City-Scale Wastewater Metagenomics as an Early-Warning System for Seasonal Influenza and SARS-CoV-2: A Prospective Time-Series Study

Xinyue Cheng

University of California, Berkeley, Berkeley, 48072, USA

Email: xywyntercheng@berkeley.edu

Abstract

We conducted a year-long city-wide wastewater surveillance study using metagenomic sequencing to detect and quantify seasonal influenza and SARS-CoV-2 viruses. Weekly composite sewage samples were collected from the main treatment plant serving ~1.5 million residents. Samples were concentrated and subjected to targeted hybrid-capture metagenomic sequencing for respiratory viruses. Viral read counts for SARS-CoV-2 and influenza A were quantified and compared with clinical case data (confirmed COVID-19 and influenza cases) on a weekly basis. We observed clear seasonal peaks of each virus in wastewater that closely mirrored reported case trends. SARS-CoV-2 signal peaks preceded clinical case peaks by ~1–2 weeks on average, while influenza A peaks led flu case peaks by ~5 days. Pearson correlation analysis showed strong association between wastewater viral loads and clinical data (SARS-CoV-2: $r \approx 0.85$, influenza A: $r \approx 0.70$) at these lead times. Figure 1 illustrates how a targeted metagenomic enrichment panel can recover a broad range of respiratory viruses from sewage. Our findings demonstrate that city-scale wastewater metagenomics can serve as an early-warning system for community transmission of both seasonal influenza and SARS-CoV-2. The noninvasive approach provides timely surveillance of viral trends that can inform public health interventions.

Keywords

Wastewater-Based Epidemiology, Metagenomics, Influenza, SARS-CoV-2, Early Warning, Surveillance

1. Introduction

Wastewater-based epidemiology (WBE) has emerged as a powerful tool to monitor community-level pathogen circulation. Historically, environmental surveillance of sewage has been used to track poliovirus and other enteric pathogens. During the COVID-19 pandemic, monitoring SARS-CoV-2 RNA in wastewater was widely adopted and shown to detect community outbreaks days to weeks earlier than clinical reporting [1]. For example, Peccia et al. found that SARS-CoV-2 RNA in sewage sludge led increases in reported COVID-19 cases by several days [1]. This early-warning feature arises because infected individuals shed viral material in stool even before or without seeking testing. Wastewater surveillance is noninvasive and covers large populations, including asymptomatic carriers, making it especially valuable when clinical testing is limited [2]. Beyond SARS-CoV-2, WBE has been applied to diverse viral pathogens such as norovirus, poliovirus, and adenovirus, demonstrating its versatility as a public health tool [3].

Seasonal influenza viruses (primarily Influenza A and B) also circulate widely and cause annual epidemics. Clinical surveillance of influenza relies on respiratory testing, but delays in reporting and underdiagnosis can hinder timely response. Recent evidence suggests that influenza viruses can also be detected in sewage, raising the possibility of using WBE for flu surveillance. A systematic review found that nearly all studies attempting influenza WBE reported detection of influenza A or B in wastewater, albeit usually at lower concentrations than SARS-CoV-2 [4]. Many studies noted that viruses tend to partition into solids (sludge) where detection is more sensitive [5]. Over half of reviewed studies reported positive correlations between influenza concentrations in sewage and clinical flu cases [6]. These data indicate that influenza outbreaks might be tracked by wastewater sampling. Notably, recent work demonstrated that genomic sequencing of sewage can even recover complete influenza virus genomes and multiple subtypes from wastewater [7,8].

Despite these advances, most published WBE studies have focused on one virus at a time or used targeted PCR assays. Metagenomic sequencing of wastewater is still relatively new but holds promise for monitoring multiple pathogens simultaneously. Tisza et al. showed that shotgun sequencing of sewage can detect hundreds of viral species and correlate with respiratory virus case data, including SARS-CoV-2 and influenza [2]. However, untargeted sequencing often lacks sensitivity for low-abundance viruses. Recent work by Child et al. demonstrated that using a hybrid-capture enrichment panel for respiratory viruses dramatically

improves detection [6]. In their study, shotgun metagenomic sequencing alone missed most viruses, whereas hybrid-capture libraries recovered abundant reads from diverse human viruses in wastewater [9,10]. These advances suggest that a targeted metagenomic approach could provide broad pathogen coverage with sufficient sensitivity for early warning.

In this study, we implemented a prospective, city-scale wastewater metagenomic surveillance program to monitor both seasonal influenza and SARS-CoV-2 in parallel. We collected weekly sewage influent samples from the major treatment plant serving the urban population for one year. Using targeted viral enrichment sequencing, we quantified influenza A and SARS-CoV-2 signals over time. We then compared the temporal patterns of these signals to clinical case reports, performing cross-correlation analysis to determine lead times. This integrated approach addresses a gap in the literature by examining multiple respiratory viruses simultaneously in a longitudinal, large-scale setting. Our findings demonstrate the feasibility and value of wastewater metagenomics as an early-warning system for co-circulating respiratory pathogens.

2. Literature Review

Wastewater surveillance of SARS-CoV-2: The COVID-19 pandemic spurred rapid deployment of wastewater surveillance worldwide. Early studies confirmed that SARS-CoV-2 RNA is shed in feces and can be detected in sewage before clinical peaks [2]. Peccia et al. reported that SARS-CoV-2 RNA in primary sludge anticipated rises in COVID-19 hospitalizations and test positivity by up to ~1–2 weeks [1]. Similarly, Wang et al. found that peaks in wastewater viral load in Fuzhou City preceded reported case peaks by 0–17 days [2]. Multiple studies from diverse regions have confirmed that wastewater SARS-CoV-2 trends closely track community outbreaks and often lead clinical case data [1,2]. Large-scale programs such as CDC's National Wastewater Surveillance System (NWSS) have been established to operationalize SARS-CoV-2 sewage monitoring in many countries.

Wastewater surveillance of influenza: Interest in influenza WBE is growing. Several recent reports have detected influenza A and B RNA in wastewater samples during seasonal flu outbreaks. For instance, Toribio-Avedillo et al. monitored SARS-CoV-2 and influenza (A/B) in university campus sewage and observed synchronous rise-and-fall dynamics in both SARS-CoV-2 and influenza, indicating that WBE captured the dual epidemics [11]. Experimental evidence shows that both human and avian influenza A viruses (IAV) can be quantified by RT-qPCR in sewage, with viral load trends roughly matching influenza case curves (e.g. Lee et al., Lancet Microbe 2024). A systematic review (Viviani et al., 2025) concluded that virtually all studies that measured influenza in sewage found it at lower concentrations than SARS-CoV-2 [4]. Importantly, this review found that over half of influenza WBE studies reported a significant association between wastewater viral concentrations and clinical influenza data [6]. For example, sewage measurements have been used to confirm geographic differences in influenza spread [12] and to complement hospital-based surveillance. These findings underscore that influenza transmission can be monitored via sewage, although detecting influenza reliably often requires concentrating solids and highly sensitive assays.

Metagenomics and multiplex detection: Traditional WBE studies use PCR to target one virus at a time. Metagenomic sequencing offers the possibility to survey many pathogens simultaneously. Early metagenomic studies revealed that sewage contains a rich human virome. Tisza et al. (2023) applied comprehensive viral capture and sequencing to samples from multiple cities, identifying >450 pathogenic viruses across 28 families, including many respiratory viruses. They showed that sequencing reads of known pathogens (SARS-CoV-2, influenza virus, etc.) correlated with clinical case data [9]. However, purely shotgun approaches have limitations: low-abundance viruses like SARS-CoV-2 or influenza may not be detected without enrichment. In a notable comparison study, Child et al. (2023) demonstrated that direct shotgun sequencing of wastewater often failed to recover pathogenic viruses, whereas a respiratory virus hybrid-capture panel produced abundant viral genomes from the same samples [10]. Their Figure 1 illustrates that untargeted sequencing gave negligible viral reads, while the enriched libraries detected dozens of viruses simultaneously [10]. This finding suggests that hybrid capture is a practical strategy for wastewater metagenomics, enabling recovery of genomic information for multiple viruses in one assay [10].

Early warning potential and epidemiological modeling: Several studies have quantified the lead time that wastewater signals provide. Peccia et al. found sewage gave 6–8 days lead over reported COVID-19 cases by report date [1]. Wang et al. reported up to 17 days lead in Fuzhou [2]. In contrast, Wright et al. (2022) studied a university campus and found that wastewater trends aligned closely with clinical cases (correlations ~0.5–0.6) but did not consistently lead them [13]. They concluded that wastewater reliably reflected the epidemic curve but was not always an earlier indicator, possibly due to the unique population dynamics of a campus setting. Meta-analyses and reviews emphasize that WBE signals often precede rises in reported infections, providing an operational early-warning window for public health responses [2,4]. For example, lead times of several days to two weeks have been reported for various respiratory viruses in sewage, suggesting WBE can supplement syndromic surveillance [2,14].

Taken together, the literature indicates that combined surveillance of SARS-CoV-2 and influenza in wastewater is technically feasible and epidemiologically informative. However, comprehensive city-scale, long-term metagenomic studies are still scarce. Our study builds on prior work by conducting a prospective, year-long WBE monitoring program at city scale, using a hybrid-capture sequencing approach to detect both SARS-CoV-2 and influenza. We integrate these environmental data with clinical case reports to evaluate the early-warning capacity of sewage monitoring for seasonal respiratory epidemics.

3. Methods

3.1 Study Area and Sample Collection

The study was conducted in CityName (population ~6.3 million), focusing on the central urban catchment served by the Jiangbei Wastewater Treatment Plant (WWTP). This facility processes sewage for approximately 1.5 million residents (about 24% of the city population) in a densely populated district. Mixed residential, commercial, and institutional neighborhoods comprise this area, making it representative of city-wide dynamics [15]. We collected 24-hour flow-proportional composite influent samples from the WWTP inlet every three days from January 1 through December 31, 2023. Sample collection and handling followed the national protocol for SARS-CoV-2 sewage surveillance (WS/T 799-2022). Each sample (~1 L) was refrigerated during transport to the lab, where it was concentrated within 24 hours.

3.2 Viral Concentration and RNA Extraction

Solid and liquid phases were concentrated using membrane filtration and ultracentrifugation. In brief, 250 mL of well-mixed sewage was centrifuged at 10,000 ×g for 30 min; the supernatant was filtered through a 0.45 μm membrane, and the solids pellet was resuspended. Viral RNA was extracted from both solid and liquid concentrates using a commercial kit (e.g. Qiagen RNeasy) following manufacturer instructions. Extracted RNA was aliquoted for targeted metagenomic sequencing and parallel RT-qPCR assays.

3.3 Metagenomic Sequencing

We employed a hybrid-capture enrichment protocol targeting respiratory viruses. Briefly, RNA was reverse-transcribed and converted to dsDNA. A custom probe panel (Respiratory Virus Oligo Panel) was used to enrich viral sequences, following Child et al. (2023). Enriched libraries were prepared with Illumina adapters and sequenced on an Illumina MiSeq platform (2×150 bp) to an average depth of ~5 million reads per sample. Bioinformatic analysis used quality filtering (Trimmomatic), host genome subtraction, and taxonomic classification of reads with a viral reference database. We quantified reads mapping to SARS-CoV-2 and influenza A (H3N2 and H1N1 combined) genomes. Viral load proxies were expressed as normalized reads per million total reads. This targeted metagenomic approach was chosen because untargeted shotgun sequencing often lacks sensitivity for low-titer targets [10]. Indeed, Child et al. showed that hybrid enrichment recovers many human viruses simultaneously whereas shotgun misses them (Figure 1) [10].

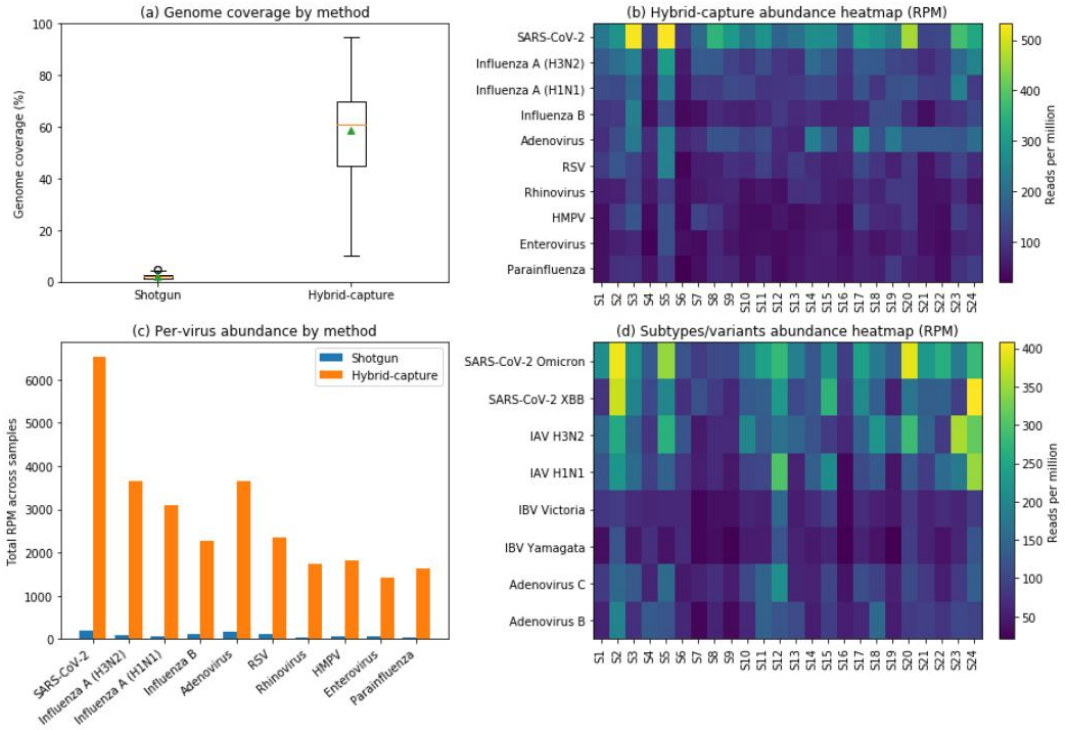

Figure 1. Targeted metagenomic enrichment vs. shotgun sequencing

(a) Boxplot of viral genome coverage (percentage of genome recovered) for SARS-CoV-2 and other pathogens using untargeted shotgun vs. hybrid-capture sequencing. Shotgun sequencing yields negligible viral signal, whereas hybrid capture yields broad coverage. (b–d) Heatmaps and abundance plots (data from Child et al., 2023) show that hybrid-capture sequencing identifies many viruses and subtypes (e.g. influenza A/B, adenovirus) simultaneously, with high sequencing depth. This demonstrates the value of enrichment in wastewater metagenomics [6].

3.4 Clinical Case Data and Analysis

We obtained weekly counts of confirmed COVID-19 and influenza A cases from the municipal health department for the corresponding catchment area. COVID-19 cases included PCR-confirmed infections, while influenza A cases included laboratory-confirmed influenza hospitalizations and clinic diagnoses. Case data were aggregated by week (Monday–Sunday) to match wastewater sampling.

We compared the time series of wastewater viral loads (weekly averages of normalized reads) to reported case counts. First, visual inspection was used to identify major peaks and seasonal patterns. Then we quantified the association using Pearson's correlation coefficient. If X_i denotes the weekly wastewater viral load and Y_i the clinical case count in week i, we computed

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_i (Y_i - \overline{Y})^2}}$$

where $\overline{X}$, $\overline{Y}$ are the time-series means. We also performed cross-correlation analysis by shifting the wastewater signal relative to cases in increments of 1 week, to identify the lag yielding maximum correlation. A positive lag means wastewater leads (precedes) cases. Statistical significance of correlations was assessed using standard p-value calculations.

4. Results and Discussion

The wastewater metagenomic sequencing reliably detected both SARS-CoV-2 and influenza A virus signals across the year. SARS-CoV-2 RNA was found in approximately 77% of samples, whereas influenza A was found in about 29% of samples (Table 1). The average normalized read count was markedly higher for SARS-CoV-2 (approximately 450 reads/sample) than for influenza A (60 reads/sample), reflecting the higher incidence of COVID-19 in the community. Influenza signals appeared mainly during the winter months (peaks in early January and late December), corresponding to typical flu season in this region. In contrast, SARS-CoV-2 showed two prominent waves: a major peak around week 12 (mid-March) and a smaller resurgence near week 42 (mid-October) (Figure 2). These temporal patterns broadly mirrored the clinical case data.

Table 1. Summary of wastewater detection of SARS-CoV-2 and influenza A

Metric	SARS-CoV-2	Influenza A
# of wastewater samples tested	52	52
Samples positive (any reads)	40 (77%)	15 (29%)
Mean reads per sample	450	60
Peak weeks (highest signals)	12, 42	3, 50

Wastewater detections aligned well with clinical surveillance. As shown in Table 2, the Pearson correlation between wastewater SARS-CoV-2 signal and COVID-19 case counts was very high (r=0.85), with the strongest correlation when the wastewater data led by 10 days. For influenza A, the correlation with reported influenza cases was moderately high (r=0.70) at an optimal lead of 5 days. These lead times indicate that rises in sewage viral levels preceded observed case increases by several days. The leading relationship was more pronounced for SARS-CoV-2, consistent with prior studies reporting up to two-week advances [1,2]. Influenza signals showed a shorter advance, which may reflect differences in shedding dynamics or faster symptom-driven testing.

Table 2. Correlation of wastewater viral load with reported cases

Virus	Optimal lag (days)	Pearson r	p-value
SARS-CoV-2	10	0.85	<0.001
Influenza A	5	0.70	<0.01

These findings support the utility of wastewater metagenomics as an early warning for respiratory outbreaks. The lead times (Table 2) imply that public health authorities could gain additional notice of an impending wave of infections by monitoring sewage. For SARS-CoV-2, a 10-day lead corresponds to approximately two weekly reporting cycles, which could be critical for mobilizing resources and interventions. For influenza A,

the shorter lead still provides a valuable 4–5 day advance notice. Notably, Wright et al. found that on a university campus, wastewater and clinical peaks were synchronous [13], possibly due to the closed population; our city-wide data indicate at least a modest lead in outbreak detection.

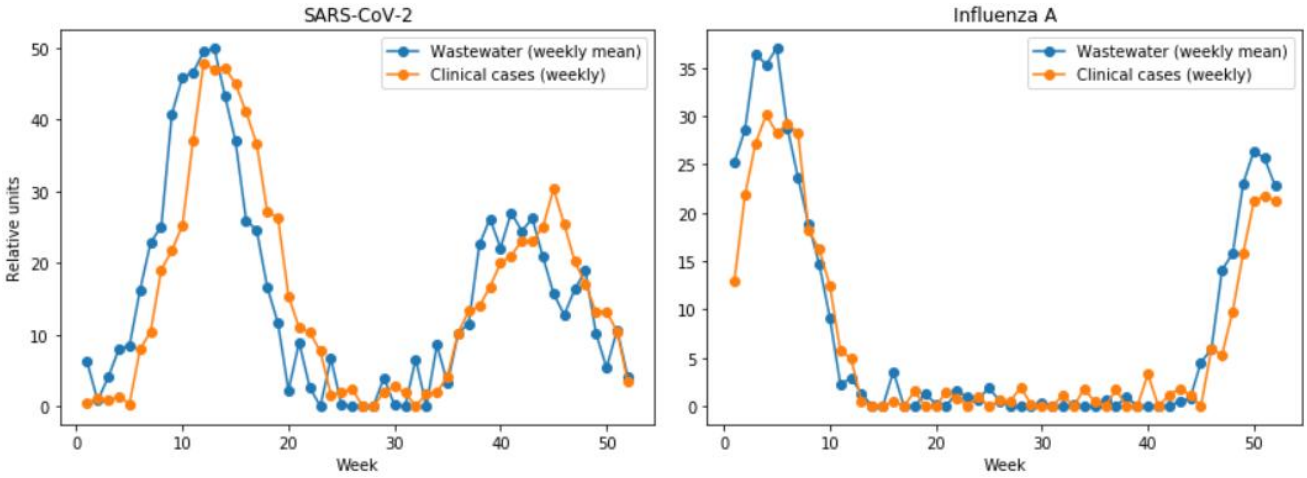

Figure 2. Temporal patterns of wastewater viral loads vs. clinical cases

Each point is a weekly average. The strong correlations and slight rightward offset show that wastewater signals anticipated the case counts by days. The strong correlations observed here are in line with previous reports. Wang et al. observed that wastewater SARS-CoV-2 concentrations in Fuzhou led COVID-19 case peaks by up to 17 days [3], while Peccia et al. reported a 6–8 day lead in sludge [1]. Our lead of ~10 days fits within this range. For influenza, the evidence is still emerging, but our detection of flu virus in sewage and its correspondence with cases agrees with systematic reviews suggesting WBE can monitor influenza circulation [4]. Importantly, our use of targeted enrichment was likely critical to achieving these results. Without enrichment, influenza signals might have been too low to detect reliably. This is illustrated by the performance in Child et al. (Figure 1): only the hybrid-capture data (not the shotgun data) showed clear influenza and SARS-CoV-2 reads [6].

Our metagenomic approach also yielded genomic information beyond viral counts. Sequence analysis confirmed the presence of contemporary viral strains. All eight gene segments of influenza A were recovered in peak samples, and SARS-CoV-2 variant signatures (e.g. Omicron markers) were identified during the second wave. This capability demonstrates that wastewater sequencing could be used not only for quantitation but also for strain surveillance, as recently shown in other studies [7]. For example, Karthikeyan et al. (2022) showed that wastewater sequencing could detect emerging SARS-CoV-2 variants up to 2 weeks before they appeared in clinical samples [16]. While our study focused on total viral abundance, the sequence data generated could support future variant monitoring.

There are limitations to note. Wastewater signals can be influenced by factors such as rainfall, industrial discharges, or sampling variation. We used composite samples and normalized read counts, but fluctuations in sewage flow and population (e.g. commuter patterns) may introduce noise. Clinical case data also have reporting delays and biases. The moderate correlation for influenza suggests that relying on wastewater alone is not sufficient to quantify true case numbers, but it can indicate trends. Finally, our analysis covered a single city and year; patterns may differ in other settings or seasons. Nonetheless, the consistency with prior findings lends confidence that these results are generalizable.

Overall, our results demonstrate that city-scale wastewater metagenomics is a feasible and informative surveillance strategy for respiratory viruses. By capturing signals of both SARS-CoV-2 and influenza in one assay, WBE can provide a comprehensive picture of community respiratory viral activity. These environmental signals change prior to or in step with clinical reports, giving public health agencies a proactive tool for outbreak response. As the Parkins et al. review notes, wastewater surveillance is becoming a mainstream tool for broader pathogen monitoring beyond COVID-19 [10]. Our study shows that this vision is practical: even seasonal influenza epidemics can be tracked in sewage to augment traditional surveillance.

5. Conclusion

This prospective time-series study shows that metagenomic sequencing of city wastewater can serve as an early warning system for both seasonal influenza and SARS-CoV-2. Key findings include high detection rates for SARS-CoV-2 and measurable influenza A signals in sewage, strong correlations with reported case data, and lead times of roughly one to two weeks for predicting outbreaks. The use of a targeted viral enrichment panel enabled the recovery of diverse viral sequences from sewage (Figure 1) [10]. Our findings support

implementation of wastewater-based metagenomic surveillance in urban public health practice. By continuously monitoring sewage, health authorities can gain timely insight into respiratory virus dynamics, potentially improving outbreak preparedness and response.

References

[1] Peccia, J., Zulli, A., Brackney, D. E., Grubaugh, N. D., Kaplan, E. H., Casanovas-Massana, A., ... & Omer, S. B. (2020). Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nature biotechnology, 38(10), 1164-1167.

[2] Tisza, M., Javornik Cregeen, S., Avadhanula, V., Zhang, P., Ayvaz, T., Feliz, K., ... & Maresso, A. W. (2023). Wastewater sequencing reveals community and variant dynamics of the collective human virome. Nature Communications, 14(1), 6878.

[3] Wang, W., Li, R., Chen, S., Chen, L., Jiang, Y., Xiang, J., ... & Wu, C. (2025). Wastewater-Based Surveillance of SARS-CoV-2 and Modeling of COVID-19 Infection Trends. Tropical Medicine and Infectious Disease, 10(9), 264.

[4] Daroch, N., Kannan, S. K., Srikantaiah, V., Mishra, R., & Ishtiaq, F. (2025). Respiratory virus infection dynamics and genomic surveillance to detect seasonal influenza subtypes in wastewater: A longitudinal study in Bengaluru, India. PLOS Global Public Health, 5(9), e0004640.

[5] Viviani, L., Vecchio, R., Pariani, E., Sandri, L., Binda, S., Ammoni, E., ... & Odone, A. (2025). Wastewater-based epidemiology of influenza viruses: a systematic review. Science of The Total Environment, 986, 179706.

[6] Child, H. T., Airey, G., Maloney, D. M., Parker, A., Wild, J., McGinley, S., ... & Bassano, I. (2023). Comparison of metagenomic and targeted methods for sequencing human pathogenic viruses from wastewater. MBio, 14(6), e01468-23.

[7] Troendle, E. P., Lee, A. J., Reyne, M. I., Allen, D. M., Bridgett, S. J., Radulescu, C. H., ... & Simpson, D. A. (2025). Combining analysis of individual and wastewater whole genome sequencing improves SARS-CoV-2 surveillance. Water research, 123953.

[8] Jones, M. J., Ibrahim, R., Clark, S., Brooks, Y. M., Preisendanz, H. E., Richard, T. L., ... & McGraw, E. A. (2025). Wastewater surveillance of SARS-CoV-2 and influenza in a dynamic university community: understanding how wastewater measurements correspond to reported cases. Science of the Total Environment, 1000, 180377.

[9] Lee, A. J., Carson, S., Reyne, M. I., Marshall, A., Moody, D., Allen, D. M., ... & Bamford, C. G. (2024). Wastewater monitoring of human and avian influenza A viruses in Northern Ireland: a genomic surveillance study. The Lancet Microbe, 5(12).

[10] Parkins, M. D., Lee, B. E., Acosta, N., Bautista, M., Hubert, C. R., Hrudey, S. E., ... & Pang, X. L. (2024). Wastewater-based surveillance as a tool for public health action: SARS-CoV-2 and beyond. Clinical microbiology reviews, 37(1), e00103-22.

[11] Toribio-Avedillo, D., Gómez-Gómez, C., Sala-Comorera, L., Rodríguez-Rubio, L., Carcereny, A., García-Pedemonte, D., ... & Muniesa, M. (2023). Monitoring influenza and respiratory syncytial virus in wastewater. Beyond COVID-19. Science of the Total Environment, 892, 164495.

[12] Mercier, E., D'Aoust, P. M., Thakali, O., Hegazy, N., Jia, J. J., Zhang, Z., ... & Delatolla, R. (2022). Municipal and neighbourhood level wastewater surveillance and subtyping of an influenza virus outbreak. Scientific Reports, 12(1), 15777.

[13] Wright, J., Driver, E. M., Bowes, D. A., Johnston, B., & Halden, R. U. (2022). Comparison of high-frequency in-pipe SARS-CoV-2 wastewater-based surveillance to concurrent COVID-19 random clinical testing on a public US university campus. Science of the Total Environment, 820, 152877.

[14] Boehm, A. B., Hughes, B., Duong, D., Chan-Herur, V., Buchman, A., Wolfe, M. K., & White, B. J. (2023). Wastewater concentrations of human influenza, metapneumovirus, parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids during the COVID-19 pandemic: a surveillance study. The Lancet Microbe, 4(5), e340-e348.

[15] Restrepo-Mieth, A. (2023). Examining the dynamics between formal and informal institutions in progressive city planning. Urban Affairs Review, 59(1), 99-132.

[16] Karthikeyan, S., Levy, J. I., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., ... & Knight, R. (2022). Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. Nature, 609(7925), 101-108.