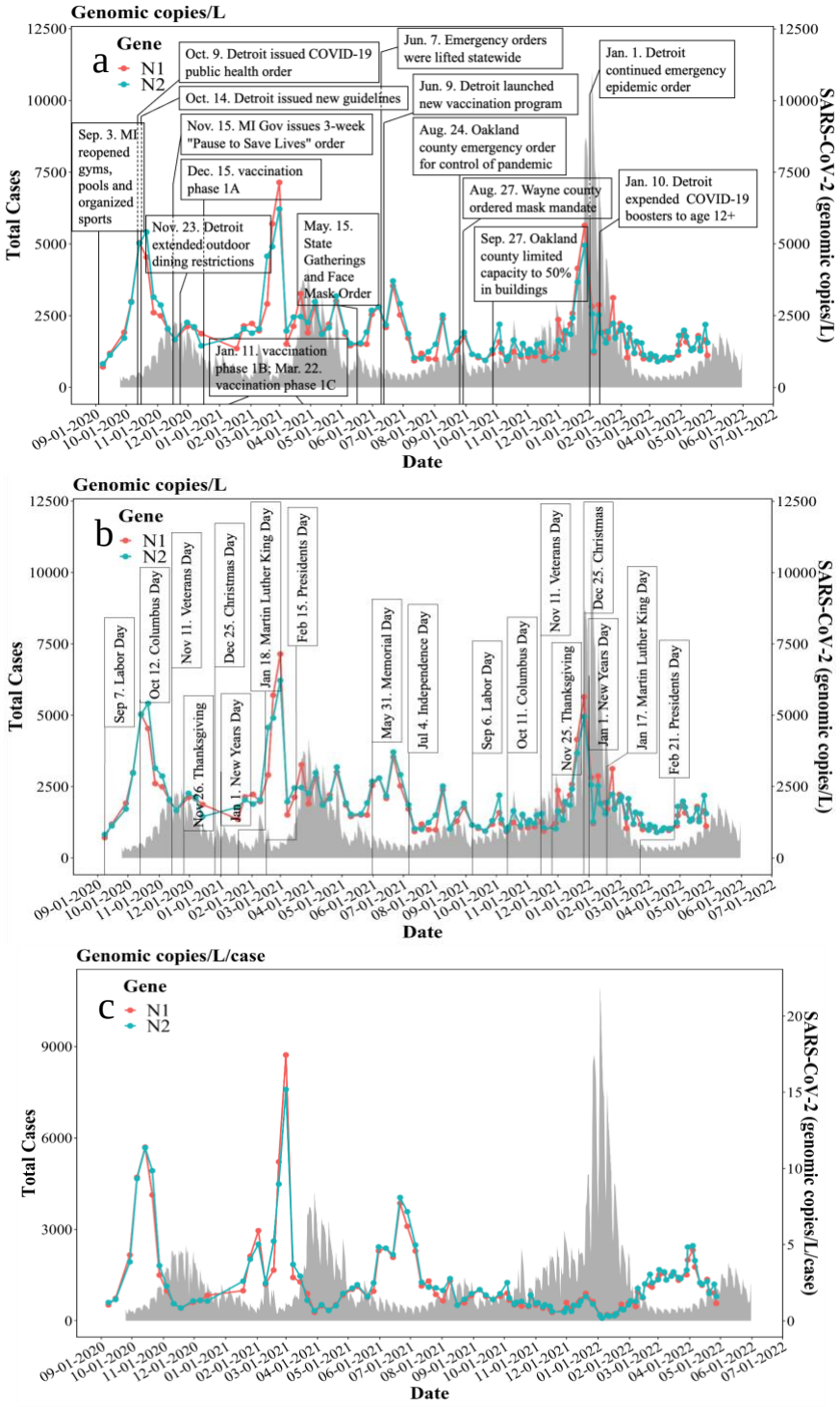


Figure 2. 3. a. Major statewide, citywide, and countywide COVID-19 public health policies in the Detroit metropolitan area; b. Major public holidays in Michigan, USA; c. w/c ratio between N1 N2 gene concentrations and 7-day moving average of total COVID-19 cases

3.2 Early warnings of COVID-19

Eight early warning methods including both real-time and post-factum methods shown in Table 2. 2. were implemented on N1 and N2 gene concentrations (gc/L) as well as w/c ratio (gc/L/case) to identify early warnings for defined peaks of clinical cases. The real-time methods include OBM, PPC, and PPCS, which were applied to direct measurements of N1 and N2 gene concentrations. The post-factum methods include MSD, VAM, SKE, KUR, and PER, which were applied to N1 and N2 gene concentrations as well as w/c ratio. The accuracy of each method was evaluated by hit rates (Table 2S. 3.).

3.2.1 Early warnings determined by real-time methods

Among real-time methods, OBM method was applied to N1 gene concentration (gc/L) and could reach 100% hit rate with method I defined peak (Figure 2. 4. a.), where all identified early warnings (shown as blue vertical lines in Figure 2. 4. a.) are in the defined-peak regions or within a five-week window ahead of the defined-peak regions (shown as gray shaded area in Figure 2. 4. a.). In other words, a 100% hit rate is representative of all identified early warnings successfully forewarning subsequent defined peaks in cases. OBM method could reach 80% hit rates with both method II and III defined peaks shown in Figures 2. 4. b. and 2. 4. c., respectively. Likewise, the application of OBM method to N2 gene concentration (gc/L) results in a 100% of hit rate with method III defined peaks (Figure 2S. 1. c.). Specifically, OBM method is based on direct measurements of N1 and N2 gene concentrations and could be immediately applied by health departments after obtaining the data to determine rapid warnings on the upcoming fluctuations of clinical cases. In addition, PPC method could also achieve a 94.44% hit rate when it is applied to N1 gene concentrations (Figure 2S. 1. l.) using method III defined peaks, and a 100% hit rate when it is applied to N2 gene concentrations with both method I (Figure 2S. 1. d.) and III defined peaks

(Figure 2S. 1. f.). The PPCS method, as applied to N1 gene concentrations, performed the best in terms of higher hit rates where it achieved 90.91%, 90.91% and 100% with method-I, -II, and -III defined peaks, respectively (Figure 2. 4.). PPCS was applied to N2 gene concentrations where it also achieved 91.67% and 100% hit rates with method I and III defined peaks, respectively (Table 2S. 3., Figures 2S. 1. g., and 2S. 1. i.). Therefore, PPCS method based on N1 gene concentrations (gc/L) is recommended as a real-time method to capture early warnings with its higher hit rates across three methods for defining clinical peaks. Method III is the recommended as the peak-defining method since it is also more conservative in terms of capturing the most surges and fluctuations of cases. All hit rates developed using real-time methods are shown in Figure 2. 6.

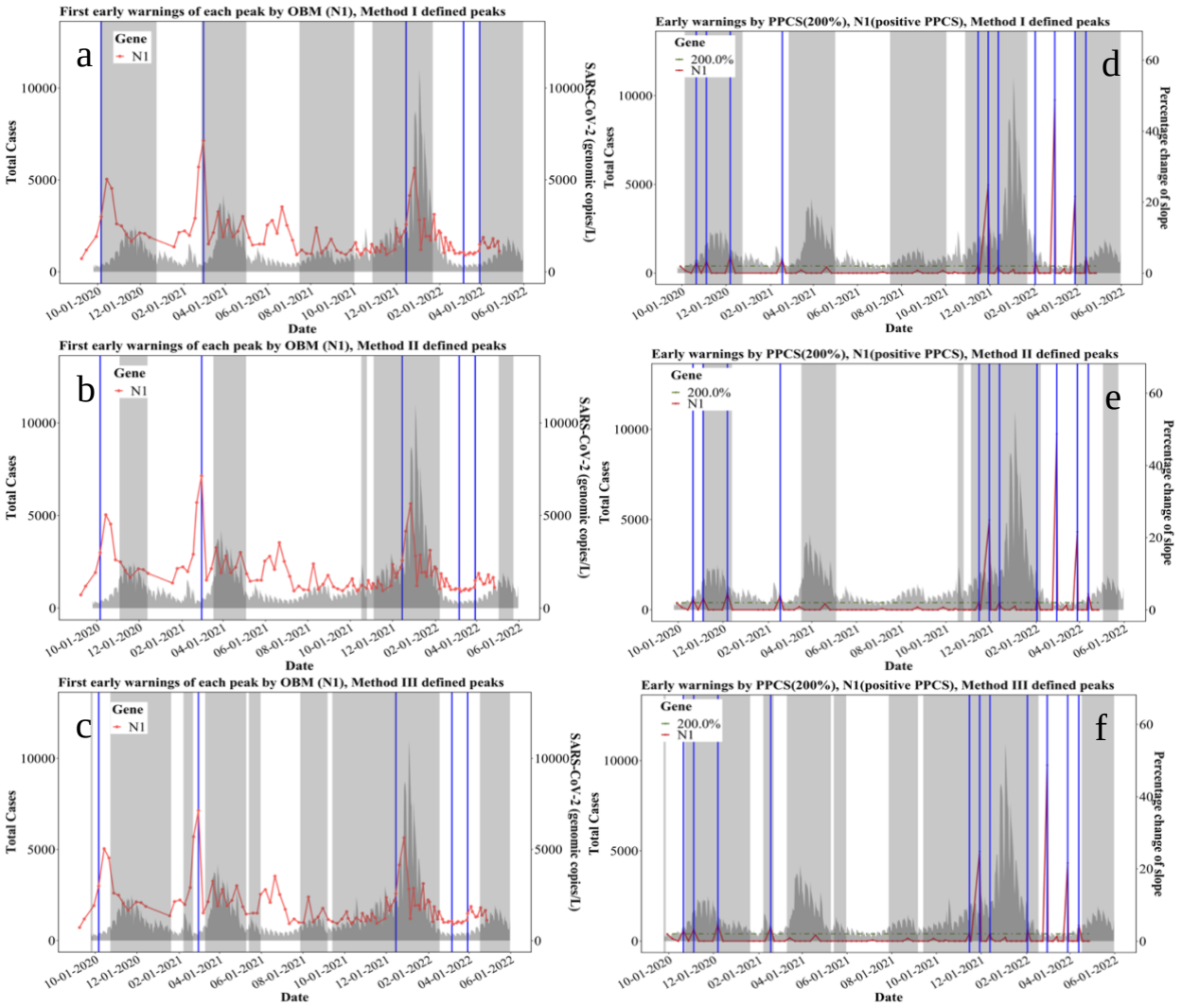


Figure 2. 4. Real-time early warning methods: OBM and PPCS based on N1 (gc/L): a. First early warnings of each peak identified by OBM (N1, gc/L) with Method I defined peaks; b. First early warnings of each peak identified by OBM (N1, gc/L) with Method II defined peaks; c. First early warnings of each peak identified by OBM (N1, gc/L) with Method III defined peaks; d. Early warnings identified by PPCS (N1, gc/L) with Method I defined peaks; e. Early warnings identified by PPCS (N1, gc/L) with Method II defined peaks; f. Early warnings identified by PPCS (N1, gc/L) with Method III defined peaks

3.2.2 Early warnings determined by post-factum methods

Post-factum methods were applied to N1 and N2 gene concentrations as well as w/c ratio. Selected methods including MSD and PER are illustrated in Figure 2. 5. MSD was applied to N1 gene concentrations and reached 100% hit rates with method I, II, and III defined peaks shown in Figures 2. 5. a., 2. 5. b., and 2. 5. c., respectively, where all identified warnings successfully forewarned the defined peaks. Likewise, PER method was applied to w/c ratio of N1 gene which are illustrated in Figures 2. 5. d., 2. 5. e., and 2. 5. f. with method I, II, and III defined peaks, where the hit rates reached 90%, 70%, and 100%, respectively (Table 2S. 3.). Notably, method II defined less peaks of cases leading to warnings identified by PER between May and July 2021 in vain (Figure 2. 5. e.). While method III defined more peaks of cases and covered more data with a wider time range thus leading to higher hit rates (Figure 2. 5. f.). From this, we conclude that clinical case peak defining approaches can affect hit rates of early warning methods to forewarn case peaks. Among all post-factum methods applied to N1 and N2 gene concentrations, MSD achieved 100% hit rates with method I, II, and III defined peaks (Table 2S. 3., Figures 2. 5., and 2S. 2.). Hence, MSD based on N1 and N2 gene concentrations (gc/L) is recommended as a post-factum method to identify early warnings. In addition, post-factum methods applied to w/c ratio, including MSD (Figure 2S. 3.), SKE (Figure 2S. 5.), KUR (Figure 2S. 6.), and PER (Figure 2S. 2.), achieved 100% hit rates except for VAM (Figure 2S. 4.), where KUR completely achieved 100% hit rates across three methods of defining peaks (Figure 2S. 6.) and MSD achieved 100% hit rates with method-I and -III defined peaks based on both w/c ratio of N1 and N2 genes. Thus, KUR and MSD based on w/c ratio are recommended for as post-factum methods to identify early warnings.

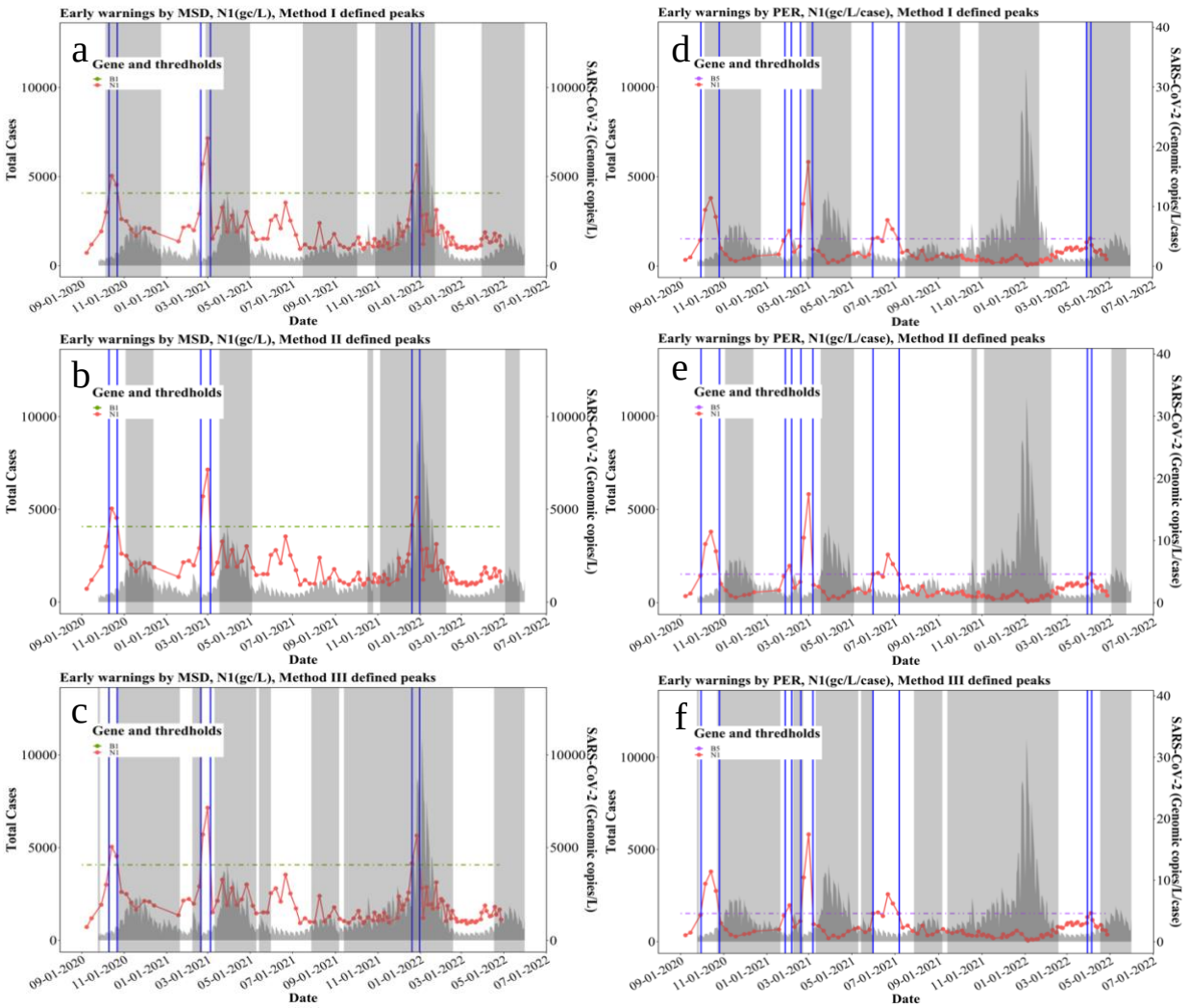


Figure 2. 5. Post-factum early warning methods MSD and PER, based on N1 (gc/L) and N1/c (gc/L/case), respectively: a. Early warnings identified by MSD (N1, gc/L) with Method I defined; b. Early warnings identified by MSD (N1, gc/L) with Method II defined; c. Early warnings identified by MSD (N1, gc/L) with Method III defined; d. Early warnings identified by PER (N1, gc/L/case) with Method I defined peaks; e. Early warnings identified by PER (N1, gc/L/case) with Method II defined peaks; f. Early warnings identified by PER (N1, gc/L/case) with Method III defined peaks

Worthy of noting, early warning methods achieved generally higher hit rate when they were implemented on N1 and N2 gene concentrations (gc/L) than on the w/c ratio (gc/L/case) dataset. Following this, w/c ratio did not significantly improve the hit rate of early warning methods in our study. Nevertheless, w/c ratio could still be used as an indicator of relationship between actual cases and testing capacity as discussed in section 3.1 and in recent studies (D'Aoust et al., 2022; Xiao et al., 2022). Among all post-factum methods, MSD method achieved higher hit rates compared with PER, VAM, and SKE methods for both N1 and N2 gene concentrations (gc/L) and w/c ratio (gc/L/case) datasets. In addition, method III could define more peak-ranges and is associated with higher hit rates for both real-time and post-factum methods when compared to method-I and-II.

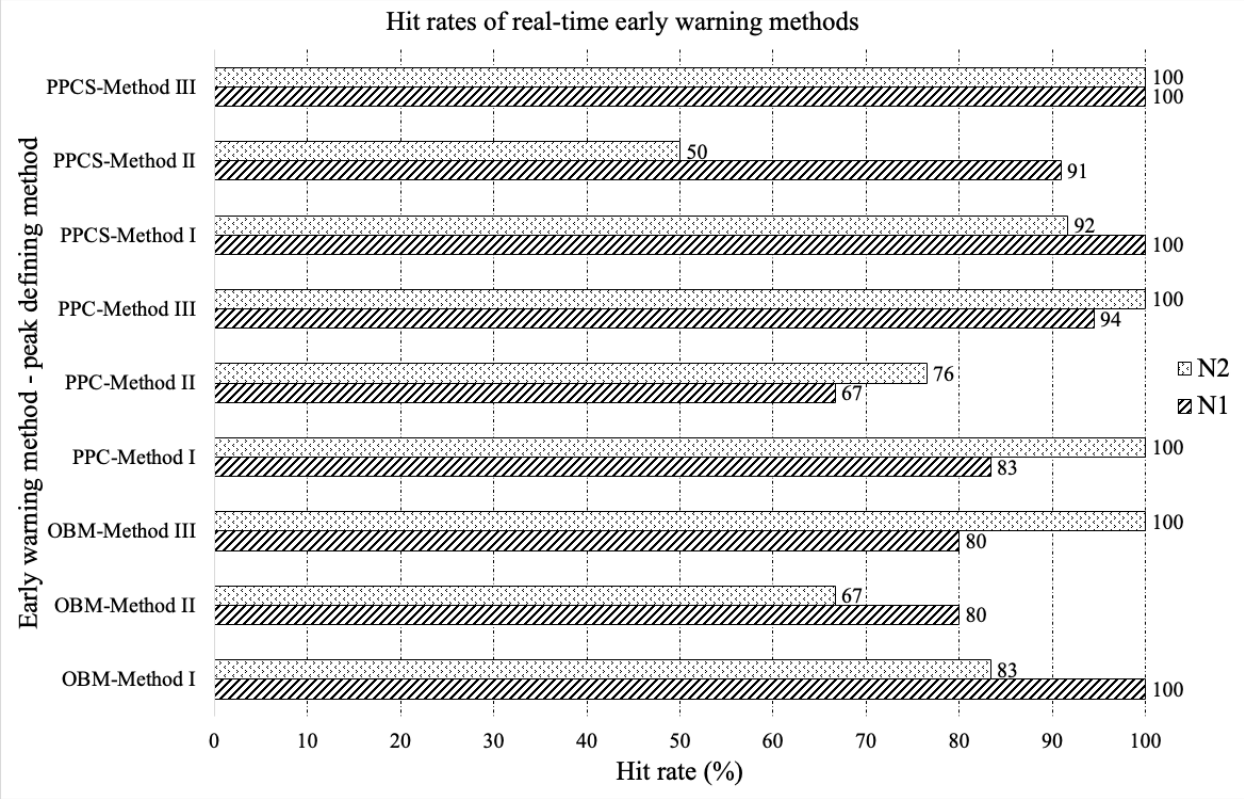


Figure 2. 6. Hit rates of real-time early warning methods: OBM, PPC, and PPCS, with three peak-defining methods: method I, method II, and method III, based on N1 and N2 gene concentrations (gc/L)

3.3 Factors affecting early warnings based on WBE and other uncertainties

Few studies have reported on straightforward, easily applied, and rapid early warning methods based on WBE for COVID-19 or other diseases (Bowman et al., 2016; Harris et al., 2020; Kostoulas et al., 2021; O'Brien et al., 2021; Prabdial-Sing et al., 2021; Zhu et al., 2021b). The early warning methods developed in the present study including both real-time and post-factum methods, effectively provide early warnings for defined peaks of COVID-19 cases in the Detroit metropolitan area. These methods can potentially be applied to other geographic regions and pathogens; however, further analysis will be necessary. None of the methods designed in this study is perfectly accurate when capturing early warnings of diseases, as numerous complex factors can affect WBE or associated reporting of clinical cases, including physiological, health system, laboratory-based and logistic factors (summarized in Table 2S. 2.) (Bibby et al., 2021; Kumar et al., 2021; Zhu et al., 2022). According to the authors of a recent WBE study conducted in Boston, Massachusetts, USA, WBE-based early warning methods should not be used in isolation, but rather in conjunction with other methods, given the complexity of factors and various unknowns (Xiao et al., 2022). Despite this recommendation, WBE-based early warning methods, could be essential in providing prompt warnings to impending epidemics, thus aiding health departments to mobilize and craft policy. Critical factors and uncertainties related to early warning methods are elucidated in the following sections.

3.3.1 Physiological factors

Shifting viral shedding cycles and dynamics may affect the accuracy of the relationship between WBE and case incidence, and therefore may affect the early warning potential of WBE (Bibby et al., 2021). Viral shedding dynamics can vary among individuals, variants, and so forth (Bibby et al., 2021; Zhao et al., 2022). Shedding of SARS-CoV-2 began for up to 7 days prior to

symptoms onset during the initial stage of the COVID-19 pandemic, but this duration declined to maximum of 3 days during the Omicron surge (Auwaerter et al., 2022; Cheng et al., 2020; Long et al., 2022; Wiersinga et al., 2020; Zhao et al., 2022). Shedding duration of SARS-CoV-2 shortened from a maximum of 24 days to 10 days during the Omicron surge (Lamers et al., 2022; Zhao et al., 2022). All of the aforementioned dynamics would affect the measured relationship between wastewater viral concentrations and clinical cases, such as the w/c ratio.

Another physiological factor is the poorly understood role of asymptomatic disease transmitters, which lead to an undercounting of true infection cases in communities that employ traditional clinical testing methods (Bibby et al., 2021). Clinical case numbers are not an ideal measure of disease prevalence in a given community, particularly when asymptomatic infections dominate. In this setting, early warning methods based on WBE may not be able to effectively provide early warnings of cases as it will be difficult to establish case-number as reference or baseline. In addition, peak defining approaches can be significantly affected by underestimations, leading to inaccuracy in early warning methods.

The proportion of infected populations who shed detectable levels of virus in their stool and viral load distribution throughout a day, are both significant factors since they have a direct impact on wastewater viral concentrations (Jones et al., 2020). Some studies estimate 48% to 67% of infected individuals shed SARS-CoV-2 in their stool (Ahmed et al., 2021). It was also demonstrated that 40% of infected populations shed SARS-CoV-2 virus RNA in their stool (Kirby et al., 2021). Viral load in wastewater throughout a given day is not evenly distributed, and therefore, some studies have suggested that optimization of sampling strategy coupled with a standardization based on toilet flushing frequency and wastewater travel time could improve the accurate detection of viral signals (Zhu et al., 2021a, 2021b).

3.3.2 Health system factors

Health system factors encompass a wide range of variables, including the delay of clinical data reporting (Torres et al., 2021; Zhao et al., 2022), prolonged duration of clinical data processing (Contreras et al., 2020), under-reporting of the true number of cases (Kronbichler et al., 2020; Salath et al., 2020), and other inconsistencies in reporting. These factors have the potential to have a prodigious effect on the accuracy of peak-defining methods, including those outlined in this study. These factors also contribute to the disparities of clinical case-numbers, which affect clinical cases used as the reference or benchmark for early warning methods, such as the w/c ratio (Figure 2. 3. c.). For instance, at the beginning of the COVID-19 pandemic, several clinical studies reported a delay of nearly 7 days from onset of symptoms to clinical testing, resulting in delayed reporting of clinical cases which already contributed to the wastewater viral concentrations (Huang et al., 2020; Zhao et al., 2022). In some cases, jurisdictions or countries accumulate delays of between 4 and 18 days (Català et al., 2021). Disparities in clinical data can furthermore be influenced by testing hesitancy or eagerness (Jimenez et al., 2021), the turnaround time for screening of cases (Larremore et al., 2021), the availability and accessibility of diagnostic tests (Olesen et al., 2021), and so forth. A shortage of testing supplies and lack of testing in some resource-constrained countries has caused a pileup of clinical samples awaiting results and delays in reporting cases, therefore, inevitably leading to inaccurate representation of actual cases (Torres et al., 2021), which could affect methods of defining peaks thus affecting the accuracy of early warning methods based on WBE.

3.3.3 Laboratory analysis and other uncertainties

Detectable viral signals in wastewater are critical to achieve early warning methods based on WBE. Therefore, potential decay of virus during wastewater transport in the municipal

wastewater collection system, sampling, and transportation of samples to laboratories may alter the accuracy of early warning methods (Ahmed et al., 2020b; Bibby et al., 2021). Some recent studies pointed out that sample transportation and laboratory processing including sample concentration, nucleic acid extraction, and PCR-based nucleic acid quantification, may take from one to three days (Bibby et al., 2021; Zhao et al., 2022), but under some scenarios this time range could be easily exceeded, leading to inevitable decay of viral signals in any of the processes. Moreover, recovery efficiency could also vary among different sampling and analytical methods (Ahmed et al., 2020c), introducing more uncertainties in the results.

Recent studies have investigated additional factors that can influence early warning potential of WBE, including sample site geographic distribution, fair sample representation of demographics/communities, uneven mixing of wastewater during sampling and laboratory analysis, dilution of viral RNA during rainfall events, and climate variability (Ahmed et al., 2020b; Bibby et al., 2021; Butler et al., 1995; Kumar et al., 2021; Zhao et al., 2022; Zhu et al., 2022). In addition, the persistence of SARS-CoV-2 in wastewater may be affected by environmental factors. These include temperature, organic matter, and microorganisms, which contribute to more uncertainties of measuring wastewater viral concentrations, ultimately affecting accuracy of early warning methods (Xiao et al., 2022). Overall, the accuracy of reported clinical data and WBE workflows are the primary influencers of effective early warning methods.

3.3.4 Discussions, limitations, and future directions

Real-time methods focus on rapid early warnings using current or recent wastewater measurements for predicting future surging of cases. Real-time methods can be implemented rapidly upon obtaining the wastewater measurements. For instance, the OBM method only requires three consecutive increasing measurements to issue an early warning. The early warning

signals determined by real-time methods precede the upcoming increase in cases. Thus, these methods can trigger alerts in real time before surging COVID-19 cases that are subsequently reported by health agencies. Real-time methods can be an effective decision-making tool for public health officials during an on-going epidemic.

Post-factum methods focus on after-the-fact analysis of wastewater and clinical data over an entire period and provide analysis of warnings for surges of cases after they have occurred. Post-factum methods can be used as complementary methodologies to real-time methods. Post-factum methods, such as the MSD method, concentrate on providing an overall picture of early warnings for an entire period of investigations which can identify all early warnings for an entire time (Bowman et al., 2016; Prabdial-Sing et al., 2021). Post-factum methods can be useful to health agencies to plan and design health measures for the next potential epidemic.

Admittedly, this study has various limitations. First, eight early warning methods and three peak-defining methods were successfully applied to WBE data collected from the Detroit metropolitan area, but these methods have not been validated in other regions. More in-depth investigations are warranted to develop and apply early warning methods based on this study. Secondly, the w/c ratio was adopted from a study that assumed that viral shedding did not change significantly over the course of the entire pandemic (Xiao et al., 2022). However, viral shedding patterns and dynamics changed during the study period for the Detroit area where the dominant variant changed from Alpha to Beta, Gamma, Delta, and Omicron variants and so forth (Xiao et al., 2022; Zhao et al., 2022). For example, the lag time of wastewater surveillance preceding the clinical testing declined from five weeks to two weeks during the Omicron surge owning predominately to the changing viral shedding dynamics (Zhao et al., 2022), potentially affecting the performance of early warning methods. Third, the w/c ratio did not capture the third major

peak of clinical cases in late December of 2021 and the beginning of January 2022, which perhaps was a consequence of changing shedding dynamics associated with Omicron, as well as the rapidly increasing testing capacity in December 2021 in the State of Michigan leading redundant counting of clinical cases, which we discussed thoroughly in 3.1. Finally, the approaches of defining peaks of COVID-19 cases were potentially specific to our sampling demographic and geographic sampling distribution in the Detroit metropolitan area in this study. Exploration of other methods for defining peaks of clinical cases for other regions seems warranted.

Overall, numerous prospects extended from this study could inspire applications of WBE data and development of early warning methods of WBE for public health benefits. The eight early warning methods described here are straightforward and easily applied, and could forewarn defined peaks with high hit rates, especially the real-time methods OBM and PPCS. The real-time methods require merely the direct measurements of N1 and N2 gene concentrations as well as simple statistical calculations, which are easily applied tools for public health departments to apply on WBE datasets to determine early warnings rapidly. Three methods of defining peaks are easily applied as well. The early warning methods and peak-defining methods proposed in this study attempt to provide rapid and straightforward approaches to determine early warnings for health departments, partner agencies, and the public, instead of applying intricate and sophisticated models. Additionally, this study demonstrates the impact of public policies on wastewater viral concentrations and subsequent clinical cases in Detroit metropolitan area for approximately two years. Combining wastewater data with clinical cases for the Detroit area, such as application of w/c ratio, could allow health departments to understand the actual infections and testing conditions in communities. However, more studies are warranted to establish a standard framework for defining peaks of clinical cases, apply and develop early warning methods that are easily applied

by health departments, to use early warnings in a timely manner. This study highlights the impact of public health policy on measured wastewater viral concentrations and clinical COVID-19 cases in Detroit. Future research should integrate public policy at a granular level.

4. Conclusions

This study introduced eight (three real-time and five post-factum) early-warning methods based on wastewater surveillance data and three peak-defining methods based on clinical data that can be easily implemented by public health departments and partner agencies to warn of viral disease fluctuations. Hit rates were calculated to evaluate the efficacy of early warning methods in predicting clinical case surges. Applying these methods to a 21-month WBE data set in the Detroit metropolitan area in Michigan amid the COVID-19 pandemic, we conclude that wastewater viral signals preceded the reported clinical cases. Both viral signal and clinical cases corresponded to social events and reflected implementation of public health policies. The early-warning methods based on WBE were proven to be efficient during the study period, as evinced by hit rates. Hit rates for early warning methods were affected by the method for defining peaks in clinical cases. Method III for defining peaks (peak defined as clinical data values higher than mean – 0.5 standard deviation of all values) identified most peaks in clinical cases and was associated with higher hit rates across all WBE based early-warning methods. Among all real-time methods, PPCS method (positive percentage change of slope greater than 200%) achieved higher hit rates. Among all post-factum methods, KUR (values greater than kurtosis) and MSD (values greater than mean + 2 standard deviation) methods achieved higher hit rates.

5. Acknowledgements

We thank the Michigan Department of Health and Human Services (MDHHS) and the Great Lakes Water Authority (GLWA) for funding this research.

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. v., & Mueller, J. F. (2020a). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment*, *728*, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Ahmed, W., Bertsch, P. M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., Gyawali, P., Korajkic, A., Riddell, S., Sherchan, S. P., Simpson, S. L., Sirikanchana, K., Symonds, E. M., Verhagen, R., Vasan, S. S., Kitajima, M., & Bivins, A. (2020b). Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environmental Research*, *191*, 110092. https://doi.org/10.1016/j.envres.2020.110092

Ahmed, W., Bertsch, P. M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B. R., Mueller, J. F., Simpson, S. L., Smith, W. J. M., Symonds, E. M., Thomas, K. v., Verhagen, R., & Kitajima, M. (2020c). Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total Environment*, *739*, 139960. https://doi.org/10.1016/j.scitotenv.2020.139960

Ahmed, W., Bivins, A., Simpson, S. L., Bertsch, P. M., Ehret, J., Hosegood, I., Metcalfe, S. S., Smith, W. J. M., Thomas, K. v., Tynan, J., & Mueller, J. F. (2022a). Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. *Environment International*, *158*, 106938. https://doi.org/10.1016/j.envint.2021.106938

Ahmed, W., Bivins, A., Metcalfe, S., Smith, W. J. M., Verbyla, M. E., Symonds, E. M., & Simpson, S. L. (2022b). Evaluation of process limit of detection and quantification variation of SARS-CoV-2 RT-qPCR and RT-dPCR assays for wastewater surveillance. *Water Research*, *213*, 118132. https://doi.org/10.1016/j.watres.2022.118132

Ahmed, W., Tscharke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T. M. H., O'Brien, J. W., Simpson, S. L., Sherman, P., Thomas, K. v., Verhagen, R., Zaugg, J., & Mueller, J. F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Science of the Total Environment*, *761*, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Auwaerter M.D., P. (2022). *Coronavirus COVID-19 (SARS-CoV-2)*. The Johns Hopkins University. https://www.hopkinsguides.com/hopkins/view/Johns_Hopkins_ABX_Guide/540747/all/Coronavirus_COVID_19__SARS_CoV_2_

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J.,

Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Bibby, K., Bivins, A., Wu, Z., & North, D. (2021). Making waves: Plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research*, *202*(July), 117438. https://doi.org/10.1016/j.watres.2021.117438

Bivins, A., & Bibby, K. (2021). Wastewater Surveillance during Mass COVID-19 Vaccination on a College Campus. *Environmental Science & Technology Letters*, *8*(9), 792–798. https://doi.org/10.1021/acs.estlett.1c00519

Bowman, L. R., Tejeda, G. S., Coelho, G. E., Sulaiman, L. H., Gill, B. S., McCall, P. J., Olliaro, P. L., Ranzinger, S. R., Quang, L. C., Ramm, R. S., Kroeger, A., & Petzold, M. G. (2016). Alarm Variables for Dengue Outbreaks: A Multi-Centre Study in Asia and Latin America. *PLOS ONE*, *11*(6), e0157971. https://doi.org/10.1371/journal.pone.0157971

Butler, D., Friedler, E., & Gatt, K. (1995). Characterising the quantity and quality of domestic wastewater inflows. *Water Science and Technology*, *31*(7), 13–24. https://doi.org/https://doi.org/10.1016/0273-1223(95)00318-H

Cao, Y., & Francis, R. (2021). On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Science of The Total Environment, 786*, 147451. https://doi.org/10.1016/j.scitotenv.2021.147451

Català, M., Pino, D., Marchena, M., Palacios, P., Urdiales, T., Cardona, P.-J., Alonso, S., López-Codina, D., Prats, C., & Alvarez-Lacalle, E. (2021). Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers. *PLOS ONE*, *16*(1), e0243701. https://doi.org/10.1371/journal.pone.0243701

Cheng, H.-Y., Jian, S.-W., Liu, D.-P., Ng, T.-C., Huang, W.-T., Lin, H.-H., & Team, T. C.-19 O. I. (2020). Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. *JAMA Internal Medicine*, *180*(9), 1156–1163. https://doi.org/10.1001/jamainternmed.2020.2020

Contreras, S., Biron-Lattes, J. P., Villavicencio, H. A., Medina-Ortiz, D., Llanovarced-Kawles, N., & Olivera-Nappa, Á. (2020). Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. *Chaos, Solitons & Fractals*, *139*, 110087. https://doi.org/https://doi.org/10.1016/j.chaos.2020.110087

Corchis-Scott, R., Geng, Q., Seth, R., Ray, R., Beg, M., Biswas, N., Charron, L., Drouillard, K. D., D'Souza, R., Heath, D. D., Houser, C., Lawal, F., McGinlay, J., Menard, S. L., Porter, L. A., Rawlings, D., Scholl, M. L., Siu, K. W. M., Tong, Y., … McKay, R. M. L. (2021). Averting an Outbreak of SARS-CoV-2 in a University Residence Hall through Wastewater Surveillance. *Microbiology Spectrum*, *9*(2). https://doi.org/10.1128/Spectrum.00792-21

D'Aoust, P. M., Tian, X., Towhid, S. T., Xiao, A., Mercier, E., Hegazy, N., Jia, J.-J., Wan, S., Kabir, M. P., Fang, W., Fuzzen, M., Hasing, M., Yang, M. I., Sun, J., Plaza-Diaz, J., Zhang, Z., Cowan, A., Eid, W., Stephenson, S., … Delatolla, R. (2022). Wastewater to clinical case (WC) ratio of COVID-19 identifies insufficient clinical testing, onset of new variants of concern and population immunity in urban communities. *Science of The Total Environment*, *853*, 158547. https://doi.org/10.1016/j.scitotenv.2022.158547

Drake, J. M., & Griffen, B. D. (2010). Early warning signals of extinction in deteriorating environments. *Nature*, *467*(7314), 456–459.

Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Adamopoulos, P. G., Peccia, J., Thompson, D. C., Kontou, A., Karagiannidis, A., Lianidou, E. S., Avgeris, M., Paraskevis, D., Tsiodras, S., Scorilas, A., Vasiliou, V., Dimopoulos, M.-A., & Thomaidis, N. S. (2022). SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Science of The Total Environment*, *804*, 150151. https://doi.org/10.1016/j.scitotenv.2021.150151

Gao, Q., Liu, Z., Xiang, J., Tong, M., Zhang, Y., Wang, S., Zhang, Y., Lu, L., Jiang, B., & Bi, P. (2021). Forecast and early warning of hand, foot, and mouth disease based on meteorological factors: Evidence from a multicity study of 11 meteorological geographical divisions in mainland China. *Environmental Research*, *192*, 110301. https://doi.org/10.1016/j.envres.2020.110301

Harris, M. J., Hay, S. I., & Drake, J. M. (2020). Early warning signals of malaria resurgence in Kericho, Kenya. *Biology Letters*, *16*(3), 20190713. https://doi.org/10.1098/rsbl.2019.0713

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Jiang, G., Wu, J., Weidhaas, J., Li, X., Chen, Y., Mueller, J., Li, J., Kumar, M., Zhou, X., Arora, S., Haramoto, E., Sherchan, S., Orive, G., Lertxundi, U., Honda, R., Kitajima, M., & Jackson, G. (2022). Artificial neural network-based estimation of COVID-19 case numbers and effective reproduction rate using wastewater-based epidemiology. *Water Research*, *218*, 118451. https://doi.org/10.1016/j.watres.2022.118451

Jimenez, M. E., Rivera-Núñez, Z., Crabtree, B. F., Hill, D., Pellerano, M. B., Devance, D., Macenat, M., Lima, D., Martinez Alcaraz, E., Ferrante, J. M., Barrett, E. S., Blaser, M. J., Panettieri Jr, R. A., & Hudson, S. v. (2021). Black and Latinx Community Perspectives on COVID-19 Mitigation Behaviors, Testing, and Vaccines. *JAMA Network Open*, *4*(7), e2117074–e2117074. https://doi.org/10.1001/jamanetworkopen.2021.17074

Jones, D. L., Baluja, M. Q., Graham, D. W., Corbishley, A., McDonald, J. E., Malham, S. K., Hillary, L. S., Connor, T. R., Gaze, W. H., Moura, I. B., Wilcox, M. H., & Farkas, K. (2020). Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. *Science of The Total*

Environment, *749*, 141364. https://doi.org/10.1016/j.scitotenv.2020.141364

Kantor, R. S., Nelson, K. L., Greenwald, H. D., & Kennedy, L. C. (2021). Challenges in Measuring the Recovery of SARS-CoV-2 from Wastewater. *Environmental Science & Technology*, *55*(6), 3514–3519. https://doi.org/10.1021/acs.est.0c08210

Kirby, A. E., Walters, M. S., Jennings, W. C., Fugitt, R., LaCross, N., Mattioli, M., Marsh, Z. A., Roberts, V. A., Mercante, J. W., Yoder, J., & Hill, V. R. (2021). Using Wastewater Surveillance Data to Support the COVID-19 Response — United States, 2020–2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*(36), 1242–1244. https://doi.org/10.15585/mmwr.mm7036a2

Kostoulas, P., Meletis, E., Pateras, K., Eusebi, P., Kostoulas, T., Furuya-Kanamori, L., Speybroeck, N., Denwood, M., Doi, S. A. R., Althaus, C. L., Kirkeby, C., Rohani, P., Dhand, N. K., Peñalvo, J. L., Thabane, L., BenMiled, S., Sharifi, H., & Walter, S. D. (2021). The epidemic volatility index, a novel early warning tool for identifying new waves in an epidemic. *Scientific Reports*, *11*(1), 23775. https://doi.org/10.1038/s41598-021-02622-3

Koureas, M., Amoutzias, G. D., Vontas, A., Kyritsi, M., Pinaka, O., Papakonstantinou, A., Dadouli, K., Hatzinikou, M., Koutsolioutsou, A., Mouchtouri, V. A., Speletas, M., Tsiodras, S., & Hadjichristodoulou, C. (2021). Wastewater monitoring as a supplementary surveillance tool for capturing SARS-COV-2 community spread. A case study in two Greek municipalities. *Environmental Research*, *200*, 111749. https://doi.org/10.1016/j.envres.2021.111749

Kronbichler, A., Kresse, D., Yoon, S., Lee, K. H., Effenberger, M., & Shin, J. il. (2020). Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis. *International Journal of Infectious Diseases*, *98*, 180–186. https://doi.org/10.1016/j.ijid.2020.06.052

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

Kumar, M., Joshi, M., Patel, A. K., & Joshi, C. G. (2021). Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environmental Research*, *196*(December 2020), 110946. https://doi.org/10.1016/j.envres.2021.110946

Lamers, M. M., Mykytyn, A. Z., Breugem, T. I., Groen, N., Knoops, K., Schipper, D., van Acker, R., van den Doel, P. B., Bestebroer, T., Koopman, C. D., Reusken, C., Muraro, M. J., GeurtsvanKessel, C. H., van Royen, M. E., Peters, P. J., Zhang, J., & Haagmans, B. L. (2022). SARS-CoV-2 Omicron efficiently infects human airway, but not alveolar epithelium. *BioRxiv*, 2022.01.19.476898. https://doi.org/10.1101/2022.01.19.476898

Larremore, D. B., Wilder, B., Lester, E., Shehata, S., Burke, J. M., Hay, J. A., Tambe, M., Mina,

M. J., & Parker, R. (2021). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Science Advances*, *7*(1). https://doi.org/10.1126/sciadv.abd5393

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

Li, X., Kulandaivelu, J., Zhang, S., Shi, J., Sivakumar, M., Mueller, J., Luby, S., Ahmed, W., Coin, L., & Jiang, G. (2021). Data-driven estimation of COVID-19 community prevalence through wastewater-based epidemiology. *Science of The Total Environment*, *789*, 147947. https://doi.org/10.1016/j.scitotenv.2021.147947

Li, Y., Miyani, B., Zhao, L., Spooner, M., Gentry, Z., Zou, Y., Rhodes, G., Li, H., Kaye, A., Norton, J., & Xagoraraki, I. (2022). Surveillance of SARS-CoV-2 in nine neighborhood sewersheds in Detroit Tri-County area, United States: Assessing per capita SARS-CoV-2 estimations and COVID-19 incidence. *Science of The Total Environment*, *851*, 158350. https://doi.org/10.1016/j.scitotenv.2022.158350

Long, B., Carius, B. M., Chavez, S., Liang, S. Y., Brady, W. J., Koyfman, A., & Gottlieb, M. (2022). Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *The American Journal of Emergency Medicine*, *54*, 46–57. https://doi.org/https://doi.org/10.1016/j.ajem.2022.01.028

MatLab, P. (2018). 9.7. 0.1190202 (R2019b). *MathWorks Inc Natick MA USA*.

Matheri, A. N., Belaid, M., Njenga, C. K., & Ngila, J. C. (2022). Water and wastewater digital surveillance for monitoring and early detection of the COVID-19 hotspot: industry 4.0. *International Journal of Environmental Science and Technology*. https://doi.org/10.1007/s13762-022-03982-7

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

Menkir, T. F., Chin, T., Hay, J. A., Surface, E. D., de Salazar, P. M., Buckee, C. O., Watts, A., Khan, K., Sherbo, R., Yan, A. W. C., Mina, M. J., Lipsitch, M., & Niehus, R. (2021). Estimating internationally imported cases during the early COVID-19 pandemic. *Nature Communications*, *12*(1), 311. https://doi.org/10.1038/s41467-020-20219-8

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., ... & Xagoraraki, I. (2021). Early Warnings of COVID-19 Second Wave in Detroit. Journal of Environmental Engineering, 147(8), 06021004.

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004.

Murhekar, M. v, Bhatnagar, T., Selvaraju, S., Saravanakumar, V., Thangaraj, J. W. V., Shah, N., Kumar, M. S., Rade, K., Sabarinathan, R., Asthana, S., Balachandar, R., Bangar, S. D., Bansal, A. K., Bhat, J., Chopra, V., Das, D., Deb, A. K., Devi, K. R., Dwivedi, G. R., … Zaman, K. (2021). SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey. *The Lancet Global Health*, *9*(3), e257–e266. https://doi.org/10.1016/S2214-109X(20)30544-1

O'Brien, D. A., & Clements, C. F. (2021). Early warning signal reliability varies with COVID-19 waves. *Biology Letters*, *17*(12). https://doi.org/10.1098/rsbl.2021.0487

Olesen, S., Imakaev, M., & Duvallet, C. (2021). Making Waves: Defining the lead time of wastewater-based epidemiology for COVID-19. *Water Research*, 117433.

Prabdial-Sing, N., Motaze, V., Manamela, J., McCarthy, K., & Suchard, M. (2021). Establishment of Outbreak Thresholds for Hepatitis A in South Africa Using Laboratory Surveillance, 2017–2020. *Viruses*, *13*(12), 2470. https://doi.org/10.3390/v13122470

Safford, H. R. , Karen, S., & Heather, B. N. (2022). Wastewater analysis can be a powerful public health tool—if it's done sensibly. *Proceedings of the National Academy of Sciences*, *119*(6), e2119600119. https://doi.org/10.1073/pnas.2119600119

Salath, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., Zwahlen, M., Senti, G., Battegay, M., Wilder-Smith, A., Eckerle, I., Egger, M., & Low, N. (2020). COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Medical Weekly*. https://doi.org/10.4414/smw.2020.20225

Sherchan, S. P., Shahin, S., Ward, L. M., Tandukar, S., Aw, T. G., Schmitz, B., Ahmed, W., & Kitajima, M. (2020). First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA. *Science of The Total Environment*, *743*, 140621. https://doi.org/10.1016/j.scitotenv.2020.140621

Team, R. C. (2022). *R: A Language and Environment for Statistical Computing. R version 4.1. 3*.

Torres, I., Sippy, R., & Sacoto, F. (2021). Assessing critical gaps in COVID-19 testing capacity: the case of delayed results in Ecuador. *BMC Public Health*, *21*(1), 637. https://doi.org/10.1186/s12889-021-10715-x

Vallejo, J. A., Trigo-Tasende, N., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Barbeito, I., Vaamonde, M., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B. K., Nasser-Ali, M., Cid, Á., Veiga, M., Acevedo, A., Lamora, C., Bou, G., Cao, R., & Poza, M. (2022). Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. *Science of The Total Environment*, *811*, 152334. https://doi.org/10.1016/j.scitotenv.2021.152334

Wang, R., Jiang, Y., Michael, E., & Zhao, G. (2017). How to select a proper early warning threshold to detect infectious disease outbreaks based on the China infectious disease automated alert and response system (CIDARS). *BMC Public Health*, *17*(1), 570.

https://doi.org/10.1186/s12889-017-4488-0

Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., & Prescott, H. C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, *324*(8), 782–793. https://doi.org/10.1001/jama.2020.12839

Xagoraraki, I. (2020a). Can we predict viral outbreaks using wastewater surveillance?. *Journal of Environmental Engineering*, *146*(11), 01820003.

Xagoraraki, I., & O'Brien, E. (2020b). Wastewater-based epidemiology for early detection of viral outbreaks. In Women in water quality (pp. 75-97). Springer, Cham.

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

Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science and Technology*, *50*(10), 5077–5085. https://doi.org/10.1021/acs.est.6b00876

Zhang, J., Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen, X., Li, M., Zheng, W., Yi, L., Chen, X., Wu, Q., Liang, Y., Wang, X., Yang, J., Sun, K., Longini, I. M., Halloran, M. E., Wu, P., Cowling, B. J., … Yu, H. (2020). Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*, *20*(7), 793–802. https://doi.org/https://doi.org/10.1016/S1473-3099(20)30230-9

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarland, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, 157040. https://doi.org/10.1016/j.scitotenv.2022.157040

Zhu, Y., Oishi, W., Maruo, C., Bandara, S., Lin, M., Saito, M., Kitajima, M., & Sano, D. (2022). COVID-19 case prediction via wastewater surveillance in a low-prevalence urban community: a modeling approach. *Journal of Water and Health*, *20*(2), 459–470. https://doi.org/10.2166/wh.2022.183

Zhu, Y., Oishi, W., Maruo, C., Saito, M., Chen, R., Kitajima, M., & Sano, D. (2021a). Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Science of The Total Environment*, *767*, 145124. https://doi.org/10.1016/j.scitotenv.2021.145124

Zhu, Y., Oishi, W., Saito, M., Kitajima, M., & Sano, D. (2021b). Early Warning of COVID-19 in Tokyo via Wastewater-based Epidemiology: How Feasible It Really Is? *Journal of Water and Environment Technology*, *19*(3), 170–183. https://doi.org/10.2965/jwet.21-024

APPENDIX

Detailed Procedure: Sampling and Virus Elution

Viruses were collected and isolated from untreated wastewater using electropositive NanoCeram column filters (Argonide, Sanford, FL) based on the EPA Virus Adsorption-Elution (VIRADEL) method (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022). Depending on the suspended solids of wastewater, approximately 10 to 50 L of untreated wastewater passed through NanoCeram electropositive cartridge filters at a rate not more than 11 L/min using a previously described method (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022). Flow meter readings were recorded at the inception and termination of each sampling event. After sampling, all NanoCeram column filters were placed in sealed plastic bags on ice, then transported to the Environmental Virology Laboratory at Michigan State University in East Lansing, Michigan, USA, within 24 h for downstream analysis. Viruses were eluted within 24 h after sampling based on a previously described method (Miyani et al., 2020, 2021; Zhao et al., 2022).

Detailed Procedure: RNA Extraction

Viral RNA was extracted using a Viral RNA QIAGEN kit (QIAGEN, Germantown, Maryland), following the manufacturer's protocol with the modification described previously (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022).

Detailed Procedure: RT-ddPCR

RT-ddPCR was performed using a QX200 AutoDG Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA), using a One-step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA). According to the CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel for SARS-CoV-2 detection (cdc.gov/coronavirus/2019-ncov/lab), the primers and probe targeting N1 and N2 of SARS-CoV-2 are shown below. These were shown to be the

most sensitive assays for identifying SARS-CoV-2 (Ahmed et al., 2022). Oligonucleotide sequences (5’ to 3’) of primers and probes are shown below:

2019-nCoV_N1-F: GAC CCC AAA ATC AGC GAA AT

2019-nCoV_N1-R: TCT GGT TAC TGC CAG TTG AAT CTG

2019-nCoV_N1-P: FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1

2019-nCoV_N2-F: TTA CAA ACA TTG GCC GCA AA

2019-nCoV_N2-R: GCG CGA CAT TCC GAA GAA

2019-nCoV_N2-P: FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1.

Samples were then run on a C1000 Touch Thermal Cycler (Bio-Rad, Hercules, CA, USA) using the conditions shown in Zhao et al., 2022, following fluorescence measurement on a QX200 Droplet Reader (Bio-Rad, Hercules, CA, USA). N1 N2 gene Duplex Assay Reaction Mixture was also adopted from Zhao et al., 2022.

Table 2S. 1. Prediction approaches using advanced models of COVID-19 cases based on wastewater surveillance

Location	Prediction method	Model input	Prediction output	References
Michigan, USA	automatic regression integrated moving average (ARIMA) and vector autoregressive (VAR) models	SARS-CoV-2 concentrations, clinical cases	COVID-19 cases	(Zhao et al., 2022)
Utah, USA	artificial neural network (ANN)	WBE data, weather, clinical testing and vaccine coverage	COVID-19 incidence, prevalence rate, COVID-19 effective reproduction rate	(Jiang et al., 2022)
Sendai, Miyagi, Japan	generalized linear model, ANN, and random forest (RF)	positive rates from consecutive calculation windows	COVID-19 cases	(Zhu et al., 2022)
Galicia, Spain	COVIDBENS models, quadratic LOESS model, simple linear regression model	viral load, time, and other variables	COVID-19 cases	(Vallejo et al., 2022)
Published literature	multiple linear regression, ANN, and adaptive neuro fuzzy inference system	SARS-CoV-2 concentrations, wastewater/air temperature, population, average daily water consumption, sampling technique, and precipitation	prevalence of COVID-19 cases	(X. Li et al., 2022)
Attica, Greece	distributed/fixed lag modelling, linear regression, and ANN	SARS-CoV-2 concentrations, efficiency of PCR method, quantification cycle	pandemic health indicators, admission rates to hospitals	(Galani et al., 2022)
Pennsylvania, USA	VAR models	SARS-CoV-2 concentrations and COVID-19 cases	COVID-19 cases	(Cao & Francis, 2021)
South Carolina, USA	susceptible-exposed-infectious-recovered model	mass rate of RNA copies released per day, reproductive number, viral half-life, and sewage temperature	numbers of infected individuals	(McMahan et al., 2021)
Larissa and Volos, Greece	linear regression, RF models were trained and tested with machine learning	SARS-CoV-2 concentrations and normalized data	COVID-19 cases	(Koureas et al., 2021)
Johannesburg, South Africa	ARIMA models	confirmed COVID-19 cases	COVID-19 cases	(Matheri et al., 2022)

Table 2S. 2. Factors affecting early warning of emerging diseases based on WBE

Category	Factors	References
Physiological /biological	viral shedding dynamics; the role of asymptomatic transmitters of disease; decay of viral signals during wastewater transport in collection systems	(Bibby et al., 2021)
Public health	delay in WBE data reporting	(Bibby et al., 2021)
	delay in clinical data reporting	(Català et al., 2021; Menkir et al., 2021; Torres et al., 2021; Zhang et al., 2020)
	clinical data processing time	(Contreras et al., 2020)
	underreporting of the real number of infections	(Kronbichler et al., 2020; Salath et al., 2020)
Laboratory analysis and/or logistics	the sensitivity of clinical and WBE analytical workflows	(Bibby et al., 2021)
	sampling and analytical methods	(Kumar et al., 2021)
	uneven mixing of wastewater	(Zhu et al., 2021)
	viral RNA can be highly diluted	
	viral load is not uniformly distributed throughout a given day	(Butler et al., 1995)
	the sample might not be representative	
Other	climatic variability	(Kumar et al., 2022)
	WWTP characteristics	
	population demographics	
	the proportion of infected people who shed detectable levels of virus in their stool	(Ahmed et al., 2021; Kirby et al., 2021)

Table 2S. 3. Hit rates of early warning methods

Analysis Type	Method Abbreviation	Method of defining the peaks	Data type	Number of possible early warning dates identified	Number of early warning dates that captured peaks (within 5 weeks ahead of the beginning of the peaks)	Hit rate % (accuracy)
Real-time	OBM	Method I	N1	5	5	100.00
		Method II		5	4	80.00
		Method III		5	4	80.00
		Method I	N2	6	5	83.33
		Method II		6	4	66.67
		Method III		6	6	100.00
	PPC	Method I	N1	18	15	83.33
		Method II		18	12	66.67
		Method III		18	17	94.44
		Method I	N2	17	17	100.00
		Method II		17	13	76.47
		Method III		17	17	100.00
	PPCS	Method I	N1	11	10	90.91
		Method II		11	10	90.91
		Method III		11	11	100.00
		Method I	N2	12	11	91.67
		Method II		12	6	50.00
		Method III		12	12	100.00
Post-factum	MSD	Method I	N1	6	6	100.00
			N2	6	6	100.00
		Method II	N1	6	6	100.00
			N2	6	6	100.00
		Method III	N1	6	6	100.00
			N2	6	6	100.00
	PER	Method I	N1	14	12	85.71
			N2	11	11	100.00
		Method II	N1	14	12	85.71
			N2	11	9	81.82
		Method III	N1	14	14	100.00
			N2	11	11	100.00
	MSD	Method I	N1/c	6	6	100.00
			N2/c	6	6	100.00
		Method II	N1/c	6	4	66.67
			N2/c	6	4	66.67
		Method III	N1/c	6	6	100.00
			N2/c	6	6	100.00
	VAM	Method I	N1/c	16	15	93.75
			N2/c	18	16	88.89
		Method II	N1/c	16	6	37.50
			N2/c	18	8	44.44

Table 2S. 3. (cont'd)

Post-factum	SKE	Method I	N1/c	14	13	92.86
			N2/c	12	12	100.00
		Method II	N1/c	14	8	57.14
			N2/c	12	6	50.00
		Method III	N1/c	14	12	85.71
			N2/c	12	12	100.00
	KUR	Method I	N1/c	2	2	100.00
			N2/c	2	2	100.00
		Method II	N1/c	2	2	100.00
			N2/c	2	2	100.00
		Method III	N1/c	2	2	100.00
			N2/c	2	2	100.00
	PER	Method I	N1/c	10	9	90.00
			N2/c	10	10	100.00
		Method II	N1/c	10	7	70.00
			N2/c	10	8	80.00
		Method III	N1/c	10	10	100.00
			N2/c	10	10	100.00
	Mean-0.5SD	Method I	N1	25	20	80.00
			N2	29	25	86.21
		Method II	N1	25	14	56.00
			N2	29	20	68.97
		Method III	N1	25	24	96.00
			N2	29	28	96.55
		Method I	N1/c	18	15	83.33
			N2/c	14	14	100.00
		Method II	N1/c	18	15	83.33
			N2/c	14	10	71.43
		Method III	N1/c	18	18	100.00
			N2/c	14	14	100.00

Note: "c" represents daily 7-day moving average of clinical cases.

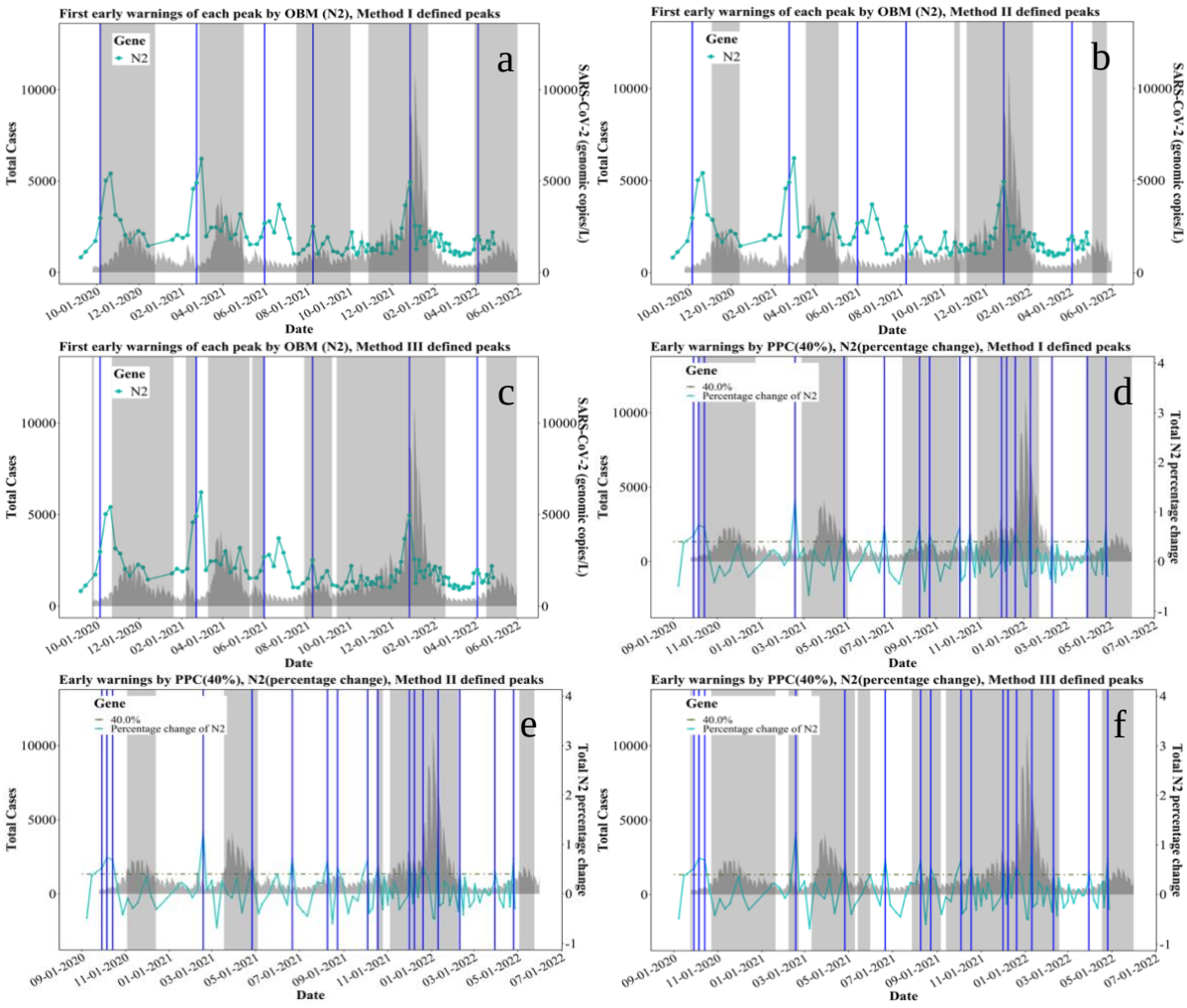
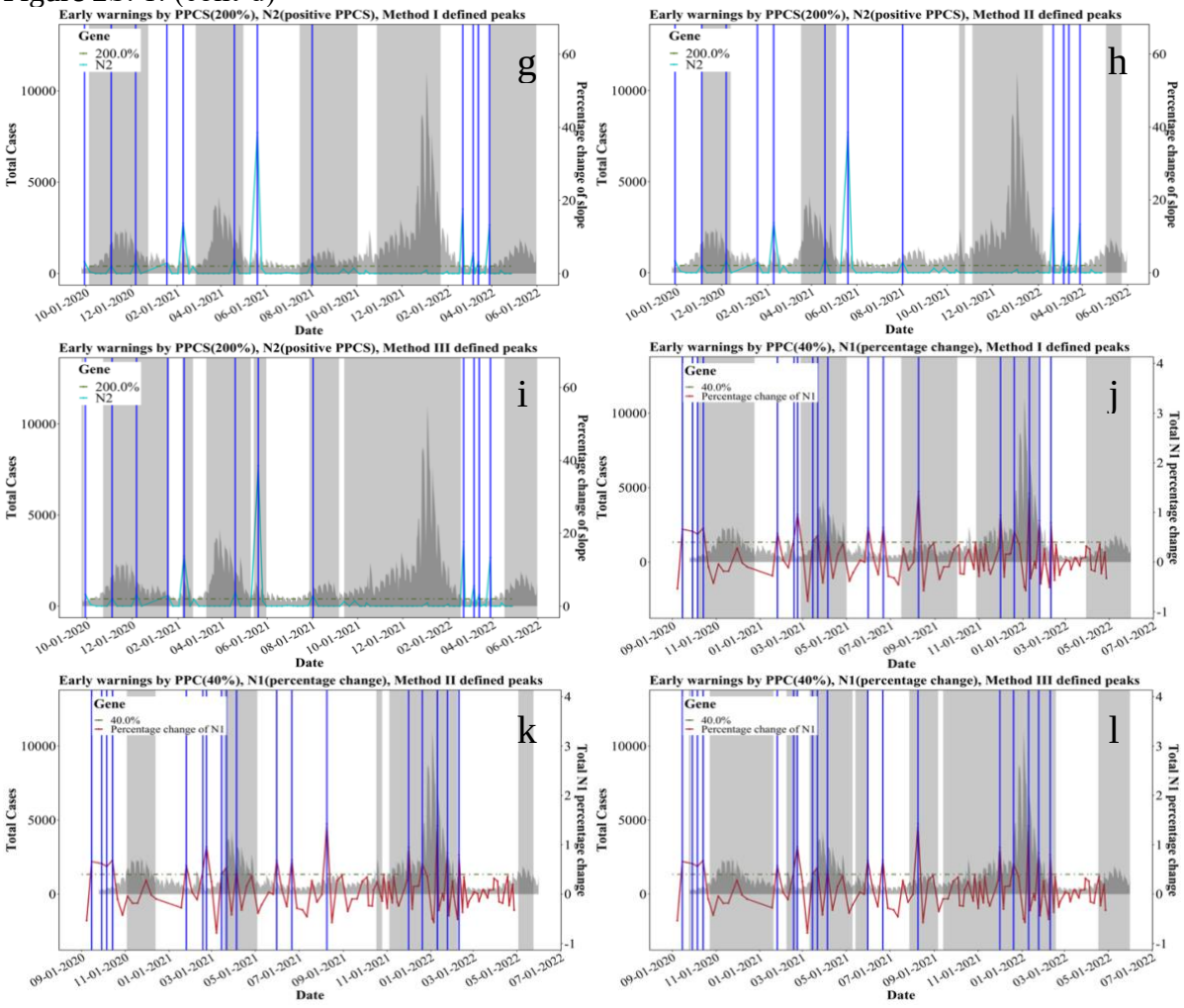


Figure 2S. 1. Real-time early warning methods based on N2: OBM, PPC, and PPCS:
a. First early warnings of each peak identified by OBM (N2, gc/L) with Method I defined peaks
b. First early warnings of each peak identified by OBM (N2, gc/L) with Method II defined peaks
c. First early warnings of each peak identified by OBM (N2, gc/L) with Method III defined peaks
d. Early warnings identified by PPC (N2, gc/L) with Method I defined peaks
e. Early warnings identified by PPC (N2, gc/L) with Method II defined peaks
f. Early warnings identified by PPC (N2, gc/L) with Method III defined peaks
g. Early warnings identified by PPCS (N2, gc/L) with Method I defined peaks
h. Early warnings identified by PPCS (N2, gc/L) with Method II defined peaks
i. Early warnings identified by PPCS (N2, gc/L) with Method III defined peaks
j. Early warnings identified by PPC (N1, gc/L) with Method I defined peaks
k. Early warnings identified by PPC (N1, gc/L) with Method II defined peaks
l. Early warnings identified by PPC (N1, gc/L) with Method III defined peaks

Figure 2S. 1. (cont'd)



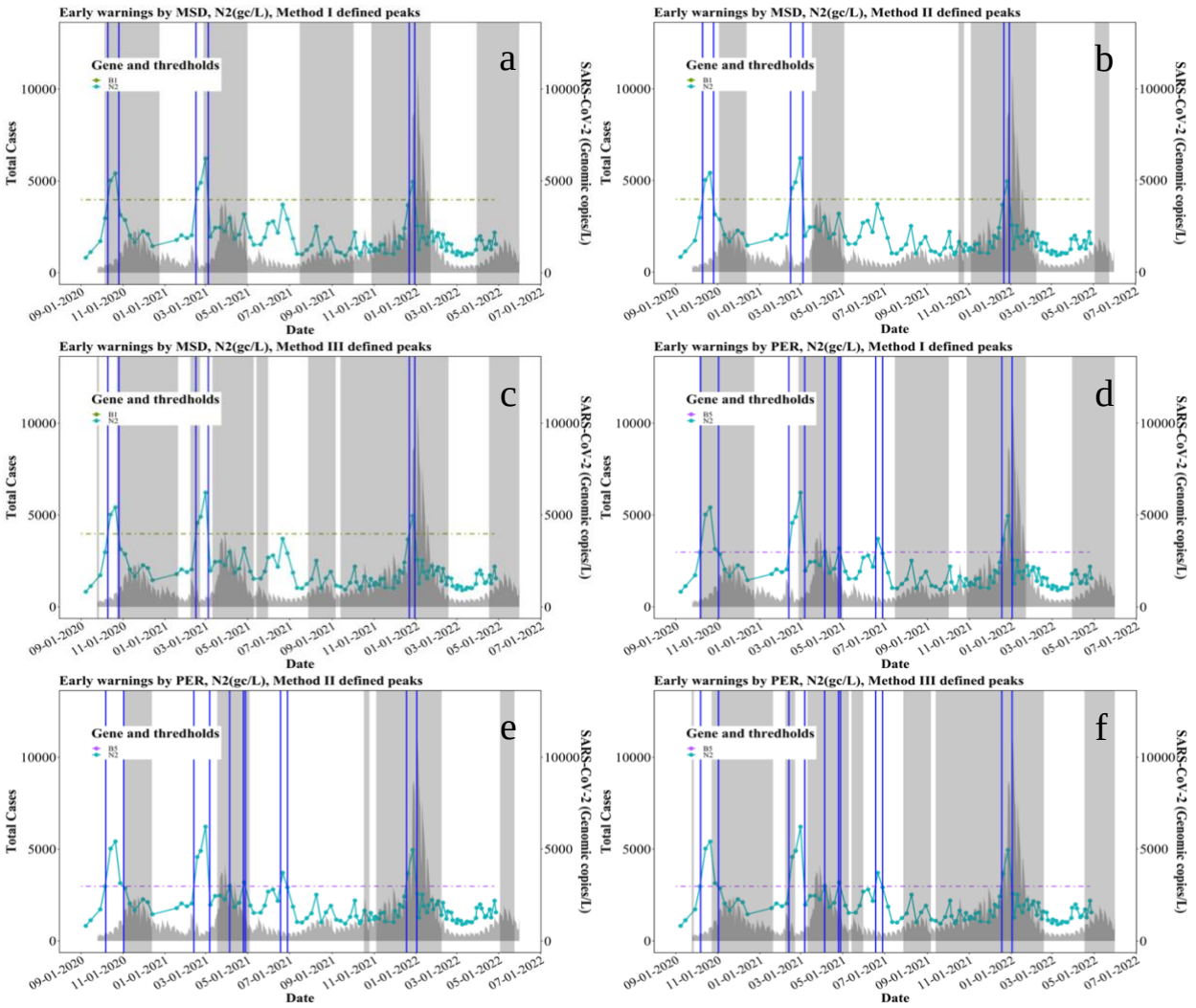
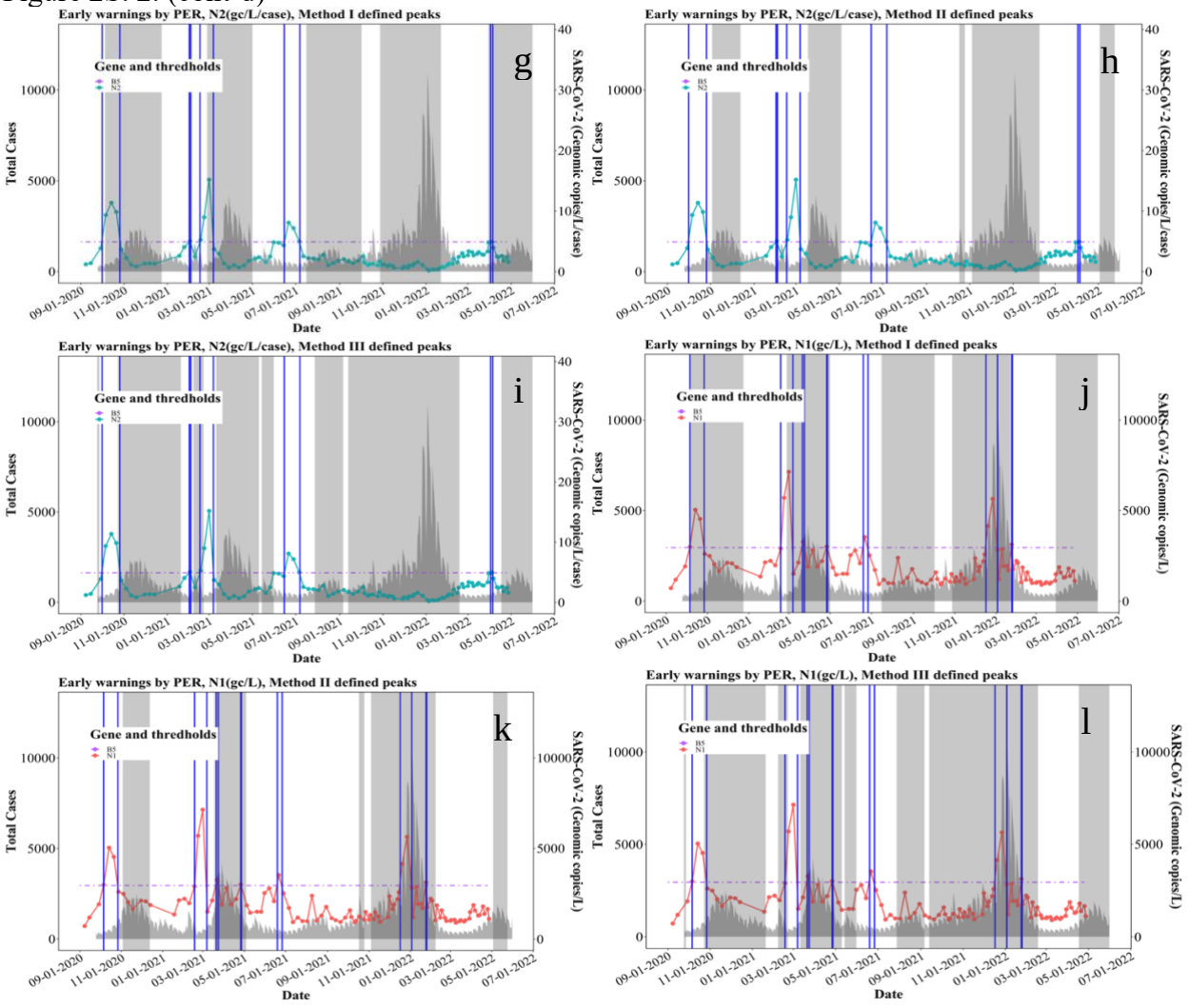


Figure 2S. 2. Post-factum early warning methods based on N2: MSD, VAM, and PER:
a. Early warnings identified by MSD (N2, gc/L) with Method I defined
b. Early warnings identified by MSD (N2, gc/L) with Method II defined
c. Early warnings identified by MSD (N2, gc/L) with Method III defined
d. Early warnings identified by PER (N2, gc/L) with Method I defined peaks
e. Early warnings identified by PER (N2, gc/L) with Method II defined peaks
f. Early warnings identified by PER (N2, gc/L) with Method III defined peaks
g. Early warnings identified by PER (N2, gc/L/case) with Method I defined peaks
h. Early warnings identified by PER (N2, gc/L/case) with Method II defined peaks
i. Early warnings identified by PER (N2, gc/L/case) with Method III defined peaks
j. Early warnings identified by PER (N1, gc/L) with Method I defined peaks
k. Early warnings identified by PER (N1, gc/L) with Method II defined peaks
l. Early warnings identified by PER (N1, gc/L) with Method III defined peaks

Figure 2S. 2. (cont'd)



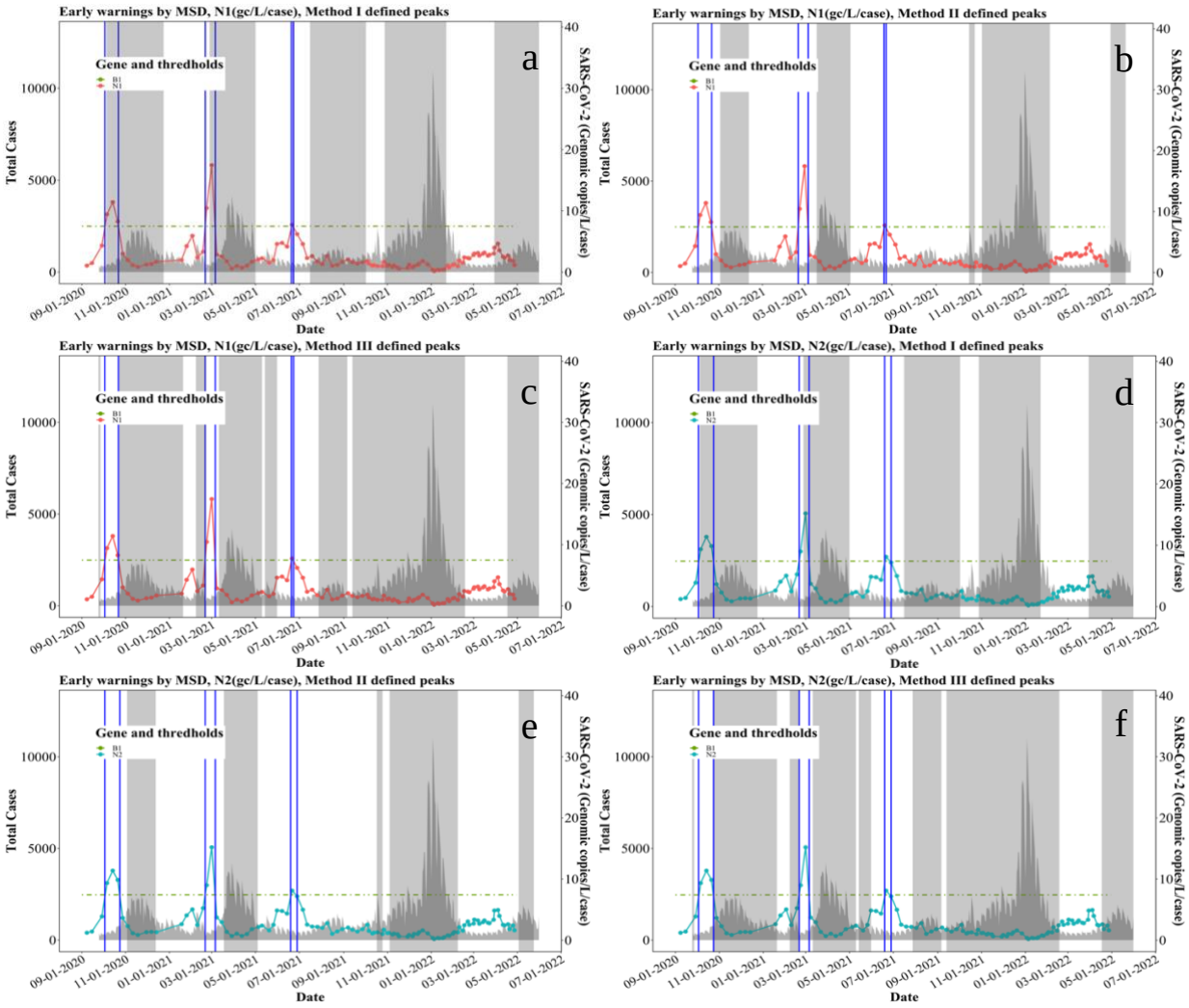


Figure 2S. 3. Early warnings by post-factum method: MSD

a. Early warnings identified by MSD (N1/c, gc/L/case) with Method I defined

b. Early warnings identified by MSD (N1/c, gc/L/case) with Method II defined

c. Early warnings identified by MSD (N1/c, gc/L/case) with Method III defined

d. Early warnings identified by MSD (N2/c, gc/L/case) with Method I defined

e. Early warnings identified by MSD (N2/c, gc/L/case) with Method II defined

f. Early warnings identified by MSD (N2/c, gc/L/case) with Method III defined

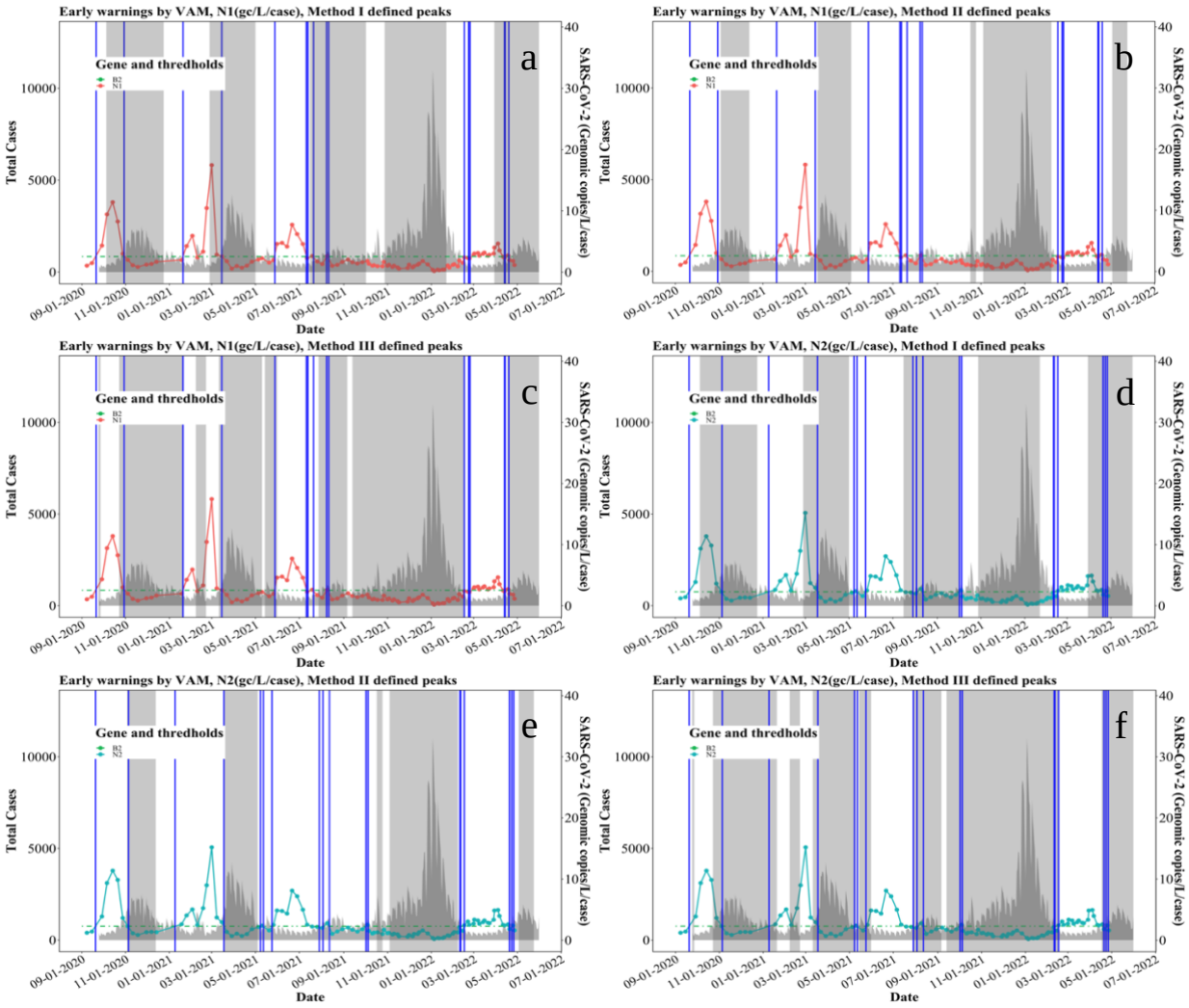


Figure 2S. 4. Early warnings by post-factum method: VAM

a. Early warnings identified by VAM (N1/c, gc/L/case) with Method I defined

b. Early warnings identified by VAM (N1/c, gc/L/case) with Method II defined

c. Early warnings identified by VAM (N1/c, gc/L/case) with Method III defined

d. Early warnings identified by VAM (N2/c, gc/L/case) with Method I defined

e. Early warnings identified by VAM (N2/c, gc/L/case) with Method II defined

f. Early warnings identified by VAM (N2/c, gc/L/case) with Method III defined

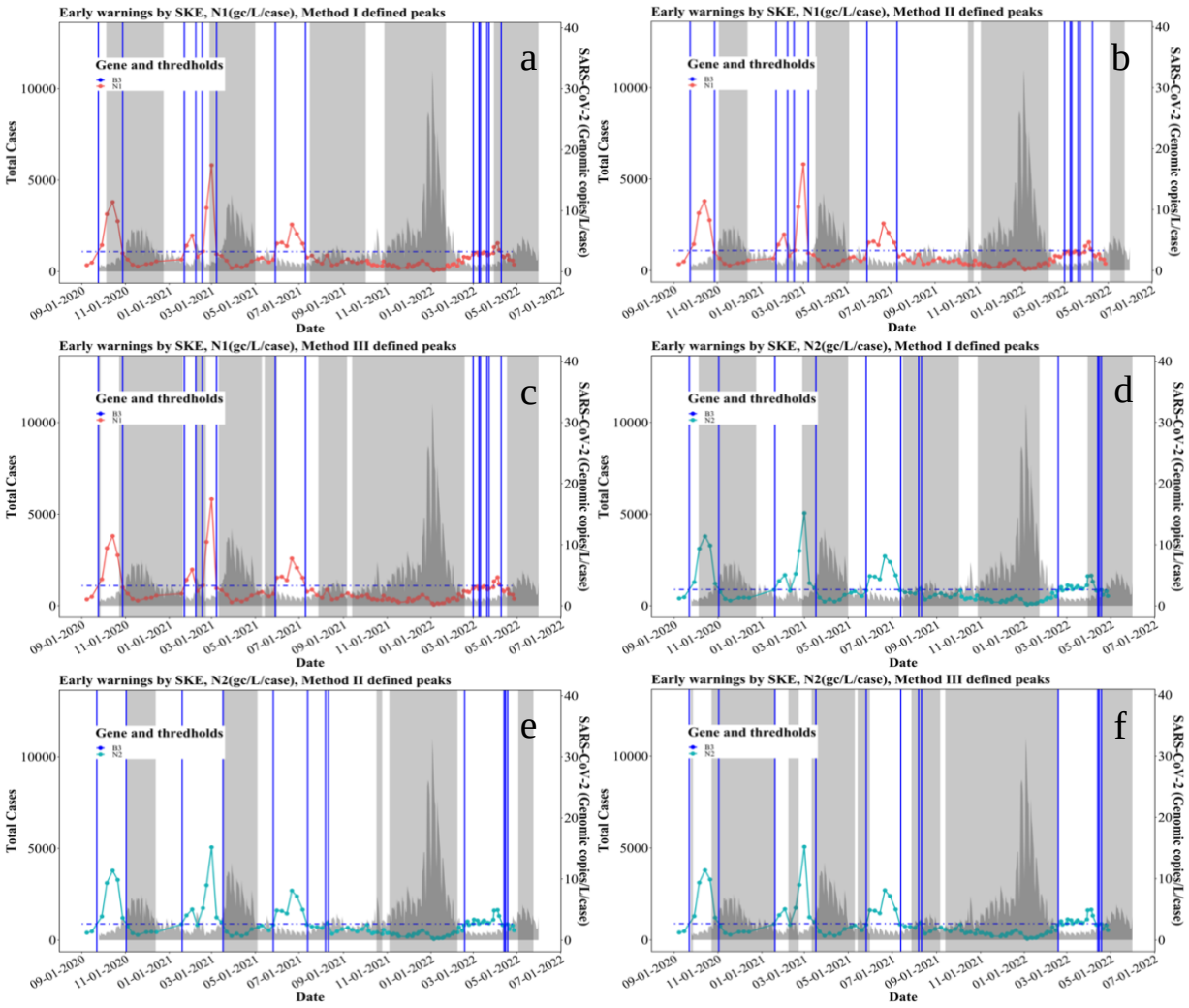


Figure 2S. 5. Early warnings by post-factum method: SKE

a. Early warnings identified by SKE (N1/c, gc/L/case) with Method I defined

b. Early warnings identified by SKE (N1/c, gc/L/case) with Method II defined

c. Early warnings identified by SKE (N1/c, gc/L/case) with Method III defined

d. Early warnings identified by SKE (N2/c, gc/L/case) with Method I defined

e. Early warnings identified by SKE (N2/c, gc/L/case) with Method II defined

f. Early warnings identified by SKE (N2/c, gc/L/case) with Method III defined

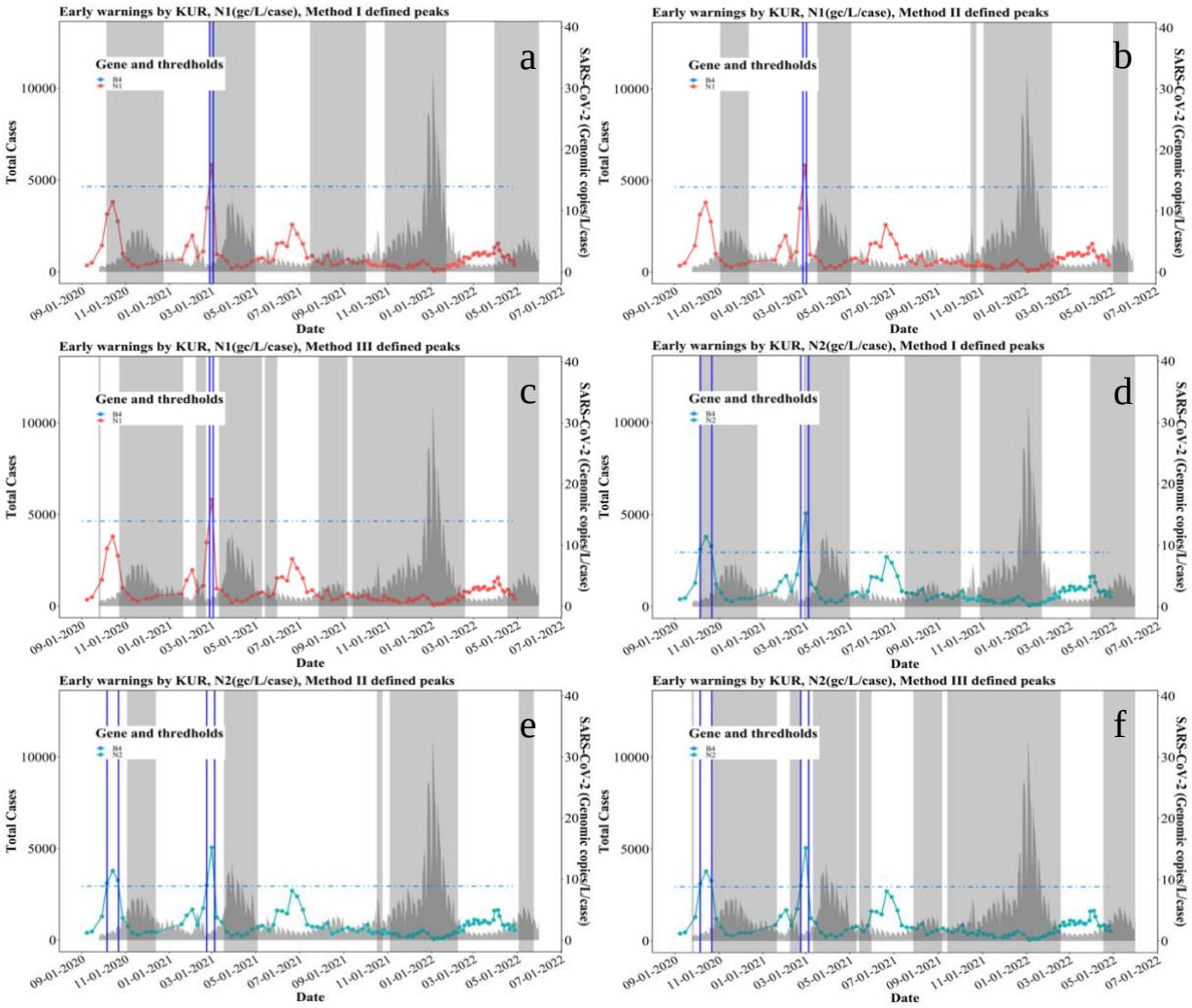


Figure 2S. 6. Early warnings by post-factum method: KUR

a. Early warnings identified by KUR (N1/c, gc/L/case) with Method I defined

b. Early warnings identified by KUR (N1/c, gc/L/case) with Method II defined

c. Early warnings identified by KUR (N1/c, gc/L/case) with Method III defined

d. Early warnings identified by KUR (N2/c, gc/L/case) with Method I defined

e. Early warnings identified by KUR (N2/c, gc/L/case) with Method II defined

f. Early warnings identified by KUR (N2/c, gc/L/case) with Method III defined

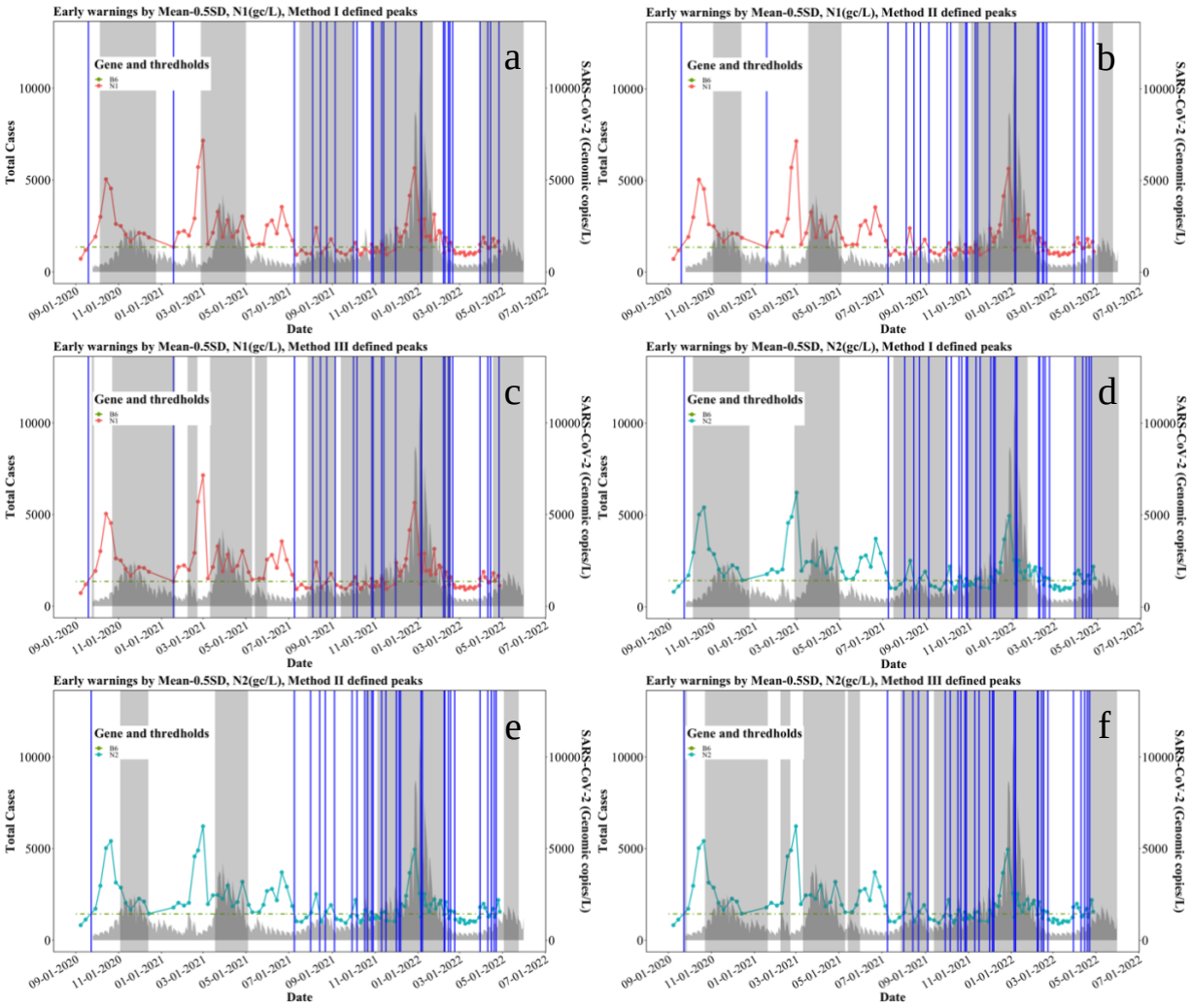


Figure 2S. 7. Early warnings by post-factum method: Mean-0.5SD

a. Early warnings identified by Mean-0.5SD (N1, gc/L) with Method I defined

b. Early warnings identified by Mean-0.5SD (N1, gc/L) with Method II defined

c. Early warnings identified by Mean-0.5SD (N1, gc/L) with Method III defined

d. Early warnings identified by Mean-0.5SD (N2, gc/L) with Method I defined

e. Early warnings identified by Mean-0.5SD (N2, gc/L) with Method II defined

f. Early warnings identified by Mean-0.5SD (N2, gc/L) with Method III defined

g. Early warnings identified by Mean-0.5SD (N1/c, gc/L/case) with Method I defined

h. Early warnings identified by Mean-0.5SD (N1/c, gc/L/case) with Method II defined

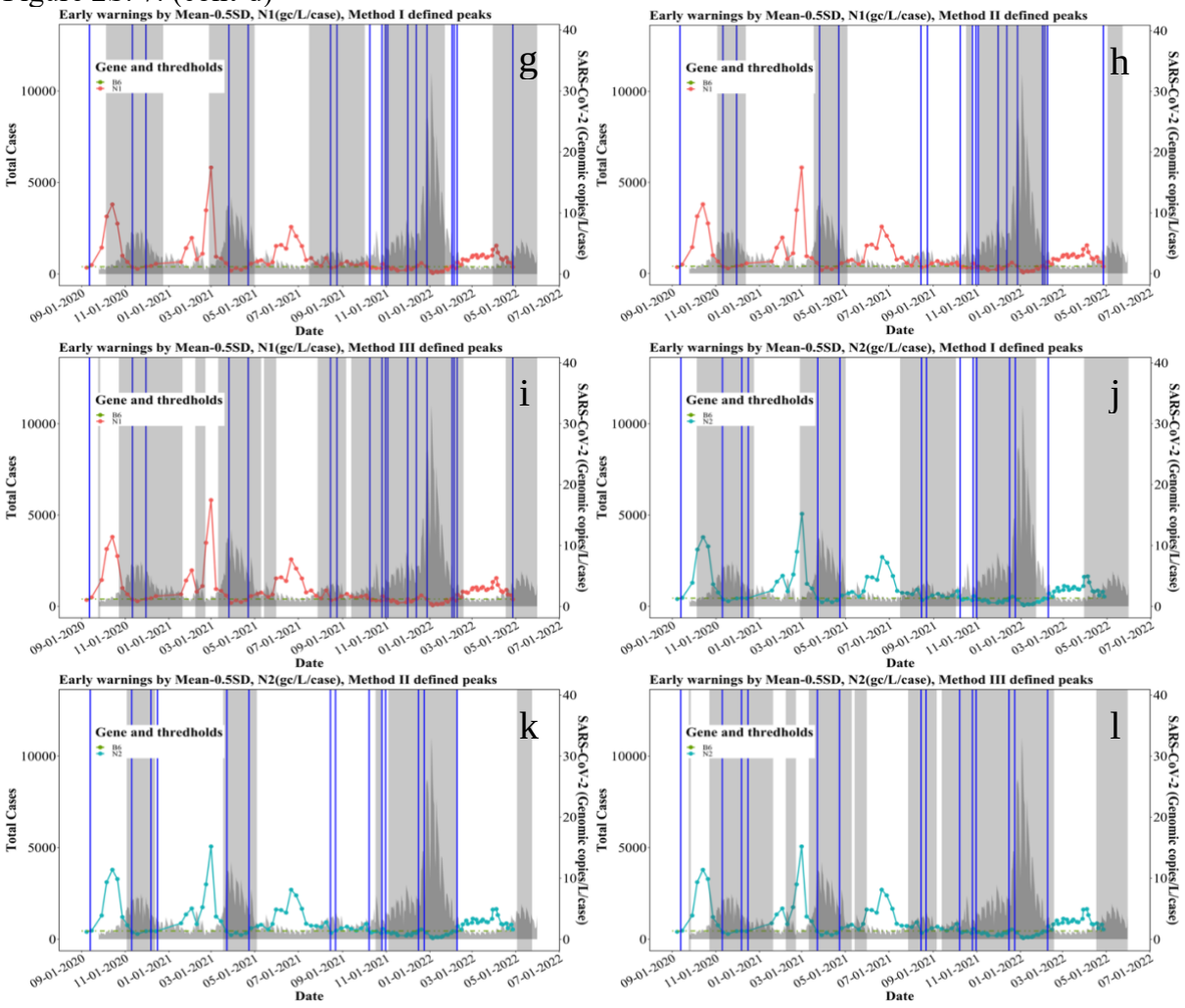
i. Early warnings identified by Mean-0.5SD (N1/c, gc/L/case) with Method III defined

j. Early warnings identified by Mean-0.5SD (N2/c, gc/L/case) with Method I defined

k. Early warnings identified by Mean-0.5SD (N2/c, gc/L/case) with Method II defined

l. Early warnings identified by Mean-0.5SD (N2/c, gc/L/case) with Method III defined

Figure 2S. 7. (cont'd)



CHAPTER 3: TARGETING A FREE VIRAL FRACTION ENHANCES THE EARLY ALERT POTENTIAL OF WASTEWATER SURVEILLANCE FOR SARS-COV-2: A METHODS COMPARISON SPANNING THE TRANSITION BETWEEN DELTA AND OMICRON VARIANTS IN A LARGE URBAN CENTER

Published in *Frontiers in Public Health*:

Zhao, L., Geng, Q., Corchis-Scott, R., McKay, R. M., Norton, J., & Xagoraraki, I. (2023). Targeting a free viral fraction enhances the early alert potential of wastewater surveillance for SARS-CoV-2: a methods comparison spanning the transition between delta and omicron variants in a large urban center. *Frontiers in Public Health*, *11*, 1140441.

Abstract

Wastewater surveillance has proven to be a valuable approach to monitoring the spread of SARS-CoV-2, the virus that causes Coronavirus disease 2019 (COVID-19). Recognizing the benefits of wastewater surveillance as a tool to support public health in tracking SARS-CoV-2 and other respiratory pathogens, numerous wastewater virus sampling and concentration methods have been tested for appropriate applications as well as their significance for actionability by public health practices. Here, we present a 34-week long wastewater surveillance study that covers nearly 4 million residents of the Detroit (MI, United States) metropolitan area. Three primary concentration methods were compared with respect to recovery of SARS-CoV-2 from wastewater: Virus Adsorption-Elution (VIRADEL), polyethylene glycol precipitation (PEG), and polysulfone (PES) filtration. Wastewater viral concentrations were normalized using various parameters (flow rate, population, total suspended solids) to account for variations in flow. Three analytical approaches were implemented to compare wastewater viral concentrations across the three primary concentration methods to COVID-19 clinical data for both normalized and non-

normalized data: Pearson and Spearman correlations, Dynamic Time Warping (DTW), and Time Lagged Cross Correlation (TLCC) and peak synchrony. It was found that VIRADEL, which captures free and suspended virus from supernatant wastewater, was a leading indicator of COVID-19 cases within the region, whereas PEG and PES filtration, which target particle-associated virus, each lagged behind the early alert potential of VIRADEL. PEG and PES methods may potentially capture previously shed and accumulated SARS-CoV-2 resuspended from sediments in the interceptors. These results indicate that the VIRADEL method can be used to enhance the early-warning potential of wastewater surveillance applications although drawbacks include the need to process large volumes of wastewater to concentrate sufficiently free and suspended virus for detection. While lagging the VIRADEL method for early-alert potential, both PEG and PES filtration can be used for routine COVID-19 wastewater monitoring since they allow a large number of samples to be processed concurrently while being more cost-effective and with rapid turn-around yielding results same day as collection.

1. Introduction

Wastewater surveillance has been widely adopted by researchers and health agencies as an effective tool for tracking Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in wastewater amid the Coronavirus Disease 2019 (COVID-19) pandemic (Ahmed et al., 2020a, 2022; Ai et al., 2021; Bivins & Bibby, 2021; Chik et al., 2021; Kaya et al., 2022; Li et al., 2022; Miyani et al., 2020, 2021; Sherchan et al., 2020; Xiao et al., 2022; Zhao et al., 2022a, 2022b). SARS-CoV-2 was first identified in Wuhan, Hubei, China, and was designated a Public Health Emergency of International Concern on January 30th, 2020, by the World Health Organization (WHO). COVID-19 was later declared a pandemic on March 11th, 2020 (who.int). Numerous studies have demonstrated that SARS-CoV-2 can be shed from the gastrointestinal tract of infected

individuals and its viral RNA can persist and be detected in wastewater (Boucau et al., 2022; Haramoto et al., 2020; He et al., 2020; Jones et al., 2020; Medema et al., 2020). To increase the sensitivity of the assay used to detect viral RNA in wastewater, samples are routinely concentrated prior to quantification (Farkas et al., 2020; Xagoraraki, 2020; Xagoraraki et al., 2014).

Methods used in published studies to recover and concentrate SARS-CoV-2 viral RNA from wastewater encompass a wide range of techniques including Virus Adsorption-Elution (VIRADEL), polyethylene glycol precipitation (PEG), ultrafiltration, ultracentrifugation, concentrating pipette, filtration and so forth. Some of the methods, such as VIRADEL, exclude large solids and focus on free and suspended viral particles in supernatant wastewater. Other methods, such as PEG precipitation and filtration, target particulate matter and the associated viruses that are sorbed onto solids. Notably, this fraction may preferentially settle within the sewer when flow is reduced and likewise is susceptible to resuspension when flows are elevated (Flood et al., 2021; Zhao et al., 2022b).

The recovery efficiencies of concentration methods are variable, differing between method, virus type and conditioning of the wastewater sample. Notably, VIRADEL was found to be effective for concentrating viruses from water samples with recovery efficiencies of more than 90% for poliovirus (Jakubowski et al., 1975; Wallis et al., 1972), 54.4% for murine norovirus (MNV) (Lee et al., 2011), 51% for echovirus (Hill et al., 2009), 35% for enteric virus (Black et al., 2007), and 4.7% for adenovirus (Francy et al., 2013). Likewise, PEG was found to be effective for concentrating viruses in water samples, with recovery efficiencies of 89.5% for echovirus (Ye et al., 2016), 86% for hepatitis A virus (Michael-Kordatou et al., 2020), 68% for poliovirus (Michael-Kordatou et al., 2020), and 56.7% (Sapula et al., 2021) and 26.4% (Barril et al., 2021) for SARS-CoV-2. Filtration was reported to recover virus from wastewater samples with recovery

efficiencies ranging from 26.7% to 65.7% for murine hepatitis virus (Ahmed et al., 2020c), and 90% for human betacoronavirus OC43 (Pecson et al., 2021).

Applying different concentration methods can achieve different goals. For instance, use of VIRADEL to concentrate SARS-CoV-2 can provide early warnings of impending COVID-19 cases (Miyani et al., 2021; Zhao et al., 2022a, 2022b). PEG precipitation is an economical and widely adopted method that allows a large number of samples to be processed concurrently and it is suitable for routine COVID-19 wastewater monitoring (Flood et al., 2021; Lu et al., 2020a). Likewise, filtration presents a cost-effective and simple approach commonly applied to recover cells and viral particles from environmental samples for nucleic acid extraction (McKindles et al., 2020), which has also been applied to recovery of SARS-CoV-2 from wastewater (Chik et al., 2021; Corchis-Scott et al., 2021; Gonçalves et al., 2021; Lu et al., 2020a).

Here we present a comparison of three primary concentration methods (VIRADEL, PEG and filtration) to detect SARS-CoV-2 viral RNA in wastewater, in relation to COVID-19 cases amid the transition from Delta to Omicron Variants of Concerns (VOCs) circulating in the Detroit, MI metropolitan area. Similarities and correlations were examined among the three concentration methods with both normalized and non-normalized data. The lead/lag time of each method in relation to the total COVID-19 cases was also assessed. The results presented in this study will assist researchers and public health practitioners to select appropriate primary concentration methods for the recovery of SARS-CoV-2 from wastewater for different wastewater surveillance practices.

2. Materials and Methods

Untreated wastewater samples were collected weekly from the Water Resource Recovery Facility (WRRF) of the Great Lakes Water Authority (GLWA) located in Detroit, MI, USA,

between October 1, 2021, and May 31, 2022. The WRRF serves the needs of Detroit and 76 area communities with a service area of more than 2450 square kilometers serving nearly 4 million people. WRRF collects and treats stormwater, as well as residential, industrial, and commercial waste, depending on service areas, with its semi-combined sewershed system. WRRF receives wastewater via three main interceptors including the Detroit River Interceptor (DRI), the North Interceptor-East Arm (NIEA), and the Oakwood-Northwest-Wayne County Interceptor (ONWI) (see chapter 1, Figure 1. 2.), serving the City of Detroit as well as the three largest Michigan counties by population: Wayne, Oakland, and Macomb. Composite samples collected over 24-h were used to compare the polyethylene glycol (PEG) precipitation and filtration methods, however, the larger volumes required by the virus adsorption-elution (VIRADEL) method necessitated a targeted approach with samples collected between 15:30 to 18:00 each afternoon. The samples were collected from the three interceptors at the point of discharge into the WRRF and maintained chilled on ice during transport to the lab for primary concentration and sample analyses.

2.1 Virus adsorption-elution (VIRADEL) method

The United States Environmental Protection Agency virus adsorption-elution (VIRADEL) method employing electropositive or electronegative filters was reported to recover and concentrate viruses from wastewater samples previously (Li et al., 2022; Lu et al., 2020a; Miyani et al., 2020, 2021; Xagoraraki et al., 2014; Zhao et al., 2022a, 2022b). Electronegative filters require preconditioning such as adjusting the pH, prior to downstream concentration processes. Electropositive filters do not require any preconditioning (Lu et al., 2020a; Xagoraraki et al., 2014). In this study, depending on the quantity of suspended solids in the wastewater, 10 to 50 L of untreated wastewater (grab sample) was passed through NanoCeram electropositive cartridge

filters (Argonide, Sanford, FL, USA) at a rate less than 11 L/min using a previously described method (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022b). Flow meter readings were tracked at the beginning and end of each sampling event to measure the total volume of wastewater passing through the filters. Following sampling, the NanoCeram filters were transported on ice to the lab for sample analyses within 24 h. The elution process releases viral particles captured by the filters (Xagoraraki et al., 2014). Viruses were eluted using 1.5% beef extract containing 0.05 M glycine, based on a previously described method (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022b). Subsequently, the eluates containing viruses were flocculated by adjusting the pH, following multiple centrifugations and resuspension of particles in sodium phosphate. Afterwards, supernatants containing viruses were separated by adjusting the pH and centrifugation. Finally, the supernatants containing viruses were passed through 0.45 μm and 0.22 μm Millipore filters (MilliporeSigma, Burlington, MA, USA), which were followed by aliquoting and storage of the final aliquots at -80 ℃ for downstream molecular analyses (Li et al., 2022; Miyani et al., 2020, 2021; Xagoraraki et al., 2014; Zhao et al., 2022b). Bacteriophage Phi6 was applied as a proxy virus to estimate the recovery efficiency during virus concentration (Kantor et al., 2021; Ye et al., 2016; Zhao et al., 2022b). Figure 3. 1. demonstrates the workflow of the VIRADEL method.

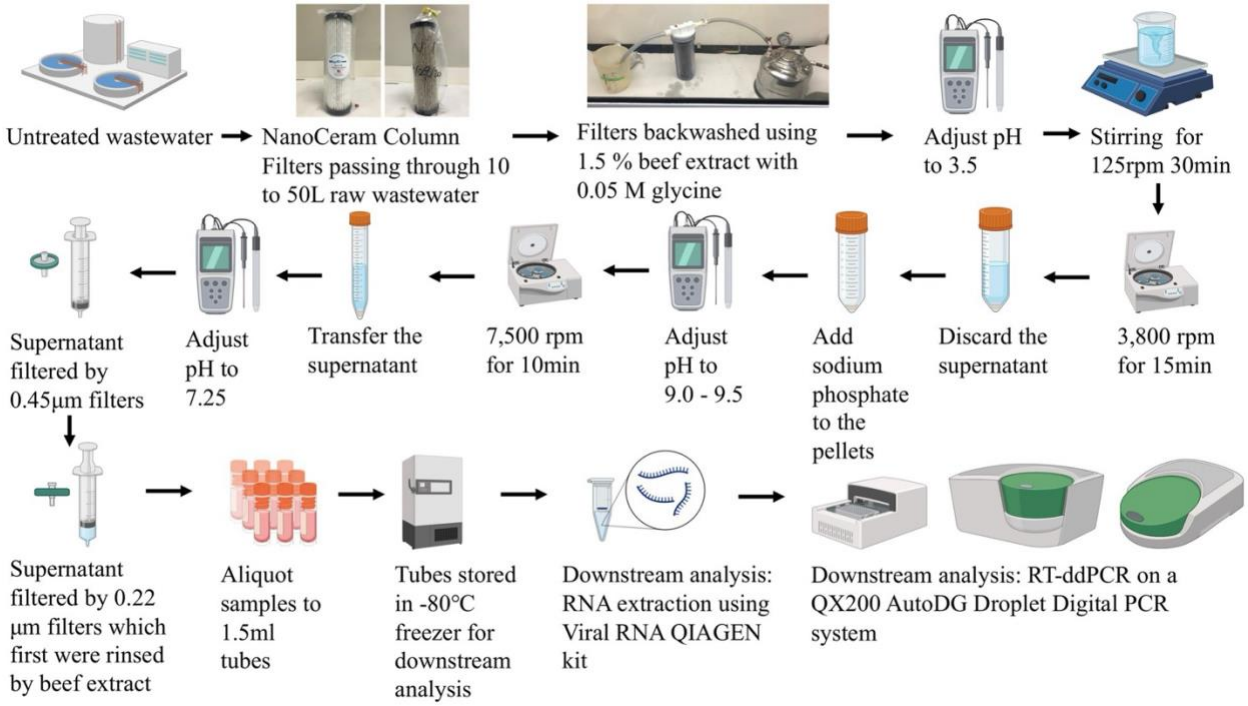


Figure 3. 1. Illustrative flowchart of the VIRADEL method and subsequent analysis

2.2 Polyethylene glycol precipitation (PEG) method

From a 24-h composite sample of untreated wastewater collected in a 1 L Nalgene bottle, 100 mL samples were mixed with 0.2 M sodium chloride and 8% polyethylene glycol (w/v). Samples were mixed gently on a magnetic stirrer at 4 °C for 2 h, followed by centrifugation at 4700×g for 45 min at 4 °C. The supernatant was removed, and the pellet was resuspended in the remaining liquid (approximately 2-3 mL). The final concentrate volumes were between 1 to 6 mL. All sample concentrates were then subjected to downstream analyses including RNA extraction and RT-ddPCR (Figure 3. 2.).

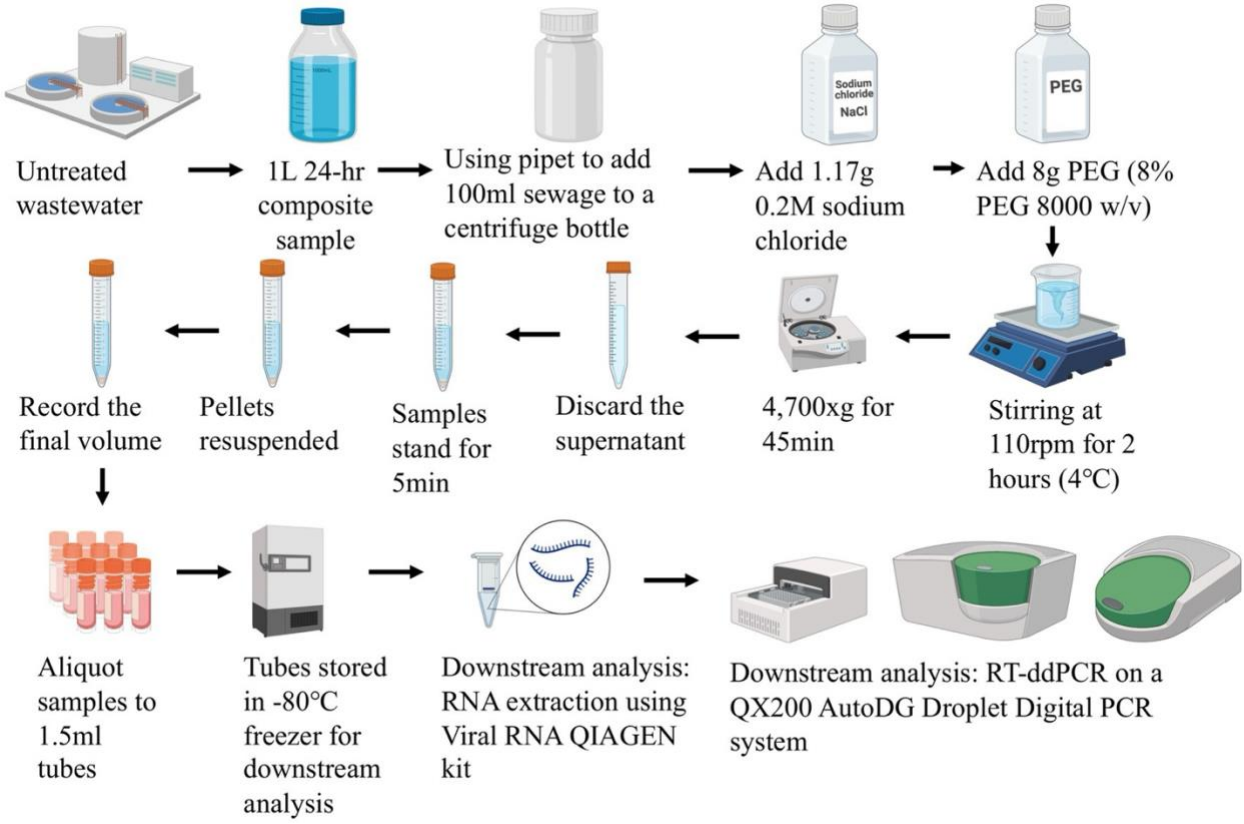


Figure 3. 2. Illustrative flowchart of the PEG method and subsequent analysis

2.3 Filtration method

Composite samples of raw wastewater collected as for the PEG method were concentrated by filtering 50-120 mL through 0.22 µm Sterivex PES cartridge filters (MilliporeSigma, Burlington, MA, USA) using a 50 mL syringe fitted into a caulking gun. Immediately following filtration, the filters were sealed and flash-frozen through immersion in liquid nitrogen as described previously (Corchis-Scott et al., 2021). Subsequently, filters were subjected to downstream processes including RNA extraction and RT-qPCR (Figure 3. 3.).

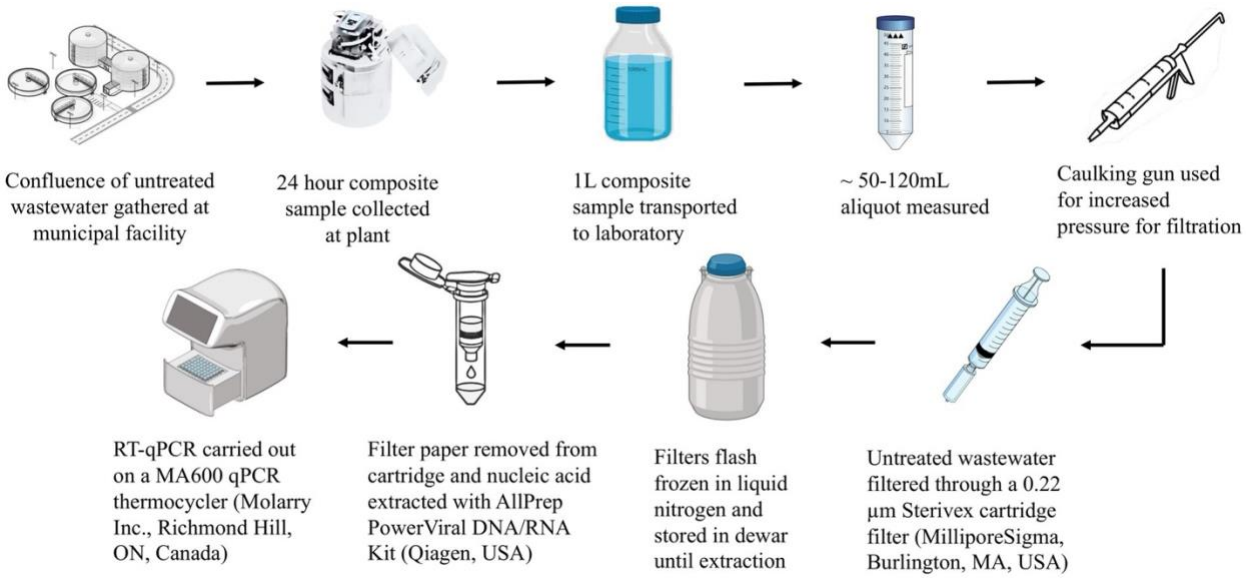


Figure 3. 3. Illustrative flowchart of the filtration method and subsequent analysis

2.4 RNA extraction, RT-ddPCR, RT-qPCR

Following VIRADEL and PEG methods, viral RNA was extracted using the QIAamp Viral RNA kit (Qiagen, Germantown, MD, USA), following the manufacturer's protocol modified by use of 140 µL elution buffer to extract the viral RNA (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022b). RT-ddPCR was performed on a QX200 AutoDG Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA), using the One-step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA) as described previously (Li et al., 2022; Zhao et al., 2022b). United States Centers for Disease Control and Prevention (US CDC) primers and probes that target the N1 and N2 genes of SARS-CoV-2 were used (Li et al., 2022; Zhao et al., 2022a, 2022b). N1 N2 gene Duplex Assay Reaction Mixture was reported previously (Li et al., 2022; Zhao et al., 2022a, 2022b). Following the preparation of the Duplex Mixture and oil droplets generation, samples were run on a C1000 Touch Thermal Cycler (Bio-Rad, Hercules, CA, USA) using the thermocycling conditions which were reported previously (Li et al., 2022; Zhao et al., 2022a,

2022b). Subsequently, the measurement of fluorescence was performed on a QX200 Droplet Reader (Bio-Rad, Hercules, CA, USA). For each RT-ddPCR run, positive controls (PTCs), negative controls (NTCs), and process negative controls were included, which were described previously (Zhao et al., 2022a). All samples were run in triplicate. The Limit of Detection (LOD) and Limit of Blank (LOB) for RT-ddPCR were described and determined previously (Li et al., 2022; Zhao et al., 2022a, 2022b).

Following the filtration method, filters were thawed, and RNA was extracted from the filters using the AllPrep PowerViral DNA/RNA kit (Qiagen, Germantown, MD, USA) modified by addition of 5% 2-mercaptoethanol (v/v). RNA was eluted in 50 µL of RNAse free water. Samples were not treated with DNase upon extraction. Assays for SARS-CoV-2 targeted regions of the nucleocapsid (N) gene using US CDC primers and probes for the N1 and N2 regions (Lu et al., 2020b). Reagents were supplied by Integrated DNA Technologies (Coralville, IA, USA). Reactions contained 5 µL of RNA template mixed with 10 µL of 2 × RT-qPCR master mix (Takyon TM Dry One-Step RT Probe MasterMix No Rox, Eurogentec, Liège, Belgium) and primers and probes in a final reaction volume of 20 µL. Reaction inhibition was assessed using VetMAX XENO Internal Positive Control RNA (Applied Biosystems Corp., Waltham, MA, USA). Due to repeated incidence of inhibition with wastewater samples processed by filtration, template was diluted 1:5 in all reactions. Technical triplicates were run for detection of gene targets. Thermal cycling was performed using a MA6000 qPCR thermocycler (Sansure Biotech, Changsha, China). RT was performed at 48 °C for 10 min, followed by polymerase activation at 95 °C for 3 min, and 50 cycles of denaturation, annealing/extension at 95 °C for 10 sec, then 60 °C for 45 sec, respectively. The EDX SARS-CoV-2 synthetic RNA standard (Exact Diagnostics, Fort Worth, TX, USA) was used to create a 7-point standard curve to quantify N1 and N2 gene

targets. No template controls yielded no amplification, and we report a limit of detection of 5 gene copies of N1 and N2 per reaction containing 5 µL of template RNA for RT-qPCR.

2.5 COVID-19 clinical data

Publicly available clinical data were accessed on August 22, 2022, for the period between October 1, 2021, and May 31, 2022, for the city of Detroit, as well as Wayne, Macomb, and Oakland counties (Michigan.gov) (Figure 3. 4. a). Clinical data with a 7-day moving average (Barua et al., 2022; Menkir et al., 2021; Zhao et al., 2022a) was used for further statistical analysis (Figure 3. 4. b.). COVID-19 clinical data were only available per city/county for the Detroit metropolitan area. Each interceptor received wastewater from portions of each city/county. Therefore, only the total SARS-CoV-2 concentrations can be correlated to the total COVID-19 cases of each city/county (Zhao et al., 2022a, 2022b).

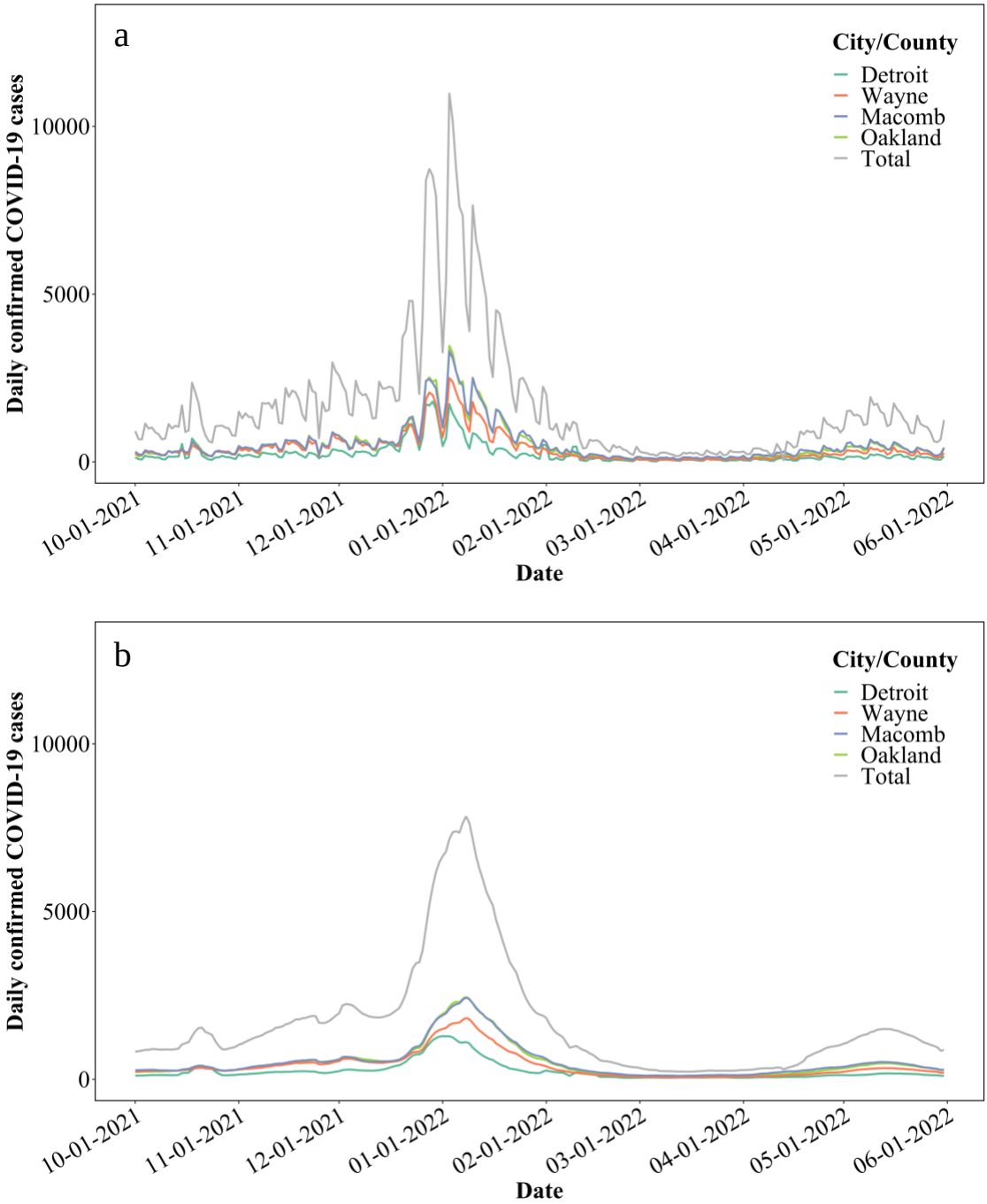


Figure 3. 4. a. COVID-19 cases in the City of Detroit, as well as Wayne, Macomb, and Oakland counties; b. 7-day moving average of the COVID-19 cases

2.6 Data analyses and visualization

Data were tracked and organized using Microsoft Excel version 16.66.1. R version 4.1.3 (2022-03-10) was applied to perform data analysis including Pearson and Spearman correlations, Dynamic Time Warping (DTW), Time Lagged Cross Correlation (TLCC) and peak synchrony, depending primarily on ggplot2 package for visualization, and packages including dtw, synchrony, dplyr, and ggpubr. Missing data from samples were filled using linear interpolation for further analysis (Lepot et al., 2017; Zhao et al., 2022a, 2022b). For VIRADEL samples, 128 gene concentrations were measured for both N1 and N2 gene between 10/1/21 and 5/31/22. For PEG samples, 88 gene concentrations were measured for both N1 and N2 gene between 10/1/21 and 5/31/22. For filtration samples, 66 gene concentrations were measured for both N1 and N2 gene between 10/1/21 and 5/31/22. To perform correlation analyses between weekly gene concentrations and daily clinical cases, linear interpolation was conducted to generate daily data based on weekly measurements. The number of interpolated daily gene concentrations were 179, 199, and 210 for VIRADEL, PEG, and filtration, respectively.

To account for the changing flow in wastewater, dilution events, and variability in the solids portion of the wastewater, four approaches (flow rate, flow rate/population, TSS, flow rate×TSS) of normalizing the N1 and N2 gene concentrations (gc/L) were implemented using Eq. (1), Eq. (2), Eq. (3), and Eq. (4) (Hopkins et al., 2023; Terry & Meschke, 2022; Zhao et al., 2022a). TSS, or “Total Suspended Solids”, is an estimate of the entire solids in wastewater in contrast to the liquid fraction or dissolved matter (Terry & Meschke, 2022). In addition, other parameters, including sanitary percentage and Biological Oxygen Demand (BOD), were proved ineffective for normalizing N1 and N2 gene concentrations for the Detroit area and other areas, thus, they were not considered in the current study (Ai et al., 2021; Zhao et al., 2022a). SARS-CoV-2 gene

concentrations measured in the wastewater following VIRADEL, PEG, and filtration methods are reported as gene copies per L (gc/L). The units after normalization using flow rate, flow rate/population, TSS, and flow rate×TSS, are gene copies per day (gc/day), gene copies per day per person (gc/day/person), gene copies per mg TSS (gc/mg TSS), and gene copies per L per pounds/day (gc/(L(pounds/day))), respectively.

$$C_{(1)} = C_{N1 \text{ or } N2 \text{ gene}} \times V \times f \quad (1)$$

$$C_{(2)} = C_{(1)} / P \quad (2)$$

$$C_{(3)} = C_{N1 \text{ or } N2 \text{ gene}} / TSS \quad (3)$$

$$C_{(4)} = C_{N1 \text{ or } N2 \text{ gene}} / (V \times f \times k \times TSS) \quad (4)$$

where $C_{(1)}$ is the normalized concentration of SARS-CoV-2 in gc/day. $C_{(2)}$ is the normalized concentration of SARS-CoV-2 in gc/day/person. $C_{(3)}$ is the normalized concentration of SARS-CoV-2 in gc/mg TSS. $C_{(4)}$ is the normalized concentration of SARS-CoV-2 in gc/(L(pounds/day)). V is the volume of wastewater flowing into WWTP interceptors during sampling events (MGD). f is the conversion factor between L and MGD. k is the conversion factor between mg and pounds. P is the total population in the Detroit metropolitan area served by WRRF's interceptors including ONWI, NIEA, and DRI. TSS represents the total suspended solids (mg/L).

2.6.1 Correlations among N1 and N2 gene concentrations by VIRADEL, PEG, and filtration

Multiple studies investigated the applications of both Pearson and Spearman correlations on analyzing the relationship between wastewater viral concentrations of SARS-CoV-2 and COVID-19 clinical cases as well as the relationship among wastewater viral concentrations by different genes or methods (Ai et al., 2021; D'Aoust et al., 2021b; Vadde et al., 2022). In this study, Pearson and Spearman correlations were performed among N1 and N2 gene concentrations

(gc/L, gc/day, gc/day/person, gc/mg TSS, gc/(L(pounds/day))) by VIRADEL, PEG, and filtration methods. The Pearson correlation measures how two time series among VIRADEL, PEG, and filtration gene concentrations covary during the study period, and indicate their linear relationships. The Spearman correlation coefficient is a simple and straightforward approach to analyze the degree of associations between two time series (Ye et al., 2015).

2.6.2 Dynamic Time Warping (DTW)

One commonly used algorithm for quantifying the similarities/dissimilarities between time series data is the Euclidean distance (ED), but numerous studies demonstrated that ED is insensitive to time shifting and patterns between time series since it compares the data points of time series in a settled sequence and cannot consider time shifting or patterns (Izakian et al., 2015; Keogh & Ratanamahatana, 2005). Dynamic time warping (DTW) is a well-established algorithm that circumvents the limitations of ED and compares two time series by computing dynamic distances between them considering regional distortions, time shifting, and the optimal warping that best aligns the time series between each other (Giorgino, 2009; Izakian et al., 2015). Therefore, similar patterns that occur at different times between time series can be considered as matching, thus, the similarity of time series can be evaluated considering their time shifting and shapes by DTW algorithm (Izakian et al., 2015). The DTW algorithm was proposed previously (Giorgino, 2009).

The outcome of DTW analysis indicates two time series with the most similar patterns by calculating the minimum overall dissimilarity or the DTW minimum distance where data points on one time series best align data points on another time series (Giorgino, 2009). Multiple studies investigated the similarities between time series using DTW algorithm (Izakian et al., 2015; Jeong et al., 2011; Rakthanmanon et al., 2012). However, to the best knowledge of the authors, this is

the first study to apply DTW algorithm to compare the similarities between wastewater gene concentrations data by three concentration methods (VIRADEL, PEG, and filtration), as well as comparing the similarities between wastewater gene concentrations data and COVID-19 clinical data. In this study, package dtw and related packages in R (version 4.1.3) were implemented to calculate DTW for the normalized (gc/day, gc/day/person, gc/mg TSS, and gc/(L(pounds/day))) and non-normalized (gc/L) data to analyze the similarities/dissimilarities between VIRADEL, PEG, and filtration methods.

One limitation is that the minimum DTW distance can be affected by the scaling factor of time series data. For instance, the minimum DTW distance between PEG (gc/day/person) and COVID-19 cases can be smaller than the distance between VIRADEL (gc/day/person) and cases, indicating that PEG presents higher similarity to cases than VIRADEL. However, this was affected by the population factor which is a constant number but is not dynamic time series data. Using flow/population normalization including a constant factor intentionally changed the similarities among time series data. Therefore, the minimum DTW distance with flow/population normalized data was not considered for further discussions.

2.6.3 TLCC and peak synchrony

To estimate the leading or lagging relationships between wastewater viral concentrations by three concentration methods (VIRADEL, PEG, and filtration) and total COVID-19 cases, TLCC and peak synchrony were performed where the total COVID-19 cases were shifted over time and correlated with wastewater viral concentrations for each concentration method. TLCC refers to correlations between two time series shifted relatively in time. It can identify the direction and relationship between two time series, for instance, a leader-follower relationship, where the leader time series develop a pattern which is repeated by the follower time series (Shen, 2015).

TLCC is widely applied in analyzing time series especially delay, lead/lag time, and lagged cross correlation and so forth (Hopkins et al., 2023; Li et al., 2007; Mei et al., 2009; Shen, 2015). TLCC is an effective approach to estimate the dynamic relationships between two time series and demonstrate how they shift over time (Hopkins et al., 2023).

In this study, TLCC is measured by gradually shifting total COVID-19 cases between -20 days (lagging) and +20 days (leading), and constantly calculating the Pearson's correlation coefficients between two time series for each shifting. Peak synchrony occurs when the peak correlation is observed. For instance, if the peak correlation is observed at the center where the lag time or offset is 0 day, this condition indicates that the time series are most synchronized at day 0 demonstrating no shifting or lag time. However, the peak correlation can be at a different offset if one time series is leading or lagging another one. R package "synchrony", "devtools", and related packages were implemented to calculate the TLCC and peak synchrony between gene concentrations (both normalized and non-normalized data, by VIRADEL, PEG, and filtration methods) and 7-day moving average total COVID-19 cases.

3. Results

3.1 SARS-CoV-2 viral RNA concentrations in wastewater derived by three concentration methods spanning the transition between Delta and Omicron VOCs

RT-ddPCR (VIRADEL and PEG samples) and RT-qPCR (filtration samples) targeting the N1 and N2 genes was used to quantify SARS-CoV-2 RNA concentrations in wastewater samples collected at GLWA's WRRF over 34 weeks. The study period captured the third major resurgence of COVID-19 cases in the region corresponding to the transition from SARS-CoV-2 Delta (B.1.617.2) variant to Omicron (B.1.1.529) variant (Hopkins et al., 2023; Zhao et al., 2022b).

Filtered samples yielded N1 and N2 gene concentrations higher than those of VIRADEL but lower than those of PEG, for both normalized and non-normalized data (Table 3. 1.). Filtered samples yielded mean N1 and N2 gene concentrations of 3.22E+04 and 1.50E+04 gc/L, respectively. VIRADEL samples yielded mean N1 and N2 gene concentrations of 1.61E+03 and 1.63E+03 gc/L, respectively. PEG samples yielded mean N1 and N2 gene concentrations of 1.61E+05 and 1.50E+05 gc/L, respectively. The overall observed trends of the VIRADEL total N1 and N2 gene concentrations increased steeply from early December 2021 and reached a peak in late December 2021 (Figure 3. 5. a.), which heralded the major wave of COVID-19 cases in late December 2021 and early January 2022. Likewise, VIRADEL N1 and N2 gene concentrations increased in early April 2022, which preceded a resurgence of COVID-19 cases later in mid-April 2022.

Table 3. 1. Total N1 and N2 gene concentrations measured in wastewater samples by VIRADEL, PEG, and filtration methods

Gene	Concentrations	VIRADEL	PEG	Filtration
N1 (gc/L)	Maximum	5.64E+03	7.02E+05	1.12E+05
	Minimum	9.01E+02	3.18E+04	5.12E+02
	Mean	1.61E+03	1.61E+05	3.22E+04
N2 (gc/L)	Maximum	4.95E+03	5.48E+05	7.34E+04
	Minimum	9.01E+02	2.97E+04	3.13E+02
	Mean	1.63E+03	1.50E+05	1.50E+04
N1 (gc/day)	Maximum	5.24E+12	4.07E+14	7.40E+13
	Minimum	5.39E+11	2.36E+13	4.12E+11
	Mean	1.35E+12	1.18E+14	2.52E+13
N2 (gc/day)	Maximum	4.62E+12	3.18E+14	4.77E+13
	Minimum	5.84E+11	2.21E+13	2.93E+11
	Mean	1.37E+12	1.12E+14	1.14E+13
N1 (gc/day/person)	Maximum	1.69E+00	1.31E+02	2.38E+01
	Minimum	1.74E-01	7.58E+00	1.32E-01
	Mean	4.34E-01	3.79E+01	8.11E+00
N2 (gc/day/person)	Maximum	1.49E+00	1.02E+02	1.54E+01
	Minimum	1.88E-01	7.12E+00	9.43E-02
	Mean	4.41E-01	3.59E+01	3.65E+00
N1 (gc/mg TSS)	Maximum	5.82E+01	5.72E+03	1.19E+03
	Minimum	6.85E+00	2.20E+02	2.91E+00
	Mean	1.69E+01	1.53E+03	2.97E+02
N2 (gc/mg TSS)	Maximum	5.21E+01	4.66E+03	6.60E+02
	Minimum	6.71E+00	1.99E+02	3.84E+00
	Mean	1.69E+01	1.44E+03	1.41E+02
N1 (gc/(L(pounds/day)))	Maximum	2.94E-02	4.96E+00	8.83E-01
	Minimum	2.00E-03	6.48E-02	1.77E-03
	Mean	9.83E-03	1.01E+00	1.89E-01
N2 (gc/(L(pounds/day)))	Maximum	2.70E-02	4.01E+00	4.84E-01
	Minimum	1.96E-03	5.90E-02	2.06E-03
	Mean	9.84E-03	9.37E-01	9.27E-02

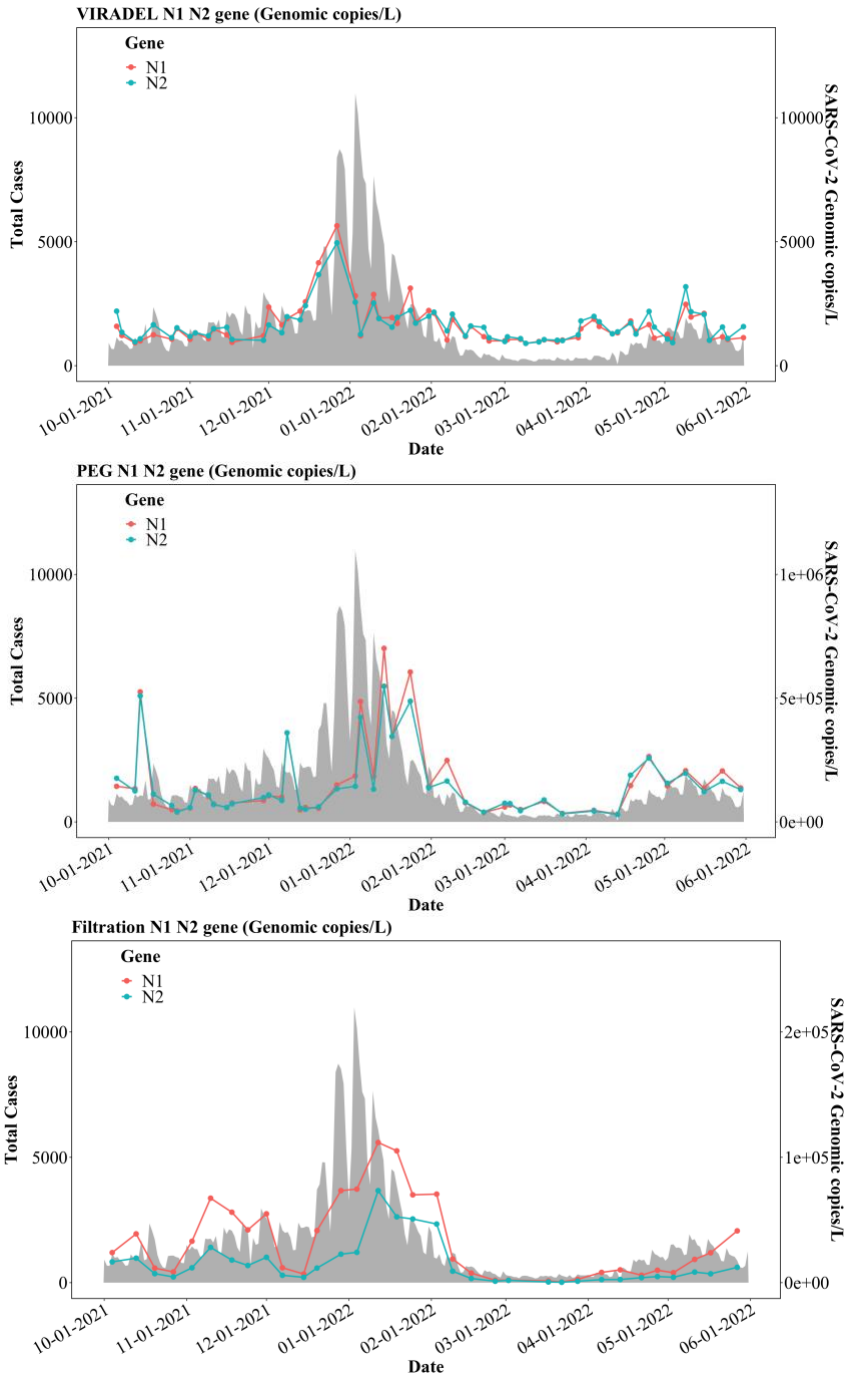


Figure 3. 5. N1 and N2 gene concentrations (gc/L) by three concentration methods: VIRADEL, PEG, Filtration, as well as total COVID-19 cases

Previous reports have demonstrated that the VIRADEL method can serve as a leading indicator of COVID-19 cases (Miyani et al., 2021; Zhao et al., 2022a, 2022b). By contrast, PEG measured N1 and N2 gene concentrations were more variable and increased significantly in January 2022, lagging the major wave of COVID-19 infections (Figure 3. 5. b.). PEG N1 and N2 gene concentrations increased simultaneously with the surge of COVID-19 cases in mid-April 2022, into May 2022. N1 and N2 gene concentrations yielded by the filtration approach increased in early November 2021 and decreased in early December 2021. Thereafter, gene concentrations rapidly increased starting in mid-December 2021, peaking in mid-January 2022, which later significantly decreased to a low level in February 2022 (Figure 3. 5. c.). Notably, the peak in SARS-CoV-2 measured in wastewater by this approach was staggered, lagging the major wave of COVID-19 cases.

3.2 Correlations and similarity analyses among three concentration methods

3.2.1 Correlations of N1 and N2 gene concentrations among three concentration methods

Multiple studies have applied Pearson and Spearman correlations to analyze the relationships between wastewater SARS-CoV-2 gene concentrations and COVID-19 cases (Ai et al., 2021; D'Aoust et al., 2021b; Zhao et al., 2022b), as well as the relationships among gene concentrations of SARS-CoV-2 in wastewater (Lancaster et al., 2022; Vadde et al., 2022). In this study, we tested the Pearson and Spearman correlations among N1 and N2 gene concentrations by VIRADEL, PEG, and filtration with normalized and non-normalized data (Table 3. 2.). A p-value that is less than 0.05 is considered statistically significant. For the non-normalized data (gc/L), the highest correlation was observed between PEG and filtration with N2 gene concentration (Pearson's $r = 0.67$, Spearman's $r = 0.6$). The lowest correlation was found between VIRADEL and PEG for N2 gene concentration (Pearson's $r = 0.12$, Spearman's $r = 0.34$). For non-normalized

data (gc/L), the correlations between PEG and filtration were stronger than the correlations between VIRADEL and filtration, which in turn was stronger than the correlations between VIRADEL and PEG. For normalized data, the highest correlation was found between PEG and filtration for N1 (Pearson's r = 0.73, Spearman's r = 0.66) and N2 (Pearson's r = 0.76, Spearman's r = 0.64) gene concentrations in gc/(L(pounds/day)). Significant correlations (Pearson coefficient > 0.63, Spearman coefficient > 0.6) were observed between PEG and filtration in gc/L, gc/mg TSS and gc/(L(pounds/day)) (Table 3. 2.). VIRADEL has stronger correlation to filtration than to PEG for both normalized and non-normalized data.

Table 3. 2. Correlation coefficients among gene concentrations by VIRADEL, PEG, and filtration methods

Methods (Units)	N1 (Pearson)	N1 (Spearman)	N2 (Pearson)	N2 (Spearman)
V-P (gc/L)	0.17	0.36	0.12	0.34
V-P (gc/day)	0.10	0.17	0.11	0.13
V-P (gc/day/person)	0.10	0.17	0.11	0.13
V-P (gc/mg TSS)	0.29	0.41	0.27	0.46
V-P (gc/(L(pounds/day)))	0.46	0.58	0.43	0.62
V-F (gc/L)	0.41	0.46	0.23	0.40
V-F (gc/day)	0.26	0.13	0.04	0.05
V-F (gc/day/person)	0.26	0.13	0.04	0.05
V-F (gc/mg TSS)	0.49	0.47	0.27	0.39
V-F (gc/(L(pounds/day)))	0.59	0.64	0.41	0.60
P-F (gc/L)	0.63	0.60	0.67	0.60
P-F (gc/day)	0.46	0.51	0.45	0.50
P-F (gc/day/person)	0.46	0.51	0.45	0.50
P-F (gc/mg TSS)	0.67	0.63	0.68	0.60
P-F (gc/(L(pounds/day)))	0.73	0.66	0.76	0.64

Note: V represents VIRADEL, P represents PEG, F represents filtration.

Normalizations using flow rate and flow rate/population reduced the correlations of gene concentrations among VIRADEL, PEG, and filtration compared to the correlations using the non-normalized data (gc/L) (Table 3. 2.). For instance, both Pearson and Spearman correlation coefficients between PEG and filtration were reduced from 0.67 (N2, Pearson, gc/L) and 0.6 (N2, Spearman, gc/L) to 0.45 (N2, Pearson, gc/day) and 0.5 (N2, Spearman, gc/day), respectively

(Table 3. 2.). Conversely, normalizations using TSS and flow rate×TSS enhanced the correlations of gene concentrations among the three methods. For instance, higher correlation coefficients (Pearson's r ranged from 0.73 (N1 gene) to 0.76 (N2 gene), Spearman's r ranged from 0.64 (N2 gene) to 0.66 (N1 gene), all $p < 0.05$) were observed between PEG and filtration gene concentrations after normalization using flow rate×TSS compared to the correlation coefficients for non-normalized data (gc/L) (Pearson's r ranged from 0.63 (N1 gene) to 0.67 (N2 gene), Spearman's $r = 0.6$ (both N1 and N2 gene), all $p < 0.05$).

3.2.2 Dynamic time warping (DTW) of N1 and N2 gene concentrations among three concentration methods

Detecting patterns and comparing similarities of gene concentration time series data are critical for comparing the concentration methods. Dynamic time warping (DTW) identifies the most similar patterns and the optimal warping match between two time series by calculating the minimum DTW distance (Berndt & Clifford, 1994; Giorgino, 2009; Jeong et al., 2011). Shorter DTW distances indicate higher degree of similarity in patterns/shapes between two time series (Guan et al., 2016; Liu et al., 2018). Table 3. 3. presents the DTW minimum distances among the N1 and N2 gene concentrations by VIRADEL, PEG, and filtration methods. Smallest DTW distances were observed between VIRADEL and filtration for both non-normalized and normalized data, which indicated that VIRADEL has a higher degree of similarity with filtration than with PEG. Largest DTW distances were observed between VIRADEL and PEG for both non-normalized and normalized data, indicating that VIRADEL and PEG have the least similarity. This finding was consistent with the sampling and concentration mechanisms since VIRADEL targets free and suspended viral particles in the dissolved phase of wastewater, whereas PEG targets particle-associated viruses, some of which may represent previously shed and accumulated viruses

in the sewer stream (Flood et al., 2021; Zhao et al., 2022b).

Normalization using flow rate decreased the similarity among methods. For instance, the DTW distance between VIRADEL and filtration increased significantly after normalizing using flow rate (gc/day), indicating that the similarity between VIRADEL and filtration was reduced after normalization (Table 3. 3.). Conversely, normalization using TSS and flow rate×TSS strengthened the similarity among methods. For instance, the DTW distances decreased in gc/mg TSS and gc/(L(pounds/day)) comparing to the DTW distance in gc/L among the methods, indicating the improvement of similarity among methods after normalization (Table 3. 3.).

Table 3. 3. Dynamic time warping (DTW) minimum distances among gene concentrations by VIRADEL, PEG, and filtration methods

Methods (Units)	N1	N2
V-P (gc/L)	4.37E+07	4.07E+07
V-P (gc/day)	3.23E+16	3.06E+16
V-P (gc/day/person)	1.04E+04	9.83E+03
V-P (gc/mg TSS)	3.93E+05	3.68E+05
V-P (gc/(L(pounds/day)))	2.42E+02	2.23E+02
V-F (gc/L)	7.51E+06	3.14E+06
V-F (gc/day)	5.87E+15	2.35E+15
V-F (gc/day/person)	1.89E+03	7.56E+02
V-F (gc/mg TSS)	6.74E+04	2.84E+04
V-F (gc/(L(pounds/day)))	4.33E+01	1.92E+01
P-F (gc/L)	2.60E+07	2.85E+07
P-F (gc/day)	1.83E+16	2.27E+16
P-F (gc/day/person)	5.89E+03	7.30E+03
P-F (gc/mg TSS)	2.45E+05	2.94E+05
P-F (gc/(L(pounds/day)))	1.46E+02	1.66E+02

Note: V represents VIRADEL, P represents PEG, F represents filtration.

3.3 Similarity and TLCC analyses between three concentration methods and COVID-19 cases

3.3.1 Dynamic time warping between three concentration methods and COVID-19 cases

Wastewater surveillance data for COVID-19 primarily contain temporal data of viral gene concentrations and clinical cases. DTW analyses were performed between gene concentrations

derived from the three concentration methods (VIRADEL, PEG, and filtration) and the 7-day moving average of total COVID-19 cases for both normalized and non-normalized data. For non-normalized data (gc/L), the smallest DTW distance was found between VIRADEL and total COVID-19 cases (Table 3. 4.). This indicates that VIRADEL (gc/L) has the highest similarity to total COVID-19 cases among the three concentration methods tested. The largest DTW distance was found between PEG (gc/L) and total COVID-19 cases, indicating the PEG method for concentration yields the least similarity to clinical cases. Normalizing gene concentration data using flow (gc/day) demonstrated similar findings. Conversely, normalization using TSS and flow×TSS can significantly increase the similarity between PEG and total COVID-19 cases but concurrently decrease the similarity between VIRADEL and total COVID-19 cases. Specifically, for normalized data (gc/mg TSS, gc/L(pounds/day)), the smallest DTW distance was identified between PEG and total COVID-19 cases, indicating the PEG has the highest similarity to total COVID-19 cases. The largest DTW distance was identified between VIRADEL and COVID-19 cases, indicating that VIRADEL has the lowest similarity to total COVID-19 cases.

Table 3. 4. Dynamic time warping (DTW) minimum distances between gene concentrations by VIRADEL, PEG, as well as filtration methods and total COVID-19 cases

Methods (Units)	N1	N2
V-cases (gc/L)	1.04E+05	1.28E+05
V-cases (gc/day)	4.72E+14	4.86E+14
V-cases (gc/day/person)	4.61E+05	4.61E+05
V-cases (gc/mg TSS)	4.55E+05	4.54E+05
V-cases (gc/(L(pounds/day)))	4.61E+05	4.61E+05
P-cases (gc/L)	4.39E+07	4.08E+07
P-cases (gc/day)	3.30E+16	3.14E+16
P-cases (gc/day/person)	4.43E+05	4.42E+05
P-cases (gc/mg TSS)	9.87E+04	1.14E+05
P-cases (gc/(L(pounds/day)))	4.61E+05	4.61E+05
F-cases (gc/L)	7.35E+06	2.95E+06
F-cases (gc/day)	6.20E+15	2.82E+15
F-cases (gc/day/person)	4.57E+05	4.59E+05
F-cases (gc/mg TSS)	2.87E+05	3.92E+05
F-cases (gc/(L(pounds/day)))	4.61E+05	4.61E+05

Note: V represents VIRADEL, P represents PEG, F represents filtration, cases represents total 7-day-moving-average clinical cases.

3.3.2 TLCC and peak synchrony between three concentration methods and COVID-19 cases

The relative timing of the wastewater gene concentrations (gc/L, gc/day, gc/day/person, gc/mg TSS, and gc/(L(pounds/day))) of VIRADEL, PEG and filtration were compared to the total COVID-19 cases using TLCC and peak synchrony. To evaluate if wastewater viral concentrations of the three methods lead or lag COVID-19 cases, the total COVID-19 case data were shifted by a period of −20 (lagging) to +20 days (leading) and the Pearson's correlation coefficients were calculated between cases and wastewater viral gene concentration for each shift. The leading or lagging metric varied for each method, which was determined by comparing the strongest Pearson's correlation coefficient.

For the VIRADEL method, both N1 and N2 gene concentrations (gc/L) were strongly correlated with COVID-19 cases, covering shifting windows between −20 and +20 days (Figure 3. 6. a.). The highest correlation coefficient was observed when offset is +12 days (Figure 3. 6. a.), indicating that SARS-CoV-2 gene concentrations (gc/L) in wastewater by the VIRADEL

method lead COVID-19 cases by 12 days, which concurred with previous findings of a 35-day lead time of gene concentrations preceding total COVID-19 cases prior to the Omicron surge (Zhao et al., 2022b). For both non-normalized and normalized data, VIRADEL always led COVID-19 cases with a variety of lead times (Table 3. 5.).

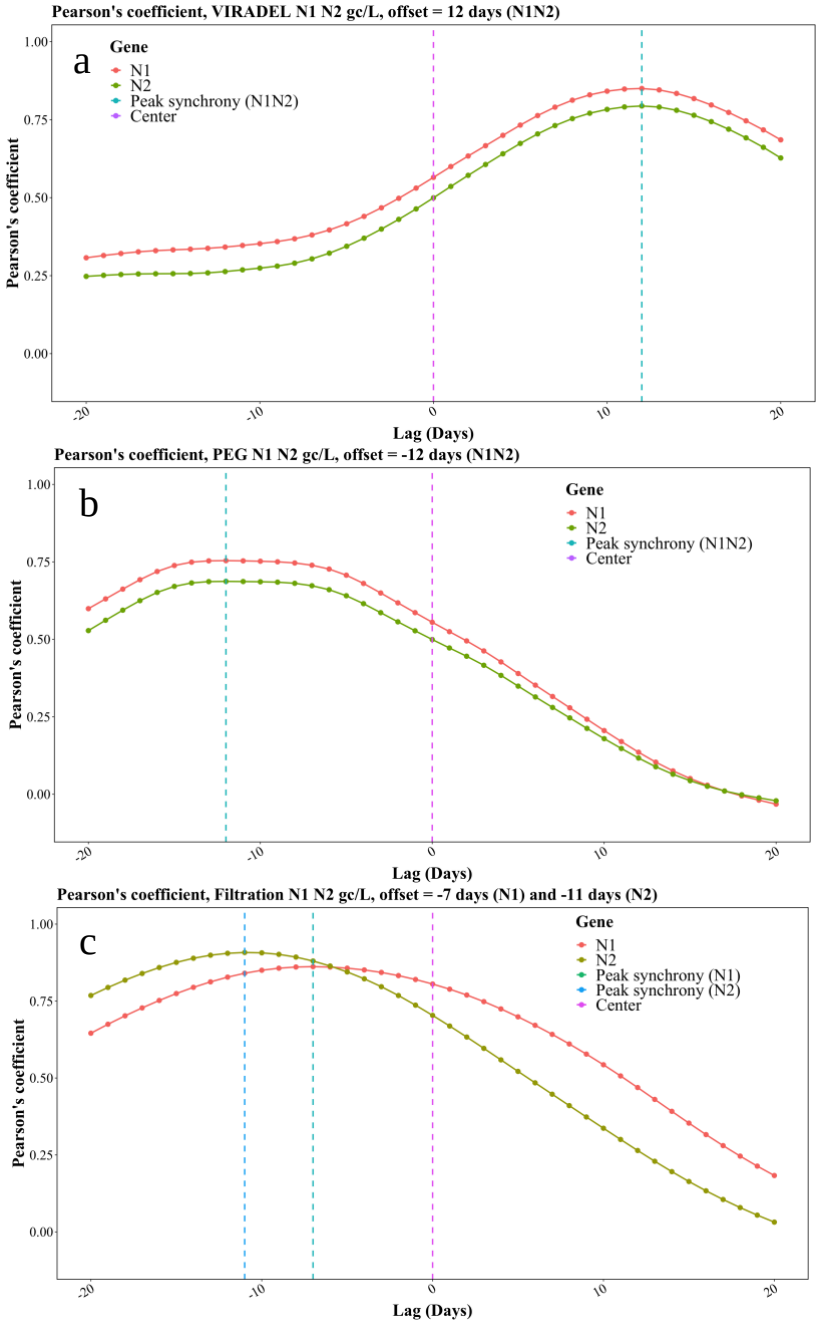


Figure 3. 6. Pearson correlation coefficients for TLCC and peak synchrony between wastewater viral concentrations and COVID-19 cases with offsets between −20 (lagging) and +20 (leading) days for the three methods, including VIRADEL (a), PEG (b), and Filtration (c)

Table 3. 5. Lead/lag time between wastewater viral concentrations by VIRADEL, PEG, as well as filtration methods and total COVID-19 cases

Units	Method (Gene)					
	V (N1)	V (N2)	P (N1)	P (N2)	F (N1)	F (N2)
gc/L*	+12	+12	-12	-12	-7	-11
gc/day	+13	+13	-6	-6	-2	-10
gc/day/person	+13	+13	-6	-6	-2	-10
gc/mg TSS	+11	+11	-9	-9	-7	-12
gc/(L(pounds/day))	+9	+9	-14	-14	-11	-13

Note: V represents VIRADEL, P represents PEG, F represents filtration, * was demonstrated in Figure 3. 6., + indicates lead time, – indicates lag time.

For the PEG method (gc/L), the strongest correlation coefficients were observed with an offset of -12 days, indicating that SARS-CoV-2 gene concentrations by the PEG method lagged reported COVID-19 cases by 12 days during the study period (Figure 3. 6. b.).

For the filtration method (gc/L), the highest correlation coefficient was observed with an offset of -7 days for the N1 gene and -11 days for the N2 gene, indicating that SARS-CoV-2 gene concentrations in wastewater lagged reported COVID-19 cases for 7 days (N1) and 11 days (N2), respectively (Figure 3. 6. c.). Likewise, similar observations were found for normalized data where the filtration method yielded data that lagged clinical cases (Table 3. 5.). Table 3. 5. summarized the lead/lag time between VIRADEL, PEG, and filtration methods and total COVID-19 cases. The length of the leading or lagging time differed with dissimilar normalizations. However, the leading or lagging pattern of each method did not change, where VIRADEL measurements were always leading COVID-19 cases, whereas PEG and filtration measurements routinely lagged COVID-19 cases.

4. Discussion

There is an ongoing effort to optimize methods to recover and concentrate SARS-CoV-2

from wastewater in support of actionable public health outcomes (Ahmed et al., 2020c; Kitajima et al., 2020). In this study, three concentration methods were evaluated for concentrating SARS-CoV-2 from wastewater, spanning the transition between Delta and Omicron variants circulating in the Detroit, MI metropolitan area. The three methods share common characteristics, especially downstream where they follow similar procedures of nucleic acid extraction and quantification such as RT-ddPCR or RT-qPCR. Likewise, their recovery efficiencies are reported as comparable (Li et al., 2022; Zhao et al., 2022b; Flood et al., 2021; Pecson et al., 2021; Torii et al., 2021).

4.1 VIRADEL: Opportunities and obstacles

Several studies have previously adopted VIRADEL as the concentration method for SARS-CoV-2 in wastewater (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022a, 2022b). An attribute of the VIRADEL method is the ability to process large volumes (10 – 50 L) of wastewater, thus facilitating capture of free and suspended viral particles that are arguably most representative of viruses shed by recently infected individuals (Lu et al., 2020a; Zhao et al., 2022b). This establishes VIRADEL as a concentration method capable to provide early warning that leads case reporting (Miyani et al., 2021; Zhao et al., 2022b), which was also verified by TLCC analyses in this study (Table 3. 5.). Limiting widescale adoption of VIRADEL is labor-intensive preparation of sampling units which require extensive washing and disinfection prior to use. VIRADEL (Bivins et al., 2022) also requires access to large volumes of wastewater which may not be available to all researchers. Further, the required large volumes may necessitate use of grab samples which typically yield higher variability than composite samples which is the sampling method of choice for many wastewater surveillance efforts (Bivins et al., 2022). VIRADEL requires trained personnel for comparatively laborious work with limited samples (n=15) processed over a relatively long time (4-6 h). VIRADEL also requires multiple large

centrifuges as well as expensive and at times, supply chain-limited consumables. Therefore, VIRADEL may not be an ideal choice for routine wastewater surveillance projects in common microbiology laboratories. However, it was clear from the comparative analyses conducted that VIRADEL has clear potential to be implemented as a tool to provide early warning to inform public health actions (Miyani et al., 2021; Zhao et al., 2022a, 2022b).

4.2 PEG: Opportunities and obstacles

Apart from requiring access to a centrifuge, the consumables required are widely available and relatively inexpensive, lending itself as one of the most broadly applied concentration methods for routine wastewater surveillance (Ahmed et al., 2020c; Flood et al., 2021; Hata et al., 2021; Sapula et al., 2021; Torii et al., 2021; Zhao et al., 2022b). On the other hand, PEG is restricted to processing smaller volumes of wastewater (usually 0.05 to 2 L) and only a portion of the sample pellet is used to recover and extract RNA, which can be affected by the variation of samples and representation of all viruses in wastewater (Ahmed et al., 2020c; Flood et al., 2021; Lu et al., 2020a; Zhao et al., 2022b).

Unlike VIRADEL, PEG targets particle-associated viruses consistent with reports that identify solids as the phase yielding highest SARS-CoV-2 concentrations in wastewater (Torii et al., 2021). While a fraction of these particles will represent recently deposited SARS-CoV-2, the majority may represent previously shed and accumulated viruses in the sewer stream and later resuspended during flow fluctuations, thus providing a mechanism for the method to yield data lagging clinical COVID-19 cases. Though the exact mechanism of PEG is not well understood, several studies proposed that it captures viruses that are sorbed to larger precipitates and solids, consistent with a high quantity of TSS in wastewater (Flood et al., 2021; Zhao et al., 2022b). In this study, through the DTW analyses, PEG yielded data were normalized using TSS and

flow×TSS, which increased the degree of similarity between PEG and total COVID-19 cases (Table 3. 4.). This demonstrated that PEG yielded data were largely affected by the presence of TSS. VIRADEL, instead, captured free and suspended viruses in the supernatant wastewater. Thus, normalizing the VIRADEL data using TSS and flow×TSS decreased the similarity between VIRADEL and cases (Table 3. 4.).

Through the TLCC analyses, this study also demonstrated that PEG gene concentrations lagged COVID-19 cases (Table 3. 5.), which embraced the aforementioned sampling mechanism of PEG (Flood et al., 2021). PEG method did not provide an early warning (leading window) for COVID-19 cases which was concurred with our previous findings, whereas VIRADEL provided early warnings ahead of clinical cases while PEG lagged clinical cases for the Detroit area (Zhao et al., 2022b).

However, several studies using PEG provided early warnings of impending COVID-19 cases (D'Aoust et al., 2021a). Notably, in these studies, PEG was applied to different types of samples such as primary sewage sludge, which is a different sample matrix from untreated wastewater samples, thus needing more investigation on the impact of sample types on early warnings (D'Aoust et al., 2021a). Kumar et al., identified early warnings using PEG in the early stage of the pandemic in August 2020 in India (Kumar et al., 2021). PEG and other concentration methods (such as ultrafiltration (Hasan et al., 2021; Medema et al., 2020) and adsorption-precipitation (Randazzo et al., 2020)) identified early warnings in the early stage of the pandemic when testing capacities were largely limited, and societal responses to the pandemic and clinical data reporting were significantly delayed (Bibby et al., 2021; Zhao et al., 2022b). In addition, earlier prevalent COVID-19 variants including Alpha, Beta and Gamma were reported with longer incubation times than Delta and Omicron variants, leading to prolonged early warning potentials

of wastewater surveillance in the early stages of the pandemic (Wu et al., 2022).

Though PEG was reported to provide early warnings, it may have a shorter early warning window than VIRADEL due to the fundamental disparity of their targets, that being newly contributed free and suspended viral particles versus particle-attached virus, some of which may be considered previously shed and accumulated and subsequently resuspended from sediment (Zhao et al., 2022b). In the current study, PEG was shown to lag clinical cases while VIRADEL was leading clinical cases for both normalized and non-normalized data (Table 3. 5.). Overall, the early warning potential of PEG needs further investigations in terms of sample types, sampling mechanisms and locations, stage of the epidemic, amongst other factors.

4.3 Filtration: Opportunities and obstacles

Filtration is commonly applied to recover and concentrate viral RNA in water samples. It achieves generally good recovery efficiencies, is relatively inexpensive using commonly available lab equipment and simple protocols and provides consistent performance and inclusive measurement since it captures viruses from both solids and liquid fractions by nature of forcing free viral particles across trapped solids (Ahmed et al., 2020c; Gonçalves et al., 2021). However, filtration has several drawbacks. First, the number of available filtration units restricts the number of samples that can be processed concurrently (Ahmed et al., 2020c). Meanwhile, clogging of filters can occur due to high variations of turbidity in wastewater. While this can be offset in part by use of a caulking gun to exert more pressure on the sample being filtered, in reality, volumes are limited to ~0.1 L. Additionally, filtration measurements lagged the COVID-19 clinical cases in the current study, thus, its ability to provide early warnings for impending cases is called into question. The recovery efficiencies also differ with different filters (Ahmed et al., 2020c).

4.4 Future directions

The mechanism and implications of primarily collecting viruses attached to solids that may have settled and resuspended before sampling, such as by the PEG, needs further investigations. Notably, multiple studies have reported that the integrity of SARS-CoV-2 RNA was higher when sorbed to suspended solids, organic matter, and large bio-solids which provide protection from predation and inactivation. This can be explained by the hydrophobicity of SARS-CoV-2 viral particles, leading to their adherence to wastewater solids and longer persistence compared to free viruses in the supernatant wastewater (Abu Ali et al., 2021; Gundy et al., 2009; Panchal et al., 2021).

The implications of seasonal variations in SARS-CoV-2 persistence in wastewater needs further investigations. Seasonal variations of wastewater temperature and pH are reported to affect the persistence of viral RNA in wastewater (Hart & Halden, 2020). However, SARS-CoV-2 RNA was shown to be highly stable at 4 °C aqueous environment or in a wide pH range at room temperature (Ahmed et al., 2020b; Chin et al., 2020). Multiple studies reported the detectability and persistence of SARS-CoV-2 RNA in untreated wastewater solids samples. For instance, researchers found that SARS-CoV-2 RNA was consistently detected for 29 days and 64 days at 4 °C and -20 °C, respectively in wastewater solids pelleted by centrifugation, (Hokajärvi et al., 2021). Another study indicated that only minimal reduction of SARS-CoV-2 RNA was observed for wastewater solids samples after 100 days (Simpson et al., 2021). Additionally, researchers established models to indicate that viral RNA can be detected in wastewater even with long sewer travel time (100 h), especially with lower average wastewater temperature in northern cities such as Detroit (Hart & Halden, 2020). Recent study also indicated that biofilms could mediate the fate of SARS-CoV-2 in wastewater, especially leading the viral RNA to prolonged presence (Li et al.,

2023).

The effect of varying sampling volumes needs further investigation. Some studies indicated that a larger sampling volume can increase the sensitivity of the sampling method, suggesting that it will detect lower levels of viral RNA in wastewater samples (McMinn et al., 2021). Similarly, researchers suggested that processing of larger sample volumes may help to lower the method detection limits (Hart & Halden, 2020). But at the same time, keeping the required samples sizes low can lead to inexpensive shipping between sampling location and the analytical laboratory as well as limited spare for storage (Hart & Halden, 2020). Other researchers indicated that detection sensitivity can be improved by increasing the sample volume from 100 ml to 500 ml wastewater for testing SARS-CoV-2 (Ahmed et al., 2020a).

However, other researchers presented that large-volume sampling did not significantly enhance the sensitivity of methods (Zheng et al., 2022). For instance, Zheng et al., found that wastewater concentration methods (they used ultracentrifugation) using less volume of wastewater was preferable than larger volume of wastewater in terms of sensitivity for testing SARS-CoV-2. The study revealed that when using the same concentration methods, no significant difference was observed in the viral RNA concentrations between experiments conducted with a larger volume of wastewater and those conducted with a smaller volume (Zheng et al., 2022). Some studies indicated that a larger sampling volume may also dilute the wastewater sample, which can lead to a lower viral RNA concentration (Bertels et al., 2022).

Overall, the sampling volume for wastewater surveillance of SARS-CoV-2 using different concentration methods will depend on several factors, including the sensitivity of the method, the concentration of viral RNA in the wastewater, and the size of the population being monitored. It is critical to consider and address these factors when analyzing wastewater surveillance data and

more in-depth research on how the sampling volume affect statistical results are needed.

The time of sampling may potentially affect results in sewershed sampling. The effect of sampling time in large interceptors, like the ones sampled in this study, is less significant, since the interceptor wastewater is mixed at the pumping stations. A few studies have reported gene concentration varying on an hourly basis (Bivins et al., 2021; Li et al., 2021) although the temporal variability of SARS-CoV-2 concentrations in wastewater remains ambiguous (Bivins et al., 2021; Ahmed et al., 2020c). It has been suggested that composite samples may circumvent the within-day variation of viral concentrations (Bivins et al., 2021). Whereas both the PEG and filtration methods used composite samples, the large volume required for VIRADEL necessitated separate sampling which was conducted over a period of several hours to help reduce temporal variability. Further, considering the vast sewersheds and population of nearly 4 million people that GLWA's three interceptors serve, the concentrations of SARS-CoV-2 in wastewater may be highly diluted and within-day variations can be negligible. Future studies are called to examine within-day variation of SARS-CoV-2.

Admittedly, there are caveats to the current study that should be considered and discussed. The study period was limited to the transition between Delta and Omicron VOCs that occurred between fall 2021 and winter 2022. With each successive resurgence of COVID-19, differences are reported related to disease trajectory including incubation time, shedding dynamics and disease severity (Baker et al., 2022; Boucau et al., 2022). For instance, the incubation time was shorter during the Omicron surge compared to the previous variants, inevitably reducing the early warning potentials of wastewater surveillance in the later stage of the pandemic (Baker et al., 2022; Zhao et al., 2022b). Further, the changing viral shedding dynamics, viral decay kinetics, and shedding duration of the Omicron variant are not well understood and many uncertainties remain (Boucau

et al., 2022; Kandel et al., 2022). As such, the lead and lag times reported here cannot be extrapolated to past or future SARS-CoV-2 variants. In addition, sampling frequency was limited to weekly samples and thus less informative for establishing time series or less likely to depict accurately the actual fluctuations of wastewater viral concentrations (cdc.gov). Feng et al., (2021) proposed a minimum of two samples collected weekly to establish the time series data of wastewater viral concentrations for continuous trend analysis (Feng et al., 2021). Some researchers have even suggested daily or very frequent sampling, if the laboratory is capable of handling increased numbers of samples, considering rapid resurgence of COVID-19 cases (Zhu et al., 2021). Indeed, the filtration method has been used to analyze samples 5 days weekly since the emergence of the Omicron VOC as part of Ontario's Wastewater Surveillance Initiative in the Windsor-Essex region located across the international border with Detroit (Q. Geng, R. Corchis-Scott, R.M. McKay, unpublished). While SARS-CoV-2 signal intensity derived from this approach does not provide a clear early warning of clinical cases, preliminary analysis supports its use as a leading indicator of COVID-19-related hospitalizations in the region (Q. Geng, R. Corchis-Scott, R.M. McKay, unpublished). This is important considering that clinical testing capacity across North America was overwhelmed by infections attributed to Omicron and is thus no longer a reliable indicator of disease prevalence (Lawal et al., 2022).

5. Conclusions

This study is among the first to implement, evaluate, and compare commonly applied wastewater virus concentration methodologies to recover and concentrate SARS-CoV-2 from wastewater amid the transition between Delta and Omicron VOCs. Analytical approaches, including Pearson and Spearman correlations, Dynamic Time Warping (DTW), and Time Lagged Cross Correlation (TLCC) and peak synchrony, were performed to analyze the relations among

three methods as well as the relations between methods and COVID-19 cases. To our knowledge, this is the only study to implement Dynamic Time Warping to compare wastewater surveillance time series data and successfully identify the similarities/dissimilarities among the methods and between methods and clinical data. The analytical approach used can be applied to different sample processing and concentration methods under various pandemic scenarios to evaluate method efficacy for different public health practices.

6. Acknowledgements

The authors acknowledge the support from operators, the laboratory team and managers of the Wastewater Resource Recovery Facility at the Great Lakes Water Authority. They also thank CDM Smith for supporting this work.

REFERENCES

Abu Ali, H., Yaniv, K., Bar-Zeev, E., Chaudhury, S., Shagan, M., Lakkakula, S., Ronen, Z., Kushmaro, A., & Nir, O. (2021). Tracking SARS-CoV-2 RNA through the Wastewater Treatment Process. *ACS ES&T Water*, *1*(5), 1161–1167. https://doi.org/10.1021/acsestwater.0c00216

Ahmed, W., Bertsch, P. M., Angel, N., Bibby, K., Bivins, A., Dierens, L., Edson, J., Ehret, J., Gyawali, P., Hamilton, K. A., Hosegood, I., Hugenholtz, P., Jiang, G., Kitajima, M., Sichani, H. T., Shi, J., Shimko, K. M., Simpson, S. L., Smith, W. J. M., … Mueller, J. F. (2020a). Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: a surveillance tool for assessing the presence of COVID-19 infected travellers. *Journal of Travel Medicine*, *27*(5). https://doi.org/10.1093/jtm/taaa116

Ahmed, W., Bertsch, P. M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., Gyawali, P., Korajkic, A., Riddell, S., Sherchan, S. P., Simpson, S. L., Sirikanchana, K., Symonds, E. M., Verhagen, R., Vasan, S. S., Kitajima, M., & Bivins, A. (2020b). Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environmental Research*, *191*, 110092. https://doi.org/10.1016/j.envres.2020.110092

Ahmed, W., Bertsch, P. M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B. R., Mueller, J. F., Simpson, S. L., Smith, W. J. M., Symonds, E. M., Thomas, K. v., Verhagen, R., & Kitajima, M. (2020c). Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total Environment*, *739*, 139960. https://doi.org/10.1016/j.scitotenv.2020.139960

Ahmed, W., Bivins, A., Simpson, S. L., Bertsch, P. M., Ehret, J., Hosegood, I., Metcalfe, S. S., Smith, W. J. M., Thomas, K. V., Tynan, J., & Mueller, J. F. (2022). Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. *Environment International*, *158*, 106938. https://doi.org/10.1016/j.envint.2021.106938

Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ru, P., Pan, X., Bohrerova, Z., & Lee, J. (2021). Wastewater SARS-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States. *Science of The Total Environment*, *801*, 149757. https://doi.org/10.1016/j.scitotenv.2021.149757

Baker, M. A., Rhee, C., Tucker, R., Badwaik, A., Coughlin, C., Holtzman, M. A., Hsieh, C., Maguire, A., Mermel Blaeser, E., Seetharaman, S., Solem, O., Vaidya, V., & Klompas, M. (2022). Rapid control of hospital-based SARS-CoV-2 Omicron clusters through daily testing and universal use of N95 respirators. *Clinical Infectious Diseases*, ciac113. https://doi.org/10.1093/cid/ciac113

Barril, P. A., Pianciola, L. A., Mazzeo, M., Ousset, M. J., Jaureguiberry, M. V., Alessandrello, M., Sánchez, G., & Oteiza, J. M. (2021). Evaluation of viral concentration methods for SARS-

CoV-2 recovery from wastewaters. *Science of The Total Environment*, *756*, 144105. https://doi.org/10.1016/j.scitotenv.2020.144105

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD Workshop*, *10*(16), 359–370.

Bertels, X., Demeyer, P., van den Bogaert, S., Boogaerts, T., van Nuijs, A. L. N., Delputte, P., & Lahousse, L. (2022). Factors influencing SARS-CoV-2 RNA concentrations in wastewater up to the sampling stage: A systematic review. *Science of The Total Environment*, *820*, 153290. https://doi.org/10.1016/j.scitotenv.2022.153290

Bibby, K., Bivins, A., Wu, Z., & North, D. (2021). Making waves: Plausible lead time for wastewater based epidemiology as an early warning system for COVID-19. *Water Research*, *202*(July), 117438. https://doi.org/10.1016/j.watres.2021.117438

Bivins, A., & Bibby, K. (2021). Wastewater Surveillance during Mass COVID-19 Vaccination on a College Campus. *Environmental Science & Technology Letters*, *8*(9), 792–798. https://doi.org/10.1021/acs.estlett.1c00519

Bivins, A., Kaya, D., Ahmed, W., Brown, J., Butler, C., Greaves, J., Leal, R., Maas, K., Rao, G., Sherchan, S., Sills, D., Sinclair, R., Wheeler, R. T., & Mansfeldt, C. (2022). Passive sampling to scale wastewater surveillance of infectious disease: Lessons learned from COVID-19. *Science of The Total Environment*, *835*, 155347. https://doi.org/10.1016/j.scitotenv.2022.155347

Black, L. E., Brion, G. M., & Freitas, S. J. (2007). Multivariate Logistic Regression for Predicting Total Culturable Virus Presence at the Intake of a Potable-Water Treatment Plant: Novel Application of the Atypical Coliform/Total Coliform Ratio. *Applied and Environmental Microbiology*, *73*(12), 3965–3974. https://doi.org/10.1128/AEM.02780-06

Boucau, J., Marino, C., Regan, J., Uddin, R., Choudhary, M. C., Flynn, J. P., Chen, G., Stuckwisch, A. M., Mathews, J., Liew, M. Y., Singh, A., Lipiner, T., Kittilson, A., Melberg, M., Li, Y., Gilbert, R. F., Reynolds, Z., Iyer, S. L., Chamberlin, G. C., … Barczak, A. K. (2022). Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection. *New England Journal of Medicine*, *387*(3), 275–277. https://doi.org/10.1056/NEJMc2202092

Chik, A. H. S., Glier, M. B., Servos, M., Mangat, C. S., Pang, X.-L., Qiu, Y., D'Aoust, P. M., Burnet, J.-B., Delatolla, R., Dorner, S., Geng, Q., Giesy, J. P., McKay, R. M., Mulvey, M. R., Prystajecky, N., Srikanthan, N., Xie, Y., Conant, B., & Hrudey, S. E. (2021). Comparison of approaches to quantify SARS-CoV-2 in wastewater using RT-qPCR: Results and implications from a collaborative inter-laboratory study in Canada. *Journal of Environmental Sciences*, *107*,

218–229. https://doi.org/10.1016/j.jes.2021.01.029

Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H.-L., Chan, M. C. W., Peiris, M., & Poon, L. L. M. (2020). Stability of SARS-CoV-2 in different environmental conditions. *The Lancet Microbe*, *1*(1), e10. https://doi.org/10.1016/S2666-5247(20)30003-3

Corchis-Scott, R., Geng, Q., Seth, R., Ray, R., Beg, M., Biswas, N., Charron, L., Drouillard, K. D., D'Souza, R., Heath, D. D., Houser, C., Lawal, F., McGinlay, J., Menard, S. L., Porter, L. A., Rawlings, D., Scholl, M. L., Siu, K. W. M., Tong, Y., … McKay, R. M. L. (2021). Averting an Outbreak of SARS-CoV-2 in a University Residence Hall through Wastewater Surveillance. *Microbiology Spectrum*, *9*(2). https://doi.org/10.1128/Spectrum.00792-21

D'Aoust, P. M., Graber, T. E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Servos, M. R., Srikanthan, N., MacKenzie, M., Figeys, D., Manuel, D., Jüni, P., MacKenzie, A. E., & Delatolla, R. (2021a). Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Science of the Total Environment*, *770*. https://doi.org/10.1016/j.scitotenv.2021.145319

D'Aoust, P. M., Mercier, E., Montpetit, D., Jia, J.-J., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Langlois, M.-A., Servos, M. R., MacKenzie, M., Figeys, D., MacKenzie, A. E., Graber, T. E., & Delatolla, R. (2021b). Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Research*, *188*, 116560. https://doi.org/https://doi.org/10.1016/j.watres.2020.116560

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

Feng, S., Roguet, A., McClary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G., & McLellan, S. L. (2021). Evaluation of Sampling, Analysis, and Normalization Methods for SARS-CoV-2 Concentrations in Wastewater to Assess COVID-19 Burdens in Wisconsin Communities. *ACS ES&T Water*, *1*(8), 1955–1965. https://doi.org/10.1021/acsestwater.1c00160

Flood, M. T., D'Souza, N., Rose, J. B., & Aw, T. G. (2021). Methods Evaluation for Rapid Concentration and Quantification of SARS-CoV-2 in Raw Wastewater Using Droplet Digital and Quantitative RT-PCR. *Food and Environmental Virology*, *13*(3), 303–315. https://doi.org/10.1007/s12560-021-09488-8

Francy, D. S., Stelzer, E. A., Brady, A. M. G., Huitger, C., Bushon, R. N., Ip, H. S., Ware, M. W., Villegas, E. N., Gallardo, V., & Lindquist, H. D. A. (2013). Comparison of Filters for Concentrating Microbial Indicators and Pathogens in Lake Water Samples. *Applied and Environmental Microbiology*, *79*(4), 1342–1352. https://doi.org/10.1128/AEM.03117-12

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, *31*(7), 1–24. https://doi.org/10.18637/jss.v031.i07

Gonçalves, J., Koritnik, T., Mioč, V., Trkov, M., Bolješič, M., Berginc, N., Prosenc, K., Kotar, T., & Paragi, M. (2021). Detection of SARS-CoV-2 RNA in hospital wastewater from a low COVID-19 disease prevalence area. *Science of The Total Environment, 755*, 143226. https://doi.org/10.1016/j.scitotenv.2020.143226

Guan, X., Huang, C., Liu, G., Meng, X., & Liu, Q. (2016). Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance. *Remote Sensing*, *8*(1), 19. https://doi.org/10.3390/rs8010019

Gundy, P. M., Gerba, C. P., & Pepper, I. L. (2009). Survival of Coronaviruses in Water and Wastewater. *Food and Environmental Virology*, *1*(1), 10. https://doi.org/10.1007/s12560-008-9001-6

Haramoto, E., Malla, B., Thakali, O., & Kitajima, M. (2020). First environmental surveillance for the presence of SARS-CoV-2 RNA in wastewater and river water in Japan. *Science of The Total Environment*, *737*, 140405. https://doi.org/10.1016/j.scitotenv.2020.140405

Hart, O. E., & Halden, R. U. (2020). Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of The Total Environment*, *730*, 138875. https://doi.org/10.1016/j.scitotenv.2020.138875

Hasan, S. W., Ibrahim, Y., Daou, M., Kannout, H., Jan, N., Lopes, A., Alsafar, H., & Yousef, A. F. (2021). Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: Surveillance of COVID-19 epidemic in the United Arab Emirates. *Science of The Total Environment*, *764*, 142929. https://doi.org/10.1016/j.scitotenv.2020.142929

Hata, A., Hara-Yamamura, H., Meuchi, Y., Imai, S., & Honda, R. (2021). Detection of SARS-CoV-2 in wastewater in Japan during a COVID-19 outbreak. *Science of The Total Environment*, *758*, 143578. https://doi.org/10.1016/j.scitotenv.2020.143578

He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., … Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, *26*(5), 672–675. https://doi.org/10.1038/s41591-020-0869-5

Hill, V. R., Polaczyk, A. L., Kahler, A. M., Cromeans, T. L., Hahn, D., & Amburgey, J. E. (2009). Comparison of Hollow-Fiber Ultrafiltration to the USEPA VIRADEL Technique and USEPA Method 1623. *Journal of Environmental Quality*, *38*(2), 822–825. https://doi.org/https://doi.org/10.2134/jeq2008.0152

Hokajärvi, A.-M., Rytkönen, A., Tiwari, A., Kauppinen, A., Oikarinen, S., Lehto, K.-M., Kankaanpää, A., Gunnar, T., Al-Hello, H., Blomqvist, S., Miettinen, I. T., Savolainen-Kopra, C., & Pitkänen, T. (2021). The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland. *Science of The Total Environment*, *770*, 145274. https://doi.org/10.1016/j.scitotenv.2021.145274

Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., & Stadler, L. B. (2023). Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves. *Science of The Total Environment*, *855*, 158967. https://doi.org/10.1016/j.scitotenv.2022.158967

Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, *39*, 235–244. https://doi.org/10.1016/j.engappai.2014.12.015

Jakubowski, W., Hill, W. F., & Clarke, N. A. (1975). Comparative Study of Four Microporous Filters for Concentrating Viruses from Drinking Water. *Applied Microbiology*, *30*(1), 58–65. https://doi.org/10.1128/am.30.1.58-65.1975

Jeong, Y.-S., Jeong, M. K., & Omitaomu, O. A. (2011). Weighted dynamic time warping for time series classification. *Pattern Recognition*, *44*(9), 2231–2240. https://doi.org/10.1016/j.patcog.2010.09.022

Jones, D. L., Baluja, M. Q., Graham, D. W., Corbishley, A., McDonald, J. E., Malham, S. K., Hillary, L. S., Connor, T. R., Gaze, W. H., Moura, I. B., Wilcox, M. H., & Farkas, K. (2020). Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. *Science of The Total Environment*, *749*, 141364. https://doi.org/10.1016/j.scitotenv.2020.141364

Kandel, C., Lee, Y., Taylor, M., Llanes, A., McCready, J., Crowl, G., Powis, J., Li, A. X., Shigayeva, A., Yip, L., Katz, K., Kozak, R., Mubareka, S., & McGeer, A. (2022). Viral dynamics of the SARS-CoV-2 Omicron Variant among household contacts with 2 or 3 COVID-19 vaccine doses. *Journal of Infection*, *85*(6), 666–670. https://doi.org/10.1016/j.jinf.2022.10.027

Kantor, R. S., Nelson, K. L., Greenwald, H. D., & Kennedy, L. C. (2021). Challenges in Measuring the Recovery of SARS-CoV-2 from Wastewater. *Environmental Science & Technology*, *55*(6), 3514–3519. https://doi.org/10.1021/acs.est.0c08210

Kaya, D., Niemeier, D., Ahmed, W., & Kjellerup, B. v. (2022). Evaluation of multiple analytical methods for SARS-CoV-2 surveillance in wastewater samples. *Science of The Total Environment*, *808*, 152033. https://doi.org/10.1016/j.scitotenv.2021.152033

Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, *7*(3), 358–386. https://doi.org/10.1007/s10115-004-0154-9

Kitajima, M., Ahmed, W., Bibby, K., Carducci, A., Gerba, C. P., Hamilton, K. A., Haramoto, E., & Rose, J. B. (2020). SARS-CoV-2 in wastewater: State of the knowledge and research needs. *Science of The Total Environment*, 139076. https://doi.org/10.1016/j.scitotenv.2020.139076

Kumar, M., Joshi, M., Patel, A. K., & Joshi, C. G. (2021). Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environmental Research*, *196*(December 2020), 110946.

https://doi.org/10.1016/j.envres.2021.110946

Lancaster, E., Byrd, K., Ai, Y., & Lee, J. (2022). Socioeconomic status correlations with confirmed COVID-19 cases and SARS-CoV-2 wastewater concentrations in small-medium sized communities. *Environmental Research*, *215*, 114290. https://doi.org/10.1016/j.envres.2022.114290

Lawal, O. U., Zhang, L., Parreira, V. R., Brown, R. S., Chettleburgh, C., Dannah, N., Delatolla, R., Gilbride, K. A., Graber, T. E., Islam, G., Knockleby, J., Ma, S., McDougall, H., McKay, R. M., Mloszewska, A., Oswald, C., Servos, M., Swinwood-Sky, M., Ybazeta, G., … Goodridge, L. (2022). Metagenomics of Wastewater Influent from Wastewater Treatment Facilities across Ontario in the Era of Emerging SARS-CoV-2 Variants of Concern. *Microbiology Resource Announcements*, *11*(7). https://doi.org/10.1128/mra.00362-22

Lee, H., Kim, M., Paik, S.-Y., Lee, C. H., Jheong, W.-H., Kim, J., & Ko, G. (2011). Evaluation of electropositive filtration for recovering norovirus in water. *Journal of Water and Health*, *9*(1), 27–36. https://doi.org/10.2166/wh.2010.190

Lepot, M., Aubin, J.-B., & Clemens, F. (2017). Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water*, *9*(10), 796. https://doi.org/10.3390/w9100796

Li, H., Futch, S. H., & Syvertsen, J. P. (2007). Cross-correlation patterns of air and soil temperatures, rainfall and Diaprepes abbreviatus root weevil in citrus. *Pest Management Science*, *63*(11), 1116–1123. https://doi.org/https://doi.org/10.1002/ps.1431

Li, J., Ahmed, W., Metcalfe, S., Smith, W. J. M., Choi, P. M., Jackson, G., Cen, X., Zheng, M., Simpson, S. L., Thomas, K. V., Mueller, J. F., & Thai, P. K. (2023). Impact of sewer biofilms on fate of SARS-CoV-2 RNA and wastewater surveillance. *Nature Water*. https://doi.org/10.1038/s44221-023-00033-4

Li, Y., Miyani, B., Zhao, L., Spooner, M., Gentry, Z., Zou, Y., Rhodes, G., Li, H., Kaye, A., Norton, J., & Xagoraraki, I. (2022). Surveillance of SARS-CoV-2 in nine neighborhood sewersheds in Detroit Tri-County area, United States: Assessing per capita SARS-CoV-2 estimations and COVID-19 incidence. *Science of The Total Environment*, *851*, 158350. https://doi.org/10.1016/j.scitotenv.2022.158350

Liu, L., Li, W., & Jia, H. (2018). Method of time series similarity measurement based on dynamic time warping. *Comput. Mater. Contin*, *57*, 97–106.

Lu, D., Huang, Z., Luo, J., Zhang, X., & Sha, S. (2020a). 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*, 141245. https://doi.org/10.1016/j.scitotenv.2020.141245

Lu, X., Wang, L., Sakthivel, S. K., Whitaker, B., Murray, J., Kamili, S., Lynch, B., Malapati, L., Burke, S. A., Harcourt, J., Tamin, A., Thornburg, N. J., Villanueva, J. M., & Lindstrom, S.

(2020b). US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, *26*(8), 1654–1665. https://doi.org/10.3201/eid2608.201246

McKindles, K. M., Manes, M. A., DeMarco, J. R., McClure, A., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2020). Dissolved Microcystin Release Coincident with Lysis of a Bloom Dominated by *Microcystis* spp. in Western Lake Erie Attributed to a Novel Cyanophage. *Applied and Environmental Microbiology*, *86*(22). https://doi.org/10.1128/AEM.01397-20

McMinn, B. R., Korajkic, A., Kelleher, J., Herrmann, M. P., Pemberton, A. C., Ahmed, W., Villegas, E. N., & Oshima, K. (2021). Development of a large volume concentration method for recovery of coronavirus from wastewater. *Science of The Total Environment*, *774*, 145727. https://doi.org/10.1016/j.scitotenv.2021.145727

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 and Technology Letters*, *7*(7), 511–516. https://doi.org/10.1021/acs.estlett.0c00357

Mei, D. C., Du, L. C., & Wang, C. J. (2009). The Effects of Time Delay on Stochastic Resonance in a Bistable System with Correlated Noises. *Journal of Statistical Physics*, *137*(4), 625–638. https://doi.org/10.1007/s10955-009-9864-4

Menkir, T. F., Chin, T., Hay, J. A., Surface, E. D., de Salazar, P. M., Buckee, C. O., Watts, A., Khan, K., Sherbo, R., Yan, A. W. C., Mina, M. J., Lipsitch, M., & Niehus, R. (2021). Estimating internationally imported cases during the early COVID-19 pandemic. *Nature Communications*, *12*(1), 311. https://doi.org/10.1038/s41467-020-20219-8

Michael-Kordatou, I., Karaolia, P., & Fatta-Kassinos, D. (2020). Sewage analysis as a tool for the COVID-19 pandemic response and management: the urgent need for optimised protocols for SARS-CoV-2 detection and quantification. *Journal of Environmental Chemical Engineering*, *8*(5), 104306. https://doi.org/10.1016/j.jece.2020.104306

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., Norton, J., & Xagoraraki, I. (2021). Early Warnings of COVID-19 Second Wave in Detroit. *Journal of Environmental Engineering*, *147*(8), 06021004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001907

Panchal, D., Prakash, O., Bobde, P., & Pal, S. (2021). SARS-CoV-2: sewage surveillance as an early warning system and challenges in developing countries. *Environmental Science and Pollution Research*, *28*(18), 22221–22240. https://doi.org/10.1007/s11356-021-13170-8

Pecson, B. M., Darby, E., Haas, C. N., Amha, Y. M., Bartolo, M., Danielson, R., Dearborn, Y., di Giovanni, G., Ferguson, C., Fevig, S., Gaddis, E., Gray, D., Lukasik, G., Mull, B., Olivas, L.,

Olivieri, A., Qu, Y., & SARS-CoV-2 Interlaboratory Consortium. (2021). Reproducibility and sensitivity of 36 methods to quantify the SARS-CoV-2 genetic signal in raw wastewater: findings from an interlaboratory methods evaluation in the U.S. *Environmental Science: Water Research & Technology*, *7*(3), 504–520. https://doi.org/10.1039/D0EW00946F

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., & Keogh, E. (2012). Searching and mining trillions of time series subsequences under dynamic time warping. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, 262. https://doi.org/10.1145/2339530.2339576

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*, *181*. https://doi.org/10.1016/j.watres.2020.115942

Sapula, S. A., Whittall, J. J., Pandopulos, A. J., Gerber, C., & Venter, H. (2021). An optimized and robust PEG precipitation method for detection of SARS-CoV-2 in wastewater. *Science of The Total Environment*, *785*, 147270. https://doi.org/10.1016/j.scitotenv.2021.147270

Shen, C. (2015). Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A*, *379*(7), 680–687. https://doi.org/10.1016/j.physleta.2014.12.036

Sherchan, S. P., Shahin, S., Ward, L. M., Tandukar, S., Aw, T. G., Schmitz, B., Ahmed, W., & Kitajima, M. (2020). First detection of SARS-CoV-2 RNA in wastewater in North America: A study in Louisiana, USA. *Science of The Total Environment*, *743*, 140621. https://doi.org/10.1016/j.scitotenv.2020.140621

Simpson, A., Topol, A., White, B. J., Wolfe, M. K., Wigginton, K. R., & Boehm, A. B. (2021). Effect of storage conditions on SARS-CoV-2 RNA quantification in wastewater solids. *PeerJ*, *9*, e11933.

Terry, I. B., & Meschke, J. S. (2022). Variability in Normalization Methods of COVID-19 Wastewater Surveillance. In *ProQuest Dissertations and Theses*. http://ezproxy.msu.edu/login?url=https://www.proquest.com/dissertations-theses/variability-normalization-methods-covid-19/docview/2695037740/se-2?accountid=12598

Torii, S., Furumai, H., & Katayama, H. (2021). Applicability of polyethylene glycol precipitation followed by acid guanidinium thiocyanate-phenol-chloroform extraction for the detection of SARS-CoV-2 RNA from municipal wastewater. *Science of The Total Environment*, *756*, 143067. https://doi.org/10.1016/j.scitotenv.2020.143067

Vadde, K. K., Al-Duroobi, H., Phan, D. C., Jafarzadeh, A., Moghadam, S. v, Matta, A., & Kapoor, V. (2022). Assessment of Concentration, Recovery, and Normalization of SARS-CoV-2 RNA from Two Wastewater Treatment Plants in Texas and Correlation with COVID-19 Cases in the Community. *ACS ES&T Water*. https://doi.org/10.1021/acsestwater.2c00054

Wallis, C., Henderson, M., & Melnick, J. L. (1972). Enterovirus Concentration on Cellulose Membranes. *Applied Microbiology*, *23*(3), 476–480. https://doi.org/10.1128/am.23.3.476-

480.1972

Wu, Y., Kang, L., Guo, Z., Liu, J., Liu, M., & Liang, W. (2022). Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains. *JAMA Network Open*, *5*(8), e2228008. https://doi.org/10.1001/jamanetworkopen.2022.28008

Xagoraraki, I. (2020). Can We Predict Viral Outbreaks Using Wastewater Surveillance? *Journal of Environmental Engineering*, *146*(11), 01820003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001831

Xagoraraki, I., Yin, Z., & Svambayev, Z. (2014). Fate of Viruses in Water Systems. *Journal of Environmental Engineering*, *140*(7), 04014020. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000827

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

Ye, J., Xiao, C., Esteves, R. M., & Rong, C. (2015). *Time Series Similarity Evaluation Based on Spearman's Correlation Coefficients and Distance Measures* (pp. 319–331). https://doi.org/10.1007/978-3-319-28430-9_24

Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science and Technology*, *50*(10), 5077–5085. https://doi.org/10.1021/acs.est.6b00876

Zhao, L., Zou, Y., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Norton, J., & Xagoraraki, I. (2022a). Simple methods for early warnings of COVID-19 surges: Lessons learned from 21 months of wastewater and clinical data collection in Detroit, Michigan, United States. *Science of The Total Environment*, 161152. https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.161152

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarland, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022b). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, 157040. https://doi.org/10.1016/j.scitotenv.2022.157040

Zheng, X., Deng, Y., Xu, X., Li, S., Zhang, Y., Ding, J., On, H. Y., Lai, J. C. C., In Yau, C., Chin, A. W. H., Poon, L. L. M., Tun, H. M., & Zhang, T. (2022). Comparison of virus concentration methods and RNA extraction methods for SARS-CoV-2 wastewater surveillance. *Science of The Total Environment*, *824*, 153687. https://doi.org/10.1016/j.scitotenv.2022.153687

Zhu, Y., Oishi, W., Maruo, C., Saito, M., Chen, R., Kitajima, M., & Sano, D. (2021). Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Science of*

The Total Environment, *767*, 145124. https://doi.org/10.1016/j.scitotenv.2021.145124

CHAPTER 4: TRACKING THE TIME LAG BETWEEN SARS-COV-2 WASTEWATER CONCENTRATIONS AND THREE COVID-19 CLINICAL METRICS: A 21-MONTH CASE STUDY IN DETROIT AREA, MICHIGAN

Published in *Journal of Environmental Engineering*:

Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2024). Tracking the Time Lag between SARS-CoV-2 Wastewater Concentrations and Three COVID-19 Clinical Metrics: A 21-Month Case Study in the Tricounty Detroit Area, Michigan. *Journal of Environmental Engineering*, *150*(1), 06023004.

Abstract

Wastewater surveillance has been widely implemented to monitor COVID-19 incidences in communities worldwide. One notable application of wastewater surveillance is for providing early warnings of disease outbreaks. Many studies have reported time lags between SARS-CoV-2 wastewater concentrations and confirmed clinical COVID-19 cases. Only a few studies, to date, have explored time lags between SARS-CoV-2 wastewater concentrations and other clinical metrics. In this study, we investigated time lags between SARS-CoV-2 wastewater concentrations and three COVID-19 clinical metrics: confirmed clinical cases, hospitalizations, and ICU admissions, in the Tri-county Detroit Area, Michigan, USA. The COVID-19 clinical metrics were dated between September 1, 2020, and October 31, 2022, and were collected from public data sources. SARS-CoV-2 N1 and N2 gene concentrations between September 1, 2020, and May 31, 2022, were generated using two sampling and concentration methods: virus adsorption-elution (VIRADEL) and polyethylene glycol precipitation (PEG). The data were collected from our recently published study. Time lagged cross correlation was implemented to estimate time lags between gene concentrations and the three clinical metrics. Original gene concentrations were

normalized by wastewater flow parameters through nine approaches to estimate the impact of wastewater flow on time lags. Vector autoregression models were established to analyze the relationship between gene concentrations and clinical metrics. The results indicate that VIRADEL gene concentrations in wastewater preceded all clinical metrics prior to the COVID-19 Omicron surge, for instance, 32 days, 47 days, and 51 days preceding confirmed cases, hospitalizations, and ICU admissions, respectively (gene concentrations unit: gc/day). When translated to a public health context, these time lags become critical lead times for officials to prepare and react. During the Omicron surge, there were significant reductions in time lags, with VIRADEL measurements trailing total ICU admissions. PEG measurements lagged behind the three clinical metrics and did not provide early warnings of disease surges.

1. Introduction

Wastewater surveillance has gained immense attention since the inception of the COVID-19 pandemic and has been widely utilized to monitor the disease globally (Ahmed et al., 2020; Ahmed et al., 2021; Bivins & Bibby, 2021; Galani et al., 2022; Gentry et al., 2023; Hopkins et al., 2023; Li et al., 2022; Miyani et al., 2020, 2021; Peccia et al., 2020; Saguti et al., 2021; Schenk et al., 2023; Zhao et al., 2022, 2023a). One of the most intensely studied and prominent applications of wastewater surveillance is the determination of the time lag between SARS-CoV-2 wastewater concentrations and COVID-19 clinical metrics, primarily confirmed COVID-19 cases (Miyani et al., 2021; Peccia et al., 2020; Zhao et al., 2022). Recently, a few studies investigated the time lag between SARS-CoV-2 wastewater surveillance and other COVID-19 clinical metrics, including COVID-19 hospitalizations and ICU admissions. Researchers found that SARS-CoV-2 gene concentrations in untreated wastewater preceded COVID-19 hospitalizations by at least 3 to 9 days in Amsterdam and Utrecht, the Netherlands (Stephens et al., 2022); 8 days in Attica, Greece

(Galani et al., 2022); 7 to 13 days in Houston, Texas, USA (Hopkins et al., 2023); and 19 to 21 days in Gothenburg, Sweden (Saguti et al., 2021). Additionally, other studies indicated that SARS-CoV-2 concentrations in wastewater sludge preceded local hospitalizations by 1 to 4 days in New Haven, Connecticut, USA (Peccia et al., 2020), and 4 days in Ottawa, Ontario, Canada (D'Aoust et al., 2021). Likewise, a few studies have also reported that SARS-CoV-2 gene concentrations in untreated wastewater preceded COVID-19 ICU admissions. Galani et al. found this time lag to be 9 days, exceeding the time lag between SARS-CoV-2 wastewater concentrations and hospitalizations by 1 day (Galani et al., 2022). Other researchers have reported a longer time lag between SARS-CoV-2 wastewater concentrations and ICU admissions, including 10 to 16 days in Houston, Texas, USA (Hopkins et al., 2023), and a maximum of 17.7 days across Australia (Schenk et al., 2023).

In this study, "time lag" was defined as the duration between peaks in measured SARS-CoV-2 wastewater concentrations and peaks in reported clinical metrics (Zhao et al., 2022). We investigated the time lag between SARS-CoV-2 N1 and N2 gene concentrations in wastewater and three clinical metrics including confirmed cases, hospitalizations, and ICU admissions, all within the Tri-county Detroit area (TCDA), Michigan, USA, between September 1, 2020, and May 31, 2022. The study period included three major surges of COVID-19 cases, associated with different SARS-CoV-2 variants such as Alpha, Delta, and Omicron shown in Tables 4S. 1. and 4S. 2. Thus, we analyzed a 21-month dataset of SARS-CoV-2 N1 and N2 gene wastewater concentrations using two sampling and concentration methods in the TCDA. These methods included the United States Environmental Protection Agency's (U.S. EPA) Virus Adsorption-Elution (VIRADEL) method and the polyethylene glycol precipitation (PEG) method. The time lag between SARS-CoV-2 wastewater concentrations and the three clinical metrics were estimated

using time lagged cross correlation (TLCC). Nine approaches of normalizing SARS-CoV-2 concentrations were performed to investigate the impact of flow parameters on time lag. We demonstrate the early warning potential of the VIRADEL method, which provided lead times for all three clinical metrics prior to the Omicron surge. During the Omicron surge, we observed that VIRADEL N1 and N2 gene concentrations preceded confirmed COVID-19 cases and hospitalizations under both normalized and non-normalized conditions. PEG SARS-CoV-2 concentrations lagged behind all three clinical metrics during the Omicron surge for both normalized and non-normalized conditions. Our results also indicate that normalizations using flow parameters can affect time lag.

2. Materials and Methods

2.1 Study Area and Sample Collection

Untreated wastewater samples were collected weekly from the Great Lakes Water Authority (GLWA) Water Resource Recovery Facility (WRRF) located in southeastern Michigan, USA, between September 1, 2020, and May 31, 2022. The WRRF operates a semi-combined sewershed system, which treats stormwater, residential, industrial, and commercial waste, depending on service areas (Zhao et al., 2023b). The WRRF serves the TCDA, including the city of Detroit, as well as Wayne, Macomb, and Oakland Counties (see chapter 1 Figure 1. 2.). The WRRF receives wastewater via three main interceptors, including the Oakwood-Northwest-Wayne County Interceptor (ONWI), the North Interceptor-East Arm (NIEA), and the Detroit River Interceptor (DRI) (see chapter 1 Figure 1. 3.). As of 2020, population served by ONWI, NIEA, and DRI are 840600, 1482000, and 492000, respectively (Miyani et al., 2021). As mentioned, two sampling and concentration methods were used, VIRADEL (Li et al., 2022; Miyani et al., 2020, 2021; Zhao et al., 2022) and PEG (D'Aoust et al., 2021; Kaya et al., 2022; Zhao et al., 2022).

Briefly, for VIRADEL, approximately 10 gallons of untreated wastewater passed through NanoCeram electropositive cartridge filters with a less than 11 L/min rate (Miyani et al., 2020, 2021; Zhao et al., 2022). For PEG, 1 L of 24-hr composite samples were collected in sterilized polythene plastic bottles weekly. Subsequently, the filters were placed in sealed plastic bags and transported together with the bottles on ice to the Environmental Virology Laboratory at Michigan State University within 24 hours for downstream processes.

2.2 Clinical Metrics and Wastewater Surveillance Dataset

Analyzed COVID-19 clinical metrics included total confirmed cases, total hospitalizations, and total ICU admissions. The total confirmed COVID-19 cases for the TCDA were accessed on December 28, 2022, for the period between September 1, 2020, and October 31, 2022, from a publicly available source (michigan.gov/coronavirus). Total COVID-19 hospitalizations and total ICU admissions for the TCDA were accessed on December 28, 2022, for the period between September 1, 2020, and October 31, 2022, also from a publicly available source (covidactnow.org/us/michigan-mi). To mitigate the impact of outliers and to ensure a more accurate representation of data, 7-day moving averages of clinical metrics (Figure 4. 1. a.) were used for downstream statistical analyses (Menkir et al., 2021; Zhao et al., 2022). Available COVID-19 clinical metrics were limited to the city or county level within the TCDA. Additionally, wastewater captured by each interceptor represented portions of each city or county in the TCDA. Consequently, only the total N1 and N2 gene concentrations could be associated with total COVID-19 clinical metrics within each jurisdiction (Miyani et al., 2020, 2021; Zhao et al., 2022). Other detailed descriptions of COVID-19 clinical metrics are provided in the Supplementary Materials.

The SARS-CoV-2 N1 and N2 gene concentration data using VIRADEL and PEG methods

were collected throughout our previously published wastewater surveillance study shown in Figures 4. 1. b. and 4. 1. c. (Zhao et al., 2023a). Missing data from samples were addressed using linear interpolation for downstream analyses (Lepot et al., 2017; Zhao et al., 2022, 2023a). Detailed information on portions of data using linear interpolation are presented in the Supplementary Materials.

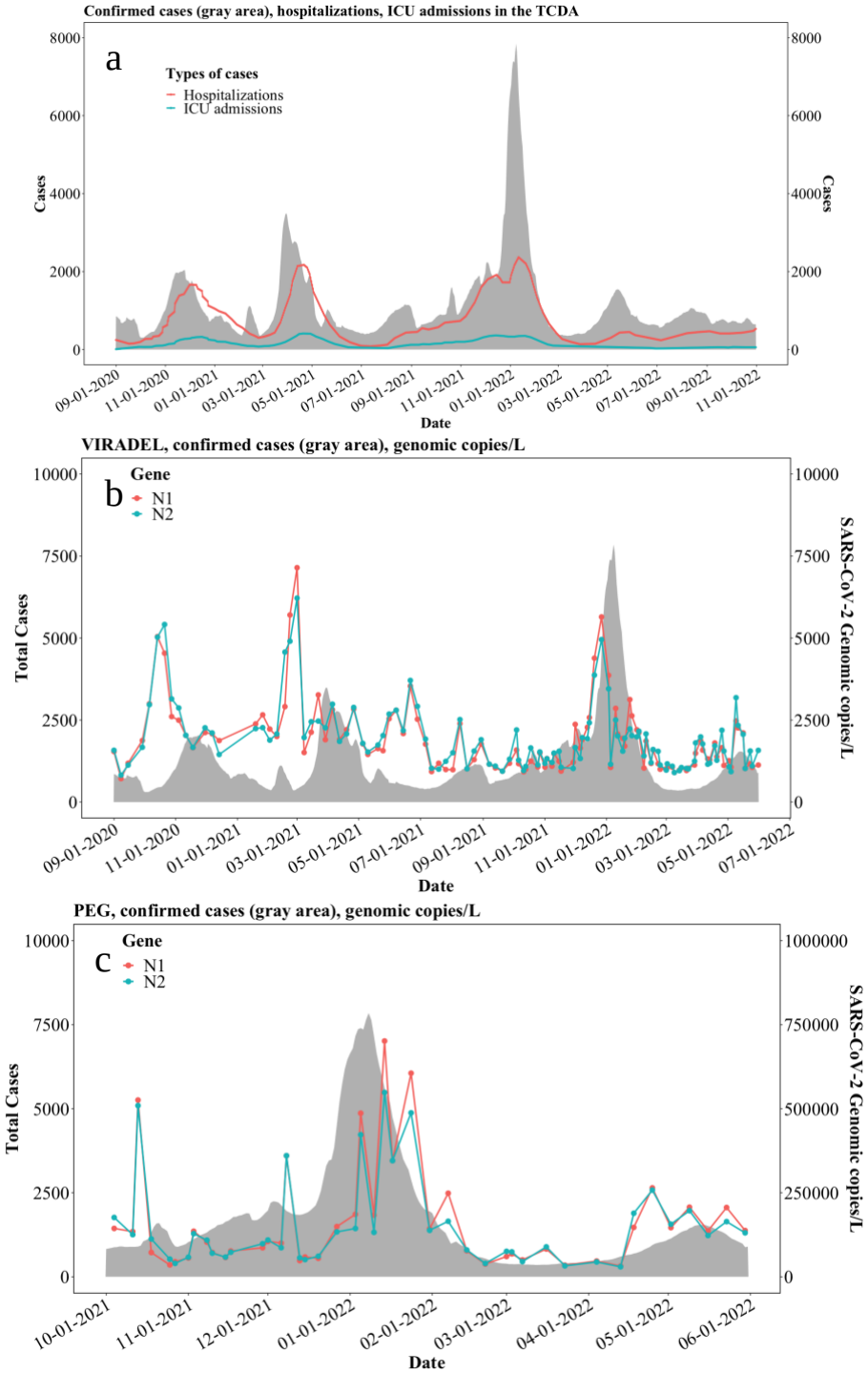


Figure 4. 1. a. Clinical COVID-19 metrics: total confirmed cases, total hospitalizations, and total ICU admissions in the TCDA (7-day moving average data). b. N1 and N2 gene concentrations (gc/L) by VIRADEL. c. N1 and N2 gene concentrations (gc/L) by PEG

2.3 Interceptor Flow Data and Normalizations

The population contributing to each interceptor for each jurisdiction in the TCDA was estimated from 2020 calculations (Miyani et al., 2021). Daily flow parameters included flow rate (million gallons/day), sanitary percentage of wastewater (%), TSS (mg/L), CBOD (mg/L), and TP (mg/L) and they were collected from the WRRF. To investigate the influence of flow and dilution of wastewater on time lag, nine approaches to normalizing wastewater N1 and N2 gene concentrations were performed. Flow rate was consistently applied to normalize wastewater SARS-CoV-2 concentrations to account for dilution effects (Hopkins et al., 2023; Zhao et al., 2022). Wastewater quality parameters of CBOD and TP were easily measured in WRRF and were applied for normalization previously (Maal-Bared et al., 2023; Isaksson et al., 2022; Zhao et al., 2022). Maal-Bared et al. indicated that using easily measured wastewater quality parameters such as TP could provide comparable advantages to the application of Pepper Mild Mottle Virus normalization (Maal-Bared et al., 2023). Sanitary percentage and TSS were frequently applied for normalization of wastewater SARS-CoV-2 concentrations (Maal-Bared et al., 2023; Zhao et al., 2022, 2023b). Especially, TSS might be useful to normalize wastewater SARS-CoV-2 concentrations by the solids-contained PEG method (Flood et al., 2021). Detailed calculations of normalization approaches are described in the Supplementary Materials. Both non-normalized and normalized N1 and N2 gene concentrations were used for downstream statistical analyses.

2.4 Data Analyses

Data were tracked and organized using Microsoft Excel version 16.73. R version 4.2.3 (2023-03-15) was utilized for performing data analyses. To estimate the time lag between N1 and N2 gene concentrations and clinical metrics, time lagged cross correlation (TLCC) was performed. Original N1 and N2 gene concentrations (gc/L) were normalized using flow parameters into the

following units: gc/day, gc/day/person, gc/L of sanitary flow percentage, gc/day of sanitary flow, gc/mg TSS, gc/mg BOD, gc/L of TSS:BOD ratio, gc/mg TP, and gc/(L*(pounds/day)). TLCC has been widely used for analyzing time lags between time series, providing an effective approach to estimate the temporal relationship between two time series and to illustrate their temporal shift (Hopkins et al., 2023; Mei et al., 2009; Shen, 2015). R packages "synchrony", "devtools", "ggplot2", "ggpubr", and other related packages were used to estimate and visualize time lag using TLCC. Prior to the Omicron surge (09/01/2020 and 08/31/2021), the clinical metrics were shifted by a period of 0 and +100 days and the Pearson's coefficients were calculated between VIRADEL gene concentrations and clinical metrics for each shift. During the Omicron surge (10/01/2021 and 05/31/2022), the clinical metrics were shifted by a period of -30 and +100 days and the Pearson's coefficients were calculated between VIRADEL as well as PEG gene concentrations and clinical metrics for each shift. The lead/lag pattern varied between the VIRADEL and PEG methods, which were determined by assessing the strongest Pearson's correlation coefficient (Hopkins et al., 2023; Zhao et al., 2022; 2023b). Figure 4. 2. demonstrates an example of time lag estimation using TLCC between VIRADEL N1 and N2 concentrations (gc/L) and total confirmed COVID-19 cases, hospitalizations, and ICU admissions, prior to the Omicron surge. Peak synchrony indicates the strongest correlation. For instance, time lag between VIRADEL N1 gene concentrations (gc/L) and confirmed COVID-19 cases is estimated as 34 days, when the strongest Pearson's correlation between the time series is observed and the time series are most synchronized (Figure 4. 2. a.). Additionally, a vector autoregressive (VAR) model was employed to estimate the relationship between wastewater gene concentrations and clinical metrics. VAR was proven to be effective in evaluating relationships between gene concentrations in wastewater and COVID-19 clinical metrics (Cao & Francis, 2021; Zhao et al., 2022).

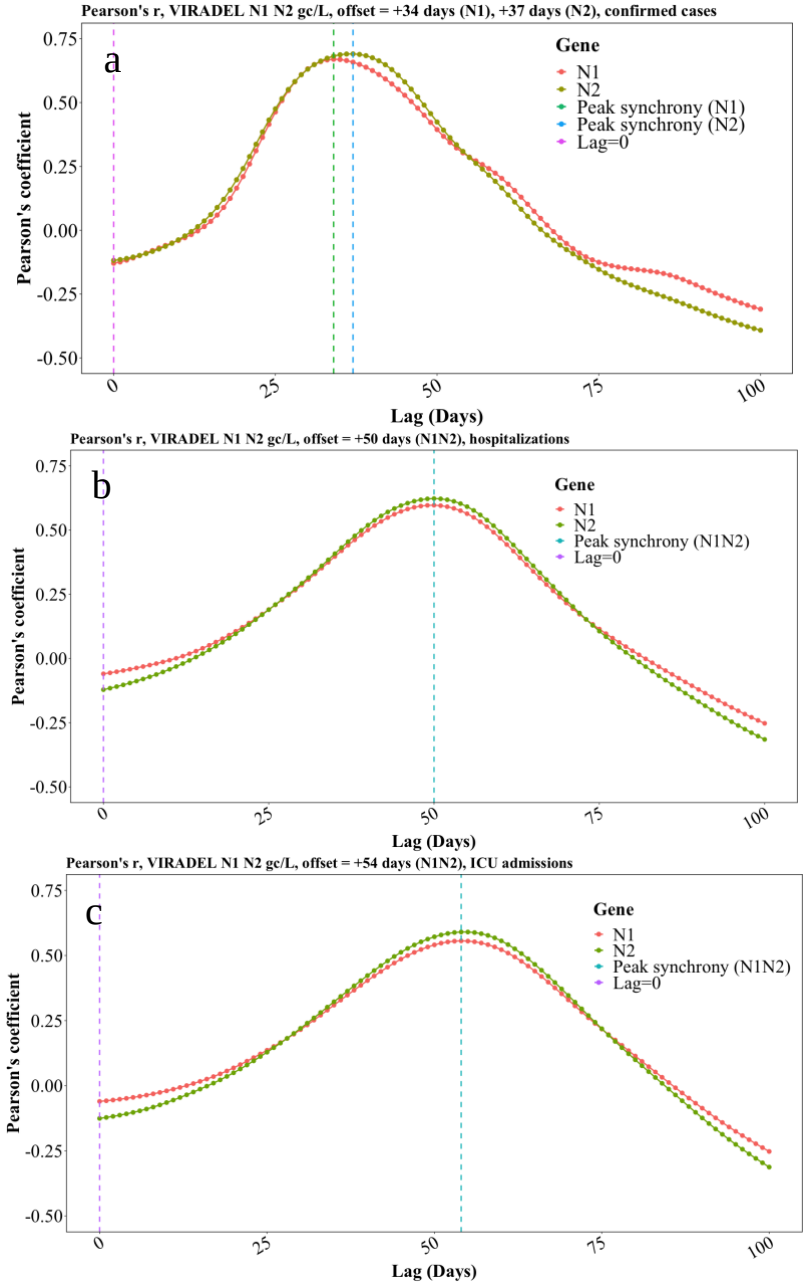


Figure 4. 2. TLCC results between N1 N2 gene concentrations (gc/L) by VIRADEL and (a) total confirmed COVID-19 cases (b) total hospitalizations, and (c) total ICU admissions

3. Results and Discussion

For the VIRADEL method, both normalized and non-normalized N1 and N2 gene concentrations were strongly correlated with clinical metrics, where time lags were estimated (Table 4. 1.). Prior to the Omicron surge, N1 and N2 gene concentrations (gc/L) preceded total confirmed cases by 34 and 37 days, respectively. Both N1 and N2 gene concentrations (gc/L) preceded total hospitalizations and total ICU admissions by 50 and 54 days, respectively. Time lags with normalized N1 and N2 gene concentrations were observed with similar lead patterns (Table 4. 1.). The time lags between gene concentrations and ICU admissions (e.g., gc/L: 54 days for both N1 and N2) were longer than the time lags between gene concentrations and hospitalizations (e.g., gc/L: 50 days for both N1 and N2), a trend that was also observed in other studies (Galani et al., 2022; Hopkins et al., 2023). The VIRADEL method was previously reported to focus on supernatant portions of viruses in wastewater, thereby explaining its significant potential in providing early warnings of COVID-19 case surges (Miyani et al., 2020, 2021; Zhao et al., 2022, 2023a, 2023b). Especially, Zhao et al. compared VIRADEL and PEG concentration methods, which indicated that VIRADEL captures free and suspended virus from supernatant wastewater while PEG targets particle-associated viruses that sorbed onto solids (Zhao et al., 2023b). Furthermore, in this study, VIRADEL demonstrated its capability of providing early warnings of hospitalizations and ICU admissions, prior to the Omicron surge.

Table 4. 1. Time lag, prior to the Omicron surge (09/01/2020 to 08/31/2021) using the VIRADEL method

Time lag (days)	Gene conc. and confirmed cases		Gene conc. and hospitalizations		Gene conc. and ICU admissions	
Gene	N1	N2	N1	N2	N1	N2
gc/L	+34	+37	+50	+50	+54	+54
gc/day	+32	+32	+47	+47	+51	+51
gc/day/person	+32	+32	+47	+47	+51	+51
gc/L of sanitary flow percentage	+32	+32	+49	+49	+54	+53
gc/day of sanitary flow	+35	+37	+47	+47	+52	+52
gc/mg TSS	+30	+30	+45	+45	+49	+49
gc/mg BOD	+30	+30	+44	+44	+47	+47
gc/L of TSS:BOD ratio	+32	+38	+47	+49	+49	+53
gc/mg TP	+32	+33	+48	+48	+53	+52
gc/(L*(pounds/day))	+30	+30	+46	+48	+50	+52

During the Omicron surge, for the VIRADEL method, N1 and N2 gene concentrations (gc/L) preceded total confirmed cases by 11 days (Table 4. 2.). Time lags, however, became shorter (11 days for both N1 and N2 [gc/L]) than before the Omicron surge (34 and 37 days for N1 and N2 [gc/L], respectively), which supports previous findings that time lag was reduced during the Omicron surge (Hopkins et al., 2023; Zhao et al., 2022). VIRADEL N1 and N2 gene concentrations lagged total ICU admissions for all normalized and non-normalized conditions (Table 4. 2.).

During the Omicron surge, the relationship between gene concentrations and total hospitalizations did not present a consistent lead or lag relationship among normalized and non-normalized conditions (Table 4. 2.). For example, non-normalized gene concentrations (gc/L) preceded total hospitalizations (14 and 16 days for N1 and N2, respectively), while normalized gene concentrations lagged total hospitalizations, including by gc/day (-12 days for both N1 and

N2), gc/day/person (-12 days for both N1 and N2), gc/day of sanitary flow (-12 days for both N1 and N2), and gc/mg BOD (-8 days for both N1 and N2). This demonstrates that different approaches of normalization may also affect time lag. This effect was also investigated in a recent study in Sweden, where Isaksson et al. found that time lags of normalized SARS-CoV-2 concentrations preceding clinical cases varied comparing to non-normalized scenarios (Isaksson et al., 2022). However, the researchers concluded that the impact of normalizations on time lags were heretofore uncertain (Isaksson et al., 2022). Future exploration is required to fully investigate and understand the influence of normalization approaches on time lags.

Table 4. 2. Time lag, during the Omicron surge using the VIRADEL method

Time lag (days)	Gene conc. and confirmed cases		Gene conc. and hospitalizations		Gene conc. and ICU admissions	
Gene	N1	N2	N1	N2	N1	N2
gc/L	+11	+11	+14	+16	-9	-12
gc/day	+8	+8	-12	-12	-9	-10
gc/day/person	+8	+8	-12	-12	-9	-10
gc/L of sanitary flow percentage	+14	+15	+21	+24	-7	-10
gc/day of sanitary flow	+7	+7	-12	-12	-10	-10
gc/mg TSS	+11	+11	+17	+20	-5	-7
gc/mg BOD	-12	-12	-8	-8	-5	-6
gc/L of TSS:BOD ratio	+9	+9	+14	+15	-10	-13
gc/mg TP	+15	+15	+21	+23	-5	-7
gc/(L*(pounds/day))	+14	+13	+21	+22	-5	-8

Shifting SARS-CoV-2 variant frequencies from Delta to Omicron, and their differing epidemiological characteristics may have played a critical role in the time lag (Ali et al., 2020; Galani et al., 2022; Hopkins et al., 2023; Zhao et al., 2022). Researchers have discovered that time lag can be affected by the dominant variants circulating during peaks in clinical metrics data (Hopkins et al., 2023). Moreover, with the continual evolution of SARS-CoV-2, changes in

epidemiological characteristics, such as the severity of disease, and developments in medical treatments, such as vaccination, influenced hospitalization and ICU admission rates over the course of the pandemic (Hopkins et al., 2023; Peng et al., 2023). This can lead to uncertainty of clinical metrics at a given time, affecting time lag calculations. In the latter stages of the pandemic, particularly during the Omicron surge, there was an increase in testing resources available and a widespread adoption of at-home testing both in the TCDA, and nationwide (Gupta et al., 2023). This could have resulted in reduced time lags and diminished early warning potential of wastewater surveillance since clinical metrics can be affected by repeatedly counted cases and massive at-home testing (Hopkins et al., 2023).

For the PEG method, N1 and N2 gene concentrations lagged clinical metrics for nearly all normalized and non-normalized conditions during the Omicron surge (Table 4. 3.). This can be potentially explained by the sampling mechanism of the PEG method, as well as the aforementioned uncertainties across clinical metrics during the Omicron surge. The PEG method incorporates and thus takes measurements from viruses attached onto larger solid particles. These particles generally settle on the sewer system floor but are then typically resuspended during events that increase flow (Flood et al., 2021; Zhao et al., 2022). Such viruses would have been excreted into the sewer system, and served as indicators of past infections within communities, creating an elongated time lag with clinical metrics when compared to the VIRADEL method. Worth noting, the VAR model was effective in establishing relationships between all three COVID-19 clinical metrics and N1 and N2 gene concentrations, both before and during the Omicron surge (Tables 4S. 3. and 4S. 4.), which supported previous findings (Cao & Francis, 2021; Zhao et al., 2022).

Table 4. 3. Time lag, during the Omicron surge using the PEG method

Time lag (days)	Gene conc. and confirmed cases		Gene conc. and hospitalizations		Gene conc. and ICU admissions	
Gene	N1	N2	N1	N2	N1	N2
gc/L	-12	-12	-3	-3	-10	-3
gc/day	-15	-15	-8	-7	-8	-6
gc/day/person	-15	-15	-8	-7	-8	-6
gc/L of sanitary flow percentage	-6	-6	-1	0	-3	-2
gc/day of sanitary flow	-16	-16	-9	-8	-9	-7
gc/mg TSS	-9	-9	-3	-2	-9	-4
gc/mg BOD	-6	-6	0	+2	-3	-2
gc/L of TSS:BOD ratio	-14	-14	-7	-6	-11	-7
gc/mg TP	-7	-7	-1	0	-4	-2
gc/(L*(pounds/day))	-6	-7	+2	+4	-13	-2

The determination of time lag can be affected by multiple factors, all of which varied across the study period, including duration and frequency of the major circulating SARS-CoV-2 variants (Tables 4S. 1. and 4S. 2.) (and corresponding shifts in epidemiological and clinical characteristics, especially varying incubation time and shedding dynamics), testing availability, clinical data consistency (such as onsite testing or at-home testing), wastewater sampling frequency, and wastewater sampling and concentration methodology (such as VIRADEL or PEG), presenting a challenge to systematically incorporating all factors into time lag calculations. Figure 4. 3. provides a comprehensive overview of potential influencers on time lag. Recent studies demonstrated the impacts on time lags by environmental factors, such as wastewater transportation within the sewer network and temperature (Ahmed et al., 2020; Bertels et al., 2022; Galani et al., 2022). Researchers found that SARS-CoV-2 RNA decayed during the long travel of wastewater within the sewer network, leading to uncertainties between the detected viral RNA at sampling points and the actual RNA excreted from hosts (Bertels et al., 2022). Ahmed et al. indicated that

temperature affected SARS-CoV-2 RNA decay albeit persist presence of its viral RNA in wastewater allowing successful detection (Ahmed et al., 2020). More research is called for investigating and integrating these factors into time lag analyses. Nonetheless, and perhaps for this very reason, it is critical to present this study as evidence that time lag changes as SARS-CoV-2 evolves, wastewater sampling and concentration methods progress, clinical metrics data vary, normalization approaches change, and so forth. This study is one of a limited number of investigations that evaluates time lags between SARS-CoV-2 wastewater concentrations and complex COVID-19 clinical data metrics, especially hospitalizations and ICU admissions (D'Aoust et al., 2021; Galani et al., 2022; Hopkins et al., 2023; Peccia et al., 2020; Saguti et al., 2021; Stephens et al., 2022). Encompassing three major waves of COVID-19 in a large metropolitan area, this study makes valuable and unique contribution to understanding temporal relationships between wastewater viral measurements and clinical data metrics. This study moreover demonstrates wastewater surveillance's meaningful potential for providing early warnings of fluctuations in clinical COVID-19 metrics data including, confirmed cases, hospitalizations, and ICU admissions, using the VIRADEL method.

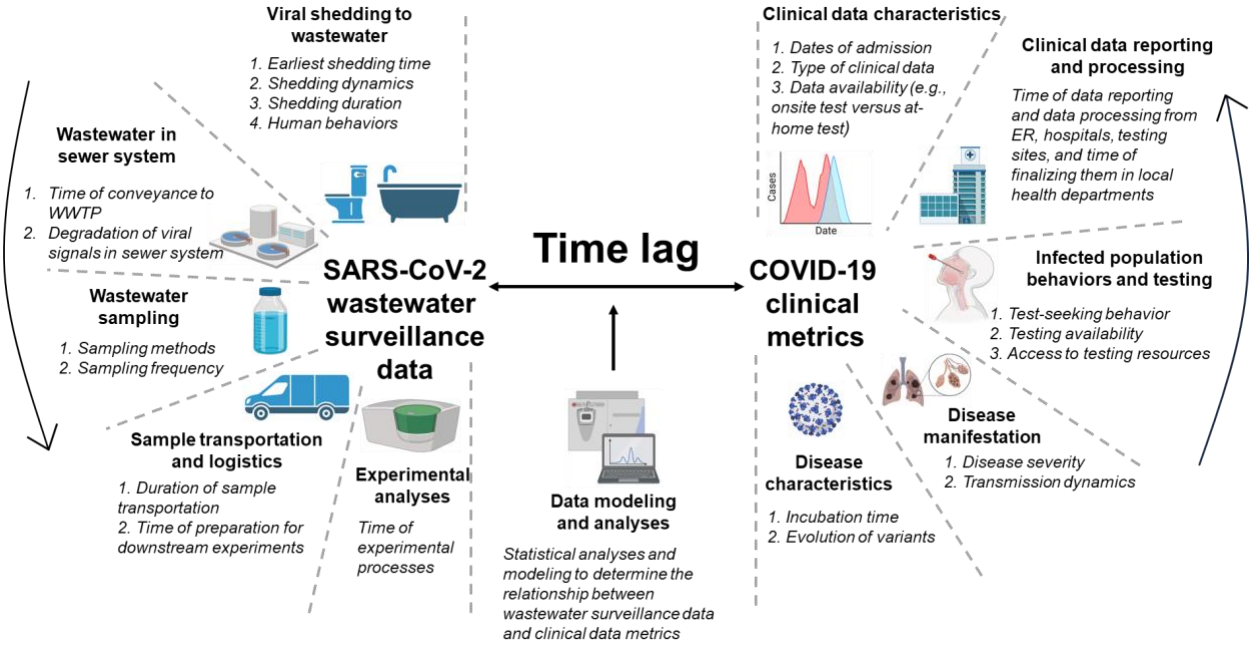


Figure 4. 3. Parameters that potentially affect the time lag

It is important to recognize that this study possesses various limitations and uncertainties. First and foremost, hospitalizations and ICU admissions were accessed from publicly available online data sources as 7-day moving average data. Unfortunately, raw data of these two clinical metrics were not accessible to the researchers. Future in-depth research is called to investigate the raw clinical data and their impact on time lags. Additionally, in this study, VIRADEL and PEG samples were collected twice and once weekly, respectively. More frequent sampling is encouraged to track nuanced temporal fluctuations of wastewater SARS-CoV-2 concentrations. However, researchers indicated that collecting a minimum of two samples on a weekly basis is sufficient to maintain the accuracy required for trend analysis (Feng et al., 2021). Besides, the current study did not perform any testing to identify the presence of different SARS-CoV-2 variants throughout the study period. Future studies should identify variants and establish relationship between time lags and variants.

4. Conclusions

This study contributes valuable insights to the field of wastewater-based epidemiology by estimating the time lag between SARS-CoV-2 concentrations and clinical metrics (confirmed cases, hospitalizations, and ICU admissions), before and during the COVID-19 Omicron surge in a large metropolitan area. By performing TLCC analyses, we were able to compare time lags between VIRADEL and PEG methodologies, between N1 and N2 gene concentrations and three key clinical data points over a 21-monthperiod. A total of nine normalization approaches were performed and evaluated for their potential impact on time lags. PEG did not provide early warnings for three clinical metrics for nearly all non-normalized and normalized conditions during the Omicron surge. VIRADEL demonstrated its potential for early warnings of total confirmed cases, hospitalizations, and ICU admissions, with lead times provided before the Omicron surge. During the Omicron surge, VIRADEL's time lags were reduced, and the early warning potential of ICU admissions was diminished. The resulting lead time can provide a critical window for hospital systems and public health entities to properly prepare for pending disease outbreaks. This study underscores the importance of a robust understanding of the temporal relationship between wastewater viral concentrations and various clinical metrics. Such an understanding can improve the effective translation of wastewater surveillance data, improve wastewater-based epidemiology models, and ultimately, enhance public health preparedness.

5. Acknowledgements

We thank the Michigan Department of Health and Human Services (MDHHS) and the Great Lakes Water Authority (GLWA) for funding this work.

REFERENCES

Ahmed, W., Tscharke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T. M. H., O'Brien, J. W., Simpson, S. L., Sherman, P., Thomas, K. V., Verhagen, R., Zaugg, J., & Mueller, J. F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Science of the Total Environment, 761,* 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Ahmed, W., Bertsch, P. M., Bibby, K., Haramoto, E., Hewitt, J., Huygens, F., ... & Bivins, A. (2020). Decay of SARS-CoV-2 and surrogate murine hepatitis virus RNA in untreated wastewater to inform application in wastewater-based epidemiology. *Environmental research, 191,* 110092. https://doi.org/10.1016/j.envres.2020.110092

Ali, I., & Alharbi, O. M. (2020). COVID-19: Disease, management, treatment, and social impact. *Science of the total Environment*, *728,* 138861. https://doi.org/10.1016/j.scitotenv.2020.138861

Bertels, X., Demeyer, P., Van den Bogaert, S., Boogaerts, T., van Nuijs, A. L., Delputte, P., & Lahousse, L. (2022). Factors influencing SARS-CoV-2 RNA concentrations in wastewater up to the sampling stage: A systematic review. *Science of The Total Environment*, *820,* 153290. https://doi.org/10.1016/j.scitotenv.2022.153290

Bivins, A., & Bibby, K. (2021). Wastewater Surveillance during Mass COVID-19 Vaccination on a College Campus. *Environmental Science & Technology Letters*, *8*(9), 792–798. https://doi.org/10.1021/acs.estlett.1c00519

Cao, Y., & Francis, R. (2021). On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Science of The Total Environment, 786*, 147451. https://doi.org/10.1016/j.scitotenv.2021.147451

D'Aoust, P. M., Graber, T. E., Mercier, E., Montpetit, D., Alexandrov, I., Neault, N., Baig, A. T., Mayne, J., Zhang, X., Alain, T., Servos, M. R., Srikanthan, N., MacKenzie, M., Figeys, D., Manuel, D., Jüni, P., MacKenzie, A. E., & Delatolla, R. (2021). Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Science of the Total Environment*, *770*. https://doi.org/10.1016/j.scitotenv.2021.145319

Feng, S., Roguet, A., McClary-Gutierrez, J. S., Newton, R. J., Kloczko, N., Meiman, J. G., & McLellan, S. L. (2021). Evaluation of sampling, analysis, and normalization methods for SARS-CoV-2 concentrations in wastewater to assess COVID-19 burdens in Wisconsin communities. *Acs Es&T Water*, *1*(8), 1955-1965. https://doi.org/10.1021/acsestwater.1c00160

Flood, M. T., D'Souza, N., Rose, J. B., & Aw, T. G. (2021). Methods Evaluation for Rapid Concentration and Quantification of SARS-CoV-2 in Raw Wastewater Using Droplet Digital and Quantitative RT-PCR. *Food and Environmental Virology*, *13*(3), 303–315. https://doi.org/10.1007/s12560-021-09488-8

Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Adamopoulos, P. G., Peccia, J., Thompson, D. C., Kontou, A., Karagiannidis, A., Lianidou, E. S., Avgeris, M., Paraskevis, D., Tsiodras, S., Scorilas, A., Vasiliou, V., Dimopoulos, M.-A., & Thomaidis, N. S. (2022). SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Science of The Total Environment*, *804*, 150151. https://doi.org/10.1016/j.scitotenv.2021.150151

Gentry, Z., Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2023). Wastewater surveillance beyond COVID-19: a ranking system for communicable disease testing in the tri-county Detroit area, Michigan, USA. *Frontiers in Public Health*, *11*, 1178515.

Gupta, P., Liao, S., Ezekiel, M., Novak, N., Rossi, A., LaCross, N., Oakeson, K., & Rohrwasser, A. (2023). Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiology Spectrum*. https://doi.org/10.1128/spectrum.00391-23

Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., & Stadler, L. B. (2023). Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves. *Science of The Total Environment*, *855*, 158967. https://doi.org/10.1016/j.scitotenv.2022.158967

Isaksson, F., Lundy, L., Hedström, A., Székely, A. J., & Mohamed, N. (2022). Evaluating the Use of Alternative Normalization Approaches on SARS-CoV-2 Concentrations in Wastewater: Experiences from Two Catchments in Northern Sweden. *Environments*, *9*(3), 39. https://doi.org/10.3390/environments9030039

Kaya, D., Niemeier, D., Ahmed, W., & Kjellerup, B. V. (2022). Evaluation of multiple analytical methods for SARS-CoV-2 surveillance in wastewater samples. *Science of The Total Environment*, *808*, 152033. https://doi.org/10.1016/j.scitotenv.2021.152033

Lepot, M., Aubin, J.-B., & Clemens, F. (2017). Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water*, *9*(10), 796. https://doi.org/10.3390/w9100796

Li, Y., Miyani, B., Zhao, L., Spooner, M., Gentry, Z., Zou, Y., Rhodes, G., Li, H., Kaye, A., Norton, J., & Xagoraraki, I. (2022). Surveillance of SARS-CoV-2 in nine neighborhood sewersheds in Detroit Tri-County area, United States: Assessing per capita SARS-CoV-2 estimations and COVID-19 incidence. *Science of The Total Environment*, *851*, 158350. https://doi.org/10.1016/j.scitotenv.2022.158350

Maal-Bared, R., Qiu, Y., Li, Q., Gao, T., Hrudey, S. E., Bhavanam, S., ... & Pang, X. (2023). Does normalization of SARS-CoV-2 concentrations by Pepper Mild Mottle Virus improve correlations and lead time between wastewater surveillance and clinical data in Alberta (Canada): comparing twelve SARS-CoV-2 normalization approaches. *Science of The Total Environment*, *856*, 158964. https://doi.org/10.1016/j.scitotenv.2022.158964

Mei, D. C., Du, L. C., & Wang, C. J. (2009). The Effects of Time Delay on Stochastic Resonance in a Bistable System with Correlated Noises. *Journal of Statistical Physics*, *137*(4),

625–638. https://doi.org/10.1007/s10955-009-9864-4

Menkir, T. F., Chin, T., Hay, J. A., Surface, E. D., De Salazar, P. M., Buckee, C. O., Watts, A., Khan, K., Sherbo, R., Yan, A. W. C., Mina, M. J., Lipsitch, M., & Niehus, R. (2021). Estimating internationally imported cases during the early COVID-19 pandemic. *Nature Communications*, *12*(1), 311. https://doi.org/10.1038/s41467-020-20219-8

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., Norton, J., & Xagoraraki, I. (2021). Early Warnings of COVID-19 Second Wave in Detroit. *Journal of Environmental Engineering*, *147*(8), 06021004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001907

Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., Ko, A. I., Malik, A. A., Wang, D., Wang, M., Warren, J. L., 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. https://doi.org/10.1038/s41587-020-0684-z

Peng, K. K., Renouf, E. M., Dean, C. B., Hu, X. J., Delatolla, R., & Manuel, D. G. (2023). An exploration of the relationship between wastewater viral signals and COVID-19 hospitalizations in Ottawa, Canada. *Infectious Disease Modelling*. https://doi.org/10.1016/j.idm.2023.05.011

Saguti, F., Magnil, E., Enache, L., Churqui, M. P., Johansson, A., Lumley, D., Davidsson, F., Dotevall, L., Mattsson, A., Trybala, E., Lagging, M., Lindh, M., Gisslén, M., Brezicka, T., Nyström, K., & Norder, H. (2021). Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Research*, *189*. https://doi.org/10.1016/j.watres.2020.116620

Schenk, H., Heidinger, P., Insam, H., Kreuzinger, N., Markt, R., Nägele, F., Oberacher, H., Scheffknecht, C., Steinlechner, M., Vogl, G., Wagner, A. O., & Rauch, W. (2023). Prediction of hospitalisations based on wastewater-based SARS-CoV-2 epidemiology. *Science of The Total Environment*, *873*, 162149. https://doi.org/10.1016/j.scitotenv.2023.162149

Shen, C. (2015). Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A*, *379*(7), 680–687. https://doi.org/10.1016/j.physleta.2014.12.036

Stephens, N., Béen, F., & Savic, D. (2022). An Analysis of SARS-CoV-2 in Wastewater to Evaluate the Effectiveness of Nonpharmaceutical Interventions against COVID-19 in The Netherlands. *ACS ES&T Water*, *2*(11), 2158–2166. https://doi.org/10.1021/acsestwater.2c00071

Zhao, L., Zou, Y., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Norton, J., & Xagoraraki, I. (2023a). Simple methods for early warnings of COVID-19 surges: Lessons learned from 21 months of wastewater and clinical data collection in Detroit, Michigan, United States. *Science of The Total Environment*, *864*, 161152.

https://doi.org/10.1016/j.scitotenv.2022.161152

Zhao, L., Geng, Q., Corchis-Scott, R., McKay, R. M., Norton, J., & Xagoraraki, I. (2023b). Targeting a free viral fraction enhances the early alert potential of wastewater surveillance for SARS-CoV-2: a methods comparison spanning the transition between Delta and Omicron Variants in a large urban center. *Frontiers in Public Health*, *11*, 1140441. https://doi.org/10.3389/fpubh.2023.1140441

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarland, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, 157040. https://doi.org/10.1016/j.scitotenv.2022.157040

APPENDIX

Detailed Procedure: Additional information on clinical metrics

Confirmed cases data source: https://www.michigan.gov/coronavirus/stats. The data description is as follows:

1. County is based on the county of residence.

2. Cases are aggregated by the date of onset of COVID-19 symptoms, if known, otherwise by laboratory specimen date if known, otherwise by case referral date.

3. Confirmed cases only include individuals who have had a positive diagnostic laboratory test for COVID-19.

Hospitalizations and ICU admissions data source: https://covidactnow.org/us/michigan-mi

The data description is as follows:

1. A Health Service area (HSA) is used when calculating county-level hospital metrics in order to correct for instances where an individual county does not have any or has few healthcare facilities within its own borders.

2. Hospitalizations consider data of:

- Current staffed acute bed capacity (capacity).
- Total number of acute beds currently in use (currentUsageTotal).
- Number of acute beds currently in use by COVID patients (currentUsageCovid).
- Number of COVID patients admitted in the past week (weeklyCovidAdmissions).

3. ICU hospitalizations consider data of:

- Current staffed ICU bed capacity (capacity).
- Total number of ICU beds currently in use (currentUsageTotal).
- Number of ICU beds currently in use by COVID patients (currentUsageCovid).

Detailed Procedure: Normalization approaches

Normalization by flow: N1N2 gene conc. gc/L × flow MGD/day = gc/day (1)

Normalization by flow and population: N1N2 gene conc. gc/day / population = gc/day/person (2)

Normalization by TSS: N1N2 gene conc. gc/L / TSS mg/L = gc/mg TSS (3)

Normalization by flow and TSS: N1N2 gene conc. gc/L / (Flow × TSS pounds/day) = gc/(L(pounds/day)) (4)

Normalization by sanitary percentage: N1N2 gene conc. gc/L / sanitary % = gc/L of of sanitary flow percentage (5)

Normalization by flow and sanitary percentage: N1N2 gene conc. gc/day × sanitary % = gc/day of sanitary flow (6)

Normalization by BOD: N1N2 gene conc. gc/L / BOD mg/L = gc/mg BOD (7)

Normalization by TSS:BOD ratio: N1N2 gene conc. gc/L / TSS:BOD ratio = gc/L of TSS:BOD ratio (8)

Normalization by TP: N1N2 gene conc. gc/L / TP mg/L = gc/mg TP (9)

Detailed Procedure: Portions of data using linear interpolation

For VIRADEL samples, 230 gene concentrations were measured for both the N1 and N2 genes between 09/01/2020 and 05/31/2022. For PEG samples, 88 gene concentrations were measured for both the N1 and N2 genes between 10/01/2021 and 05/31/2022. To perform TLCC analyses between weekly gene concentrations and daily clinical metrics data, linear interpolation was conducted to generate daily data based on weekly measurements. The number of interpolated daily gene concentrations were 408 and 155 for VIRADEL and PEG, respectively.

Table 4S. 1. Timeline of major SARS-CoV-2 variants of concern (VOC) in the United States

Major variants	Lineage	Duration	References
Alpha	B.1.1.7	January 2021 – July 2021	cdc.gov
Beta	B.1.351	January 2021 – June 2021	
Gamma	P.1	April 2021 – July 2021	
Delta	B.1.617.2	May 2021 – December 2021	
Omicron	B.1.1.529	December 2021 – May 2022	

Table 4S. 2. Time of first detection of major SARS-CoV-2 variants of concern in MI, USA

Major variants	Lineage	First detected in MI, USA	References
Alpha	B.1.1.7	01/16/2021	michigan.gov
Beta	B.1.351	03/08/2021	
Gamma	P.1	03/31/2021	
Delta	B.1.617.2	05/09/2021	
Omicron	B.1.1.529	12/03/2021	

Table 4S. 3. Vector autoregression models between N1 gene conc. and clinical metrics

Sampling and concentration methods	Clinical data metrics	Equation (N1, gc/L)	RMSE
VIRADEL (09/01/2020 – 05/31/2022)	Total confirmed cases	$y_t = 1.8385y_{t-1} - 0.8454y_{t-2} + 0.0269x_{t-1} - 0.0216x_{t-2} - 1.9001$	75.4245
	Total hospitalizations	$y_t = 1.8872y_{t-1} - 0.8894y_{t-2} - 0.0021x_{t-1} + 0.0034x_{t-2} + 0.6991$	16.8205
	Total ICU admissions	$y_t = 1.9120y_{t-1} - 0.9136y_{t-2} + 0.0004x_{t-1} - 0.0003x_{t-2} + 0.3808$	1.5033
PEG (10/01/2021 – 05/31/2022)	Total confirmed cases	$y_t = 1.8551y_{t-1} - 0.8619y_{t-2} + 5.9532e\text{-}05x_{t-1} - 5.6699e\text{-}05x_{t-2} – 11.4032$	106.5390
	Total hospitalizations	$y_t = 1.9649y_{t-1} - 0.9656y_{t-2} - 1.7695e\text{-}05x_{t-1} + 1.4276e\text{-}06x_{t-2} - 3.7552$	9.5298
	Total ICU admissions	$y_t = 1.9826y_{t-1} - 0.9832y_{t-2} - 4.6127e\text{-}06x_{t-1} + 3.6819e\text{-}06x_{t-2} – 0.3486$	1.3460

Note: In Tables S4. 3., and S4. 4., *X* represents measured SARS-CoV-2 concentrations in wastewater, while *Y* represents COVID-19 clinical metrics data. Pearson's correlation and root mean square error (RMSE) are calculated between the actual clinical metrics data and predicted clinical metrics data.

Table 4S. 4. Vector autoregression models between N2 gene conc. and clinical metrics

Sampling and concentration methods	Clinical data metrics	Equation (N2, gc/L)	RMSE
VIRADEL (09/01/2020 – 05/31/2022)	Total confirmed cases	$y_t = 1.8407y_{t-1} - 0.8473y_{t-2} + 0.0244x_{t-1} - 0.0200\ x_{t-2} – 0.4719$	75.3310
	Total hospitalizations	$y_t = 1.8849\ y_{t-1} - 0.8870\ y_{t-2} - 0.0084\ x_{t-1} + 0.0098\ x_{t-2} + 0.3739$	16.8063
	Total ICU admissions	$y_t = 1.9114\ y_{t-1} - 0.9130\ y_{t-2} - 0.0002\ x_{t-1} + 0.0003x_{t-2} + 0.3397$	1.4941
PEG (10/01/2021 – 05/31/2022)	Total confirmed cases	$y_t = 1.8560y_{t-1} - 0.8630y_{t-2} + 7.0905e\text{-}05x_{t-1} - 6.3612e\text{-}05x_{t-2} – 10.9896$	106.5174
	Total hospitalizations	$y_t = 1.9666y_{t-1} - 0.9675y_{t-2} - 1.5212e\text{-}05x_{t-1} - 1.5978e\text{-}06x_{t-2} - 4.0006$	9.9537
	Total ICU admissions	$y_t = 1.9838y_{t-1} - 0.9845y_{t-2} - 5.0867e\text{-}06x_{t-1} + 4.1023e\text{-}06x_{t-2} – 0.3637$	8.8621

CHAPTER 5: COMPARATIVE ANALYSES OF SARS-COV-2 RNA CONCENTRATIONS IN DETROIT WASTEWATER QUANTIFIED WITH CDC N1, N2, AND SC2 ASSAYS REVEAL OPTIMAL TARGET FOR PREDICTING COVID-19 CASES

Published in *Science of The Total Environment*:

Zhao, L., Guzman, H. P., & Xagoraraki, I. (2024). Comparative analyses of SARS-CoV-2 RNA concentrations in Detroit wastewater quantified with CDC N1, N2, and SC2 assays reveal optimal target for predicting COVID-19 cases. *Science of The Total Environment, 945*, 174140.

Abstract

To monitor COVID-19 through wastewater surveillance, global researchers dedicated significant endeavors and resources to develop and implement diverse RT-qPCR or RT-ddPCR assays targeting different genes of SARS-CoV-2. Effective wastewater surveillance hinges on the appropriate selection of the most suitable assay, especially for resource-constrained regions where scant technical and socioeconomic resources restrict the options for testing with multiple assays. Further research is imperative to evaluate the existing assays through comprehensive comparative analyses. Such analyses are crucial for health agencies and wastewater surveillance practitioners in the selection of appropriate methods for monitoring COVID-19. In this study, untreated wastewater samples were collected weekly from the Detroit wastewater treatment plant, Michigan, USA, between January and December 2023. Polyethylene glycol precipitation (PEG) was applied to concentrate the samples followed by RNA extraction and RT-ddPCR. Three assays including N1, N2 (US CDC Real-Time Reverse Transcription PCR Panel for Detection of SARS-CoV-2), and SC2 assay (US CDC Influenza SARS-CoV-2 Multiplex Assay) were implemented to detect SARS-CoV-2 in wastewater. The limit of blank and limit of detection for the three assays were

experimentally determined. SARS-CoV-2 RNA concentrations were evaluated and compared through three statistical approaches, including Pearson and Spearman's rank correlations, Dynamic Time Warping, and vector autoregressive models. N1 and N2 demonstrated the highest correlation and most similar time series patterns. Conversely, N2 and SC2 assay demonstrated the lowest correlation and least similar time series patterns. N2 was identified as the optimal target to predict COVID-19 cases. This study presents a rigorous effort in evaluating and comparing SARS-CoV-2 RNA concentrations quantified with N1, N2, and SC2 assays and their interrelations and correlations with clinical cases. This study provides valuable insights into identifying the optimal target for monitoring COVID-19 through wastewater surveillance.

1. Introduction

Wastewater surveillance of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has been rapidly deployed as a disease monitoring practice worldwide since the start of the coronavirus disease 2019 (COVID-19) pandemic (Ahmed et al., 2020, 2022; Barua et al., 2022; Bivins & Bibby, 2021; Boehm et al., 2023; Gentry et al., 2023; Li et al., 2022, 2024; Miyani et al., 2020, 2021; Xagoraraki, 2020; Zhao et al., 2022, 2024). Multiple RT-qPCR and RT-ddPCR assays have been developed and implemented such as assays targeting SARS-CoV-2 N, ORF1ab, E, S, M, RdRp genes, and so on (Bivins et al., 2021; Calderón-Franco et al., 2022; Saththasivam et al., 2021; Shah et al., 2022; Tamáš et al., 2022). Bivin et al., summarized 208 RT-qPCR assays and found that US CDC recommended assays targeting the N1 and N2 genes were the most frequently applied assays, including 45 % of the RT-qPCR assays utilized to detect SARS-CoV-2 in wastewater (Bivins et al., 2021). Likewise, another study found that the N gene (including N1, N2, and N3 genes) was the most frequently targeted while the S gene demonstrated the most positive samples (Shah et al., 2022). Recently, researchers found that the SC2 assay of the CDC's

Influenza SARS-CoV-2 Multiplex Assay demonstrated advantages including higher throughput, fewer reagents, fewer freeze-thaw cycles, and fewer chances of pipetting errors for detecting SARS-CoV-2 with high-level of sensitivity and specificity, compared to N1 and N2 genes (Shu et al., 2021). The SC2 assay strictly detects the 3' region from the carboxy terminus of the N gene into the 3' untranslated region of the SARS-CoV-2's genome (Shu et al., 2021). It does not detect any other respiratory pathogens or other human coronaviruses such as SARS-CoV or MERS-CoV (Shu et al., 2021). Researchers also successfully applied the SC2 assay to detect B.1.1.7, B.1.351, and P.1 SARS-CoV-2 variants in clinical samples (Nörz et al., 2021; Shu et al., 2021). The current study implemented the N1, N2 genes and SC2 assay to detect SARS-CoV-2 RNA concentrations in Detroit wastewater for the entire 2023.

Multiple studies evaluated and compared assays targeting the N1 and N2 genes in terms of their specificity and sensitivity. For instance, many researchers found that N1 gene was more appropriate for SARS-CoV-2 testing in wastewater, since it presented higher sensitivity, detection, and quantification than N2 gene (Hong et al., 2021; Lanzarini et al., 2023). Similarly, Vogels et al., compared N1 and N2 genes testing results for COVID-19 clinical samples, including nasopharyngeal swabs, saliva, urine, etc., and they identified N1 as a more sensitive gene to target for detecting SARS-CoV-2 with more apparent distinction between positive and negative values (Vogels et al., 2020). Besides, N1 gene was found to have greater sensitivity, especially when concentrations were close to the limit of detection (Grube et al., 2023). However, other investigations demonstrated the opposite outcome, where N2 gene outperformed N1 gene. Scott et al., found that N2 gene targeting assays presented higher precision and reliability in terms of estimating COVID-19 cases in a dormitory on a university campus than N1 gene (Scott et al., 2021). Likewise, Gonzalez et al., demonstrated that N2 gene presented higher sensitivity using

RT-ddPCR (Gonzalez et al., 2020). Notably, the US CDC recommended the N2 gene for SARS-CoV-2 detection since it presented lower mismatches compared to other assays, which was also demonstrated elsewhere (Rahman et al., 2021). Recently, Rao et al., compared N2 and SC2 RT-qPCR and ddPCR assays in terms of the Ct (cycle threshold) values as indirect indicators of viral concentrations and they found that the Ct values for N2 and SC2 were very similar, indicating both assays were equivalent on this basis (Rao et al., 2023).

Despite the aforementioned investigations regarding the comparison among N1, N2 genes and SC2 assay, few studies directly compared their efficiency in correlating with and predicting clinical cases. To the best of our knowledge, published studies have only conducted limited comparative analyses of SARS-CoV-2 RNA concentrations quantified with N1, N2 genes and SC2 assay as well as clinical cases. Hence, during a 12-month study in Detroit, Michigan, USA, we aimed to compare the correlations and examine the similarities/dissimilarities among N1, N2 genes and SC2 assay through comprehensive statistical approaches, establish models to predict COVID-19 cases based on SARS-CoV-2 RNA concentrations quantified with the three assays, thereby identifying the optimal assay for monitoring COVID-19. Specifically, the statistical approaches include Pearson and Spearman's rank correlations, Dynamic Time Warping, and vector autoregressive models. Besides, thorough experimental procedures of determining the Limit of Blank and Limit of Detection for the three assays using RT-ddPCR were summarized.

2. Materials and Methods

2.1 Wastewater treatment facility and sample collection

The Great Lakes Water Authority (GLWA) Water Resource Recovery Facility (WRRF), located in southeastern Michigan, is the second-largest single-site wastewater treatment facility in North America and the largest of its kind in the United States (Norton et al., 2022). The WRRF

serves an area of more than 946 square miles covering the City of Detroit and the three largest counties in Michigan, including Wayne, Macomb and Oakland counties, with a total population of over three million. Three interceptors convey raw wastewater to the WRRF, including the Oakwood-Northwest-Wayne County Interceptor (ONWI), the North Interceptor-East Arm (NIEA), and the Detroit River Interceptor (DRI) (see chapter 1 Figure 1. 2.). Untreated 1 L 24-hour composite wastewater samples were collected weekly from all three interceptors between January 1 and December 31, 2023. A total of 147 samples were collected including 49 samples from each interceptor. The samples were collected using sterilized Nalgene bottles which were then enclosed in sealed plastic bags and placed on ice. The samples were then transported to the Environmental Virology Laboratory at Michigan State University for downstream analyses within 24 hours.

2.2 Sample concentration, RNA extraction, and RT-ddPCR

Samples were concentrated using a previously described polyethylene glycol precipitation (PEG) method (Zhao et al., 2022). PEG was commonly used in concentrating wastewater samples for the detection of SARS-CoV-2 and other viruses, including influenza A, rotavirus, norovirus, measles virus, and human coronavirus (Borchardt et al., 2017; Dimitrakopoulos et al., 2022; Farkas et al., 2022). The concentrated samples were aliquoted into 2 mL tubes and stored at -80 °C for downstream analyses. Viral RNA was extracted using QIAamp Viral RNA QIAGEN kits (QIAGEN, Germantown, MD, USA) based on a previously described method (Miyani et al., 2020, 2021). Bacteriophage Phi6 was utilized as a standard measure to estimate the recovery of RNA during the processes, where the observed recoveries ranged from 10.37% to 58.96% (Ye et al., 2016; Zhao et al., 2022).

The Reverse transcription droplet digital PCR (RT-ddPCR) was performed on a QX200

AutoDG Droplet Digital PCR system, including an Automated Droplet Generator, a PTC Tempo Thermal Cycler, a Droplet Reader, and the QuantaSoft Software (Bio-Rad, Hercules, CA, USA). Primers and probes of N1, N2 genes, and SC2 assay were applied to identify SARS-CoV-2 and quantify the viral RNA concentrations in samples (Table 5. 1.). The One-step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA) was utilized. Briefly, for each sample (5.5 μL), 5.5 μL 1-Step RT-ddPCR Supermix (20x), 2.2 μL Reverse Transcriptase, 1.1 μL 300mM DTT, 3.3 μL N1 and N2 primer-probe mix or 1.1 μL SC2 primer-probe mix were added to the Mastermix, achieving a total of 22 μL for each well. The target primer and probe concentrations for all assays were 900 nM and 250 nM, respectively. Each RT-ddPCR run includes positive controls for SARS-CoV-2, negative controls using nuclease-free water, and process controls of bacteriophage Phi6, as per a previously established method (Li et al., 2022; Zhao et al., 2022, 2023b). Each sample was tested in triplicates by RT-ddPCR, and the three concentrations were averaged, and standard deviations were calculated.

Table 5. 1. Sequences of the primers and probes for the CDC recommended N1, N2, and SC2 assays

Target	Name of primers and probes	Oligonucleotide Sequence (5'>3')	References
N1 gene	2019-nCoV_N1-F	GAC CCC AAA ATC AGC GAA AT	(Lu et al., 2020)
	2019-nCoV_N1-R	TCT GGT TAC TGC CAG TTG AAT CTG	
	2019-nCoV_N1-P	FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1	
N2 gene	2019-nCoV_N2-F	TTA CAA ACA TTG GCC GCA AA	
	2019-nCoV_N2-R	GCG CGA CAT TCC GAA GAA	
	2019-nCoV_N2-P	FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1	
SC2 assay	SC2_F	CTG CAG ATT TGG ATG ATT TCT CC	(Shu et al., 2021; Xu et al., 2021)
	SC2_R	CCT TGT GTG GTC TGC ATG AGT TTA G	
	SC2_P	ATT GCA ACA /TAO/ ATC CAT GAG CAG TGC TGA CTC	

2.3 Determination of Limit of Blank and Limit of Detection

Limit of Blank (LOB) and Limit of Detection (LOD) were determined as per the protocol "A Practical Guide for Evaluating Detection Capability Using Droplet Digital PCR" provided by

the manufacturer Bio-Rad for assessing the analytical sensitivity and validating RT-ddPCR assays. The Bio-Rad protocol was developed as per the Clinical and Laboratory Standards Institute (CLSI) protocol "EP17 Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures" (Pierson-Perry et al., 2012). It is crucial to indicate that our previously published study presented partial discussions of LOB and LOD for N1 and N2 genes (Zhao et al., 2022). However, this study provided a holistic protocol for determining LOB and LOD for all three assays including SC2.

LOB indicates the highest template concentration that can be measured using a particular assay in a blank sample, which should mimic the actual sample matrix as closely as possible without containing the target DNA or RNA sequence, with a defined probability (α) (Armbruster & Pry, 2008). A value of 0.05 for α was chosen, indicating that any unknown measurement would only have a 5 % chance of producing a false positive. Here, four types of samples were selected for LOB tests to represent blank samples, including prior-to-COVID-19 pandemic samples that were collected on February 18, 2018, from the same interceptors, nuclease-free water, negative process control samples from concentration and extraction processes, and autoclaved wastewater samples during the study period. The selection of sample types for LOB tests relied on the Bio-Rad protocol and previous studies (Beattie et al., 2022; Ciesielski et al., 2021; Zhao et al., 2022). LOB was tested across four consecutive days for N1 and N2 genes, and two consecutive days for SC2 assay. This approach scrutinizes unnoticeable impacts caused by tests performed across multiple days. Besides, the CLSI EP17 document recommended a minimum of 60 replicate templates for LOB tests (Pierson-Perry et al., 2012). Here, 96 replicate templates were performed for different types of samples to determine LOB for N1, N2 genes and SC2 assay. The data produced from all LOB tests presented non-normal distributions, which led to the nonparametric

or rank-order method to determine the final LOBs for N1, N2 genes and SC2 assay (Linnet & Kondratovich, 2004; Milbury et al., 2014). Briefly, 96 values were ranked from lowest to highest and the 92nd-position value was determined as the LOB for the current test, which was corresponding to the predetermined probability ($\alpha = 0.05$ or 95 % confidence). Finally, LOBs for N1 gene, N2 gene, and SC2 assay ddPCR were determined as 0.09 gc/μL, 0.08 gc/μL, and 0.13 gc/μL with 95 % confidence (Zhao et al., 2022).

LOD indicates the lowest concentration of the target nucleic acid sequence that can be detected using a particular assay with a desired probability (β), which was set as 0.05 in the current study (Linnet & Kondratovich, 2004; Milbury et al., 2014). This indicates that if a sample with a concentration equal to the LOD is constantly tested, there would be a 95 % chance of getting a positive result. A series of dilutions for SARS-CoV-2 ranging from 10^(-4) and 10^2 gc/μL using N1, N2 genes were performed across nine consecutive days and using SC2 assay across two consecutive days to determine the empirical LOD. Then, 96 replicate templates with the concentration of the empirical LOD were tested following the nonparametric trial-error method (Linnet & Kondratovich, 2004). Briefly, 96 values were ranked from the lowest to highest and if less than 5 % of measurements were below LOB, then the empirical LOD was determined as the final LOD. Otherwise, a higher concentration was selected as the empirical LOD and repeated the tests until the final LOD was determined. Eventually, LODs were determined as 0.1 gc/μL with 72.92 % confidence for the N1 gene ddPCR, 0.1 gc/μL with 81.25 % confidence for the N2 gene ddPCR, and 0.2 gc/μL with 95 % confidence for SC2 ddPCR (Zhao et al., 2022).

2.4 COVID-19 epidemiological data

Daily confirmed COVID-19 cases for the City of Detroit, as well as Wayne, Macomb, and Oakland counties were downloaded and updated on January 10, 2024, from publicly accessible

databases of Michigan (michigan.gov/coronavirus). The daily clinical data encompassed a range from January 1 to December 31, 2023. In addition, the clinical data were only available per city or county for the Detroit tri-county area and each interceptor received wastewater from the corresponding catchment areas of each city or county. Hence, only the total SARS-CoV-2 RNA concentrations could be correlated to the total COVID-19 cases in the city and counties (Zhao et al., 2022, 2024).

2.5 Statistical analyses and visualization

Data collection, organization, and initial data analyses were performed using Microsoft Excel (version 16.82). R version 2023.06.0+421 (R-project.org) was implemented to develop the vector autoregression models and perform statistical analyses including Pearson and Spearman's rank correlations and Dynamic Time Warping. The statistical analyses and visualization relied on R packages, primarily including ggplot2 (Wickham, 2016), ggpubr (Kassambara, 2018), dtw (Giorgino, 2009), tseries (Trapletti et al., 2007), forecast (Hyndman & Khandakar, 2008), vars (Pfaff, 2008). Weekly SARS-CoV-2 RNA concentrations were filled into daily data using linear interpolation, in order to compare with daily COVID-19 cases data (Zhao et al., 2022, 2024).

2.5.1 Pearson and Spearman correlations

The degree of association between total SARS-CoV-2 RNA concentrations in wastewater measured by N1, N2 genes as well as SC2 assay and total confirmed COVID-19 cases were estimated through Pearson correlation and Spearman's rank correlation coefficients. Scatter plots between time series data were created and presented in Figure 5S. 1., including correlation coefficients and significance levels.

2.5.2 Dynamic time warping (DTW)

Dynamic time warping (DTW) is a widely used algorithm that estimates the

similarities/dissimilarities between time series data by calculating and comparing the DTW distance. DTW accounts for the time series data patterns by identifying the most similar and best-matching data points on time series data regarding the shapes of data (Izakian et al., 2015). DTW had been used in numerous fields for time series analyses including time series data mining and clustering (Jeong et al., 2011; H. Li, 2021), time series classification (Kate, 2016), and multivariate time series correlation and dissimilarity analyses (Bankó & Abonyi, 2012), etc. Recently, DTW was utilized in wastewater surveillance of SARS-CoV-2. For instance, researchers applied DTW to quantify and compare similarities between time series data of SARS-CoV-2 RNA concentrations in wastewater and time series data of COVID-19 clinical cases, where the method was demonstrated to be effective in identifying the most similar time series data (Zhao et al., 2023a). Likewise, researchers investigated the similarities of wastewater testing results from varying sampling sites on a college campus using DTW to indicate potential similar trends among sites (Tang et al., 2022). In this study, DTW was employed to analyze SARS-CoV-2 RNA concentrations measured by N1, N2 genes and SC2 assays to identify the similarities/dissimilarities among these targets. The DTW distance was computed in the R packages mentioned above, and the detailed calculations and results were demonstrated in Figure 5S. 2.

2.5.3 Vector autoregressive model (VAR)

Vector autoregressive model (VAR) was one of the most commonly used methods to correlate multivariate time series data such as COVID-19 epidemiological data (Rajab et al., 2022; Zivot & Wang, 2006) and to forecast clinical cases based on wastewater measurements (Cao & Francis, 2021). VAR was proved to be more effective in estimating clinical cases based on SARS-CoV-2 RNA concentrations than other models, such as linear regression and Autoregressive

Integrated Moving Average (ARIMA) models (Zhao et al., 2022). In this study, SARS-CoV-2 RNA concentrations (measured by N1, N2 genes and SC2 assays) and clinical data were modeled as two time series data, and the relationships between them were estimated through VAR models:

$$X_{1,t} = \zeta_1 + \Phi_{11} X_{t-1,1} + \Phi_{12} X_{t-1,2} + \ldots + \omega_{t,1} \quad (1)$$

$$X_{2,t} = \zeta_2 + \Phi_{21} X_{t-1,1} + \Phi_{22} X_{t-1,2} + \ldots + \omega_{t,2} \quad (2)$$

$$X_{3,t} = \zeta_3 + \Phi_{31} X_{t-1,1} + \Phi_{32} X_{t-1,2} + \ldots + \omega_{t,3} \quad (3)$$

The three predicted time series based on N1, N2 genes and SC2 assay were denoted by $X_{1,t}$, $X_{2,t}$, and $X_{3,t}$. $X_{t-1,1}$ denotes the SARS-CoV-2 RNA concentrations at the time when t equals to one lag. $X_{t-1,2}$ denotes the clinical data at the time when t equals to one lag. $\Phi_{n,m}$ and $\omega_{t,n}$ denote constants to adjust the predictions. R squared (R^2) and Root Mean Squared Error (RMSE) values were calculated for the predicted results to evaluate accuracy.

3. Results

3.1 SARS-CoV-2 RNA concentrations and trends measured by the CDC N1, N2, and SC2 assays

Over the course of the entire 12-month study, SARS-CoV-2 RNA concentrations were measured using N1, N2 genes, and SC2 assay RT-ddPCR in untreated wastewater samples collected from the three interceptors at GLWA's WRRF. Figure 5. 1. presents the weekly averaged concentrations (gc/100mL) for N1, N2 genes, and SC2 assay in ONWI, NIEA, and DRI interceptors. SARS-CoV-2 RNA concentrations measured by all three assays reveal a surge beginning the week of January 2 and reaching a peak in the week of January 16 in 2023 for all three interceptors. Then all SARS-CoV-2 RNA concentrations decreased rapidly and fluctuated by the middle of March 2023. Subsequently, the SC2 assay could not detect any positive SARS-CoV-2 concentrations higher than its LOD for consecutive ten weeks until the end of May 2023.

During the summer of 2023 from May to the end of July, SARS-CoV-2 RNA concentrations were detected slightly higher than the LODs for all three assays. Afterwards, all SARS-CoV-2 concentrations fluctuated and experienced significant surges from mid-October to the end of December 2023. Figure 5. 2. demonstrates the total SARS-CoV-2 RNA concentrations measured by N1, N2 genes, and SC2 assay where concentrations below LODs were replaced by LODs as well as the total COVID-19 cases in the Detroit tri-county area. Notably, all SARS-CoV-2 RNA concentrations measured by three assays and the total cases reveal similar trends, especially during winter months of January-to-February and October-to-December 2023 when both wastewater viral concentrations and clinical cases experienced significant surges, as well as during summer months from May to August 2023 when both wastewater viral concentrations and clinical cases demonstrated consistent low numbers (Figure 5. 2.). These results embraced previous studies where researchers observed elevated concentrations of SARS-CoV-2 and other viruses, such as influenza A, respiratory syncytial virus, during winter months (Boehm et al., 2023). This can potentially be explicated by a plethora of contributing factors during winter months, including significantly longer time before depletion of genetic materials in wastewater than that in summers, higher proliferation of seasonal "winter virus" infections, more dense indoor gatherings, weaker immune responses of human due to insufficient daylight, and so on (Hart & Halden, 2020a, 2020b; Moriyama et al., 2020).

3.2 Correlations and similarities among total SARS-CoV-2 RNA concentrations by three assays, and clinical cases

Pearson and Spearman's rank correlations analyses were conducted among the cases and viral concentrations measured by three targets including N1, N2 genes, and SC2 assay (Table 5. 2.). Figure 5S. 1. demonstrated scatter plots for all correlations where correlation coefficients and

significance levels were presented. For correlations among the three assays, the most significant coefficients were observed between N1 and N2 (Pearson r = 0.97, Spearman rho = 0.95, both with p < 2.2e-16). The lowest coefficients were observed between N2 and SC2 (Pearson r = 0.89, Spearman rho = 0.86, both with p < 2.2e-16). Notably, both Pearson's and Spearman's coefficients demonstrated consistently strong correlations among N1, N2 genes, and SC2 assay with desired significance levels. For correlations between viral concentrations and cases, the most significant coefficients were observed between SC2 and cases (Pearson r = 0.62, Spearman rho = 0.77, with p = 1.9e-06, and p = 1.2e-10, respectively) followed by slightly lower correlation coefficients between N1 and cases (Pearson r = 0.61, Spearman rho = 0.71, with p = 2.4e-06, and p = 9.3e-09, respectively).

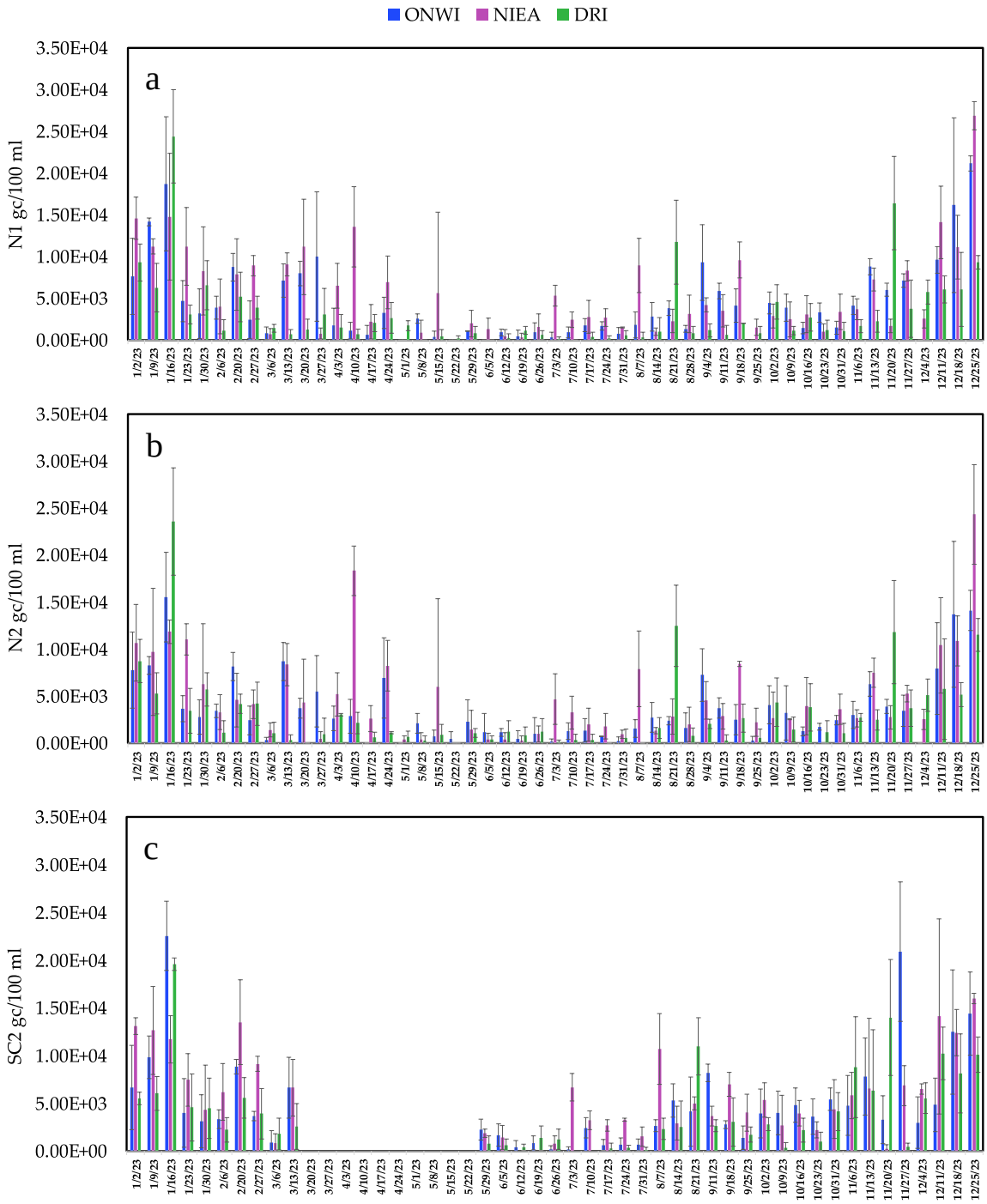


Figure 5. 1. RT-ddPCR concentrations measured by N1 (a), N2 (b), and SC2 (c) assays, in samples collected at the three main interceptors (ONWI, NIEA, DRI) feeding the Great Lakes Water Authority Water Resource Recovery Facility in 2023

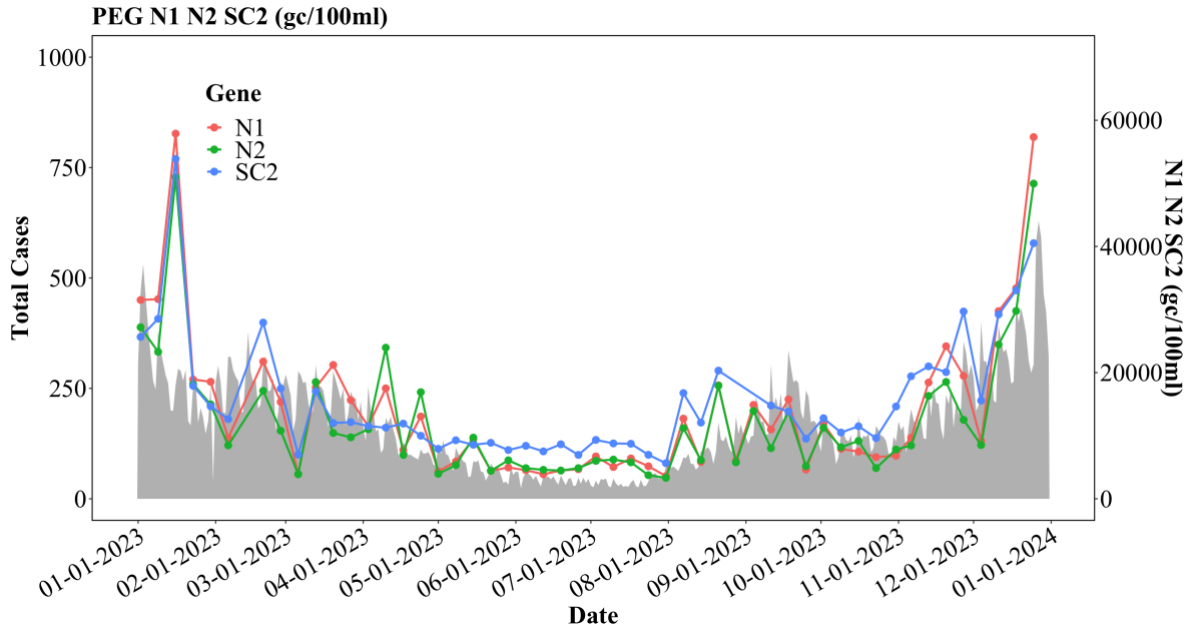


Figure 5. 2. Total (sum of all three interceptors) concentrations of SARS-CoV-2 quantified by the N1, N2, and SC2 RT-ddPCR assays (including LOD); and total confirmed COVID-19 cases in the City of Detroit, Wayne, Macomb, and Oakland counties in 2023

Table 5. 2. Pearson and Spearman correlations among total N1, N2, SC2 SARS-CoV-2 concentrations, and cases

Data	Type of correlation	Correlation coefficients	Significance
N1 and N2	Pearson	0.97	ρ<2.2e-16
	Spearman	0.95	ρ<2.2e-16
N1 and SC2	Pearson	0.93	ρ<2.2e-16
	Spearman	0.89	ρ<2.2e-16
N2 and SC2	Pearson	0.89	ρ<2.2e-16
	Spearman	0.86	ρ<2.2e-16
N1 and cases	Pearson	0.61	ρ=2.4e-06
	Spearman	0.71	ρ=9.3e-09
N2 and cases	Pearson	0.54	ρ=4.7e-05
	Spearman	0.65	ρ=2e-07
SC2 and cases	Pearson	0.62	ρ=1.9e-06
	Spearman	0.77	ρ=1.2e-10

Additionally, similarities of concentrations measured by N1, N2 genes, and SC2 assay were evaluated by DTW distances, which were computed between N1 and N2, N1 and SC2, as well as N2 and SC2 (Zhao et al., 2023a). The details of calculating DTW distances are shown in Figure 5S. 2. The most/least similar time series data patterns were identified as follows. Notably, the smallest DTW distances were observed between N1 and N2, indicating the most similar time

series patterns between the two genes. Conversely, N2 and SC2 presented the least similar time series patterns with the largest DTW distances. These findings from the DTW analyses corroborated the findings from correlation analyses. Overall, the statistical approaches above revealed that N1 and N2 exhibited the highest correlations as well as the most similar time series patterns, N2 and SC2 exhibited the lowest correlations as well as least similar time series patterns.

3.3 Vector autoregressive models potentially indicate the optimal target for estimating COVID-19 cases

Prediction of COVID-19 cases was accomplished through the establishment of vector autoregressive models, utilizing SARS-CoV-2 RNA concentrations measured by N1, N2 genes, and SC2 assay as well as clinical cases, respectively. Figure 5. 3. demonstrated the actual cases and predicted cases based on three assays, where the N2-based prediction of cases (cyan-colored line) closely aligned with the actual cases (red-colored line). This observation was reinforced by robust statistical parameters with the R-squared value of 0.76, and a root mean square error (RMSE) of 94.79, which demonstrated the effectiveness of N2-based predictive models in accurately estimating COVID-19 cases. N1-based prediction yielded stronger statistical parameters with an R-squared value of 0.34, and a RMSE of 199.93 than SC2-based predictions (Table 5S. 1.). The prediction formulas and corresponding statistics are presented in Table 5S. 1. These findings corroborated previously published results. Particularly, Scott et al., identified N2 gene as a more accurate and reliable gene to estimate COVID-19 cases than N1 gene. Moreover, their study also demonstrated that N2 gene alone served as the most effective predictor of COVID-19 cases in a college dormitory (Scott et al., 2021). Similarly, researchers demonstrated that N2 gene exhibited the optimal capability in estimating daily confirmed cases compared to N1 gene. This is potentially attributed to N2-gene's comparatively lower mismatch frequencies than other assays and its

targeting region is less susceptible to mutation (Ai et al., 2021; Rahman et al., 2021).

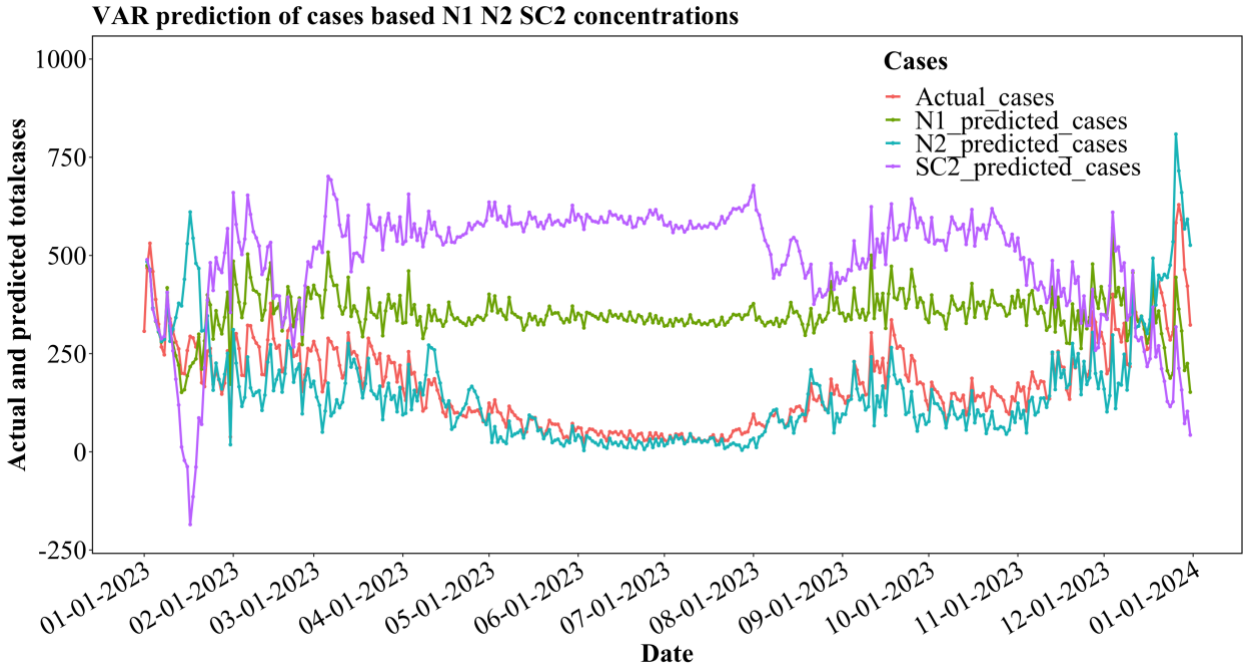


Figure 5. 3. VAR prediction of total cases based on SARS-CoV-2 concentrations

3.4 Limitations and future directions

Admittedly, this study has several limitations. First, this study only implemented, evaluated, and compared three commonly used assays to detect SARS-CoV-2 in wastewater. More research is needed to conduct comparative analyses for multiple assays with higher sensitivity and prevalent usage. For instance, a recent study compared ddPCR results for seven different primer and probe assays such as N1, E, ORF, RdRP, NSP, etc. They recommended parallel detection of multiple assays to increase the robustness of detection for SARS-CoV-2 since coronaviruses demonstrated a high potential for mutations (Ho et al., 2022). Nevertheless, this might not be practical for resource-constrained areas as previously discussed. Besides, future studies are called to investigate the mechanisms behind varying efficiencies when comparing N1, N2, and SC2 assays.

Second, it is critical to recognize that replacing non-detectable concentrations with the

LOD of the method can lead to some uncertain effects, particularly when a substantial portion of data falls below the LOD. However, in the current study, despite SC2 containing the largest portion of undetected concentrations below its LOD between March 13th and May 22nd, 2023, replacing the non-detectable values with LOD did not significantly affect the correlation analyses or prediction results (Tables 5S. 2., 5S. 3.). Using LODs for N1 or N2 did not affect the conclusions, which were also corroborated in our previous studies (Zhao et al., 2022, 2023a). Nevertheless, LOD-replaced data slightly improved the correlation coefficients among N1, N2, and SC2.

Third, this study compared the three CDC assays through examining the correlations of SARS-CoV-2 RNA concentrations in wastewater especially with confirmed cases. Future studies are called on comprehensive analyses of correlations between wastewater viral concentrations and clinical cases encompassing asymptomatic or undercounted infections. Notably, recent studies indicated that the exact percentage of asymptomatic infections in the wastewater sewersheds remains unresolved (Wu et al., 2022) and it can vary significantly during different stages of the pandemic and locations, such as 79.2% asymptomatic infections reported in a local study conducted in Arizona, USA, (Schmitz et al., 2021), and 15.6% asymptomatic infections reported in a meta-analysis of worldwide COVID-19 data (He et al., 2021). Nevertheless, the presence of asymptomatic populations can potentially lead to undercounted COVID-19 cases (He et al., 2021), therefore, certainly affecting the correlations between measured SARS-CoV-2 RNA concentrations and available reported clinical cases.

4. Conclusions

This study presents comprehensive and innovating comparative analyses in evaluating the relationship among SARS-CoV-2 RNA concentrations in wastewater quantified with the N1, N2, and SC2 RT-ddPCR assays, as well as their efficiency in correlating with and predicting clinical

cases. The major conclusions were summarized as follows:

(1) During the twelve-month study, 147 untreated 24-hour composite wastewater samples were collected and analyzed using N1, N2 , and SC2 RT-ddPCR assays. Total SARS-CoV-2 RNA concentrations in positive samples for N1 gene ranged from 3601.33 gc/100mL to 57890.67 gc/100mL, for N2 gene ranged from 3322.67 gc/100mL to 50984 gc/100mL, for SC2 gene ranged from 800 gc/100mL to 53874.67 gc/100mL.

(2) Experimental procedures of determining LOBs and LODs for the three assays were extensively elucidated.

(3) N1 and N2 genes concentrations presented the highest correlation and the most similar time series pattern. Conversely, N2 gene and SC2 assay concentrations presented the lowest correlation and the least similar time series pattern.

(4) N2 gene was identified as the best target to predict COVID-19 cases based on vector autoregressive models.

5. Acknowledgements

We thank the Michigan Department of Health and Human Services (MDHHS) for funding this research.

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tscharke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of the Total Environment*, *728*, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

Ahmed, W., Bivins, A., Smith, W. J. M., Metcalfe, S., Stephens, M., Jennison, A. V., Moore, F. A. J., Bourke, J., Schlebusch, S., McMahon, J., Hewitson, G., Nguyen, S., Barcelon, J., Jackson, G., Mueller, J. F., Ehret, J., Hosegood, I., Tian, W., Wang, H., … Simpson, S. L. (2022). Detection of the Omicron (B.1.1.529) variant of SARS-CoV-2 in aircraft wastewater. *Science of The Total Environment*, *820*, 153171. https://doi.org/10.1016/j.scitotenv.2022.153171

Ai, Y., Davis, A., Jones, D., Lemeshow, S., Tu, H., He, F., Ru, P., Pan, X., Bohrerova, Z., & Lee, J. (2021). Wastewater SARS-CoV-2 monitoring as a community-level COVID-19 trend tracker and variants in Ohio, United States. *Science of The Total Environment*, *801*, 149757. https://doi.org/10.1016/j.scitotenv.2021.149757

Armbruster, D. A., & Pry, T. (2008). Limit of blank, limit of detection and limit of quantitation. *The Clinical Biochemist Reviews*, *29*(Suppl 1), S49.

Bankó, Z., & Abonyi, J. (2012). Correlation based dynamic time warping of multivariate time series. *Expert Systems with Applications*, *39*(17), 12814–12823. https://doi.org/10.1016/j.eswa.2012.05.012

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Beattie, R. E., Blackwood, A. D., Clerkin, T., Dinga, C., & Noble, R. T. (2022). Evaluating the impact of sample storage, handling, and technical ability on the decay and recovery of SARS-CoV-2 in wastewater. *PLOS ONE*, *17*(6), e0270659. https://doi.org/10.1371/journal.pone.0270659

Bivins, A., & Bibby, K. (2021). Wastewater Surveillance during Mass COVID-19 Vaccination on a College Campus. *Environmental Science & Technology Letters*, *8*(9), 792–798. https://doi.org/10.1021/acs.estlett.1c00519

Bivins, A., Kaya, D., Bibby, K., Simpson, S. L., Bustin, S. A., Shanks, O. C., & Ahmed, W. (2021). Variability in RT-qPCR assay parameters indicates unreliable SARS-CoV-2 RNA quantification for wastewater surveillance. *Water Research*, *203*, 117516. https://doi.org/10.1016/j.watres.2021.117516

Boehm, A. B., Wolfe, M. K., White, B. J., Hughes, B., Duong, D., & Bidwell, A. (2023). More than a Tripledemic: Influenza A Virus, Respiratory Syncytial Virus, SARS-CoV-2, and Human Metapneumovirus in Wastewater during Winter 2022–2023. *Environmental Science & Technology Letters*, *10*(8), 622–627. https://doi.org/10.1021/acs.estlett.3c00385

Borchardt, M. A., Spencer, S. K., Hubbard, L. E., Firnstahl, A. D., Stokdyk, J. P., & Kolpin, D. W. (2017). Avian Influenza Virus RNA in Groundwater Wells Supplying Poultry Farms Affected by the 2015 Influenza Outbreak. *Environmental Science & Technology Letters*, *4*(7), 268–272. https://doi.org/10.1021/acs.estlett.7b00128

Calderón-Franco, D., Orschler, L., Lackner, S., Agrawal, S., & Weissbrodt, D. G. (2022). Monitoring SARS-CoV-2 in sewage: Toward sentinels with analytical accuracy. *Science of The Total Environment*, *804*, 150244. https://doi.org/10.1016/j.scitotenv.2021.150244

Cao, Y., & Francis, R. (2021). On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Science of The Total Environment, 786*, 147451. https://doi.org/10.1016/j.scitotenv.2021.147451

Ciesielski, M., Blackwood, D., Clerkin, T., Gonzalez, R., Thompson, H., Larson, A., & Noble, R. (2021). Assessing sensitivity and reproducibility of RT-ddPCR and RT-qPCR for the quantification of SARS-CoV-2 in wastewater. *Journal of Virological Methods*, *297*, 114230. https://doi.org/10.1016/j.jviromet.2021.114230

Dimitrakopoulos, L., Kontou, A., Strati, A., Galani, A., Kostakis, M., Kapes, V., Lianidou, E., Thomaidis, N., & Markou, A. (2022). Evaluation of viral concentration and extraction methods for SARS-CoV-2 recovery from wastewater using droplet digital and quantitative RT-PCR. *Case Studies in Chemical and Environmental Engineering*, *6*, 100224. https://doi.org/https://doi.org/10.1016/j.cscee.2022.100224

Farkas, K., Pellett, C., Alex-Sanders, N., Bridgman, M. T. P., Corbishley, A., Grimsley, J. M. S., Kasprzyk-Hordern, B., Kevill, J. L., Pântea, I., Richardson-O'Neill, I. S., Lambert-Slosarska, K., Woodhall, N., & Jones, D. L. (2022). Comparative Assessment of Filtration- and Precipitation-Based Methods for the Concentration of SARS-CoV-2 and Other Viruses from Wastewater. *Microbiology Spectrum*, *10*(4). https://doi.org/10.1128/spectrum.01102-22

Gentry, Z., Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2023). Wastewater surveillance beyond COVID-19: a ranking system for communicable disease testing in the tri-county Detroit area, Michigan, USA. *Frontiers in Public Health*, *11*, 1178515.

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, *31*(7), 1–24. https://doi.org/10.18637/jss.v031.i07

Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M. H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J., & Gonzalez, D. (2020). COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Research*, *186*, 116296. https://doi.org/10.1016/j.watres.2020.116296

Grube, A. M., Coleman, C. K., LaMontagne, C. D., Miller, M. E., Kothegal, N. P., Holcomb, D. A., Blackwood, A. D., Clerkin, T. J., Serre, M. L., Engel, L. S., Guidry, V. T., Noble, R. T., & Stewart, J. R. (2023). Detection of SARS-CoV-2 RNA in wastewater and comparison to COVID-19 cases in two sewersheds, North Carolina, USA. *Science of The Total Environment*, *858*, 159996. https://doi.org/10.1016/j.scitotenv.2022.159996

Hart, O. E., & Halden, R. U. (2020a). Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of The Total Environment*, *730*, 138875. https://doi.org/10.1016/j.scitotenv.2020.138875

Hart, O. E., & Halden, R. U. (2020b). Modeling wastewater temperature and attenuation of sewage-borne biomarkers globally. *Water Research*, *172*, 115473. https://doi.org/10.1016/j.watres.2020.115473

He, J., Guo, Y., Mao, R., & Zhang, J. (2021). Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of medical virology*, *93*(2), 820-830.

Ho, J., Stange, C., Suhrborg, R., Wurzbacher, C., Drewes, J. E., & Tiehm, A. (2022). SARS-CoV-2 wastewater surveillance in Germany: Long-term RT-digital droplet PCR monitoring, suitability of primer/probe combinations and biomarker stability. *Water Research*, *210*, 117977. https://doi.org/10.1016/j.watres.2021.117977

Hong, P.-Y., Rachmadi, A. T., Mantilla-Calderon, D., Alkahtani, M., Bashawri, Y. M., Al Qarni, H., O'Reilly, K. M., & Zhou, J. (2021). Estimating the minimum number of SARS-CoV-2 infected cases needed to detect viral RNA in wastewater: To what extent of the outbreak can surveillance of wastewater tell us? *Environmental Research*, *195*, 110748. https://doi.org/10.1016/j.envres.2021.110748

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*, 1–22.

Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, *39*, 235–244. https://doi.org/10.1016/j.engappai.2014.12.015

Jeong, Y.-S., Jeong, M. K., & Omitaomu, O. A. (2011). Weighted dynamic time warping for time series classification. *Pattern Recognition*, *44*(9), 2231–2240. https://doi.org/10.1016/j.patcog.2010.09.022

Kassambara, A. (2018). ggpubr:'ggplot2'based publication ready plots. *R Package Version*, 2.

Kate, R. J. (2016). Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, *30*(2), 283–312. https://doi.org/10.1007/s10618-015-0418-x

Lanzarini, N. M., Mannarino, C. F., Ribeiro, A. V. C., Prado, T., Vahia, L. S., Siqueira, M. M.,

Resende, P. C., Quintaes, B. R., & Miagostovich, M. P. (2023). SARS-CoV-2 surveillance-based on municipal solid waste leachate in Brazil. *Environmental Science and Pollution Research*, *30*(25), 67368–67377. https://doi.org/10.1007/s11356-023-27019-9

Li, H. (2021). Time works well: Dynamic time warping based on time weighting for time series data mining. *Information Sciences*, *547*, 592–608. https://doi.org/10.1016/j.ins.2020.08.089

Li, Y., Miyani, B., Faust, R. A., David, R. E., & Xagoraraki, I. (2024). A broad wastewater screening and clinical data surveillance for virus-related diseases in the metropolitan Detroit area in Michigan. *Human Genomics*, *18*(1), 14. https://doi.org/10.1186/s40246-024-00581-0

Li, Y., Miyani, B., Zhao, L., Spooner, M., Gentry, Z., Zou, Y., Rhodes, G., Li, H., Kaye, A., Norton, J., & Xagoraraki, I. (2022). Surveillance of SARS-CoV-2 in nine neighborhood sewersheds in Detroit Tri-County area, United States: Assessing per capita SARS-CoV-2 estimations and COVID-19 incidence. *Science of The Total Environment*, *851*, 158350. https://doi.org/10.1016/j.scitotenv.2022.158350

Linnet, K., & Kondratovich, M. (2004). Partly Nonparametric Approach for Determining the Limit of Detection. *Clinical Chemistry*, *50*(4), 732–740. https://doi.org/10.1373/clinchem.2003.029983

Lu, X., Wang, L., Sakthivel, S. K., Whitaker, B., Murray, J., Kamili, S., Lynch, B., Malapati, L., Burke, S. A., Harcourt, J., Tamin, A., Thornburg, N. J., Villanueva, J. M., & Lindstrom, S. (2020). US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, *26*(8), 1654–1665. https://doi.org/10.3201/eid2608.201246

Milbury, C. A., Zhong, Q., Lin, J., Williams, M., Olson, J., Link, D. R., & Hutchison, B. (2014). Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomolecular Detection and Quantification*, *1*(1), 8–22. https://doi.org/10.1016/j.bdq.2014.08.001

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., Norton, J., & Xagoraraki, I. (2021). Early Warnings of COVID-19 Second Wave in Detroit. *Journal of Environmental Engineering*, *147*(8), 06021004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001907

Moriyama, M., Hugentobler, W. J., & Iwasaki, A. (2020). Seasonality of Respiratory Viral Infections. *Annual Review of Virology*, *7*(1), 83–101. https://doi.org/10.1146/annurev-virology-012420-022445

Norton, J., Gosine, A., Busch, A., Almansa, X. F., Mehrotra, A., Sheets, J., Miyani, B., Zhao, L., & Xagoraraki, I. (2022). Development of a Wastewater Surveillance Program to Protect Public Health in Southeastern Michigan. *Utility Management Conference 2022*.

Nörz, D., Grunwald, M., Olearo, F., Fischer, N., Aepfelbacher, M., Pfefferle, S., & Lütgehetmann, M. (2021). Evaluation of a fully automated high-throughput SARS-CoV-2 multiplex qPCR assay with built-in screening functionality for del-HV69/70- and N501Y variants such as B.1.1.7. *Journal of Clinical Virology*, *141*, 104894. https://doi.org/10.1016/j.jcv.2021.104894

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.

Pierson-Perry, J. F., Vaks, J. E., Vore, T. E. K., Durham, A. P., Fischer, C., Gutenbrunner, C., Hillyard, D., Kondratovich, M. V, Ladwig, P., & Middleberg, R. A. (2012). *Evaluation of detection capability for clinical laboratory measurement procedures; approved guideline*. Clinical Laboratory Standards Institute.

Rahman, M. S., Islam, M. R., Alam, A. S. M. R. U., Islam, I., Hoque, M. N., Akter, S., Rahaman, Md. M., Sultana, M., & Hossain, M. A. (2021). Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *Journal of Medical Virology*, *93*(4), 2177–2195. https://doi.org/10.1002/jmv.26626

Rajab, K., Kamalov, F., & Cherukuri, A. K. (2022). Forecasting COVID-19: Vector Autoregression-Based Model. *Arabian Journal for Science and Engineering*, *47*(6), 6851–6860. https://doi.org/10.1007/s13369-021-06526-2

Rao, A., Lin, J., Parsons, R., Greenleaf, M., Westbrook, A., Lai, E., Bowers, H. B., McClendon, K., O'Sick, W., & Baugh, T. (2023). Standardization and Comparison of Emergency Use Authorized COVID-19 Assays and Testing Laboratories. *MedRxiv*.

Saththasivam, J., El-Malah, S. S., Gomez, T. A., Jabbar, K. A., Remanan, R., Krishnankutty, A. K., Ogunbiyi, O., Rasool, K., Ashhab, S., Rashkeev, S., Bensaad, M., Ahmed, A. A., Mohamoud, Y. A., Malek, J. A., Abu Raddad, L. J., Jeremijenko, A., Abu Halaweh, H. A., Lawler, J., & Mahmoud, K. A. (2021). COVID-19 (SARS-CoV-2) outbreak monitoring using wastewater-based epidemiology in Qatar. *Science of The Total Environment*, *774*, 145608. https://doi.org/10.1016/j.scitotenv.2021.145608

Schmitz, B. W., Innes, G. K., Prasek, S. M., Betancourt, W. Q., Stark, E. R., Foster, A. R., ... & Pepper, I. L. (2021). Enumerating asymptomatic COVID-19 cases and estimating SARS-CoV-2 fecal shedding rates via wastewater-based epidemiology. *Science of the Total Environment*, *801*, 149794.

Scott, L. C., Aubee, A., Babahaji, L., Vigil, K., Tims, S., & Aw, T. G. (2021). Targeted wastewater surveillance of SARS-CoV-2 on a university campus for COVID-19 outbreak detection and mitigation. *Environmental Research*, *200*, 111374. https://doi.org/10.1016/j.envres.2021.111374

Shah, S., Gwee, S. X. W., Ng, J. Q. X., Lau, N., Koh, J., & Pang, J. (2022). Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Science of The Total Environment*, *804*, 150060. https://doi.org/10.1016/j.scitotenv.2021.150060

Shu, B., Kirby, M. K., Davis, W. G., Warnes, C., Liddell, J., Liu, J., Wu, K.-H., Hassell, N., Benitez, A. J., Wilson, M. M., Keller, M. W., Rambo-Martin, B. L., Camara, Y., Winter, J., Kondor, R. J., Zhou, B., Spies, S., Rose, L. E., Winchell, J. M., … Barnes, J. R. (2021). Multiplex Real-Time Reverse Transcription PCR for Influenza A Virus, Influenza B Virus, and Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, *27*(7), 1821–1830. https://doi.org/10.3201/eid2707.210462

Tamáš, M., Potocarova, A., Konecna, B., Klucar, Ľ., & Mackulak, T. (2022). Wastewater Sequencing—An Innovative Method for Variant Monitoring of SARS-CoV-2 in Populations. *International Journal of Environmental Research and Public Health*, *19*(15), 9749. https://doi.org/10.3390/ijerph19159749

Tang, W., Chen, T., Slocum, Z., Lan, Y., Delmelle, E., Chen, D., Mittal, N., Rice-Boayue, J., Shukla, T., Lin, S., Akella, S., Schlueter, J., Munir, M., & Gibas, C. (2022). A Web-based Spatial Decision Support System of Wastewater Surveillance for COVID-19 Monitoring: A Case Study of a University Campus. *MedRxiv*, 2021.12.29.21268516. https://doi.org/10.1101/2021.12.29.21268516

Trapletti, A., Hornik, K., & LeBaron, B. (2007). tseries: Time series analysis and computational finance. *R Package Version 0.10-11*.

Vogels, C. B. F., Brito, A. F., Wyllie, A. L., Fauver, J. R., Ott, I. M., Kalinich, C. C., Petrone, M. E., Casanovas-Massana, A., Catherine Muenker, M., Moore, A. J., Klein, J., Lu, P., Lu-Culligan, A., Jiang, X., Kim, D. J., Kudo, E., Mao, T., Moriyama, M., Oh, J. E., … Grubaugh, N. D. (2020). Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology*, *5*(10), 1299–1305. https://doi.org/10.1038/s41564-020-0761-6

Wickham, H. (2016). *Data analysis*. Springer.

Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., ... & Alm, E. J. (2022). SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. *Science of The Total Environment*, *805*, 150121.

Xagoraraki, I. (2020). Can We Predict Viral Outbreaks Using Wastewater Surveillance? *Journal of Environmental Engineering*, *146*(11), 01820003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001831

Xu, Y., Borgaro, J. G., Ren, G., Vrtis, K. B., Patton, G. C., & Nichols, N. M. (2021). Optimized conditions for the CDC Influenza SARS-CoV-2 (Flu SC2) Multiplex Assay using Luna® One-Step RT-qPCR Reagents. *Rn*, *107*(106), 106.

Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science and Technology*, *50*(10), 5077–5085. https://doi.org/10.1021/acs.est.6b00876

Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2024). Tracking the Time Lag between SARS-CoV-2 Wastewater Concentrations and Three COVID-19 Clinical Metrics: A

21-Month Case Study in the Tricounty Detroit Area, Michigan. *Journal of Environmental Engineering*, *150*(1). https://doi.org/10.1061/JOEEDU.EEENG-7509

Zhao, L., Geng, Q., Corchis-Scott, R., McKay, R. M., Norton, J., & Xagoraraki, I. (2023a). Targeting a free viral fraction enhances the early alert potential of wastewater surveillance for SARS-CoV-2: a methods comparison spanning the transition between delta and omicron variants in a large urban center. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1140441

Zhao, L., Zou, Y., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Norton, J., & Xagoraraki, I. (2023b). Simple methods for early warnings of COVID-19 surges: Lessons learned from 21 months of wastewater and clinical data collection in Detroit, Michigan, United States. *Science of The Total Environment*, *864*, 161152. https://doi.org/10.1016/j.scitotenv.2022.161152

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarland, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, 157040. https://doi.org/10.1016/j.scitotenv.2022.157040

Zivot, E., & Wang, J. (2006). Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-PLUS®*, 385–429.

APPENDIX

Table 5S. 1. VAR models based N1, N2 genes and SC2 assay

VAR models	Formula	R^2	RMSE
N1 predicted cases	-0.0039N1 + 0.9242Cases – 0.0014N1 – 0.4439Cases + 344.7230	0.34	218.68
N2 predicted cases	-0.0045N2 + 0.8711Cases + 0.0143N2 – 0.4800Cases – 42.3500	0.76	68.03
SC2 predicted cases	0.0093SC2 + 0.8262Cases – 0.0273SC2 – 0.5153Cases + 722.6044	0.11	394.95

Table 5S. 2. Pearson and Spearman correlations among total N1, N2, SC2 SARS-CoV-2 concentrations (without LODs), and cases

Data	Type of correlation	Correlation coefficients	Significance
N1 and N2	Pearson	0.93	ρ<2.2e-16
	Spearman	0.94	ρ<2.2e-16
N1 and SC2	Pearson	0.78	ρ=1.5e-10
	Spearman	0.66	ρ=5.2e-07
N2 and SC2	Pearson	0.72	ρ=9.7e-09
	Spearman	0.65	ρ=8.3e-07
N1 and cases	Pearson	0.77	ρ=1.3e-10
	Spearman	0.73	ρ=2.8e-09
N2 and cases	Pearson	0.68	ρ=7.2e-08
	Spearman	0.65	ρ=4e-07
SC2 and cases	Pearson	0.7	ρ=4.8e-08
	Spearman	0.6	ρ=9e-06

Table 5S. 3. VAR models for prediction of COVID-19 cases based N1, N2 genes and SC2 assay (without LODs)

VAR models	Formula	R^2	RMSE
N1 predicted cases	-0.0038N1 + 0.9241Cases – 0.0013N1 – 0.4439Cases + 344.7229	0.29	225.00
N2 predicted cases	-0.0045N2 + 0.8710Cases + 0.0143N2 – 0.4800Cases – 42.3499	0.76	73.30
SC2 predicted cases	0.0114SC2 + 0.8101Cases – 0.0304SC2 – 0.5299Cases + 754.7753	0.19	489.03

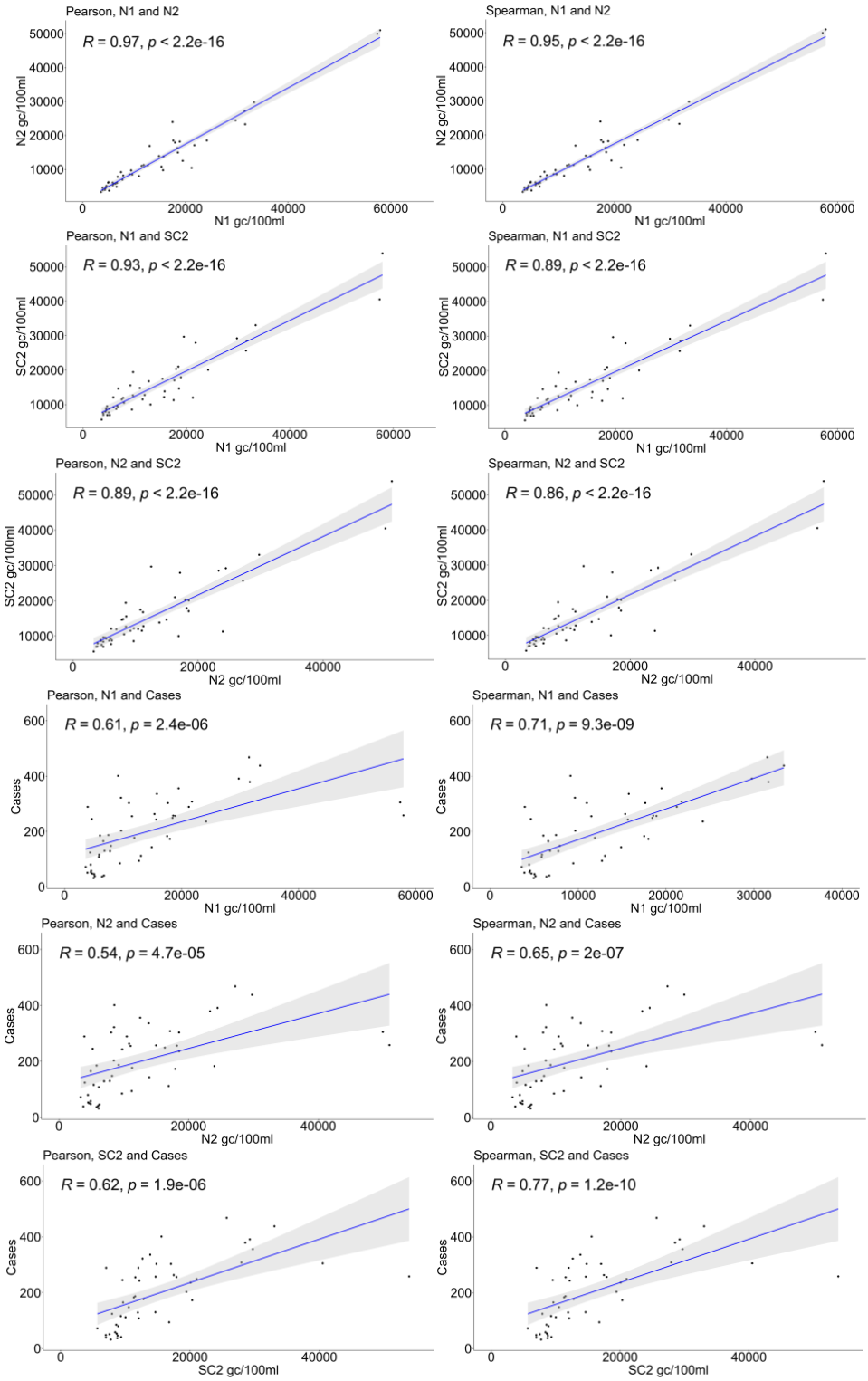


Figure 5S. 1. Scatter plots for Pearson and Spearman correlations among total N1, N2, SC2 concentrations, and cases

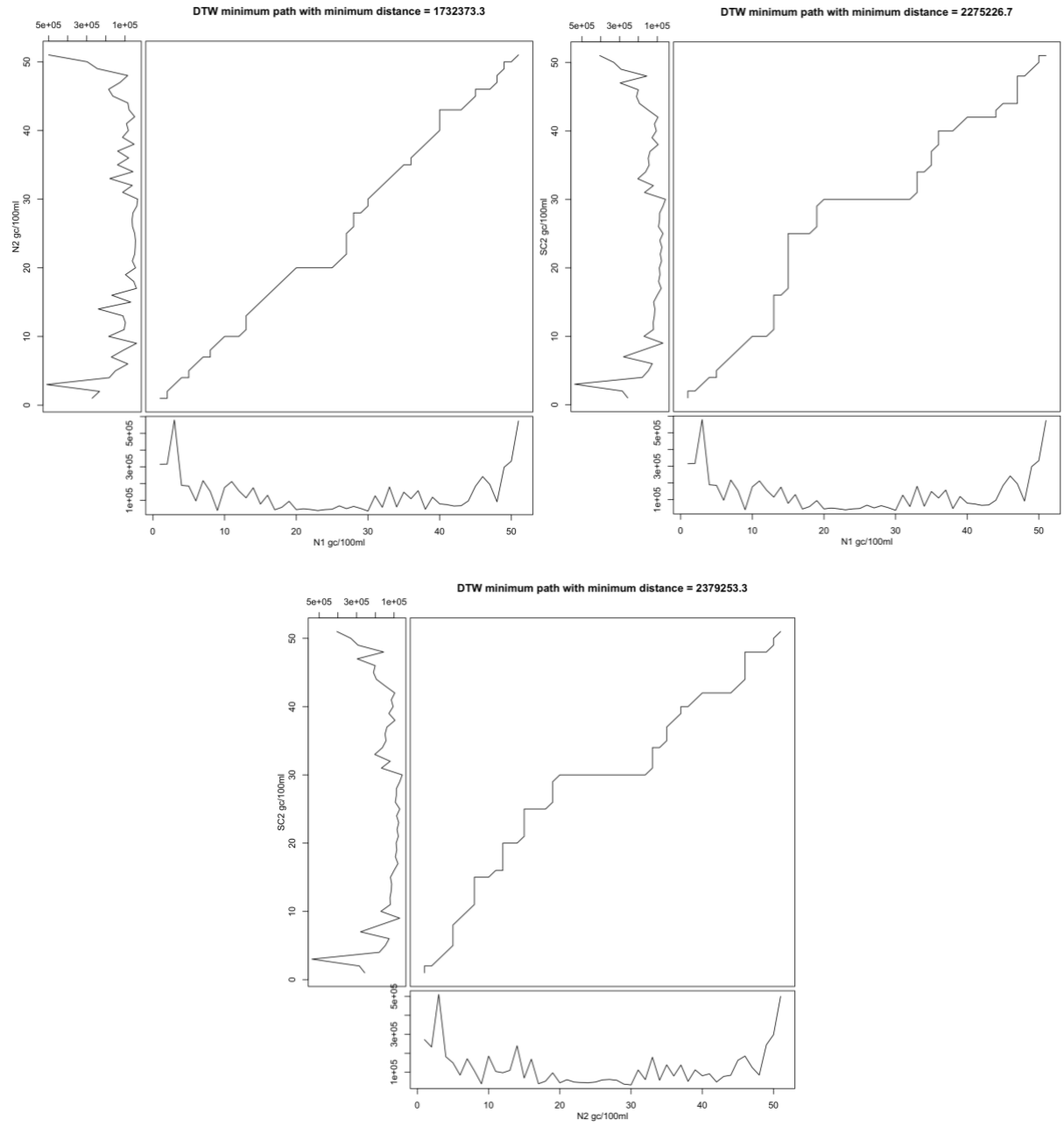


Figure 5S. 2. DTW distances among SARS-CoV-2 concentrations measured by N1, N2 genes and SC2 assays

CHAPTER 6: TRACKING CHLAMYDIA AND SYPHILIS IN THE DETROIT METRO AREA BY MOLECULAR ANALYSIS OF ENVIRONMENTAL SAMPLES

Submitted for publication: Liang Zhao, Heidy Peidro Guzman, Irene Xagoraraki

Abstract

This paper describes one of the first studies applying wastewater surveillance to monitor Chlamydia and Syphilis and back-estimate their infections based on bacterial shedding and wastewater surveillance data. Molecular biology laboratory methods were optimized, and a workflow was designed to implement wastewater surveillance tracking Chlamydia and Syphilis in the Detroit metro area (DMA), one of the most populous metropolitans in the U.S. Untreated composite wastewater samples were collected weekly from three interceptors at Great Lakes Water Authority that services the DMA, and from street manholes that service three neighborhood sewersheds in Wayne, Macomb, and Oakland counties. Centrifugation, DNA extraction, and ddPCR methods were performed and optimized targeting *Chlamydia trachomatis* and *Treponema pallidum* that cause Chlamydia and Syphilis, respectively. Limit of Blank and Limit of Detection were determined experimentally for both targets. Both targets were detected and monitored in wastewater between December 25th, 2023, and April 22nd, 2024. The magnitude of *C. trachomatis* and *T. pallidum* concentrations were observed higher in neighborhood sewersheds compared to interceptors. Infections of Chlamydia and Syphilis were back-estimated through an optimized formula based on shedding dynamics and wastewater surveillance data, which indicated potentially underreported conditions with the clinical data as a benchmark.

1. Introduction

Wastewater surveillance or wastewater-based epidemiology (WBE) has experienced significant advancements since the onset of the COVID-19 pandemic. Studies have shown that most

respiratory, enteric, vector-borne, and bloodborne disease pathogens can be detected in wastewater and other environmental samples (Gentry et al., 2023; McCall et al., 2020; O'Brien & Xagoraraki, 2019; Xagoraraki & O'Brien, 2020). Numerous investigations have implemented wastewater surveillance to monitor fluctuations of SARS-CoV-2 and explored its applications in comprehensive aspects (Barua et al., 2022; Beattie et al., 2022; Li et al., 2022a, 2022b, 2024; McCall et al., 2022; Miyani et al., 2020, 2021; Tiwari et al., 2022; Zhao et al., 2022, 2023a, 2023b). Recently, wastewater surveillance has also become recognized as an effective method for monitoring other viral diseases beyond COVID-19, such as norovirus, respiratory syncytial virus (RSV), and influenza viruses (Ammerman et al., 2024; Mercier et al., 2022, 2023). Despite significant technological, methodological, and translational advancements in wastewater surveillance, its applications have been largely limited to monitoring viral communicable diseases encompassing adenovirus, astrovirus, enterovirus, viral hepatitis, rotavirus, poliovirus, norovirus, etc., which were summarized in a systematic review (Kilaru et al., 2023). Only two bacterial targets, *Escherichia coli* and *Salmonella*, were identified for wastewater surveillance in Kilaru et al.'s study (Kilaru et al., 2023; Philo et al., 2023). Few studies have yet explored applications of wastewater surveillance on monitoring bacterial communicable diseases. Notably, researchers monitored concentrations of *Salmonella* in municipal wastewater in Hawaii, U.S., and demonstrated positive correlations between *Salmonella* concentrations and clinical cases of Salmonellosis (Yan et al., 2018). Yan et al., also observed large fluctuations and outliers of *Salmonella* concentrations that presented significant challenges and uncertainties for bacterial testing in wastewater. Likewise, Matrajt et al., identified environmental surveillance methods to monitor *Salmonella* Typhi and *Salmonella* Paratyphi which caused typhoid fever (Matrajt et al., 2020). Other researchers also conducted surveys to identify next targets for bacterial monitoring

such as *E. coli*, *Enterococci*, and *Enterobacteriaceae* in water environments (Liguori et al., 2022). To date, only one study was identified to monitor *Chlamydia trachomatis* in wastewater on a Florida's university campus (Chin Quee, 2023). Philo et al., identified four major challenges facing bacterial wastewater surveillance including prioritizing new bacterial targets, establishing relationships between wastewater data and human infections, designing and developing methodologies, as well as normalizing bacterial wastewater data (Philo et al., 2023). This study presented one of the first investigations using wastewater surveillance to monitor STIs and provided some specific solutions to the challenges and research gaps proposed by researchers previously regarding wastewater bacterial surveillance (Philo et al., 2023), including selecting new bacterial targets, relating bacterial wastewater data to human infections, and developing sampling and analytical methodologies. Recently, we developed a ranking system that prioritizes next wastewater surveillance targets among 96 communicable diseases for the Detroit metro area (Gentry et al., 2023). Among them, sexually transmitted infections (STIs) particularly Chlamydia and Syphilis were prioritized as the top 5th and 7th communicable diseases to be monitored using wastewater surveillance (Gentry et al., 2023). Chlamydia and Syphilis were caused by gram-negative bacteria, *Chlamydia trachomatis* and *Treponema pallidum*, respectively. In 2020, the World Health Organization (WHO) estimated new Chlamydia and Syphilis infections as 128.5 million and 7.1 million, respectively (WHO, 2021). In the U.S., about 1.6 million Chlamydia cases and nearly a quarter million Syphilis cases were reported in 2022 (cdc.gov).

Sexually transmitted infections (STIs) infections have been rapidly increasing in the U.S. especially since the COVID-19 pandemic. First, the COVID-19 pandemic significantly reduced staff, supplies of testing materials and medications, laboratory capacity, access to STIs services, and surveillance activities in STIs programs, which contributed to delays in STIs diagnosis and

treatment with concomitant increases in STIs transmissions and incidences (Johnson et al., 2021; Wright et al., 2022). Second, researchers reported STIs patients might be reluctant to seek testing due to fear of COVID-19 exposure, leading to underreported STIs (Johnson et al., 2021). Shifted focus and funds from STIs to COVID-19 programs also contributed to underdiagnosis or underreporting. In the State of Michigan and Detroit metro area, Chlamydia have been reported with high number of infections annually and Syphilis infections have experienced significant increases between 2013 and 2023 shown in Figure 6. 1. In particular, the confirmed incidences of Syphilis in Michigan and Detroit metro area surged during the COVID-19 pandemic from 2020 to 2021. More importantly, both Chlamydia and Syphilis has reportedly been underdiagnosed and underestimated. Currently, self-testing, opportunistic testing, and clinical testing are the commonly implemented methods to monitor Chlamydia within populations (Chin Quee, 2023). However, these monitoring methods of Chlamydia are unable to detect the majority of infections since few infected populations are likely to seek testing in clinical settings (Chin Quee, 2023). Vulnerable populations, particularly STIs infected individuals, unlikely undergo testing themselves due to lack of education and privacy (Blake et al., 2003). Balfe et al. reported the following factors including the cost of Chlamydia testing, inconvenient STIs services, long waiting times, and stigma related to STIs contributed to the difficulty for testing (Balfe et al., 2010). Likewise, Syphilis cases are potentially underreported for similar reasons (Shockman et al., 2014). To our best knowledge, limited studies yet investigated wastewater surveillance of both *C. trachomatis* and *T. pallidum* in a major metropolitan area nor to perform back-estimation of Chlamydia and Syphilis infections from bacterial shedding and wastewater measurements.

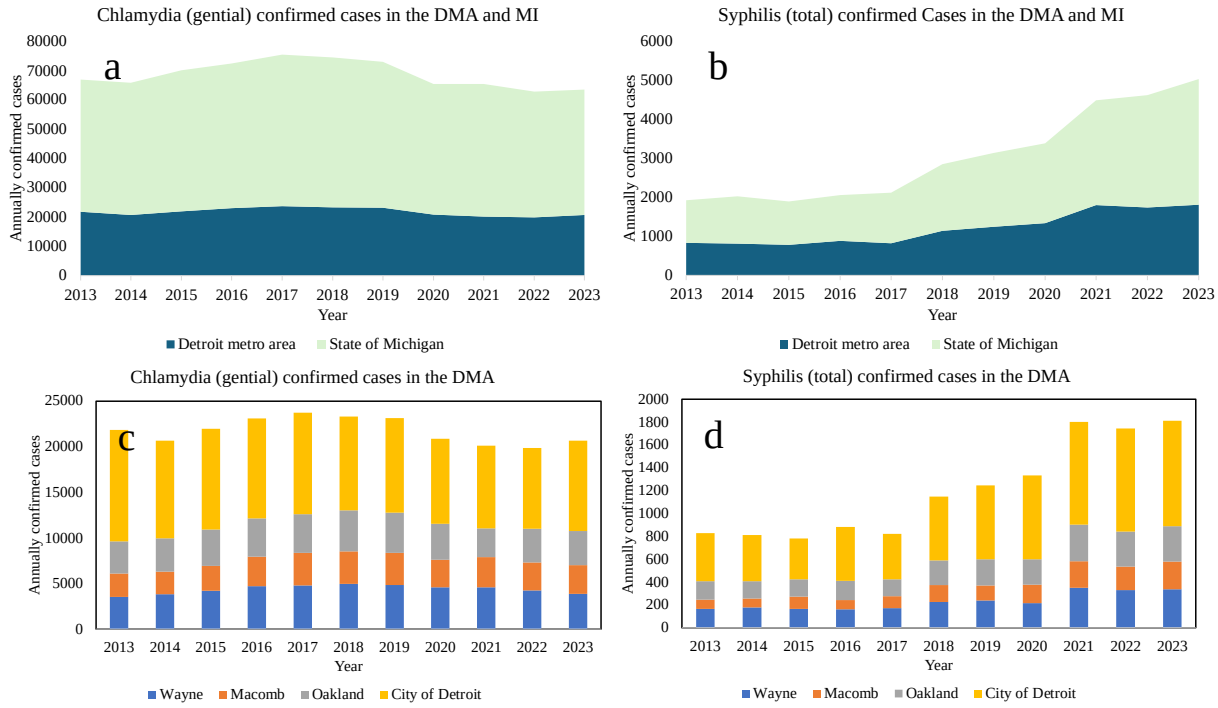


Figure 6. 1. Annually confirmed cases of Chlamydia (genital) and Syphilis (total) in the entire Detroit metro area and State of Michigan (a, b), City of Detroit, as well as Wayne, Macomb, and Oakland counties (c, d) between 2013 and 2023

In this study, untreated wastewater samples were collected weekly from three interceptors of the Great Lakes Water Authority (GLWA) in southeastern Michigan, which service the City of Detroit, Wayne, Macomb, and Oakland counties, as well as from street manholes which service three smaller neighborhood sewersheds (EP in the Macomb county, D3 in the Wayne county, and OP in the Oakland county), between December 25th, 2023, and April 22nd, 2024. We developed a workflow encompassing sampling, centrifugation, DNA extraction, and droplet digital PCR to quantify and monitor *C. trachomatis* and *T. pallidum* concentrations in wastewater samples. We optimized PCR assays, Mastermix, and thermocycling conditions targeting *C. trachomatis* and *T. pallidum*. We also investigated the bacterial shedding of *C. trachomatis* and *T. pallidum* and optimized a formula for estimating Chlamydia and Syphilis infections from bacterial shedding and wastewater measurements. Our results demonstrated the first wastewater surveillance study for bacterial STIs particularly Chlamydia and Syphilis where the bacterial wastewater surveillance

was identified as a screening tool, which can be followed by more targeted clinical testing in communities. The back-estimations of Chlamydia and Syphilis infections from bacterial shedding and wastewater measurements indicated likely underreported cases, which further enhanced the significance of bacterial wastewater surveillance on monitoring STIs.

2. Materials and Methods

2.1 Positive controls

Genomic DNA from *C. trachomatis* serovar D strain UW-3/Cx (ATCC VR-885D) was obtained from ATCC (Manassas, VA, USA). Quantitative synthetic *T. pallidum* DNA (ATCC BAA-2642SD) was obtained from ATCC (Manassas, VA, USA). Both products were immediately stored in -80°C upon arrival. The gene copy numbers of *C. trachomatis* and *T. pallidum* standard controls were determined experimentally. We thawed both vials on ice and avoided as many freeze-thaw cycles as possible to circumvent degradation of their DNA and variation in copy numbers by aliquoting the materials. Prior to experiments, we gently homogenized the vials to ensure uniform distribution of the materials and briefly centrifuge the vials to ensure all liquids are at the bottom.

2.2 Epidemiological data

Both annually and weekly reported Chlamydia and Syphilis cases for the Detroit metro area (DMA), including the City of Detroit, as well as Wayne, Macomb, and Oakland counties, were downloaded from publicly available databases of the Michigan Disease Surveillance System (MDSS) Weekly Surveillance Reports (WSR) (michigan.gov). The Morbidity and Mortality Weekly Report (MMWR) week was established by the U.S. CDC from Sunday to Saturday and is given a sequentially increasing number from the first week in January annually. The annually reported cases of Chlamydia and Syphilis for each jurisdiction were obtained from the annual

WSR between 2013 and 2023 for the MMWR week 52, which included the total cases of each disease in the year. The weekly reported cases of Chlamydia and Syphilis encompassed a range from the MMWR week 52 of 2023 (the week of December 25th, 2023) to the MMWR week 17 of 2024 (the week of April 22nd, 2024).

The Chlamydia cases reported in the WSR were denoted as "genital". The Syphilis cases reported in the WSR included nine reportable conditions in Michigan, including congenital, early latent, late latent, latent of unknow duration, late with manifestations, primary, secondary, to be determined, and unknown duration or late syphilis. And we incorporated the total cases of all reportable conditions of Syphilis in this study, which was denoted as "Syphilis (total)". In addition, the reported data were only available for each city or country in the Detroit metro area. Each interceptor of GLWA received wastewater from the corresponding sewersheds of each city or county. Epidemiological data of Chlamydia and Syphilis for smaller jurisdictions such as zip-code areas are unavailable for the study area. Hence, only the total *C. trachomatis* and *T. pallidum* concentrations of three interceptors can be compared to the total Chlamydia (genital) and Syphilis (total) cases in the Detroit metro area. The estimated infections from *C. trachomatis* and *T. pallidum* concentrations of the ONWI interceptor, the NIEA interceptor, and the DRI interceptor can be compared to confirmed cases for approximate regions of Wayne county, combined Oakland and Macomb county, and the City of Detroit, respectively.

2.3 Sampling locations and sample collection

Sampling locations encompass three interceptors in GLWA's Water Resource Recovery Facility (WRRF) and street manholes covering three neighborhood sewersheds in Wayne, Macomb and Oakland counties. The three GLWA WRRF interceptors service an area of more than 946 square miles covering the City of Detroit and the three most populous counties in

Michigan, including Wayne, Macomb and Oakland counties, with a total population of approximately three million. Three interceptors encompass the Oakwood-Northwest-Wayne County Interceptor (ONWI), the North Interceptor-East Arm (NIEA), and the Detroit River Interceptor (DRI), which services inhabitants of approximately 840600, 1482000, and 492000, respectively, by 2020 (Miyani et al., 2021; Zhao et al., 2024). The WRRF consists of a semi-combined system that collects and treats stormwater together with residential, industrial, and commercial waste, according to specific service areas (Zhao et al., 2022). Samplings at street manholes cover three neighborhood sewersheds, including East Point (EP, ZIP code: 48021) located in the Macomb county, D3 (ZIP code: 48235) located in the Wayne county, and Oak Park (OP, ZIP code: 48237) located in the Oakland county, with covered populations of 2400, 1300, and 2270 (Li et al., 2022b). A total of 108 untreated 1 L 24-hour composite wastewater samples were collected weekly from three interceptors at GLWA and street manholes of three neighborhood sewersheds in the Detroit metro area between December 25th, 2023, and April 22nd, 2024. The samples were collected using sterilized Nalgene bottles which were then enclosed in sealed plastic bags and placed on ice. The samples were then transported to the Environmental Virology Laboratory at Michigan State University for downstream analyses within 24 hours.

2.4 Centrifugation, DNA extraction, and droplet digital PCR

Samples were concentrated using a modified centrifugation method, where 1 L wastewater samples were concentrated at 12,000×*g* for 40 minutes (Fu et al., 2020; Mania-Pramanik et al., 2006; Somboonna et al., 2018; Varma et al., 2009). The concentrated samples were stored at -80 °C for downstream analyses. Bacterial DNA was extracted from the pellets using the QIAGEN DNeasy PowerLyzer PowerSoil Kit (12855-50) according to the manufacturer's protocol with slight modifications (QIAGEN, Germantown, MD, USA). As per the protocol, the mass of the

pellets was controlled approximately as 0.25 grams and was recorded for downstream calculations. A Vortex Adapter for 24 tubes (13000-V1-24) (QIAGEN, Germantown, MD, USA) was utilized at the vortex's maximum speed for 20 minutes to achieve the bead beating process. The same QIAGEN kit was utilized for extracting *C. trachomatis* DNA from wastewater samples previously and was proven to be effective and efficient (Chin Quee, 2023). Finally, 100 µL bacterial DNA was extracted and stored in -80°C for downstream analyses.

A QX200 AutoDG Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA) was employed to perform ddPCR. Primers and probes targeting *ompA* and *polA* nucleotide sequences were applied to identify *C. trachomatis* and *T. pallidum*, respectively, and quantify their DNA concentrations in wastewater samples (Table 6. 1.). We implemented the same primers and probes to identify and quantify *C. trachomatis* and *T. pallidum* as used in previous studies (Heymans et al., 2010; Koek et al., 2006; Nieuwenburg et al., 2022; Salle et al., 2023; Stevens et al., 2010). Optimized Mastermix reaction and thermocycling conditions were demonstrated in Tables 6. 2., and 6. 3., with a total of 22 µL for each well on the 96-well plate. The target primer and probe concentrations for both assays were determined as 900 nM and 250 nM, respectively. Each run includes positive controls for *C. trachomatis* and *T. pallidum*, and negative controls using nuclease-free water.

Table 6. 1. Sequences of the primers and probes for targeting *C. trachomatis* and *T. pallidum*

Organism	Target gene	Oligonucleotide Sequence (5'>3')	References
C. trachomatis	*ompA*	Forward: CATGARTGGCAAGCAAGTTTA	(Chin Quee, 2023; Stevens et al., 2010)
		Reverse: GCAATACCGCAAGATTTTCTAG	
		Probe: FAM-TGTTCACTCCYTACATTGGAGT-BHQ1	
T. pallidum	*polA*	Forward: GGTAGAAGGGAGGGCTAGTA	(Heymans et al., 2010; Koek et al., 2006; Nieuwenburg et al., 2022)
		Reverse: CTAAGATCTCTATTTTCTATAGGTATGG	
		Probe: FAM-ACACAGCACTCGTCTTCAACTCC-BHQ1	

Table 6. 2. Optimized ddPCR Mastermix reaction for both *C. trachomatis* and *T. pallidum*

Reagents	Volume (µL)
Standard solution of the Bio-Rad Kit	8.8
Forward primer	1.98
Reverse primer	1.98
Probe	1.1
PCR-grade water	2.64
Extracted DNA	5.5

Table 6. 3. Optimized thermocycling conditions

Conditions for *C. trachomatis*	Conditions for *T. pallidum*
Standard conditions of the Bio-Rad Kit: 25°C for 3 min, 50°C for 60 min, 95°C for 10 min	
55 Cycles* of 95°C for 10 sec*, 55°C for 20 sec*, 65°C for 40 sec*, and 40°C for 10 sec (Stevens et al., 2010)	50 Cycles* of 95°C for 30 sec*, 55°C for 30 sec*, 72°C for 30 sec* (Koek et al., 2006)
98°C for 10 min	
Hold at 4°C for ∞	

Note: *No hot start, 40 µL reaction, and slow ramp speed of 2°C/second are required for all.

2.5 Limit of Blank and Limit of Detection

Limit of Blank (LOB) and Limit of Detection (LOD) were determined experimentally for evaluating the analytical sensitivity and validating *C. trachomatis* and *T. pallidum* assays, according to the manufacturer's protocol (Bio-Rad, Hercules, CA, USA). Two types of samples were determined to represent blank samples in LOB tests, including nuclease-free water and autoclaved wastewater samples during the study period. The selection of sample types for LOB tests were predicated on the Bio-Rad protocol and previous studies (Barua et al., 2022; Beattie et al., 2022; Zhao et al., 2022). LOB was tested across two consecutive days for *C. trachomatis* and *T. pallidum* assays, respectively. The multiple-day testing approach examines any subtle impacts caused by tests conducted on different days (Zhao et al., 2022). The nonparametric or rank-order method listed in the manufacture's protocol was chosen to further calculate LOBs since the data generated from all LOB tests demonstrated non-normal distributions. Henceforth, 96 values were ranked from the lowest to the highest and the value at the 92nd-position (95% confidence) was

determined as the LOB for the tested assay (Linnet & Kondratovich, 2004; Milbury et al., 2014). Ultimately, LOBs for *C. trachomatis* and *T. pallidum* assays were determined as 0.07 gc/μL and 0.08 gc/μL, respectively, with 95 % confidence.

A sequence of dilutions for positive controls of *C. trachomatis* and *T. pallidum* ranging from 10^5 and 10^{-1} gc/μL were performed. Subsequently, 96 replicate templates with the concentration of each dilution were tested using the *C. trachomatis* and *T. pallidum* assays following the nonparametric trial-error method to determine LODs as per the manufacture's protocol (Linnet & Kondratovich, 2004). Briefly, 96 values were ranked from the lowest to the highest, where the template concentration was determined as the LOD if less than 5 % of measurements were below the predetermined LOB. Otherwise, a higher template concentration would be chosen to repeat the aforementioned tests until the final LOD was determined. Ultimately, LODs were determined as 0.125 gc/μL with 95 % confidence for both *C. trachomatis* and *T. pallidum* assays.

2.6 Calculations of concentrations and back-estimations of infections

Concentrations of *C. trachomatis* and *T. pallidum* bacterial DNA were calculated using the formula (1), based on modifications of a formula that we proposed and implemented previously (Li et al., 2022b; Zhao et al., 2022).

$$C_{final} = C_{ddPCR} \times S_{total} / V_C \times V_{EX} / S_{DNA} \times V_{ddPCR} / V_{DNA} \quad (1)$$

where, C_{final} is the concentration of bacterial DNA in wastewater samples (gc/L); C_{ddPCR} is the measured concentration obtained from the droplet reader (gc/μL); S_{total} is the measured weight of final total pellets after centrifugation (gram, g); V_C is the volume of wastewater samples used for centrifugation (1 L); V_{EX} is the volume of extracted bacterial DNA (100 μL); S_{DNA} is the measured weight of pellets used for DNA extraction (g); V_{ddPCR} is the total volume of ddPCR final reaction

for each sample (22 μL); V_{DNA} is the sample volume added to Mastermix (5.5 μL).

Back-estimation of Chlamydia and Syphilis infections from *C. trachomatis* and *T. pallidum* concentrations in wastewater were performed using a modified formula (2) that was proposed in recent studies (Guo et al., 2022a; Li et al., 2022a).

$$W = (C_{DNA} \times e^{\wedge}(k \times t) \times Q \times \alpha) / (P_S \times Q_S \times C_S) \quad (2)$$

$$Q = Q_{total} / N \times F \quad (3)$$

where, W is the number of back-estimated infections for Chlamydia and Syphilis; C_{DNA} is the measured DNA concentrations of *C. trachomatis* or *T. pallidum* in wastewater (gc/L); k is the decay rate of the bacterial DNA in wastewater (d^{-1}); t is the wastewater in-sewer travel time (d); α is the adjustment factor involving wastewater dilution and other uncertainties discussed below; Q is the average wastewater generated per capita in each sewershed during the study period (L/d·person); N is the population captured in each sewershed; Q_{total} is the average wastewater flowing to each sewershed during the study period (million gallons per day, mg/d); F is the unit conversion factor between mg/d and L/d, which is 3.785×10^6; P_S is the rate of positive detection of *C. trachomatis* or *T. pallidum* in urine from individuals of the suspected disease; Q_S is the daily volume of urine generated from an individual (mL/d·person); C_S is the shedding magnitude of bacterial DNA in urine (gc/mL). Notably, urine-based parameters were selected for back-estimations since unlike enteric viruses that are shed from human feces, there were limited studies reporting the shedding of both *C. trachomatis* and *T. pallidum* in human feces. Tables 6S. 5. and 6S. 6. demonstrate the positive detection rates of *C. trachomatis* and *T. pallidum* in clinical and urine samples excreted from patients with suspected diseases. We also performed the global sensitivity analysis to quantify and compare the relative importance of each parameter on the final infection estimates. R package multisensi was implemented to perform sensitivity analysis on the

output of the multivariate model (Lamboni et al., 2011). The details of model implementation using multisensi was included in the Supporting Information.

Briefly, Table 6S. 2. summarized decay rate constant k (d^{-1}) for different bacteria in the aqueous environment. Among them, the range of k (d^{-1}) for Campylobacter and Salmonella in wastewater (0.17-0.52) were determined as the approximate k (d^{-1}) for *C. trachomatis* or *T. pallidum* in wastewater (Guo et al., 2022b). The four bacteria share common characteristics where they are classified as the gram-negative bacterium with an outer membrane especially where the *ompA* gene of *C. trachomatis* is targeted (Zdrodowska-Stefanow et al., 2003). The wastewater in-sewer travel time t for interceptors were obtained from GLWA (Table 6S. 3.), including 0.51, 0.94, and 0.36 days for DRI, NIEA, and ONWI interceptors, respectively. For the three smaller neighborhood sewersheds, t was estimated as 2.4 hours or 0.1 days as per an estimation conducted in similar sewersheds of comparable size of both inhabitants and service area (McCall et al., 2022). P_S for *C. trachomatis* was determined as 8.6%, the average rate reported previously between 5.3% 14.4% (Božičević et al., 2011; Mania-Pramanik et al., 2006; Møller et al., 2008, 2010). P_S for *T. pallidum* was determined as 25%, the average rate reported previously between 12.8% and 37.1% (Nieuwenburg et al., 2022). Additionally, due to the limited understanding of the shedding rates of *C. trachomatis* and *T. pallidum*, the full spectrum of reported urine detection rates for both bacteria was considered in back-estimating infections. W_{min}, or the minimum number of back-estimated infections, was computed based on the maximum positive detection rates of *C. trachomatis* or *T. pallidum* in urine, which were presented in Tables 6S. 5., and 6S. 6., respectively. Likewise, W_{max}, or the maximum number of back-estimated infections, was computed based on the minimum positive detection rates of *C. trachomatis* or *T. pallidum* in urine, which were also presented in Tables 6S. 5., and 6S. 6., respectively. W_{ave} was computed based on the averaged

positive detection rates of *C. trachomatis* or *T. pallidum* in urine. Values marked with an asterisk (*) in Tables 6S. 5., and 6S. 6. indicate the data used for calculating positive detection rates in urine. The full ranges of reported detection rates of both bacteria in urine could help identify bounds on estimated cases. Q_S was estimated as 1400 mL/d·person, the average volume of urine generated by an individual, where the normal range for 24-hour urine generation of an individual ranges from 800 to 2000 mL/d with an average fluid intake of about 2000 mL/d (medlineplus.gov). C_S for *C. trachomatis* was previously reported as 3.72×10^3 gc/mL for urine shedding (Twin et al., 2011). Likewise, C_S for *T. pallidum*, was previously reported as 2.767×10^3 gc/mL for urine shedding (Wang et al., 2022). The wastewater flows to GLWA WRRF interceptors include stormwater, residential, industrial, and commercial wastewater. The three interceptors service large sewersheds with enormous populations where significant dilution events can occur, including dilutions caused by industrial and commercial wastewater (Zhao et al., 2022), stormwater (Guo et al., 2022b), and snowmelt infiltration in March and April in Michigan (Tiwari et al., 2022). Besides, significant decay of bacterial DNA can occur due to long in-sewer travel time for large sewersheds (McCall et al., 2022). Therefore, the adjustment factor α of 100 was chosen for interceptors. The wastewater flows to manholes, covering significantly smaller sewersheds, primarily include sanitary wastewater with intermittent stormwater. Hence, an α of 10 was chosen for their adjustment (Guo et al., 2022b). Notably, the flow data (Q_{total}) for the three neighborhood sewersheds (EP, D3, and OP) were unavailable during the study period. Thus, we utilized historical flow data for D3 for the same season between January 2021 to April 2021 to approximate the flow data in the current study period. Historic flow data of EP and OP were unavailable, and we utilized historic flow data a nearby Detroit metro community (Southfield, zip-code 48076) with comparable size and population for the same season to approximate the flow

data in EP and OP (Li et al., 2022b). Additionally, it is noteworthy that the temperature of Detroit wastewater in interceptors were estimated between 5 to 25 ^{o}C and were measured ranging from 7.6 to 21.7 ^{o}C in neighborhood sewersheds shown in Table 6S. 4.

Since the flow data for zip-code areas within the sewersheds of interceptors or selected neighborhood sewersheds were unavailable, we measured and collected the PMMoV (pepper mild mottle virus) and crAssphage (cross-assembly phage) data for the same weeks during the study period. These data were utilized to normalize *C. trachomatis ompA* and *T. pallidum polA* concentrations for comparing disparities between interceptors and selected sewersheds. The details of testing PMMoV and crAssphage were demonstrated in the Supporting Information.

3. Results

3.1 Concentrations of *C. trachomatis* and *T. pallidum* in interceptors and neighborhood sewersheds wastewater

C. trachomatis and *T. pallidum* DNA were detected in wastewater samples across three interceptors and three neighborhood sewersheds between December 25th, 2023, and April 22nd, 2024 (Figure 6. 2.). Concentrations of *C. trachomatis* DNA ranged from non-detect to 1794 gc/L (DRI) in wastewater samples collected from interceptors, and from non-detect to 20367 gc/L (D3) in wastewater samples collected from neighborhood sewersheds. Concentrations of *T. pallidum* DNA ranged from non-detect to 1929 gc/L (ONWI) in wastewater samples collected from interceptors, and from non-detect to 4664 gc/L (OP) in wastewater samples collected from neighborhood sewersheds.

For interceptors, the highest weekly concentration of *C. trachomatis* was observed in the DRI interceptor (1794 gc/L) comparing to that of ONWI (908 gc/L) and NIEA (800 gc/L) interceptors. DRI interceptor primarily services the City of Detroit, where the highest weekly

average cases of Chlamydia were reported as 171, comparing to that of 59, 45, and 64, in Wayne, Macomb, and Oakland counties. Notably, D3, which is located within the City of Detroit, demonstrated highest weekly concentration of *C. trachomatis* as 20367 gc/L comparing to that of EP (5842 gc/L) and OP (15416 gc/L) sewersheds. For *T. pallidum*, the highest weekly concentration was observed in ONWI interceptor (1929 gc/L) among three interceptors (NIEA: 1703 gc/L, DRI: 1394 gc/L). Among the three neighborhood sewersheds, the highest weekly concentration was observed in OP (4664 gc/L), comparing that of D3 (3637 gc/L) and EP (1808 gc/L).

Additionally, normalization approaches also embraced previous findings. For instance, both the average *C. trachomatis* and *T. pallidum* concentrations in DRI were observed higher than those of ONWI and NIEA after normalization using PMMoV and crAssphage. Both *C. trachomatis* and *T. pallidum* concentrations in D3 were observed higher than those of EP and OP after normalization using PMMoV. These results indicated higher normalized *C. trachomatis* and *T. pallidum* concentrations were observed in DRI among interceptors and D3 among selected sewersheds, which both located within the City of Detroit (Figures 6S. 1., 6S. 2., and 6S. 3.).

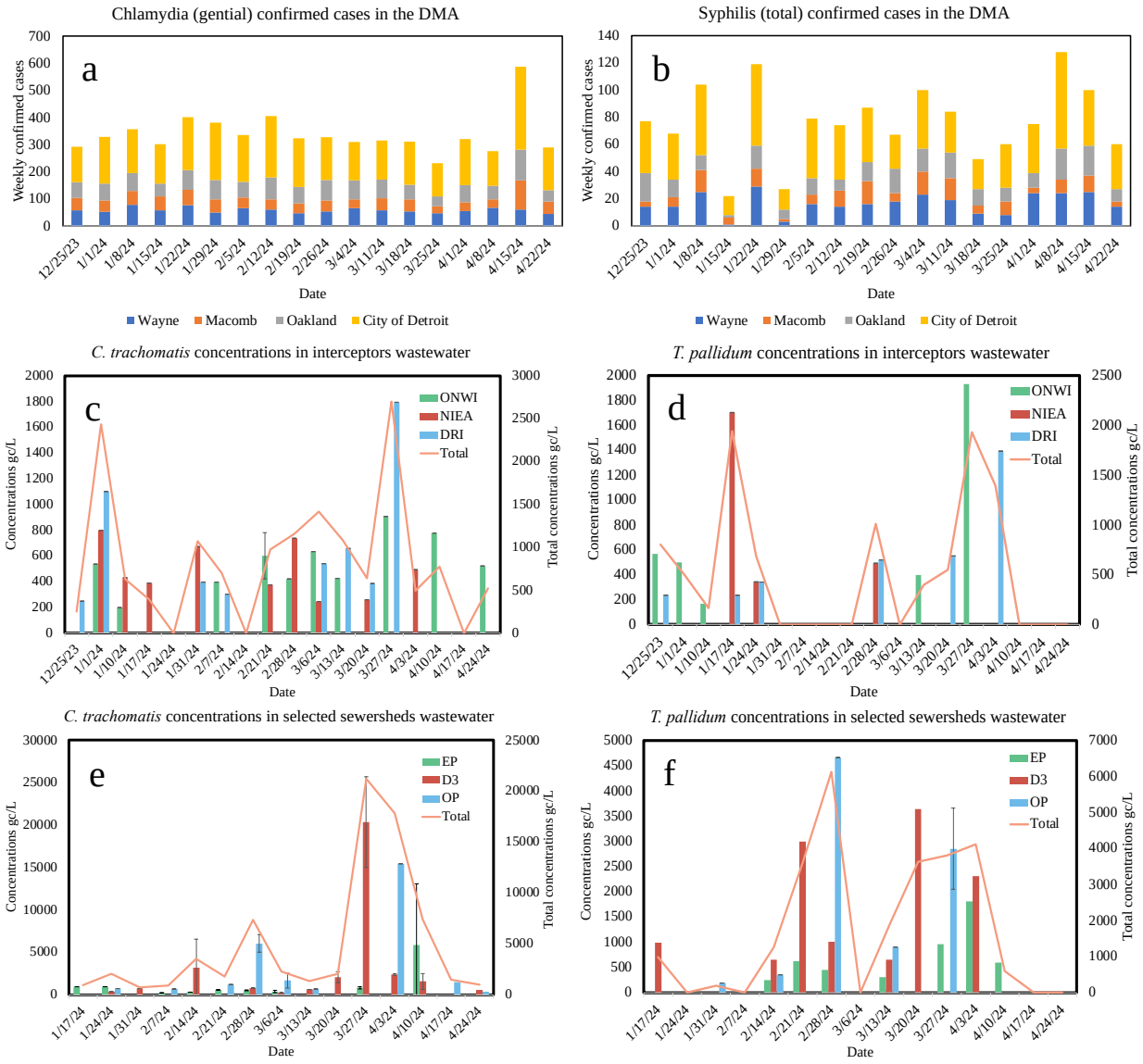


Figure 6. 2. Weekly confirmed Chlamydia cases (a) and Syphilis cases (b) in the DMA, measured *C. trachomatis* concentrations in interceptors (c) and neighborhood sewersheds wastewater (e) and *T. pallidum* concentrations in interceptors (d) and neighborhood sewersheds wastewater (f)

3.2 Bacterial shedding of *C. trachomatis* and *T. pallidum*

Extensive literature reviews were conducted to investigate the bacterial shedding of *C. trachomatis* and *T. pallidum* through human bodily fluids. For *C. trachomatis*, the positive detection rates in human bodily fluids with suspected Chlamydia were reported, including an average rate of 8.6% in urine, as well as rates ranging from 8.3% to 18.4% in clinical samples such as vaginal and genital ulcer swabs (Božičević et al., 2011; Mania-Pramanik et al., 2006; Møller et al., 2008, 2010; Ngobese et al., 2022; Pickett et al., 2021; Tshaka et al., 2022). For *T. pallidum*, the positive detection rates in human bodily fluids with suspected Syphilis were reported including an average rate of 25% in urine, as well as rates ranging from 0.3% to 47% in clinical samples such as genital, anal ulcers and ano-rectal swabs (Dubourg et al., 2015; Glatz et al., 2014; Heymans et al., 2010; Koek et al., 2006; Nieuwenburg et al., 2022; Shields et al., 2012; Tshaka et al., 2022). To our best knowledge, limited studies yet reported shedding of *C. trachomatis* and *T. pallidum* through human stool or positive detections of these bacteria in stool samples in clinical settings. The details of reported positive rates of *C. trachomatis* and *T. pallidum* in urine and clinical samples were summarized in Tables 6S. 5. and 6S. 6., respectively.

Additionally, the shedding duration of *C. trachomatis* can persist up to 10 days until the infection clears itself spontaneously (Dukers-Muijrers et al., 2022; Igietseme et al., 2001). The incubation time of *C. trachomatis* were reported with significant variabilities from 5 days (O'Connell & Ferone, 2016) to 28 days (Jones & Lopez, 2014). Likewise, the shedding duration of *T. pallidum* exhibited significant uncertainties and variabilities depending on the stages of syphilis, such as primary, secondary, and early latent. Researchers identified men shed *T. pallidum* concurrently from anal and oral routes ranging from 7 to 180 days (Towns et al., 2022). Similarly, the incubation time of *T. pallidum* were ranging from 9 to 90 days (French, 2007). For

asymptomatic infections, it is unclear how long the incubation time would be for both *C. trachomatis* and *T. pallidum*.

3.3 Back-estimations of Chlamydia and Syphilis infections based on bacterial shedding and wastewater measurements

The values of the input parameters using the formula (2) for back-estimating Chlamydia and Syphilis infections in interceptors and neighborhood sewersheds were demonstrated in Table 6. 4. Figures S6. 4., and S6. 5., both demonstrated dynamics of the sensitivity analysis indices of Chlamydia and Syphilis infection estimates, respectively. Parameters including α, Q, Ps, Cs, Qs, and e^{kt}, demonstrated descending importance on the infection estimates. Notably, catchment-specific characteristics including α, Q, and e^{kt} demonstrated relatively higher impacts for both the Chlamydia and Syphilis infection estimates (Figures S6. 4., and S6. 5.). Both parameters of α and Q are closely related to the scales of sampling and dilution effects, and both demonstrate highest sensitivity indices for infection estimates. The characteristics related to bacterial shedding also played a crucial role in infection estimates including Ps, Cs, Qs, which demonstrated relatively higher sensitivity indices.

For interceptors, it is noteworthy that the highest estimated infections of both Chlamydia and Syphilis was observed in DRI, despite DRI services the smallest number of inhabitants in its service area. For neighborhood sewersheds, the highest estimated infections of both Chlamydia and Syphilis was observed in D3, despite D3 services the lowest number of inhabitants in its sewersheds. Notably, DRI services the majority of the City of Detroit communities and the sewershed covered by D3 was located within the City of Detroit. Besides, these observations of estimated Chlamydia and Syphilis infections agree with the trends of confirmed cases of Chlamydia and Syphilis in Detroit, where the highest weekly confirmed cases were observed in

the city (Figure 6. 2.).

Additionally, the comparisons between ranges of estimated average weekly Chlamydia and Syphilis infections for interceptors and weekly confirmed cases for their approximate regions were demonstrated in Table 6. 5. Notably, the weekly confirmed cases for both Chlamydia and Syphilis for all approximate regions were potentially underreported comparing to the back-estimated weekly infections from the interceptors, with only one exception for the Chlamydia estimation at the NIEA interceptor. Ratio between ranges of estimated weekly infections and weekly confirmed cases for approximate regions were computed (Table 6. 5.). Ratio W_{ave}/C_{ave} for both Chlamydia and Syphilis are above 1 except for NIEA interceptor. For Syphilis, values of W_{ave}/C_{ave} were consistent approximately around 3 for NIEA, DRI, and total interceptors. Overall Syphilis demonstrated higher values of the three ratios (W_{ave}/C_{ave}, W_{min}/C_{ave}, and W_{max}/C_{ave}), which are all above 1, comparing to those of Chlamydia.

Table 6. 4. Back-estimation of average weekly infections using formula (2)

Bacteria	Site	N	C_{DNA}	Q_{total}	Q	t	K	e^{kt}	P_S	Q_S	C_S	α	W_{ave}	W_{min}	W_{max}
C. trachomatis	ONWI	840600	542	231	1040.13	0.36	0.17 - 0.52	1.06-1.21	ave = 8.6%, min = 5.3%, max = 14.4%	1400	3.72 × 10^3	100	133.42-152.30	79.68-90.96	216.49-247.13
	NIEA	1482000	489	218	556.77	0.94		1.17-1.63					71.12-99.08	42.48-59.18	115.40-160.78
	DRI	492000	679	193	1484.77	0.51		1.09-1.3					245.35-292.62	146.53-174.76	398.12-474.82
	EP	2400	1176	0.19	299.65	0.1		1.02-1.05				10	8.03-8.26	4.79-4.93	13.02-13.40
	D3	1300	2997	0.8	2329.23								158.98-163.65	94.94-97.74	257.96-265.55
	OP	2270	3148	0.19	316.81								22.71-23.38	13.56-13.96	36.85-37.94
T. pallidum	ONWI	840600	713	231	1040.13	0.36		1.06-1.21	ave = 25%, min = 12.8%, max = 37.1%		2.767 × 10^3	100	81.17-92.66	54.70-62.44	158.54-180.97
	NIEA	1482000	847	218	556.77	0.94		1.17-1.63					56.97-79.37	38.39-53.49	111.28-155.02
	DRI	492000	546	193	1484.77	0.51		1.09-1.3					91.24-108.82	61.48-73.33	178.21-212.54
	EP	2400	713	0.19	299.65	0.1		1.02-1.05				10	2.25-2.32	1.52-1.56	4.39-4.52
	D3	1300	1749	0.8	2329.23								42.91-44.17	28.91-29.76	83.80-86.27
	OP	2270	1793	0.19	316.81								5.98-6.16	4.03-4.15	11.69-12.03

Table 6. 5. Comparison between back-estimated weekly cases in the sewersheds serviced by interceptors and average weekly confirmed cases in the approximate regions in the Detroit metro area

Disease	Interceptors	Estimated average weekly cases (W_{ave})	Estimated minimum weekly cases (W_{min})	Estimated maximum weekly cases (W_{max})	Confirmed average weekly cases (C_{ave}) for approximate regions	Ratio W_{ave}/C_{ave}	Ratio W_{min}/C_{ave}	Ratio W_{max}/C_{ave}
Chlamydia	ONWI	133.42-152.30	79.68-90.96	216.49-247.13	Wayne county: 58.72	2.27-2.59	1.36-1.55	3.69-4.21
	NIEA	71.12-99.08	42.48-59.18	115.40-160.78	Oakland and Macomb counties: 109.06	0.65-0.91	0.39-0.54	1.06-1.47
	DRI	245.35-292.62	146.53-174.76	398.12-474.82	City of Detroit: 170.9	1.44-1.71	0.86-1.02	2.33-2.78
	Total	449.89-544.00	268.69-324.89	730.01-882.72	Total: 338.7	1.33-1.61	0.79-0.96	2.16-2.61
Syphilis	ONWI	81.17-92.66	54.70-62.44	158.54-180.97	Wayne county: 16.44	4.94-5.64	3.33-3.80	9.64-11.01
	NIEA	56.97-79.37	38.39-53.49	111.28-155.02	Oakland and Macomb counties: 23	2.48-3.45	1.67-2.33	4.84-6.74
	DRI	91.24-108.82	61.48-73.33	178.21-212.54	City of Detroit: 37.22	2.45-2.92	1.65-1.97	4.79-5.71
	Total	229.39-280.85	154.57-189.25	448.02-548.54	Total: 76.67	2.99-3.66	2.02-2.47	5.84-7.15

4. Discussion

4.1 Wastewater surveillance as a screening tool for Chlamydia and Syphilis

This study demonstrates the utility of wastewater surveillance on monitoring STIs particularly Chlamydia and Syphilis in wastewater in a large urban center as well as small neighborhood sewersheds for continuous 4 months. Unlike wastewater surveillance for SARS-CoV-2 or other viral targets, wastewater surveillance for Chlamydia and Syphilis adds its unique value to the wastewater surveillance field of being a screening tool for bacterial diseases, which can be followed by more targeted clinical testing in communities before they become widespread.

Recent studies indicated that factors including the willingness of STIs patients for clinical testing, asymptomatic infections of STIs, and shift of focus from STIs to COVID-19 programs all contributed to uncertainties in the accuracy of clinically reported cases for Chlamydia and Syphilis (Johnson et al., 2021; Wright et al., 2022). Particularly, due to the high infectivity and rapid transmission of STIs in densely populous areas such as Detroit, universal screening, regardless of asymptomatic or symptomatic conditions, in clinical settings is practically impossible. Thus, wastewater surveillance of Chlamydia and Syphilis highlights possibilities of monitoring fluctuations of STIs infections in communities, complementing clinically reported cases to represent actual community infections.

Notably, higher DNA concentrations of both *C. trachomatis* and *T. pallidum* were observed in the three smaller neighborhood sewersheds than the three interceptors that cover the entire DMA, which can be likely attributed to different scales of dilution impacts. Researchers indicated that longer in-sewer travel time in larger sewersheds led to greater variabilities and degradation of pathogenic concentrations in wastewater with potentially 50% or more signals degrading (McCall et al., 2022). In particular, the size and in-sewer travel time of interceptors

(ONWI, NIEA, and DRI) are significantly larger than those of the smaller neighborhood sewersheds (EP, D3, and OP).

4.2 Potential underreporting of Chlamydia and Syphilis incidences when wastewater surveillance estimates are benchmarked against clinical data

This study demonstrates possibilities of back-estimations on Chlamydia and Syphilis infections based on wastewater data using a modified formula (Guo et al., 2022a). Comparative analyses between ONWI-sewershed and Wayne county, NIEA-sewershed and combined counties of Oakland and Macomb, DRI-sewershed and City of Detroit were carried out and results were demonstrated in Table 6. 3. For Chlamydia infection estimates, the confirmed weekly cases were likely underreported for some sewersheds serviced by interceptors if they are benchmarked against clinical data as the ground truth, including OWNI-sewershed. For instance, the total estimated average infections (ranging from 449.9 to 544 for Chlamydia and from 229.4 to 280.9 for Syphilis) also indicated that the total confirmed weekly cases (338.7 and 76.67 for Chlamydia and Syphilis, respectively) for the DMA are likely underreported if clinical data represents the actual infection scenario. The ratio between Syphilis infection estimates and corresponding clinical data were ranging from 1.65 to 11.01, which also demonstrated higher infection estimates comparing to reported clinical data (Table 6. 3.). These results potentially embraced previous findings where researchers indicated that Chlamydia and Syphilis have been underdiagnosed and underreported due to asymptomatic infections and inadequate participation by infected individuals (Chin Quee, 2023; Lafetá et al., 2016). STIs such as Chlamydia and Syphilis remain underdiagnosed and untreated in communities, which have the potential to become widespread without a comprehensive screening method for both symptomatic and asymptomatic infections. To address the issues above, the workflow of bacterial wastewater surveillance monitoring both Chlamydia

and Syphilis proposed in this study, can be an ideal screening method.

4.3 Concentrations of *C. trachomatis* and *T. pallidum* might relate to socioeconomic demographics

Among the three neighborhood sewersheds EP, D3, and OP, it was observed that D3 presented significantly higher estimated infections of both Chlamydia (158.98-163.65) and Syphilis (42.91-44.17), despite having the lowest population. Particularly, D3 exhibited distinctive demographic characteristics among the three neighborhood sewersheds. These demographic characteristics of the D3 sewershed include the highest poverty percentage, and the highest population density, in contrast to OP and EP sewersheds (Table 6S. 8.). Likewise, researchers identified that low socioeconomic status generally relates to health care access. And STIs rates generally relate to social determinants (Hogben & Leichliter, 2008). Notably, the highest ranges of estimated infections for both Chlamydia and Syphilis were observed in D3 and DRI, both located within the City of Detroit. D3 demonstrates significant disparities in socioeconomic demographics among selected sewersheds (Table 6S. 8.). Nevertheless, the clear connection between higher infection rates of both diseases and socioeconomic demographics requires further investigation.

4.4 Limitations and future directions

This study demonstrated one of the first workflows for bacterial wastewater surveillance and presented the detection of *C. trachomatis* and *T. pallidum* in wastewater as a screening method. We adopted a centrifugation method to concentrate and isolate bacteria in wastewater. However, the limitation of this method is the potential losses of bacteria in the supernatant (Aw et al., 2012). We initially implemented a membrane filtration method using 0.2 μm filters but results were not presented due to low recoveries. Besides, we initially tested *C. trachomatis* and *T.*

pallidum in supernatant wastewater, but their recoveries were low. Researchers identified that the direct centrifugation method demonstrated a higher recovery rate when comparing to filtration coupled with centrifugation (Varma et al., 2009). Besides, bacterial cells can often be isolated using centrifugation method with a speed of more than 8000×g (Aw et al., 2012) and previous studies concentrating bacteria from environmental and clinical samples using centrifugation were summarized in Table 6S. 1.

Some parameters that were implemented in the back-estimation model still needs further investigations. For instance, there were extremely limited studies reporting the Cs parameter for urine, which was identified as a limitation of the current study. Consequently, future research is needed to investigate Cs in urine, feces, and other bodily fluids.

Albeit the wide applications of PMMoV and crAssphage as biomarkers for normalization of wastewater pathogenic concentrations, the relative recovery of their signal may differ from the recovery of gene concentrations of *C. trachomatis* and *T. pallidum*. In addition, differences in genomes (RNA versus DNA) may also affect the normalization outcomes. Therefore, in addition to these human fecal indicators, investigations on other closely related biomarkers are needed. Future research is also needed to address gaps in types of process, recovery, and inhibition controls when wastewater surveillance is expanded to monitoring bacterial targets (Philo et al., 2023).

It is critical to indicate that neither infection estimates based on wastewater surveillance nor clinically reported cases may depict the actual infections in communities. Wastewater surveillance estimates could be beneficial when clinical testing capacity is limited, asymptomatic infections is dominating, or wiliness of individuals clinical testing is low. However, wastewater surveillance estimates may not provide sufficient data for unsewered areas. Clinical testing generally screens population who is willing to get tested. Other sources of data can also be used

to compare with wastewater surveillance estimates and clinical data, including google trends, digital epidemiological data, mobility data, ground truth data include reports published by health departments or the World Health Organization, census statistics, data obtained from scientific studies, and data in news and media (Park et al., 2018). Data from all these aforementioned sources need to be integrated to generate coherent information for decision-makings.

Finally, several questions still remain to be further investigated including whether bacterial cells of *C. trachomatis* and *T. pallidum* can grow in wastewater, and can we directly relate their bacterial DNA concentrations in wastewater to Chlamydia and Syphilis infections in communities? There is little published research investigating growth of *C. trachomatis* and *T. pallidum* in wastewater. Both *C. trachomatis* and *T. pallidum* are classified as gram-negative bacteria that replicate only within the host cells but not in the environment (Elwell et al., 2016; Varma et al., 2013; Witkin et al., 2017). *C. trachomatis* cannot replicate outside human host cells. Likewise, *T. pallidum* is an obligate human pathogen, and animal reservoirs for *T. pallidum* were not reported yet, which enhances its dependence on human host cells (Powers-Fletcher, 2011; Varma et al., 2013). Besides, both *C. trachomatis* and *T. pallidum* relies entirely on the supplies of nutrients from the human host cell (Liang et al., 2018; Radolf et al., 2015). To date, there is limited research investigating on bacterial growth of *C. trachomatis* and *T. pallidum* in wastewater. Therefore, more research is called on investigating the impacts of bacterial growth in wastewater on the relationship between wastewater data and human infections.

5. Environmental Implications

This study fills multiple important knowledge gaps in the current field of wastewater surveillance. First, this study demonstrated one of the first wastewater surveillance applications in monitoring widespread STIs, particularly Chlamydia and Syphilis, in a large urban area as well as

neighborhood sewer-sheds. This information highlights the utility of bacterial wastewater surveillance as a screening tool to complement clinically reported cases of bacterial diseases. This study also established a workflow of implementing bacterial wastewater surveillance, where molecular biology laboratory methods were optimized to detect and quantify *C. trachomatis* and *T. pallidum* in wastewater. Second, the results of different concentrations of *C. trachomatis* and *T. pallidum* in wastewater demonstrated disparities of corresponding socioeconomic characteristics of sewer-sheds. Third, Chlamydia and Syphilis infections were back-estimated using a modified formula based on extensive investigations on shedding dynamics of *C. trachomatis* and *T. pallidum* in environmental and clinical samples, revealing potentially underreported cases of both diseases in the Detroit metro area.

6. Acknowledgements

We thank the Michigan Department of Health and Human Services (MDHHS) and the Great Lakes Water Authority (GLWA) for funding this research.

REFERENCES

Ammerman, M. L., Mullapudi, S., Gilbert, J., Figueroa, K., de Paula Nogueira Cruz, F., Bakker, K. M., Eisenberg, M. C., Foxman, B., & Wigginton, K. R. (2024). Norovirus GII wastewater monitoring for epidemiological surveillance. *PLOS Water*, *3*(1), e0000198. https://doi.org/10.1371/journal.pwat.0000198

Aw, T. G., Gin, K. Y.-H., Goh, S. G., & Te, S. H. (2012). Sample Preparation of Microbial Contaminants in Water. In *Comprehensive Sampling and Sample Preparation* (pp. 723–742). Elsevier. https://doi.org/10.1016/B978-0-12-381373-2.00111-3

Balfe, M., Brugha, R., O'Connell, E., McGee, H., O'Donovan, D., & Vaughan, D. (2010). Why don't young women go for Chlamydia testing? A qualitative study employing Goffman's stigma framework. *Health, Risk & Society*, *12*(2), 131–148. https://doi.org/10.1080/13698571003632437

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Beattie, R. E., Blackwood, A. D., Clerkin, T., Dinga, C., & Noble, R. T. (2022). Evaluating the impact of sample storage, handling, and technical ability on the decay and recovery of SARS-CoV-2 in wastewater. *PLOS ONE*, *17*(6), e0270659. https://doi.org/10.1371/journal.pone.0270659

Blake, D. R., Kearney, M. H., Oakes, J. M., Druker, S. K., & Bibace, R. (2003). Improving Participation in Chlamydia Screening Programs. *Archives of Pediatrics & Adolescent Medicine*, *157*(6), 523. https://doi.org/10.1001/archpedi.157.6.523

Božičević, I., Grgić, I., Židovec-Lepej, S., Čakalo, J.-I., Belak-Kovačević, S., Štulhofer, A., & Begovac, J. (2011). Urine-based testing for Chlamydia trachomatis among young adults in a population-based survey in Croatia: Feasibility and prevalence. *BMC Public Health*, *11*(1), 230. https://doi.org/10.1186/1471-2458-11-230

Chin Quee, J. E. (2023). *Using Wastewater-Based Epidemiology to Study Chlamydia Occurrence on a College Campus*.

Dubourg, G., Edouard, S., Prudent, E., Fournier, P.-E., & Raoult, D. (2015). Incidental Syphilis Diagnosed by Real-Time PCR Screening of Urine Samples. *Journal of Clinical Microbiology*, *53*(11), 3707–3708. https://doi.org/10.1128/JCM.01026-15

Dukers-Muijrers, N. H. T. M., Evers, Y. J., Hoebe, C. J. P. A., Wolffs, P. F. G., de Vries, H. J. C., Hoenderboom, B., van der Sande, M. A. B., Heijne, J., Klausner, J. D., Hocking, J. S., & van Bergen, J. (2022). Controversies and evidence on Chlamydia testing and treatment in

asymptomatic women and men who have sex with men: a narrative review. *BMC Infectious Diseases*, *22*(1), 255. https://doi.org/10.1186/s12879-022-07171-2

Elwell, C., Mirrashidi, K., & Engel, J. (2016). Chlamydia cell biology and pathogenesis. *Nature Reviews Microbiology*, *14*(6), 385–400. https://doi.org/10.1038/nrmicro.2016.30

French, P. (2007). Syphilis. *BMJ*, *334*(7585), 143–147. https://doi.org/10.1136/bmj.39085.518148.BE

Fu, Y., Ye, Z., Jia, Y., Fan, J., Hashmi, M. Z., & Shen, C. (2020). An Optimized Method to Assess Viable Escherichia coli O157:H7 in Agricultural Soil Using Combined Propidium Monoazide Staining and Quantitative PCR. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01809

Gentry, Z., Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2023). Wastewater surveillance beyond COVID-19: a ranking system for communicable disease testing in the tri-county Detroit area, Michigan, USA. *Frontiers in Public Health*, *11*, 1178515.

Glatz, M., Juricevic, N., Altwegg, M., Bruisten, S., Komericki, P., Lautenschlager, S., Weber, R., & Bosshard, P. P. (2014). A multicenter prospective trial to asses a new real-time polymerase chain reaction for detection of Treponema pallidum, herpes simplex-1/2 and Haemophilus ducreyi in genital, anal and oropharyngeal ulcers. *Clinical Microbiology and Infection*, *20*(12), O1020–O1027. https://doi.org/10.1111/1469-0691.12710

Guo, Y., Li, J., O'Brien, J., Sivakumar, M., & Jiang, G. (2022). Back-estimation of norovirus infections through wastewater-based epidemiology: A systematic review and parameter sensitivity. *Water Research*, *219*, 118610. https://doi.org/10.1016/j.watres.2022.118610

Guo, Y., Sivakumar, M., & Jiang, G. (2022). Decay of four enteric pathogens and implications to wastewater-based epidemiology: Effects of temperature and wastewater dilutions. *Science of The Total Environment*, *819*, 152000. https://doi.org/10.1016/j.scitotenv.2021.152000

Heymans, R., van der Helm, J. J., de Vries, H. J. C., Fennema, H. S. A., Coutinho, R. A., & Bruisten, S. M. (2010). Clinical Value of Treponema pallidum Real-Time PCR for Diagnosis of Syphilis. *Journal of Clinical Microbiology*, *48*(2), 497–502. https://doi.org/10.1128/JCM.00720-09

Hogben, M., & Leichliter, J. S. (2008). Social Determinants and Sexually Transmitted Disease Disparities. *Sexually Transmitted Diseases*, *35*(12), S13–S18. https://doi.org/10.1097/OLQ.0b013e31818d3cad

Igietseme, J. U., Portis, J. L., & Perry, L. L. (2001). Inflammation and Clearance of *Chlamydia trachomatis* in Enteric and Nonenteric Mucosae. *Infection and Immunity*, *69*(3), 1832–1840. https://doi.org/10.1128/IAI.69.3.1832-1840.2001

Johnson, K. A., Burghardt, N. O., Tang, E. C., Long, P., Plotzker, R., Gilson, D., Murphy, R., & Jacobson, K. (2021). Measuring the Impact of the COVID-19 Pandemic on Sexually Transmitted

Diseases Public Health Surveillance and Program Operations in the State of California. *Sexually Transmitted Diseases*, *48*(8), 606–613. https://doi.org/10.1097/OLQ.0000000000001441

Jones, R. E., & Lopez, K. H. (2014). Sexually Transmitted Diseases. In *Human Reproductive Biology* (pp. 323–347). Elsevier. https://doi.org/10.1016/B978-0-12-382184-3.00017-9

Kilaru, P., Hill, D., Anderson, K., Collins, M. B., Green, H., Kmush, B. L., & Larsen, D. A. (2023). Wastewater Surveillance for Infectious Disease: A Systematic Review. *American Journal of Epidemiology*, *192*(2), 305–322. https://doi.org/10.1093/aje/kwac175

Koek, A. G., Bruisten, S. M., Dierdorp, M., van Dam, A. P., & Templeton, K. (2006). Specific and sensitive diagnosis of syphilis using a real-time PCR for Treponema pallidum. *Clinical Microbiology and Infection*, *12*(12), 1233–1236. https://doi.org/10.1111/j.1469-0691.2006.01566.x

Lafetá, K. R. G., Martelli Júnior, H., Silveira, M. F., & Paranaíba, L. M. R. (2016). Maternal and congenital syphilis, underreported and difficult to control. *Revista Brasileira de Epidemiologia*, *19*, 63–74.

Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, *96*(4), 450–459. https://doi.org/10.1016/j.ress.2010.12.002

Li, X., Kulandaivelu, J., Guo, Y., Zhang, S., Shi, J., O'Brien, J., Arora, S., Kumar, M., Sherchan, S. P., Honda, R., Jackson, G., Luby, S. P., & Jiang, G. (2022). SARS-CoV-2 shedding sources in wastewater and implications for wastewater-based epidemiology. *Journal of Hazardous Materials*, *432*, 128667. https://doi.org/10.1016/j.jhazmat.2022.128667

Li, Y., Miyani, B., Faust, R. A., David, R. E., & Xagoraraki, I. (2024). A broad wastewater screening and clinical data surveillance for virus-related diseases in the metropolitan Detroit area in Michigan. *Human Genomics*, *18*(1), 14. https://doi.org/10.1186/s40246-024-00581-0

Li, Y., Miyani, B., Zhao, L., Spooner, M., Gentry, Z., Zou, Y., Rhodes, G., Li, H., Kaye, A., Norton, J., & Xagoraraki, I. (2022). Surveillance of SARS-CoV-2 in nine neighborhood sewersheds in Detroit Tri-County area, United States: Assessing per capita SARS-CoV-2 estimations and COVID-19 incidence. *Science of The Total Environment*, *851*, 158350. https://doi.org/10.1016/j.scitotenv.2022.158350

Liang, P., Rosas-Lemus, M., Patel, D., Fang, X., Tuz, K., & Juárez, O. (2018). Dynamic energy dependency of Chlamydia trachomatis on host cell metabolism during intracellular growth: Role of sodium-based energetics in chlamydial ATP generation. *Journal of Biological Chemistry*, *293*(2), 510–522. https://doi.org/10.1074/jbc.M117.797209

Liguori, K., Keenum, I., Davis, B. C., Calarco, J., Milligan, E., Harwood, V. J., & Pruden, A. (2022). Antimicrobial Resistance Monitoring of Water Environments: A Framework for Standardized Methods and Quality Control. *Environmental Science & Technology*, *56*(13), 9149–9160. https://doi.org/10.1021/acs.est.1c08918

Linnet, K., & Kondratovich, M. (2004). Partly Nonparametric Approach for Determining the Limit of Detection. *Clinical Chemistry*, *50*(4), 732–740. https://doi.org/10.1373/clinchem.2003.029983

Mania-Pramanik, J., Potdar, S., & Kerkar, S. (2006). Diagnosis of Chlamydia trachomatis infection. *Journal of Clinical Laboratory Analysis*, *20*(1), 8–14. https://doi.org/10.1002/jcla.20092

Matrajt, G., Lillis, L., & Meschke, J. S. (2020). Review of Methods Suitable for Environmental Surveillance of Salmonella Typhi and Paratyphi. *Clinical Infectious Diseases*, *71*(Supplement_2), S79–S83. https://doi.org/10.1093/cid/ciaa487

McCall, C., Fang, Z. N., Li, D., Czubai, A. J., Juan, A., LaTurner, Z. W., Ensor, K., Hopkins, L., Bedient, P. B., & Stadler, L. B. (2022). Modeling SARS-CoV-2 RNA degradation in small and large sewersheds. *Environmental Science: Water Research & Technology*, *8*(2), 290–300. https://doi.org/10.1039/D1EW00717C

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

Mercier, E., D'Aoust, P. M., Thakali, O., Hegazy, N., Jia, J.-J., Zhang, Z., Eid, W., Plaza-Diaz, J., Kabir, M. P., Fang, W., Cowan, A., Stephenson, S. E., Pisharody, L., MacKenzie, A. E., Graber, T. E., Wan, S., & Delatolla, R. (2022). Municipal and neighbourhood level wastewater surveillance and subtyping of an influenza virus outbreak. *Scientific Reports*, *12*(1), 15777. https://doi.org/10.1038/s41598-022-20076-z

Mercier, E., Pisharody, L., Guy, F., Wan, S., Hegazy, N., D'Aoust, P. M., Kabir, M. P., Nguyen, T. B., Eid, W., Harvey, B., Rodenburg, E., Rutherford, C., Mackenzie, A. E., Willmore, J., Hui, C., Paes, B., Delatolla, R., & Thampi, N. (2023). Wastewater-based surveillance identifies start to the pediatric respiratory syncytial virus season in two cities in Ontario, Canada. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1261165

Milbury, C. A., Zhong, Q., Lin, J., Williams, M., Olson, J., Link, D. R., & Hutchison, B. (2014). Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomolecular Detection and Quantification*, *1*(1), 8–22. https://doi.org/10.1016/j.bdq.2014.08.001

Miyani, B., Fonoll, X., Norton, J., Mehrotra, A., & Xagoraraki, I. (2020). SARS-CoV-2 in Detroit Wastewater. *Journal of Environmental Engineering*, *146*(11), 06020004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001830

Miyani, B., Zhao, L., Spooner, M., Buch, S., Gentry, Z., Mehrotra, A., Norton, J., & Xagoraraki, I. (2021). Early Warnings of COVID-19 Second Wave in Detroit. *Journal of Environmental Engineering*, *147*(8), 06021004. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001907

Møller, J. K., Pedersen, L. N., & Persson, K. (2008). Comparison of Gen-Probe Transcription-

Mediated Amplification, Abbott PCR, and Roche PCR Assays for Detection of Wild-Type and Mutant Plasmid Strains of *Chlamydia trachomatis* in Sweden. *Journal of Clinical Microbiology*, *46*(12), 3892–3895. https://doi.org/10.1128/JCM.00412-08

Møller, J. K., Pedersen, L. N., & Persson, K. (2010). Comparison of the Abbott RealTi *m* e CT New Formulation Assay with Two Other Commercial Assays for Detection of Wild-Type and New Variant Strains of *Chlamydia trachomatis*. *Journal of Clinical Microbiology*, *48*(2), 440–443. https://doi.org/10.1128/JCM.01446-09

Ngobese, B., Swe-Han, K. S., Tinarwo, P., & Abbai, N. S. (2022). Significant Associations between Chlamydia trachomatis and Neisseria gonorrhoeae Infections in Human Immunodeficiency Virus-Infected Pregnant Women. *Infectious Diseases in Obstetrics and Gynecology*, *2022*, 1–13. https://doi.org/10.1155/2022/7930567

Nieuwenburg, S. A., Zondag, H. C. A., Bruisten, S. M., Jongen, V. W., Schim van der Loeff, M. F., van Dam, A. P., & de Vries, H. J. C. (2022). Detection of Treponema pallidum DNA During Early Syphilis Stages in Peripheral Blood, Oropharynx, Ano-Rectum and Urine as a Proxy for Transmissibility. *Clinical Infectious Diseases*, *75*(6), 1054–1062. https://doi.org/10.1093/cid/ciac056

O'Brien, E., & Xagoraraki, I. (2019). A water-focused one-health approach for early detection and prevention of viral outbreaks. *One Health*, *7*, 100094. https://doi.org/10.1016/J.ONEHLT.2019.100094

O'Connell, C. M., & Ferone, M. E. (2016). Chlamydia trachomatis Genital Infections. *Microbial Cell*, *3*(9), 390–403. https://doi.org/10.15698/mic2016.09.525

Park, H.-A., Jung, H., On, J., Park, S. K., & Kang, H. (2018). Digital Epidemiology: Use of Digital Data Collected for Non-epidemiological Purposes in Epidemiological Studies. *Healthcare Informatics Research*, *24*(4), 253. https://doi.org/10.4258/hir.2018.24.4.253

Philo, S. E., De León, K. B., Noble, R. T., Zhou, N. A., Alghafri, R., Bar-Or, I., Darling, A., D'Souza, N., Hachimi, O., Kaya, D., Kim, S., Gaardbo Kuhn, K., Layton, B. A., Mansfeldt, C., Oceguera, B., Radniecki, T. S., Ram, J. L., Saunders, L. P., Shrestha, A., … Delgado Vela, J. (2023). Wastewater surveillance for bacterial targets: current challenges and future goals. *Applied and Environmental Microbiology*. https://doi.org/10.1128/aem.01428-23

Pickett, M. L., Visotcky, A., Brazauskas, R., Ledeboer, N. A., & Drendel, A. L. (2021). Can a Clean Catch Urine Sample Be Used to Diagnose Chlamydia and Gonorrhea in Adolescent Females? *Journal of Adolescent Health*, *69*(4), 574–578. https://doi.org/10.1016/j.jadohealth.2021.02.022

Powers-Fletcher, M. (2011). A Review of the Biology and the Laboratory Diagnosis of Treponema pallidum Infections. *Journal of Continuing Education Topics & Issues*, *13*(3), 80.

Radolf, J. D., Tramont, E. C., & Salazar, J. C. (2015). Syphilis (Treponema pallidum). In *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* (pp. 2684-

2709.e4). Elsevier. https://doi.org/10.1016/B978-1-4557-4801-3.00239-3

Salle, R., Mayslich, C., Grange, P. A., Leducq, V., Ollagnier, G., Heller, U., Saule, J., Martinet, P., Robert, J.-L., Benhaddou, N., Fouere, S., & Dupin, N. (2023). Specific detection of Treponema pallidum in clinical samples: validation of a qPCR assay combining two genomic targets. *Sexually Transmitted Infections*, *99*(2), 91. https://doi.org/10.1136/sextrans-2021-055364

Shields, M., Guy, R. J., Jeoffreys, N. J., Finlayson, R. J., & Donovan, B. (2012). A longitudinal evaluation of Treponema pallidumPCR testing in early syphilis. *BMC Infectious Diseases*, *12*(1), 353. https://doi.org/10.1186/1471-2334-12-353

Shockman, S., Buescher, L. S., & Stone, S. P. (2014). Syphilis in the United States. *Clinics in Dermatology*, *32*(2), 213–218. https://doi.org/10.1016/j.clindermatol.2013.08.005

Somboonna, N., Choopara, I., Arunrut, N., Sukhonpan, K., Sayasathid, J., Dean, D., & Kiatpathomchai, W. (2018). Rapid and sensitive detection of Chlamydia trachomatis sexually transmitted infections in resource-constrained settings in Thailand at the point-of-care. *PLOS Neglected Tropical Diseases*, *12*(12), e0006900. https://doi.org/10.1371/journal.pntd.0006900

Stevens, M. P., Twin, J., Fairley, C. K., Donovan, B., Tan, S. E., Yu, J., Garland, S. M., & Tabrizi, S. N. (2010). Development and Evaluation of an *ompA* Quantitative Real-Time PCR Assay for *Chlamydia trachomatis* Serovar Determination. *Journal of Clinical Microbiology*, *48*(6), 2060–2065. https://doi.org/10.1128/JCM.02308-09

Tiwari, A., Lipponen, A., Hokajärvi, A.-M., Luomala, O., Sarekoski, A., Rytkönen, A., Österlund, P., Al-Hello, H., Juutinen, A., Miettinen, I. T., Savolainen-Kopra, C., & Pitkänen, T. (2022). Detection and quantification of SARS-CoV-2 RNA in wastewater influent in relation to reported COVID-19 incidence in Finland. *Water Research*, *215*, 118220. https://doi.org/10.1016/j.watres.2022.118220

Towns, J. M., Chow, E. P. F., Wigan, R., Fairley, C. K., Williamson, D., Azzato, F., Graves, S., Zhang, L., & Chen, M. Y. (2022). Anal and oral detection of Treponema pallidum in men who have sex with men with early syphilis infection. *Sexually Transmitted Infections*, *98*(8), 570. https://doi.org/10.1136/sextrans-2021-055370

Tshaka, T. R., Singh, R., Apalata, T. R., & Mbulawa, Z. Z. A. (2022). Aetiology of genital ulcer disease and associated factors among Mthatha public clinic attendees. *Southern African Journal of Infectious Diseases*, *37*(1). https://doi.org/10.4102/sajid.v37i1.444

Twin, J., Moore, E. E., Garland, S. M., Stevens, M. P., Fairley, C. K., Donovan, B., Rawlinson, W., & Tabrizi, S. N. (2011). Chlamydia trachomatis Genotypes Among Men Who Have Sex With Men in Australia. *Sexually Transmitted Diseases*, *38*(4), 279–285. https://doi.org/10.1097/OLQ.0b013e3181fc6944

Varma, M., Field, R., Stinson, M., Rukovets, B., Wymer, L., & Haugland, R. (2009). Quantitative real-time PCR analysis of total and propidium monoazide-resistant fecal indicator bacteria in wastewater. *Water Research*, *43*(19), 4790–4801.

https://doi.org/10.1016/j.watres.2009.05.031

Varma, R., Estcourt, C., & Mindel, A. (2013). Syphilis. In *Sexually Transmitted Diseases* (pp. 427–462). Elsevier. https://doi.org/10.1016/B978-0-12-391059-2.00017-6

Wang, C., Zheng, X., Guan, Z., Zou, D., Gu, X., Lu, H., Shi, M., & Zhou, P. (2022). Quantified Detection of Treponema pallidum DNA by PCR Assays in Urine and Plasma of Syphilis Patients. *Microbiology Spectrum*, *10*(2). https://doi.org/10.1128/spectrum.01772-21

Witkin, S. S., Minis, E., Athanasiou, A., Leizer, J., & Linhares, I. M. (2017). Chlamydia trachomatis: the Persistent Pathogen. *Clinical and Vaccine Immunology*, *24*(10). https://doi.org/10.1128/CVI.00203-17

World Health Organization. (2021). *Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: accountability for the global health sector strategies 2016–2021: actions for impact*. World Health Organization.

Wright, S. S., Kreisel, K. M., Hitt, J. C., Pagaoa, M. A., Weinstock, H. S., & Thorpe, P. G. (2022). Impact of the COVID-19 Pandemic on Centers for Disease Control and Prevention–Funded Sexually Transmitted Disease Programs. *Sexually Transmitted Diseases*, *49*(4), e61–e63. https://doi.org/10.1097/OLQ.0000000000001566

Xagoraraki, I., & O'Brien, E. (2020). Wastewater-Based Epidemiology for Early Detection of Viral Outbreaks. In D. J. O'Bannon (Ed.), *Women in Water Quality: Investigations by Prominent Female Engineers* (pp. 75–97). Springer International Publishing. https://doi.org/10.1007/978-3-030-17819-2_5

Yan, T., O'Brien, P., Shelton, J. M., Whelen, A. C., & Pagaling, E. (2018). Municipal Wastewater as a Microbial Surveillance Platform for Enteric Diseases: A Case Study for Salmonella and Salmonellosis. *Environmental Science & Technology*, *52*(8), 4869–4877. https://doi.org/10.1021/acs.est.8b00163

Zdrodowska-Stefanow, B., Ostaszewska-Puchalska, I., & Puciło, K. (2003). The immunology of Chlamydia trachomatis. *Archivum Immunologiae et Therapiae Experimentalis*, *51*(5), 289–294.

Zhao, L., Faust, R. A., David, R. E., Norton, J., & Xagoraraki, I. (2024). Tracking the Time Lag between SARS-CoV-2 Wastewater Concentrations and Three COVID-19 Clinical Metrics: A 21-Month Case Study in the Tricounty Detroit Area, Michigan. *Journal of Environmental Engineering*, *150*(1). https://doi.org/10.1061/JOEEDU.EEENG-7509

Zhao, L., Geng, Q., Corchis-Scott, R., McKay, R. M., Norton, J., & Xagoraraki, I. (2023). Targeting a free viral fraction enhances the early alert potential of wastewater surveillance for SARS-CoV-2: a methods comparison spanning the transition between delta and omicron variants in a large urban center. *Frontiers in Public Health*, *11*. https://doi.org/10.3389/fpubh.2023.1140441

Zhao, L., Zou, Y., David, R. E., Withington, S., McFarlane, S., Faust, R. A., Norton, J., &

Xagoraraki, I. (2023). Simple methods for early warnings of COVID-19 surges: Lessons learned from 21 months of wastewater and clinical data collection in Detroit, Michigan, United States. *Science of The Total Environment*, *864*, 161152. https://doi.org/10.1016/j.scitotenv.2022.161152

Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R. E., Withington, S., McFarland, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley, P., Busch, A. W. U., Norton, J., & Xagoraraki, I. (2022). Five-week warning of COVID-19 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance. *Science of The Total Environment*, 157040. https://doi.org/10.1016/j.scitotenv.2022.157040

APPENDIX

Additional details on GLWA WRRF interceptors. The three interceptors (the Oakwood-Northwest-Wayne County Interceptor (ONWI), the North Interceptor-East Arm (NIEA), and the Detroit River Interceptor (DRI)) transport raw wastewater to the Water Resource Recovery Facility (WRRF). They are sewers that convey large volumes of wastewater to the wastewater treatment facilities. For instance, the DRI lasts for approximately 12 miles, and it parallels to the Detroit River from the WRRF to the eastern part of the Detroit City. The size of the interceptor ranges from 8 to 16 ft. DRI conveys most of the Detroit's sewage to the WRRF (jaydee.us/projects/repair-and-rehabilitation-of-detroit-river-interceptor).

Additional details on the positive controls. *C. trachomatis* DNA was isolated from McCoy cells (ATCC CRL-1696) infected with *C. trachomatis* serovar D strain UW-3/Cx (ATCC VR-885) as per the manufacture's descriptions. It can be utilized to evaluate analytical sensitivity and specificity of primers and probes targeting *C. trachomatis*. Likewise, the quantitative synthetic *T. pallidum* DNA control can be used for the validation and evaluation of molecular-based assays performance targeting *T. pallidum*.

Additional details on the sensitivity and specificity of primers and probes. Previous studies utilized qPCR in testing *C. trachomatis* and reported average efficiencies of 87.4% in wastewater samples (Chin Quee, 2023) and 88.9% in clinical samples (Stevens et al., 2010). Likewise, researchers utilized qPCR or real-time PCR in testing *T. pallidum* where they observed high sensitivities and specificities of 89% to 100% in clinical samples (Heymans et al., 2010; Koek et al., 2006; Nieuwenburg et al., 2022; Salle et al., 2023).

Additional details on LOB and LOD. The Bio-Rad protocol was developed predicated on the Clinical and Laboratory Standards Institute (CLSI) protocol "EP17 Evaluation of Detection

Capability for Clinical Laboratory Measurement Procedures" (Pierson-Perry et al., 2012). The CLSI EP17 document suggested at least 60 replicate templates to determine LOBs (Pierson-Perry et al., 2012). Thus, 96 replicate templates were conducted for both types of samples to determine LOBs for *C. trachomatis* and *T. pallidum* assays, respectively.

Additional details on the uncertainty analysis. We adopted a commonly used statistical method of propagation of uncertainty to estimate the effect of each variable's uncertainties on the uncertainty of the estimated infection (W) for formula (2) for interceptors. We did not include the neighborhood selected sewersheds (EP, D3, OP) since the flow data and wastewater travel time are values adopted from historical records. The uncertainty analysis for interceptors can sufficiently demonstrate the disparities of uncertainties of each parameter in formula (2) and their impacts on estimated infections. In formula (2), we first identified the uncertainty of each independent variable, they are: $\sigma(C_{DNA})$, $\sigma(k)$, $\sigma(t)$, $\sigma(Q)$, $\sigma(\alpha)$, $\sigma(Ps)$, $\sigma(Qs)$, and $\sigma(Cs)$. When calculating the propagation of uncertainty, we would neglect the uncertainties of remaining variables when we focused on the targeted variable for calculating its uncertainty for the estimated infections. Then, we computed the partial derivatives of W with respect to each variable. The following equation is an example of computing the partial derivative of W with respect to the measured concentration C_{DNA}.

$$\partial W/(\partial C(_{DNA}))=(e\text{^}(kt)\times Q\alpha)/(Ps\times Qs\times Cs)$$

$$\partial W/\partial k=(C(_{DNA})\times t\times e\text{^}(kt)\times Q\alpha)/(Ps\times Qs\times Cs)$$

$$\partial W/\partial t=(C(_{DNA})\times k\times e\text{^}(kt)\times Q\alpha)/(Ps\times Qs\times Cs)$$

$$\partial W/\partial Q=(C(_{DNA})\times e\text{^}(kt)\times \alpha)/(Ps\times Qs\times Cs)$$

$$\partial W/\partial \alpha=(C(_{DNA})\times e\text{^}(kt)\times Q)/(Ps\times Qs\times Cs)$$

$$\partial W/\partial Ps=-(C(_{DNA})\times e\text{^}(kt)\times Q\alpha)/((Ps\text{^}2)\times Qs\times Cs)$$

$$\partial W/\partial Qs=-(C(_{DNA})\times e\text{^}(kt)\times Q\alpha)/((Qs\text{^}2)\times Ps\times Cs)$$

$$\partial W/\partial Cs=-(C(_{DNA})\times e\text{^}(kt)\times Q\alpha)/((Cs\text{^}2)\times Qs\times Ps)$$

Subsequently, the uncertainties of each variable is determined as described below. σ (C_{DNA}) is determined as the LOD for both targets (0.125). σ (k) is determined as the standard deviation (0.25) of identified k values used in the formula. σ (t) is determined as the standard deviation of transportation time of wastewater for each interceptor: ONWI (0.27), NIEA (0.53), and DRI (0.43). σ (Q) is determined as the standard deviation of flow data for each interceptor: ONWI (88.39), NIEA (114.47), and DRI (80.58). Qs is regarded as a constant and its uncertainty σ (Qs) is zero. There is a lack of research on Cs in the current literature, and we have identified it as a future research need. Finally, through the computation of derivatives of W with respect to each parameter, the following two tables summarized uncertainties of each parameter based on values used in the study. It can be seen that the uncertainties of measured concentrations are negligible compared to those of other variables since the precision of measured concentrations relied on LOD of experiments. Other variables including k, t, and Q are computed via standard deviation according to a series of collected data and they presented similar levels of uncertainties. Values of Ps were adopted from literature studies and exhibited significant variations, therefore leading to the highest contribution to uncertainties of the infection estimates.

C. trachomatis σ (W)	ONWI		NIEA		DRI	
σ (C_{DNA})	0.031	0.035	0.018	0.025	0.045	0.035
σ (K)	12.01	13.71	16.71	23.28	31.28	37.31
σ (t)	12.43	14.19	13.00	8.55	36.40	43.41
σ (Q)	11.34	12.94	14.62	20.37	13.32	15.88
σ P_S	4803.40	5483.13	2560.51	3567.21	8833.06	10534.85

T. pallidum σ (W)	ONWI		NIEA		DRI	
σ (C_{DNA})	0.014	0.016	0.008	0.012	0.021	0.025
σ (K)	7.31	8.34	13.39	18.65	11.63	13.87
σ (t)	7.56	8.63	10.42	14.51	13.54	16.14
σ (Q)	6.90	7.87	11.71	16.32	4.95	5.91
σ P_S	5579.01	6368.49	3915.79	5455.33	6271.23	7479.44

Additional details on the sensitivity analysis. We adopted the global sensitivity analysis using the R package multisensi to estimate the sensitivity and relative importance of each parameter in the model on the final infection estimates (Lamboni et al., 2011). For sensitivity analysis, data ranges of parameters were described in the section 2.6. Briefly, the decay rate term e^(kt) was denoted as "E" and is ranging from 1 to 2. The data range of Q is determined from 300 to 2500. The dilution adjustment factor was denoted as "a". For *C. trachomatis*, Ps is ranging from 0.05 and 2. For *T. pallidum*, Ps is ranging from 0.128 to 0.371. Qs is ranging from 800 to 2000 as per previous descriptions. As per the values and the details described in the section 2.6 of sensitivity model inputs, dynamics of the sensitivity indices for all parameters were demonstrated in the Figures 6S. 4., and 6S. 5., for Chlamydia and Syphilis infection estimates, respectively.

Approaches to compare concentrations between interceptors and selected sewersheds. PMMoV presents in human feces primarily due to the consumption of pepper products such as hot sauces and it remains stable in wastewater with little seasonal variation (Kitajima et al., 2018). CrAssphage is a bacteriophage that pervasively infects the human gut commensal bacteria and is excreted in feces (Greenwald et al., 2021). Both PMMoV and crAssphage were proved to be the most consistent biomarkers and human fecal indicators in wastewater, which were implemented in recent studies to normalize wastewater viral (i.e., SARS-CoV-2, norovirus GI/GII, astrovirus), bacterial (i.e., *Campylobacter jejuni*, *Clostridioides difficile*, *Salmonella* spp., *Yersinia enterocolitica*), fungal (i.e., *Blastocystis* spp.), and protozoan (i.e., *Balantidium coli*) concentrations to mitigate the influence of systematical variations due to routine WWTPs

operations, reduce background noise, account for dilution effects and enhance comparability among sites (Boehm et al., 2023; Greenwald et al., 2021; Holm et al., 2022; Rao et al., 2024). Both PMMoV and crAssphage were highly associated with large solids collected by centrifugation, which is the concentration method adopted for isolation of *C. trachomatis* and *T. pallidum* in this study (Nagarkar et al., 2022). Therefore, both targets were selected for normalizing *C. trachomatis ompA* and *T. pallidum polA* concentrations for comparing disparities between interceptors and selected sewersheds.

Table 6S. 1. Centrifugation methods for bacterial concentration from environmental and clinical samples

Bacterial target	Sample matrix	Centrifugation speed and time	References
Chlamydia trachomatis	Endocervical swab sample	≥12,000×*g* for 30 minutes at 4°C	(Somboonna et al., 2018)
	Cervical and introital specimens	10,000 rpm for 15 min	(Mania-Pramanik et al., 2006)
Salmonella Typhi	Influent wastewater grab sample	1-minute 1000×g then supernatant for 15 minutes 4000×g	(Zhou et al., 2023)
Escherichia coli	Surface natural water (river) and wastewater (raw wastewater before treatment)	8000×*g*, 10 min, 22 °C	(El Boujnouni et al., 2022)
Legionella pneumophila	Tap water samples	8150×g for 15 min or 3800×g for 30 min	(Villari et al., 1998)
Enterococcus faecalis	Wastewater samples	12,000×*g* for 1 min	(Varma et al., 2009)
Leptospira	Water samples	8000×*g*, 10 min	(Lall et al., 2016)
Escherichia coli	Cultured medium	8000g, 10 min, 4 °C 12,000×g for 10 min	(Fu et al., 2020)

Table 6S. 2. Decay rate constant k (d^{-1}) for different bacteria in the aqueous environment

Bacteria	Method	Environment	Temperature	Decay rate	Reference
E. Coli	RT-qPCR	Seawater	22–24 °C	0.06-1.47	(Zhang et al., 2015)
Enterococci				0.18-0.76	
C. perfringens				0-0.77	
Enterococci (intestinal)	Membrane filtration with incubation on selective media	Seawater	4–20 °C	0.03-1.05	(Eregno et al., 2018)
E. Coli	IDEXX Colilert 18 Quanti-Tray/2000			0.05-1.13	
Campylobacter	Calculations using the Arrhenius equation based on data collected from published studies	Wastewater	20 °C	0.17-0.19	(Guo et al., 2022)
		Fresh water		1.72-1.88	
		Saline water		1.15-2.75	
Salmonella		Wastewater		0.4-0.52	
		Fresh water		0.37-2.37	
		Saline water		0.74-1.06	
Bacteroidales	Real-time PCR	Seawater	18.3-18.7 °C	0.95-1.11	(Jeanneau et al., 2012)
		Fresh water		1.37-1.41	
Bifidobacterium adolescentis		Seawater		0.62-0.64	
		Fresh water		0.62-0.7	
E. coli (culturable)	Culture	Seawater		1.24-1.4	
		Fresh water		0.39-0.43	
Enterococci (culturable)		Seawater		0.56-0.88	
		Fresh water		0.67-0.91	
E. coli (culturable)	Culture	Seawater	15.06-19.22 °C	0.5	(Mattioli et al., 2017)
Enterococci (culturable)				0.65	
Enterococci	RT-qPCR			0.3	

Table 6S. 3. Wastewater in-sewer travel time in GLWA interceptors

Interceptor (hours)	Weighted Average	Minimum	Maximum
DRI	12.3	0.2	41.8
NIEA	22.5	0.7	51.2
ONWI	8.6	0.1	25.9

Table 6S. 4. Temperature and pH of wastewater samples collected from the neighborhood sewersheds including EP, D3, and OP

Week	Sample site (duplicate samples)	Sample Date	Temperature (°C)	pH
1/8/24	OP	1/10/24	pH/Temp sensor missing	pH/Temp sensor missing
	OP			
	D3		pH/Temp sensor missing	pH/Temp sensor missing
	D3			
	EP		pH/Temp sensor missing	pH/Temp sensor missing
	EP			
1/22/24	OP	1/24/24	11.1	7.75
	OP			
	D3		12.1	7.41
	D3			
	EP		13.1	7.35
	EP			
1/29/24	OP	1/31/24	12.5	7.75
	OP			
	D3		7.6	7.72
	D3			
	EP		12.3	7.53
	EP			
2/5/24	OP	2/7/24	14.3	7.45
	OP			
	D3		14	7.57
	D3			
	EP		14.3	7.34
	EP			
2/12/24	OP	2/14/24	pH/Temp sensor missing	pH/Temp sensor missing
	OP			
	D3		pH/Temp sensor missing	pH/Temp sensor missing
	D3			
	EP		pH/Temp sensor missing	pH/Temp sensor missing
	EP			
2/19/24	OP	2/21/24	14.9	7.53
	OP			
	D3		15.8	7.62
	D3			
	EP		16.6	7.35
	EP			
2/26/24	OP	2/28/24	12.1	7.78
	OP			
	D3		10.5	7.8
	D3			
	EP		8.7	7.47
	EP			

Table 6S. 4. (cont'd)

3/4/24	OP	3/6/24	11.9	7.82
	OP			
	D3		13.1	7.77
	D3			
	EP		11.4	7.64
	EP			
3/11/24	OP	3/13/24	19	7.71
	OP			
	D3		15.8	7.53
	D3			
	EP		21.7	7.6
	EP			
3/18/24	OP	3/20/24	9.5	7.63
	OP			
	D3		9	7.61
	D3			
	EP		11	7.49
	EP			
3/25/24	OP	3/27/24	10.3	7.61
	OP			
	D3		11	7.66
	D3			
	EP		11.3	7.54
	EP			
4/1/24	OP	4/3/24	14.1	7.74
	OP			
	D3		12.9	7.65
	D3			
	EP		15.5	7.63
	EP			
4/8/24	OP	4/10/24	16.8	7.58
	OP			
	D3		17	7.66
	D3			
	EP		18.4	7.48
	EP			
4/16/24	OP	4/18/24	18.3	7.65
	OP			
	D3		16.8	7.52
	D3			
	EP		19.9	7.58
	EP			
4/22/24	OP	4/24/24	13.9	7.64
	OP			
	D3		13.7	7.6
	D3			
	EP		14.7	7.45
	EP			

Table 6S. 5. Positive detection of *C. trachomatis* in clinical samples excreted from patients with suspected diseases (* marked values are used in formula 2)

Type of samples	Presence (%)	References
Female urine	5.3*	(Božičević et al., 2011)
	6.6*	(Møller et al., 2010)
	6.0*	(Møller et al., 2008)
All gender urine	6.2*	(Božičević et al., 2011)
	9.0*	(Møller et al., 2008)
	9.1*	(Møller et al., 2010)
Male urine	13.4*	(Møller et al., 2010)
	14.4*	(Møller et al., 2008)
	7.3*	(Božičević et al., 2011)
Female urine combined with vaginal swab	8.3	(Møller et al., 2008)
Male first voided urine specimens	8.9*	(Mania-Pramanik et al., 2006)
All gender genital ulcer swabs	10.5	(Tshaka et al., 2022)
Female cervical and introital specimens	12.3	(Mania-Pramanik et al., 2006)
Female vaginal swabs	12.2	(Ngobese et al., 2022)
	18.4	(Pickett et al., 2021)

Table 6S. 6. Positive detection of *T. pallidum* in clinical samples excreted from patients with suspected diseases (* marked values are used in formula 2)

Type of samples	Presence (%)	References
Male lesion swabs	0.3	(Dubourg et al., 2015)
All gender genital ulcer swabs	6.7	(Tshaka et al., 2022)
Male urine (early latent syphilis)	12.8*	(Nieuwenburg et al., 2022)
All gender ano-rectal swabs (primary syphilis)	13.0	(Heymans et al., 2010)
All gender genital ulcer swabs	13.4	(Koek et al., 2006)
Male ano-rectal swabs (secondary syphilis)	18.6	(Nieuwenburg et al., 2022)
Male genital, anal or oral ulcers	20.8	(Shields et al., 2012)
Male ano-rectal swabs (early latent syphilis)	24.4	(Nieuwenburg et al., 2022)
All ano-rectal swabs (secondary syphilis)	25.6	(Heymans et al., 2010)
Male urine (secondary syphilis)	37.1*	(Nieuwenburg et al., 2022)
Male genital, anal or oropharyngeal ulcers	47.0	(Glatz et al., 2014)

Table 6S. 7. Incubation time of *C. trachomatis*

Incubation time	References
5-10 days	(O'Connell & Ferone, 2016)
7-21 days	lacounty.gov, ndhealth.gov, iowa.gov, epi.utah.gov, odh.ohio.gov
7-14 days	vic.gov.au, nsw.gov.au, gov.mb.ca
mean of 21 days	mass.gov
7-28 days	(Jones & Lopez, 2014)
2-60 days	ashm.org.au
1-3 weeks	oregon.gov
1-5 weeks	nj.gov
14-21 days, maximum 6 weeks	healthunit.org, gnb.ca
7-14 days for trachoma and genital infections, 3-30 days for LGV	Public Health Agency of Canada (canada.ca)

Table 6S. 8. Demographic characteristics of the neighborhood sewersheds

Sites	Zip	Population	Density	Black	Hispanic	White	Poverty	Total household income
EP	48021	2400	9	37%	5%	54%	5%	56450
D3	48235	1300	10	95%	0%	2%	44%	22100
OP	48237	2270	8	85%	3%	6%	15%	51680

Notes: units for Area, Density, and Total household income are acres, people per acre, and USD, respectively.

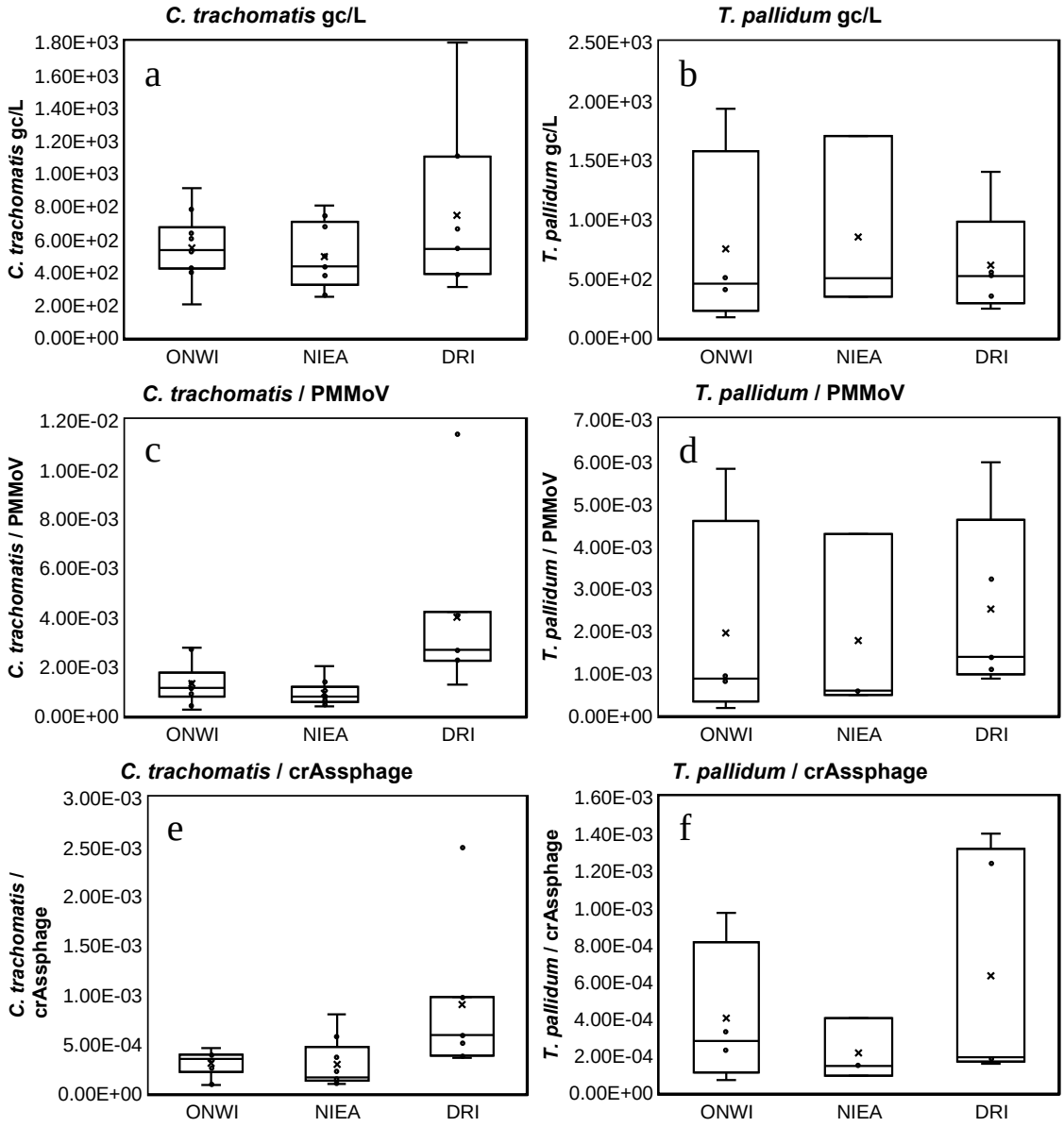


Figure S6. 1. Box plots for *C. trachomatis* and *T. pallidum* concentrations: non-normalized (a, b), normalized by PMMoV (c, d), normalized by crAssphage (e, f) for interceptors

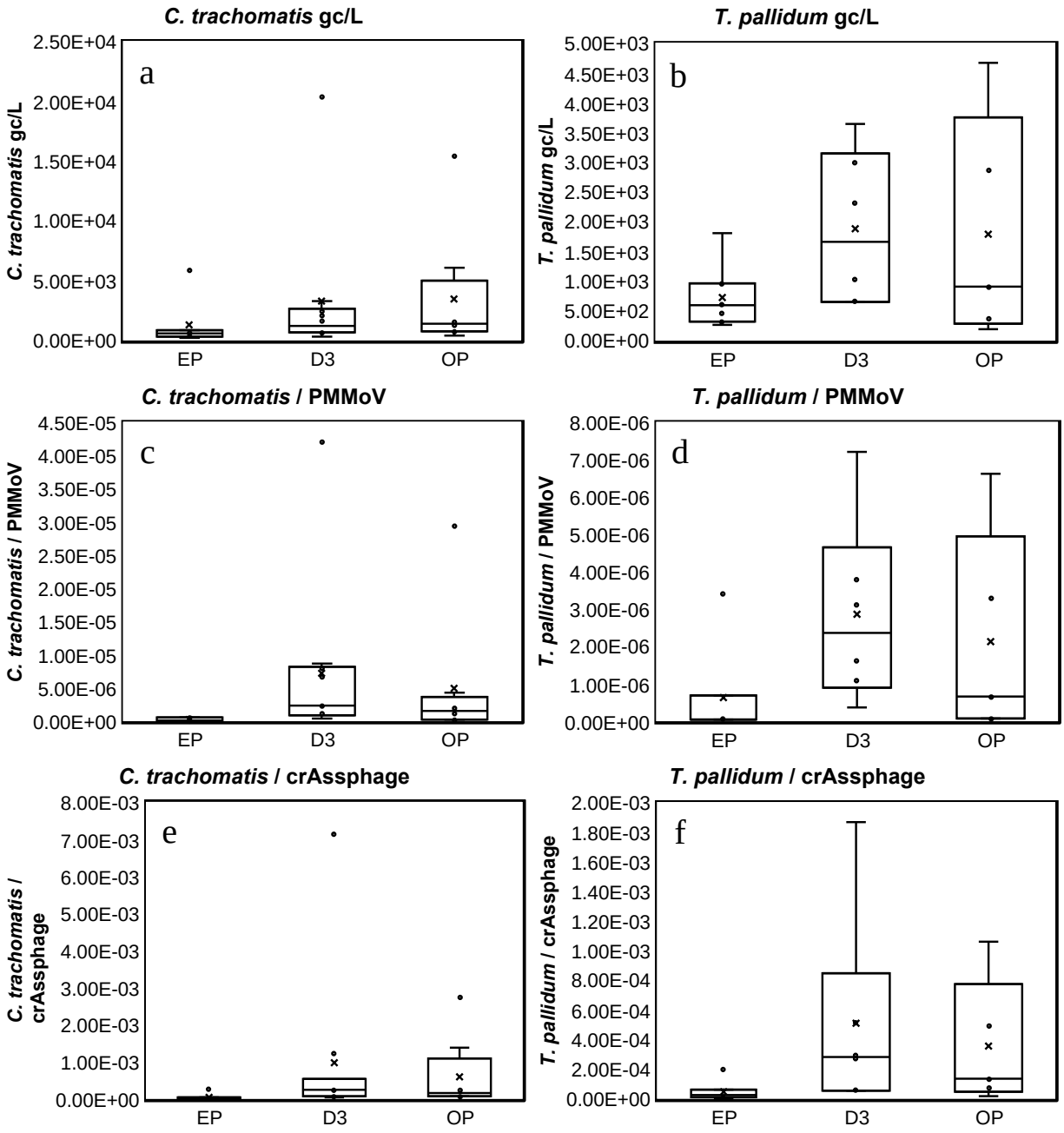


Figure S6. 2. Box plots for *C. trachomatis* and *T. pallidum* concentrations: non-normalized (a, b), normalized by PMMoV (c, d), normalized by crAssphage (e, f) for selected sewersheds

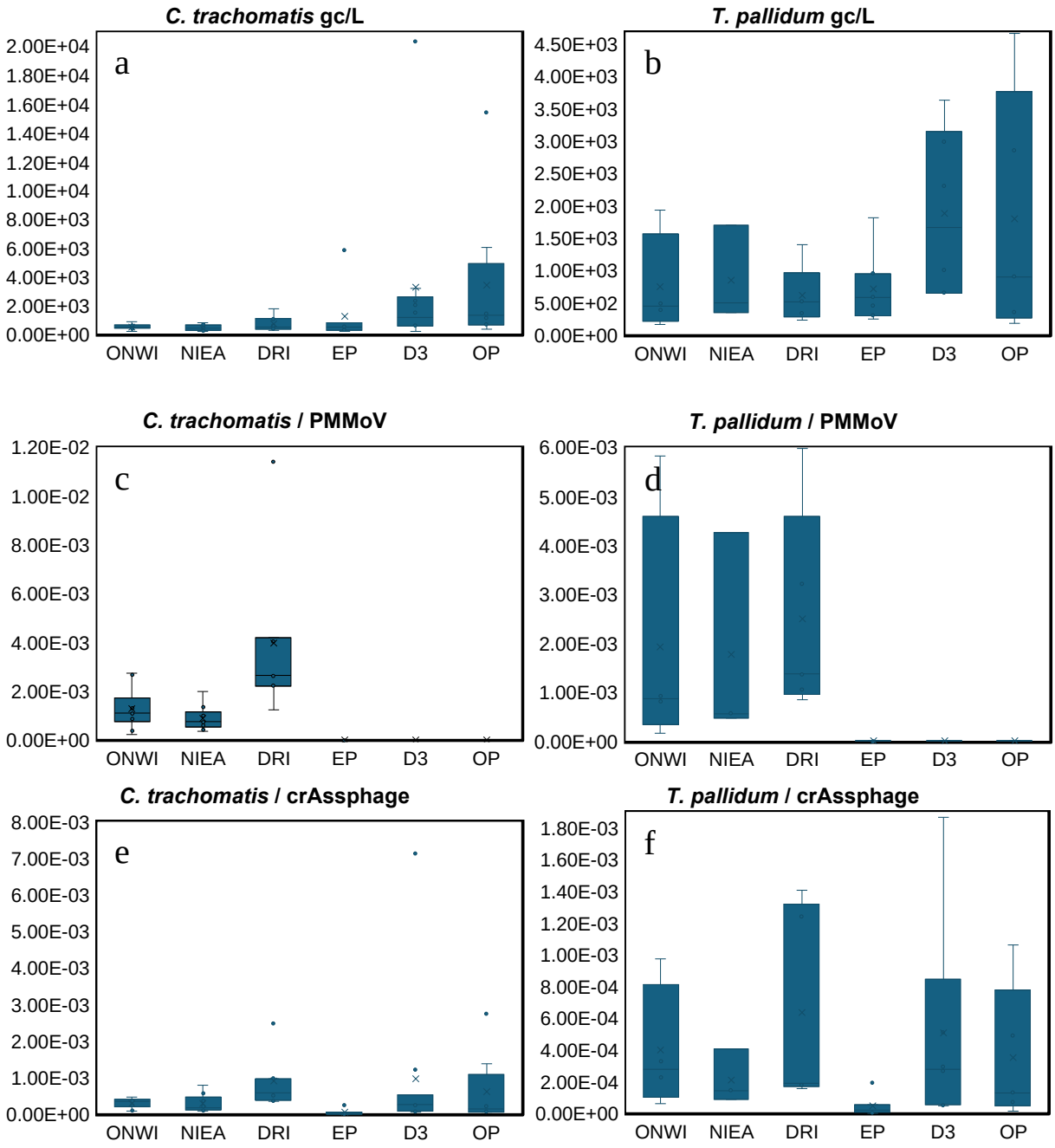


Figure S6. 3. Comparison between interceptors and selected sewersheds for *C. trachomatis* and *T. pallidum* concentrations: non-normalized (a, b), normalized by PMMoV (c, d), normalized by crAssphage (e, f)

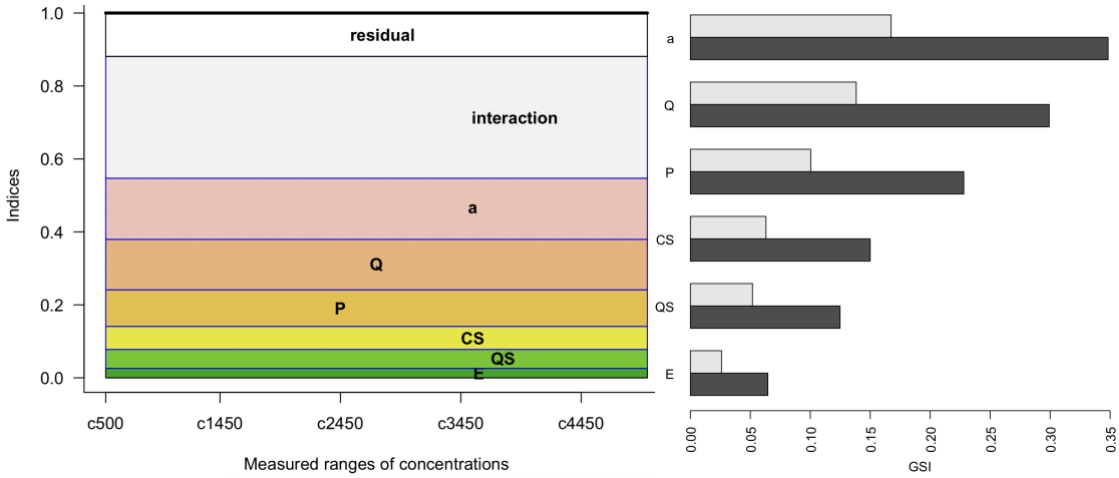


Figure S6. 4. Dynamics of the sensitivity indices of Chlamydia infection estimates with indices normalized to 1, and a bar plot of the PCA generalized sensitivity indices of the infection estimates model for Chlamydia

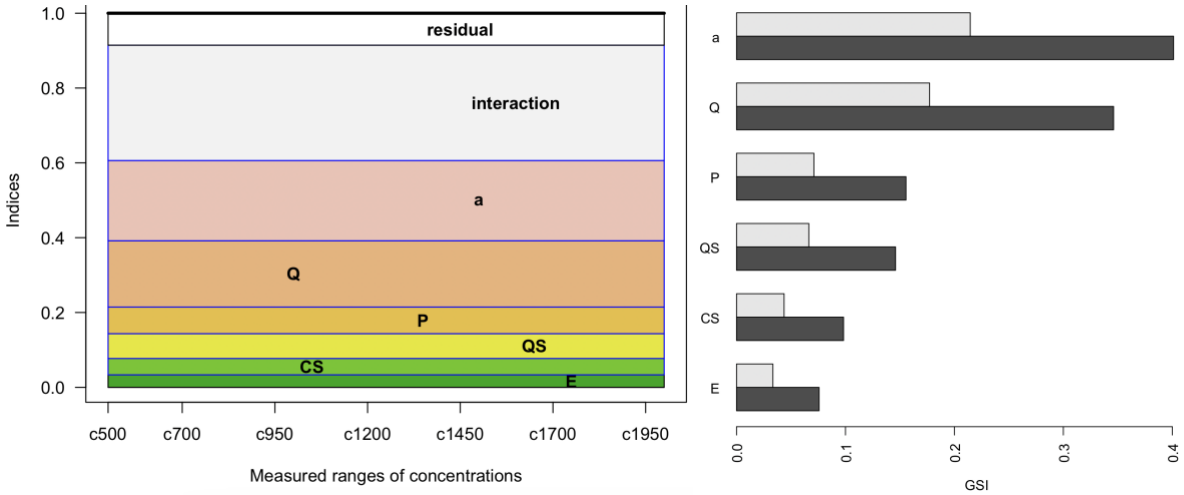


Figure S6. 5. Dynamics of the sensitivity indices of Syphilis infection estimates with indices normalized to 1, and a bar plot of the PCA generalized sensitivity indices of the infection estimates model for Syphilis

CONCLUSIONS AND SIGNIFICANCE

Wastewater surveillance or wastewater-based epidemiology (WBE) has undergone significant advancements over the past two decades. The COVID-19 pandemic, in particular, has acted as a catalyst, expediting its development and applications. This dissertation extensively explores the laboratorial, technical, and translational methodologies of implementing WBE and makes significant advancements in WBE in the following aspects:

First, in the beginning of the COVID-19 pandemic, even WBE has gradually become recognized as an effective method for the early detection of outbreaks, technological and translational advancements on predicting fluctuations of COVID-19 incidences have not yet been made. In chapter 1, we implemented the U.S. EPA Virus Adsorption-Elution (VIRADEL) method that targets the viruses in supernatant wastewater to circumvent the potential input of "old" viruses via desorption of settled viruses during high flows. Predicated on the measured SARS-CoV-2 RNA concentrations in wastewater using RT-ddPCR, we built and deployed four mathematical models that predicted the fluctuations of COVID-19 cases five weeks before clinical reporting. Autoregressive models with time effect and vector autoregressive models are two examples of models that precisely predict the fluctuations of impending COVID-19 cases. We systematically evaluated the time lag between peaks in measured concentrations of SARS-CoV-2 in wastewater and peaks in reported COVID-19 cases and, for the first time, proposed a time lag mechanistic model. Both the prediction models and the time lag mechanistic model allow researchers to accurately depict fluctuations of future disease incidences before its clinical reporting.

Second, while intricate models have been developed to determine early warnings based on wastewater surveillance data, there is an exigent need for straightforward, rapid, broadly applicable methods for health departments and partner agencies to implement. In chapter 2, we

aimed to develop and evaluate such early-warning methods and clinical-case peak-detection methods based on wastewater surveillance data to mount an informed public health response. We designed eight statistical methods to identify early warnings for surging COVID-19 incidences in the TCDA. We demonstrated the utility of these methods for providing early warnings for COVID-19 incidences in real scenarios, with their counterpart accuracies evaluated by hit rates, which can reach 100% accuracy to capture surges of COVID-19 incidences. The proposed methods were utilized by health agencies including the World Health Organization on August 30th, 2023, in its *SARS-CoV-2 Variant Risk Evaluation Framework* and local health departments in Michigan, U.S., and Ontario, Canada, to capitalize on wastewater surveillance data to assess trends of COVID-19 and RSV and implement quick public health responses to future epidemics.

Third, there is a vast number of methods used to recover and concentrate SARS-CoV-2 viral RNA from wastewater, including Virus Adsorption-Elution (VIRADEL), polyethylene glycol precipitation (PEG), ultrafiltration, ultracentrifugation, concentrating pipette, filtration, etc. However, no studies yet compared concentration methods in terms of their early warning potential. In chapter 3, we presented a comparison of three primary concentration methods (VIRADEL, PEG and PES filtration) to detect SARS-CoV-2 viral RNA in wastewater, in relation to COVID-19 cases amid the transition from Delta to Omicron Variants of Concerns (VOCs) circulating in the TCDA. We identified that VIRADEL method can be used to enhance the early-warning potential of wastewater surveillance applications albeit drawbacks may include the need to process large amount of wastewater to concentrate sufficiently free and suspended virus for detection.

Fourth, numerous studies have reported time lags between SARS-CoV-2 RNA concentrations in wastewater and confirmed clinical COVID-19 cases. Only a limited number of studies have examined the time lags between SARS-CoV-2 RNA concentrations in wastewater

and other clinical metrics. Chapter 4 contributes valuable insights into the field of WBE by estimating the time lags between SARS-CoV-2 concentrations and clinical metrics (confirmed cases, hospitalizations, and ICU admissions), before and during the COVID-19 Omicron surge in TCDA. We found that PEG did not provide early warnings for three clinical metrics for nearly all nonnormalized and normalized conditions during the Omicron surge. However, VIRADEL demonstrated its potential for early warnings of total confirmed cases, hospitalizations, and ICU admissions, with leading time lags provided before the Omicron surge. During the Omicron surge, VIRADEL's leading time lags were reduced, and the early warning potential of ICU admissions was diminished. The leading time lags can provide a critical window for hospital systems and public health authorities to properly prepare for pending disease outbreaks. Chapter 4 contributes the understanding of the temporal relationship between wastewater viral concentrations and various clinical metrics as well as the parameters potentially affecting the relationship. Such understanding can improve the effective translation of wastewater surveillance data, improve WBE models, and ultimately, enhance public health preparedness.

Fifth, despite the numerous investigations regarding the comparison among U.S. CDC N1, N2 genes and SC2 assay, few studies directly compared their efficiency in correlating with and predicting clinical cases. Published studies have only conducted limited comparative analyses of SARS-CoV-2 RNA concentrations quantified with N1, N2 genes and SC2 assay as well as clinical cases. Hence, in chapter 5, we compared the correlations and examined the similarities/dissimilarities among N1, N2 genes and SC2 assay through comprehensive statistical approaches. We established models to predict COVID-19 cases based on SARS-CoV-2 RNA concentrations quantified with the three assays and identified N2 as the optimal assay for predicting COVID-19 in TCDA.

Finally, wastewater surveillance has demonstrated its substantial potential of monitoring and predicting infections of communicable diseases. Yet, most studies implemented wastewater surveillance to monitor viral communicable diseases including COVID-19, RSV, influenza, and norovirus. Limited studies attempted to monitor widespread bacterial communicable diseases particularly sexually transmitted diseases (STDs) including Chlamydia and Syphilis despite their rapidly increasing trends and underreported infections in the U.S. Henceforth, in chapter 6, we optimized molecular biology laboratory methods including DNA extraction and ddPCR to target *Chlamydia trachomatis* and *Treponema pallidum* that cause Chlamydia and Syphilis, respectively. We also designed a workflow to implement wastewater surveillance tracking Chlamydia and Syphilis in the TCDA. The magnitude of *Chlamydia trachomatis* and *Treponema pallidum* concentrations were observed higher in neighborhood sewersheds than interceptors. Predicated on extensive investigations on bacteria shedding and wastewater surveillance data, we back-estimated the infections of Chlamydia and Syphilis where results indicated likely underreported conditions, further highlighting the values of WBE on tracking STDs. Different concentrations of *C. trachomatis* and *T. pallidum* in wastewater demonstrated disparities of varying characteristics of sewersheds and they were potentially related to socioeconomic status. In chapter 6, we developed and implemented one of the first bacterial wastewater surveillance as a screening tool to monitor STDs, and estimated bacterial infections based on bacterial shedding and wastewater surveillance data.