Spatial and temporal variation in respiratory syncytial virus (RSV) subtype RNA in wastewater, and relation to clinical specimens

Winnie Zambrana¹, ChunHong Huang², Daniel Solis², Malaya K. Sahoo², Benjamin A. Pinsky²,³, Alexandria B. Boehm¹*

¹. Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA, USA, 94305
². Department of Pathology, Stanford University School of Medicine, Stanford CA, 94305
³. Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford CA, 94305

*Corresponding author
Alexandria B. Boehm, aboehm@stanford.edu, 650-724-9128

Keywords
RSV A, RSV B, wastewater surveillance, subtype variability, subtype monitoring, respiratory syncytial virus

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Respiratory syncytial virus (RSV) causes a large burden of respiratory illness, globally. It has two subtypes, RSV A and RSV B, but little is known regarding the predominance of these subtypes during different seasons and their impact on morbidity and mortality. Using molecular methods, we quantified RSV A and RSV B RNA in wastewater solids across multiple seasons and metropolitan areas to gain insight into the predominance of RSV subtypes. We determined the predominant subtype for each group using the proportion of RSV A to total RSV (RSV A + RSV B) in each wastewater sample ($P_{A,WW}$), and conducted a comparative analysis temporally, spatially and against clinical specimens. A median $P_{A,WW}$ of 0.00 in the first season and 0.58 in the second season indicated a temporal shift in the predominant subtype. Spatially, while we observed dominance of the same subtype, $P_{A,WW}$ was higher in some areas ($P_{A,WW} = 0.58$ to 0.88). The same subtype predominated in wastewater and clinical samples, but clinical samples showed higher levels of RSV A (RSV A positivity in clinical samples = 79%, median $P_{A,WW} = 0.58$). These results suggest that wastewater, alongside clinical data, holds promise for enhanced subtype surveillance.

Importance. Respiratory syncytial virus (RSV) causes a large burden of respiratory illness, globally. It has two subtypes, RSV A and RSV B, but little is known regarding the predominance of these subtypes during different seasons and their impact on morbidity and mortality. The study illustrates that information on subtype predominance can be gleaned from wastewater. As a biological composite sample from the entire contributing population, wastewater monitoring of RSV A and B can complement clinical surveillance of RSV.
1. Introduction

Respiratory syncytial virus (RSV) is a leading cause of lower respiratory tract infection across all age groups. Infants, young children, and older adults are the groups with highest risk of developing severe complications from an RSV infection. An RSV infection typically causes flu-like symptoms such as congestion, cough and fever, and can cause bronchiolitis and pneumonia in severe cases. Each year in the United States, RSV causes between 58,000 to 80,000 children hospitalizations, and 6,000 to 10,000 older adults deaths. It is estimated that over 100,000 deaths of children under the age of five are attributed to RSV annually globally. The true burden of RSV, however, is substantially underestimated, as such an estimate relies on clinical testing of patients, often only including patients with more severe symptoms that have access to and seek medical treatment. Most older children and adults infected with RSV are excluded from these estimates, as they often only experience mild symptoms. An RSV infection can also be asymptomatic which also contributes to the incomplete understanding of the true burden.

RSV is an enveloped, single stranded, negative sense RNA virus that infects humans, and has two subtypes, RSV A and RSV B. The major difference between subtypes is the variation in the G glycoprotein (G protein) that makes up part of the envelope. The pharmaceutical interventions available to combat RSV, such as prophylactic treatments and the recently approved vaccines, consider their difference in structure and can be cross-reactive for both subtypes. Regardless, these are only available to a limited section of the population (e.g. individuals older than 60 years old, pregnant individuals, and infants younger than 19 months) and have even encountered supply issues.

Despite the known structural differences between RSV subtypes, the variability in subtype predominance remains poorly understood. The predominant subtype during an outbreak is often not characterized, as routine surveillance of RSV typically does not include subtype monitoring. Studies show that although co-circulation of both subtypes is expected, typically one subtype predominates in a particular season. Some epidemiological studies have found that seasons in which RSV A dominated started earlier in the
winter, peaked faster, and lasted longer than seasons in which RSV B dominated\textsuperscript{22}, while others found no
difference between RSV A- and RSV B-dominated seasons\textsuperscript{23}. There is also a limited understanding of the
difference in virulence between subtypes. Several epidemiological studies found that RSV A was
associated with a more severe disease than RSV B, while others reported the opposite or found no
difference between them\textsuperscript{10}. Understanding how subtype predominance varies could aid in discovering
potential differences in infection dynamics and virulence caused by each subtype, and could contribute to
a better understanding of RSV spread.

Wastewater may represent a means for understanding the variability in RSV subtype circulation.
Wastewater has recently emerged as a tool for monitoring the disease levels of a community contributing
to it, as concentrations of viral genetic material in wastewater have been shown to correlate with
positivity rates or clinical case rates for a number of different viruses\textsuperscript{24–27}. Several studies have measured
RSV RNA wastewater concentrations\textsuperscript{25,28–35}, with only three of them measuring RSV subtypes using
PCR-methods\textsuperscript{25,29,30}, and one using a sequence-based methods\textsuperscript{32}. Overall, there is limited research
investigating the variability of RSV subtype predominance in wastewater. While some epidemiological
studies have studied variations in subtype dominance\textsuperscript{22,23,36,37}, their scope has been confined to clinical
samples, typically encompassing individuals with severe RSV cases. This limitation is crucial when
considering that the majority of individuals infected with RSV do not seek medical treatment\textsuperscript{38}.
Wastewater data bypasses biases inherently included in clinical case data and, thus, could be used to
understand variations in RSV subtype predominance at a broader population scale.

The goal of this study is to investigate temporal and spatial variations in the predominant RSV subtype
detected in wastewater solids samples. This investigation involves measuring concentrations of both RSV
A and RSV B in wastewater solids samples across multiple seasons and metropolitan areas, determining
the predominant subtype for each, and conducting a comparative analysis temporally, spatially and
against clinical data.
2. Materials and Methods

Study Design

This study was carried out at different wastewater treatment facilities across the United States (Table 1), comprising 240 wastewater solids samples in total. The samples in this study are part of a larger wastewater surveillance effort which measures total RSV (RSV A + RSV B) routinely in wastewater solids, and were selected due to their high total RSV concentration (more information in SI). For each wastewater solids sample, we measured the concentrations of both RSV A and RSV B using digital droplet reverse transcription polymerase chain reaction (RT-PCR), and calculated the main outcome of this study — the proportion of RSV A to total RSV ($P_{A,WW}$):

$$P_{A,WW} = \frac{C_{RSV\ A}}{C_{RSV\ A} + C_{RSV\ B}}$$ (1)

where $C_{RSV\ A}$ is the concentration of RSV A, and $C_{RSV\ B}$ is the concentration of RSV B in the sample in units of copies/g dry weight. A proportion greater than 0.5, for example, indicates that RSV A is the predominant subtype for that sample. Equation 1 assumes that $C_{RSV\ Total} = C_{RSV\ A} + C_{RSV\ B}$ in each sample. Shedding of RSV A and RSV B by infected individuals into the wastewater stream is assumed to be similar due to the lack of information in the literature.

Table 1. Characteristics of wastewater treatment plants (WWTPs) included in the study, number of samples per WWTP, sample type, and the evaluations in which they were considered. Settled solids were collected directly from the primary clarifier (“primary clarifier”), or were collected from a raw influent sample (“influent”) and allowed to settle for 10-15 minutes in the laboratory, and later aspirated into a falcon tube using a serological pipette.

<table>
<thead>
<tr>
<th>WWTP Name</th>
<th>Closest Metropolitan Area</th>
<th>Population Served</th>
<th>Number of samples (Sample dates)</th>
<th>Sample Type</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palo Alto Regional Water Quality Control Plant (PA)</td>
<td>Santa Clara County, CA</td>
<td>236,000</td>
<td>40 (Nov 2021 - Feb 2022)</td>
<td>Primary clarifier</td>
<td>Temporal, Spatial, Clinical</td>
</tr>
<tr>
<td>City of Sunnyvale Water Pollution Control Plant (SV)</td>
<td>Santa Clara County, CA</td>
<td>153,000</td>
<td>40 (Nov 2021 - Feb 2022) 40 (Nov 2022 - Feb 2023)</td>
<td>Primary clarifier</td>
<td>Temporal, Spatial, Clinical</td>
</tr>
<tr>
<td>City of Garland Rowlett Creek WWTP (GR)</td>
<td>Dallas, TX</td>
<td>200,000</td>
<td>20 (Nov 2022 - Feb 2023)</td>
<td>Primary clarifier</td>
<td>Spatial</td>
</tr>
<tr>
<td>Duck Creek WWTP (DC)</td>
<td>Dallas, TX</td>
<td>186,000</td>
<td>20 (Nov 2022 - Feb 2023)</td>
<td>Primary clarifier</td>
<td>Spatial</td>
</tr>
<tr>
<td>RM Clayton Water Reclamation Center (RM)</td>
<td>Atlanta, GA</td>
<td>294,660</td>
<td>20 (Nov 2022 - Feb 2023)</td>
<td>Influent</td>
<td>Spatial</td>
</tr>
<tr>
<td>South River Water Reclamation Center (SR)</td>
<td>Atlanta, GA</td>
<td>105,160</td>
<td>20 (Nov 2022 - Feb 2023)</td>
<td>Influent</td>
<td>Spatial</td>
</tr>
</tbody>
</table>

**Temporal Evaluation**

We compared $P_{A,WW}$ between two different RSV seasons (2021-2022 and 2022-2023) using wastewater solids samples collected from two WWTPs in the Santa Clara County, CA area (PA and SV). Samples from Season 1 (2021-2022) were collected between November 15, 2021 and February 28, 2022, and samples from Season 2 (2022-2023) were collected between November 1, 2022, and February 28, 2023 (Table 1). The WWTPs included in this evaluation are located 20 kilometers (km) from each other and have neighboring sewersheds (Figure 1). In a secondary evaluation, we compared results between these WWTPs within each season to determine if samples from the WWTPs, located in close proximity (~20 km apart with neighboring sewersheds), yield similar results.

**Spatial Evaluation**

We compared $P_{A,WW}$ between WWTPs located across the United States using wastewater solids samples collected during a single RSV season (2022-2023). Wastewater solids samples were collected between November 1, 2022 and February 28, 2023 from three United States areas: Santa Clara County, CA, Dallas, TX, and Atlanta, GA (Table 1). These areas were selected as they are considered some of the most populous metropolitan cities in the United States, and each have two WWTPs with neighboring...
sewersheds. Each metropolitan area was represented by two different WWTPs located within 20 km from each other and with neighboring sewersheds. The distances between each metropolitan area were approximately 1,000 km, 2,000 km, and 3,000 km (Figure 1). Note that the samples from Santa Clara County, CA used for this analysis were the same ones used for the temporal variation analysis described above.

Figure 1. Wastewater treatment plants included in this study. Distance (km) between each of the three metropolitan areas evaluated (A). Distance (km) between both WWTPs selected to represent each area and their corresponding sewersheds: Santa Clara County, California (B), Dallas, Texas (C), and Atlanta, Georgia (D). WWTPs shown in Figure 1B are part of all three evaluations, and WWTPs shown in Figures 1C-1D are only included in the spatial evaluation. This figure was generated using QGIS; map layer from OpenStreetMap (openstreetmap.org/copyright).

Clinical Evaluation

We compared the relative proportion of RSV A identified using wastewater solids (PA,WW) to the fraction of clinical specimens collected from individuals residing within and adjacent to the sewersheds in Santa
Clara County, CA identified as subtype A (F<sub>A,CI</sub>). A total of 80 wastewater solids samples and 602 clinical samples were included in this evaluation (Table 1), and all were collected in the 2022-2023 RSV season. The wastewater solids samples were already described as those used in the temporal evaluation study. Clinical samples, collected between October 1, and December 31, 2022, were processed by the Stanford Clinical Virology Laboratory (Palo Alto, CA). The laboratory receives samples from patients in the Santa Clara, CA area, including patients living and/or working within the sewersheds of the selected WWTPs. This study was conducted with Stanford Institutional Review Board approval (protocol 62834), and individual consent was waived.

Using the total number of clinical samples subtyped, we calculated the fraction of clinical specimens identified as subtype A to the total number identified as RSV positive (F<sub>A,CI</sub>):

\[
F_{A,CI} = \frac{N_{RSV_A}}{N_{RSV}}
\]  \hspace{1cm} (2)

where \(N_{RSV_A}\) is the number of samples that were positive for RSV A, and \(N_{RSV}\) is the total number of samples that were subtyped and found positive for either RSV A or RSV B, \(N_{RSV} = N_{RSV_A} + N_{RSV_B}\).

F<sub>A,CI</sub> was compared to P<sub>A,WW</sub>.

Wastewater solids storage and freeze-thaw cycles impact assessment

Concentrations of pepper mild mottle virus (PMMoV) and total RSV were measured in all wastewater samples immediately after they were collected with no storage (Table S1). To assess the impact of storage and freeze-thaw cycles in our samples, a subset of randomly selected samples (n = 30) from Sunnyvale (SV) and Palo Alto (PA) WWTPs were measured a second time for PMMoV after two freeze-thaw cycles. For all samples (n = 240), the sum of RSV A and RSV B concentrations taken after two freeze-thaw cycles were compared against the measurements of total RSV in fresh samples.
Statistical Analysis

The median $P_{A,WW}$ of wastewater solids samples were used to compare the predominant subtype between different groups in the statistical analyses. Non-parametric methods were employed as the wastewater data were found to be not normally distributed (Shapiro-Wilk test, $W=0.90$, $p=9.1 \times 10^{-12}$). A Kruskal-Wallis test was used to test the null hypothesis that the $P_{A,WW}$ was not different between groups, and a post-hoc test (Conover-Iman test) was used to compare the $P_{A,WW}$ between groups.

A bootstrap method was implemented to compare the $F_{A,Cl}$ to the $P_{A,WW}$. A synthetic dataset consisting of zeros and ones was generated using the total number of clinical samples subtyped ($N_{RSV}$) and the total number of samples identified as subtype A ($N_{RSV_A}$). Next, a collection of 1000 $F_{A,Cl}$ was established by randomly selecting, with replacement, from the synthetic list and calculating it using equation 2. From the collection of 1000 $F_{A,Cl}$, a subset of 80 were randomly selected (equivalent to the number of wastewater solids samples), and subsequently used in the Kruskal-Wallis and post-hoc analyses.

Lastly, three supplementary statistical analyses were performed. Two Kruskal-Wallis tests were conducted to compare fresh sample measurements of RSV and pepper mild mottle virus (PMMoV) with measurements from samples subjected to storage and freeze-thaw cycles to assess the impact of storage on measurements in the wastewater samples. Additionally, Kendall’s $\tau$ was used to test the null hypothesis that total RSV (RSV A + RSV B) concentration levels and $P_{A,WW}$ are statistically independent.

Two hypotheses were tested (two tests: Kruskal-Wallis test and post-hoc test) for each comparison (temporal, spatial or clinical) and three hypotheses were tested for the supplementary analysis, for a total of 9 hypothesis tests. A $p$ value of 0.006 (0.05/9) for alpha = 0.05 was used to adjust for multiple comparisons (Bonferroni correction). Statistical analysis was completed using R version 4.3.0 within RStudio version 2021.09.1. Wastewater solids data from this study is available at the Stanford Digital Repository.

All rights reserved. No reuse allowed without permission.
Procedures

Wastewater solids sample collection and pre-analytical processing. Wastewater solids were either grab samples of settled solids from the primary clarifier, or 24-hr composite raw influent samples in which solids were allowed to settle for 10-15 minutes in the laboratory, and aspirated into a falcon tube using a serological pipette (Table 1). Samples were collected in sterile bottles by WWTP staff, immediately stored at 4°C, delivered to the laboratory, and processed within 6 hours upon arrival. Samples were dewatered by centrifugation and resuspended in DNA/RNA shield (Zymo Research, Irvine, CA) spiked with bovine coronavirus vaccine (BCoV, Zoetis, Calf-Guard Cattle Vaccine) to a final concentration of 75 mg/ml to minimize inhibition of measurements in solids. Resuspended solids were homogenized and centrifuged, and the supernatant was withdrawn for nucleic-acid extraction. A 0.5 g to 1 g aliquot of the dewatered solids was oven dried to measure its dry weight for its use in the dimensional analysis. Detailed description of the pre-analytical processing can be found elsewhere including in detailed protocols on protocols.io.

Wastewater solids RNA extraction and quantification. RNA extraction from the supernatant was performed using the Chemagic Viral DNA/RNA 300 kit H96 for the Perkin Elmer Chemagic 360 (Perkin Elmer, Waltham, MA). Extract was further purified using the Zymo One-Step PCR inhibitor removal columns (Zymo Research, Irvine, CA). Nuclease-free water served as a negative extraction control, and BCoV served as a process control. RNA extracts were immediately processed (no storage or freeze-thaw) for quantification of BCoV, pepper mild mottle virus (PMMoV), and total RSV (RSV A + RSV B). Additional aliquots of RNA extracts were stored at -80°C (for 0-700 days) and later subjected to two freeze-thaw cycles for RSV A and RSV B measurements. Although PMMoV was already measured using fresh samples, they were measured again on a subset of the stored and freeze thawed samples to assess the impact of storage and freeze-thaw cycles in our samples (as described in the Study Design section).
RNA extracts were used as template neat using one-step RT-ddPCR Advanced Kit for Probes. RSV A and RSV B were measured in a duplex assay using methods described below, and PMMoV, BCoV and total RSV were measured in fresh samples using methods described in detail in other publications\textsuperscript{39}. Primers and probes were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Primers and probe sequences and corresponding thermal cycling conditions were selected from published assays (Table S1)\textsuperscript{25,48–50}. The fluorescence signal was analyzed in the HEX channel for RSV A and the FAM channel for RSV B (Table S1). Nuclease-free water served as a negative PCR control, and gene blocks (Integrated DNA Technologies (IDT), Coralville, IA) served as positive controls for all targets. Non-infectious intact RSV A virus (NATRSVA-STQ, Zeptomatrix, Buffalo, NY) and non-infectious intact RSV B virus (NATRSVB-STQ, Zeptomatrix, Buffalo, NY) were used as additional positive controls for RSV A and RSV B, correspondingly. RNA target concentrations were measured via digital droplet RT-PCR (RT-ddPCR) using an Automated Droplet Generator (Bio-Rad, Hercules, CA), C1000 Touch (Bio-Rad) thermocycler, and a QX200 Droplet Reader (Bio-Rad). Droplets were analyzed using QuantaSoft and QuantaSoft Analysis Pro software. Each sample had three replicate wells. Replicate wells were merged, and were required to have at least 10,000 droplets post-merger. A sample was required to have three or more positive droplets across three merged wells to be scored as positive for a target. The lower detection limit for the RSV A and RSV B assays was approximately 2,200 gc/g dry weight, assuming three positive droplets across merged wells. A detailed description of RNA extraction of resuspended solids, quantification using digital RT-ddPCR, and dimensional analysis used to convert concentrations per reaction to copies per gram of dry weight is described in detail in other publications\textsuperscript{46,51} and on protocols.io\textsuperscript{52,53}.

Clinical samples RNA extraction and quantification. RSV-positive respiratory samples, including nasopharyngeal, mid-turbinate, and anterior nasal swabs in viral transport media, as well as bronchoalveolar lavage fluid, were extracted on the PerkinElmer Chemagic 360 instrument using the Chemagic viral DNA/RNA 300 Kit H96 according to the manufacturer's recommendations. Samples were
extracted from 300 µL and eluted in 60 µL. Samples were identified as positive for either RSV A or RSV B using reverse-transcription qPCR (RT-qPCR). The primer and probe sequences were adapted from the assay described by Wang et al. 2019, and modified to include RNAse P primers and probe to serve as an internal control (Table S2).

RT-qPCR was performed using Invitrogen Superscript III Platinum One-Step qRT-PCR kit (Invitrogen, Carlsbad, CA) on the Bio-rad CFX-96 instrument (Bio-rad, Hercules, CA). Each 25 µL reaction contained 12.5 µL of 2X Buffer, 0.5 µL of enzyme mix, 2 µL of primer/probe mix, and 10 µL of eluate. Cycling conditions were as follows: hold at 52°C for 15 min, 94°C for 2 min, then 45 cycles of 94°C for 15 sec, 55°C for 40 sec, and 68°C for 20 sec. The fluorescence signal was analyzed in the FAM (A), Cy5 (B), and HEX (internal control: RNase P) channels. Thresholds were set at 3000, 1000, and 200 relative fluorescence units (RFU), respectively. Any exponential amplification curve crossing either the FAM or Cy5 RFU thresholds were interpreted as positive for the corresponding target. Samples were considered to have failed extraction or contain inhibitory substances if RNase P did not amplify at a cycle threshold value ≤ 35 cycles. Samples were considered to be untypable if both RSV A and RSV B were not detected.

3. Results and Discussion

QA/QC.

Results are reported as suggested by the Environmental Microbiology Minimal Information (EMMI) guidelines (Figure S3-S4). Negative and positive PCR and extraction controls yielded negative and positive results, respectively. Median BCoV recovery in fresh samples was around 100% (median = 110%, IQR = 69%), and PMMoV was stable across fresh samples (median = 1.16 x10⁹ gc/g, IQR = 1.19 x10⁹ gc/g).

A comparison of total RSV and PMMoV concentrations reported in this study to concentrations measured in samples with no storage and no freeze-thaw cycles showed an effect on target quantification. The median ratios of RSV and PMMoV in stored to fresh samples were 1.29 (IQR = 0.62, n = 240) and 0.40...
(IQR = 0.41, n = 30), respectively. Although the ratios show less than an order of magnitude difference between measurements in fresh and stored samples, we found that fresh sample measurements for both RSV and PMMoV were significantly different from measurements from those measurements in stored samples (Kruskal-Wallis test, RSV comparison \( p=1.60 \times 10^{-3} \); PMMoV comparison \( p= 1.69 \times 10^{-10} \)). This suggests that storage and freeze thaw might have had an impact on target measurements.

We found no significant correlation between concentrations of total RSV and \( P_{A,WW} \) (Kendall’s \( \tau \) test, \( \tau = 0.09, z = 2.05, p=0.04 \)). This suggests that the levels of total RSV RNA detected in wastewater are not reflective of the predominance of a specific subtype.

### Overall Wastewater and Clinical Samples Results.

All wastewater solids samples were positive for RSV A and/or RSV B, with the exception of three samples that were non-detects for either RSV A or RSV B. Median RSV A concentration across all wastewater samples (\( n = 240 \)) was \( 1.31 \times 10^4 \) gc/g dry weight (IQR = \( 4.07 \times 10^4 \) gc/g), and median RSV B concentration was \( 1.90 \times 10^4 \) gc/g dry weight (IQR = \( 2.54 \times 10^4 \) gc/g) (Figure S1). Overall median \( P_{A,WW} \) across all seasons and all wastewater treatment plants was 0.47 (IQR = 0.74, \( n = 237 \)) (Figure S2). A total of 593 of 602 RSV-positive clinical samples were identified as either RSV A or RSV B, and 9 samples were considered untypable. We identified the predominant RSV subtype across multiple wastewater treatment plants and conducted a comparative analysis temporally, spatially, and against clinical specimens.

### Temporal Evaluation.

The median \( P_{A,WW} \) was significantly larger in Season 2 (2022-2023) than Season 1 (2021-2022) for both wastewater treatment plants (Palo Alto Regional Water Quality Control Plant [PA] and City of Sunnyvale Water Pollution Control Plant [SV]) (Conover-Iman test, exact \( p \)-values shown in Table S3-S4). The combined median \( P_{A,WW} \) from both WWTPs for Season 1 was 0 (PA median = 0.0, IQR = 0.11, \( n = 40 \); SV median = 0.0, IQR = 0.16, \( n = 40 \)), while combined median \( P_{A,WW} \) for Season 2 was 0.58 (PA median = 0.0, IQR = 0.21, \( n = 40 \); SV median = 0.0, IQR = 0.16, \( n = 40 \)).
The results confirm a change in the predominant subtype between two consecutive RSV seasons. RSV B was the predominant subtype during 2021-2022 with limited RSV A circulation, while in the following season (2022-2023), RSV A dominated with RSV B also circulating.

Our results align with epidemiological studies that have also reported shifts in subtype predominance between seasons.23,36,56. Observing these results in wastewater solids, which capture contributions from individuals with a wider spectrum of disease severity than clinical data, suggests that changes in subtype dominance between seasons are also observable at the population level. Tracking subtype predominance in wastewater each season could enhance our understanding of RSV dynamics at the population level, particularly when subtype monitoring is not integrated into routine clinical surveillance.20. Wastewater treatment plants with neighboring sewersheds (PA and SV) showed similar subtype patterns. For both seasons studied, we found no significant difference in the median $P_{A,WW}$ between each wastewater treatment plant (Conover-Iman test, exact $p$-values shown in Table S3-S4). This result suggests that if resources are constrained, monitoring one geographic area may provide insights into subtype predominance dynamics in adjacent areas.
Figure 2. Proportion of RSV A to total RSV (P_{A,WW}) for the temporal evaluation comparing Palo Alto Regional Water Quality Control Plant (PA) and City of Sunnyvale Water Pollution Control Plant (SV) for Season 1 (2021-2022) and Season 2 (2022-2023). Each box plot is made up of the 25th quartile, median, and 75th quartile proportion of RSV A total RSV for each WWTP per season, and length of each whisker 1.5 times the interquartile range (IQR). Boxplots are overlaid with jittered data points from each group.* = Statistically Significant per post-hoc (Conover–Iman) test with a significance level p = 0.006 accounting for the Bonferroni Correction.

Spatial Evaluation.

Across the three metropolitan areas evaluated during the 2022-2023 season, RSV subtype A predominated (Santa Clara County, CA median = 0.58, IQR = 0.26, n = 80; Dallas, TX median = 0.88, IQR = 0.34, n = 39; Atlanta, GA median = 0.68, IQR = 0.74, n = 38) (Figure 3). However, the levels at which RSV A was found varied. We found that P_{A,WW} was significantly higher in Dallas, TX than in Santa Clara County, CA and Atlanta, GA, but was similar between Santa Clara County, CA and Atlanta, GA (Conover-Iman test, exact p-values shown in Table S3-S4).

Our results, showing dominance of the same RSV subtype across the three metropolitan areas, are consistent with epidemiological studies using clinical samples that have found geographic clustering of
viral lineages in large geographical areas (e.g., Australian states)\textsuperscript{36}, and that have concluded that variants are restricted to a local level (i.e., country-wide rather than global)\textsuperscript{37}.

Despite sharing a predominant subtype, we observed a lack of association between geographic proximity and levels of one specific subtype. The areas closest to each other (~1,000 km and ~2,000 km apart) were found to have different $P_{A,WW}$, and the two areas with furthest distance from each other (~3,000 km apart) were found to follow similar subtype patterns. Factors beyond physical distance — such as environmental, demographic and transit accessibility factors — may play a crucial role in shaping the levels of subtype dominance. For instance, local climate, minimum temperature, precipitation, and population density have been shown to influence RSV dynamics including onset and timing of peak infectivity\textsuperscript{57,58}, and urban public transportation systems have been shown to facilitate transmission of respiratory illnesses\textsuperscript{59}. The three metropolitan areas included in this study experience similar temperate climates with mild winters, but experience different levels of precipitation\textsuperscript{60,61}, have different population densities\textsuperscript{62}, and have varying levels of public transportation usage\textsuperscript{63} which could have contributed to the difference in $P_{A,WW}$.
**Figure 3.** Proportion of RSV A to total RSV (P\textsubscript{A,WW}) for the spatial evaluation comparing wastewater solids samples from Santa Clara County, CA, Dallas, TX, and Atlanta for Season 2 (2022-2023). Each box plot is made up of the 25th quartile, median, and 75th quartile proportion of RSV A total RSV for two WWTPs per area, and length of each whisker 1.5 times the interquartile range (IQR). Boxplots are overlaid with jittered data points from each group.*= Statistically Significant per post-hoc (Conover–Iman) test with a significance level p= 0.006 accounting for the Bonferroni Correction.

**Clinical Evaluation.**

During the 2022-2023 season, 79% (470/593) of the clinical RSV-positive samples were subtyped as RSV A (bootstraping F\textsubscript{A,CI} median = 0.79, IQR = 0.03, n =80). Wastewater solids samples had a median P\textsubscript{A,WW} of 0.58 (IQR = 0.26, n = 80). Therefore, subtype A dominated in both sample types. However, we found the levels of subtype A in clinical samples was significantly higher than in wastewater solids samples (Conover-Iman test, p=3.88 x 10\textsuperscript{-26}) (Table S3-S4).

To our knowledge, this is the first study to confirm that wastewater data reflects the same RSV subtype dominance as clinical data. Our results highlight the validity of wastewater as an appropriate epidemiological tool to monitor subtype predominance within a community. The discrepancies in specific subtype levels observed between clinical and wastewater solids samples may stem from the broader representation in wastewater, encompassing mild and asymptomatic cases. Clinical samples, in contrast,
might have a larger representation of pediatric infections when considering that children under two are less likely contributors to wastewater solids samples. Wastewater data, in combination with clinical data, could help to provide a comprehensive understanding of the RSV burden.

**Limitations and Future Work.**

This study has several limitations. First, this study was conducted during atypical RSV periods due to the COVID-19 pandemic and the implementation of nonpharmaceutical interventions. While both seasons began earlier in the year than prepandemic levels, the 2021-2022 season lasted longer, and the 2022-2023 season exhibited a higher peak number of cases compared to typical RSV epidemics. Another limitation is that we were not able to account for any potential impact of the newly approved vaccines or prophylactic antibody on RSV dynamics, as they were approved for public usage in certain populations subsequent to the 2022-2023 season. Additionally, the temporal evaluation was conducted using data from only two RSV seasons, and the spatial evaluation was limited to three metropolitan areas. A wider range of seasons across a more extensive array of metropolitan areas would enhance our conclusions regarding spatial-temporal subtype predominance. Due to the nature of this retrospective study, another limitation is that wastewater solids samples used to detect RSV subtypes were subjected to storage and freeze thaw which might have had an impact on target measurements. An additional limitation is that the clinical data might also include individuals residing outside the sewershed from the selected WWTPs which could potentially bias the clinical evaluation results. Lastly, although wastewater captures a larger percent of the population infected with RSV than clinical data, it is likely that children under the age of two contribute minimally to wastewater solids samples; therefore, wastewater provides a comprehensive yet incomplete picture of the burden.

Despite the aforementioned limitations, the findings in this study suggest that future work regarding RSV subtype variability in several areas is warranted. Future studies could conduct finer-scale temporal evaluations of subtype predominance variability in wastewater, examining potential differences in onset...
and peak timings between seasons dominated by different subtypes. Spatial evaluations could consider differences in environmental and/or demographic conditions between areas studied that might influence subtype predominance such as climate, precipitation, population density, public transportation usage and other factors. Investigations comparing subtype predominance in wastewater to clinical data could include patient characteristics, distinguishing between pediatric and non-pediatric cases, as well as in-patient vs outpatient cases, to determine the different factors that affect subtype dominance in each signal. Lastly, wastewater monitoring could be integrated into long-term epidemiological studies investigating variability, infection dynamics, virulence characteristics of RSV subtypes, or be used to investigate the effectiveness of the newly approved vaccines at the population level. Wastewater could provide a vital comprehensive picture of cases ranging in severity, and combined with clinical data, could be robust tools to better understand these complex issues.

4. Conclusions

This study presents an analysis of RSV subtypes in wastewater across multiple seasons, metropolitan areas, and in comparison to clinical specimens. We found that the predominant RSV subtype varied temporally, but remained consistent spatially and between wastewater and clinical samples. Temporal evaluation revealed a shift in the predominant subtype between consecutive seasons. Spatially, while we observed dominance of the same subtype across metropolitan areas, the levels of one specific subtype varied, indicating that factors beyond geographic proximity may influence levels of subtype prevalence. When comparing wastewater samples to clinical samples, we found that the same subtype dominated in both sample types, but presence of a specific subtype was higher in clinical samples. Our findings suggest that wastewater, in conjunction with clinical data, holds promise for enhanced subtype surveillance, understanding subtype dynamics, and controlling RSV spread. Future research in RSV subtype predominance can investigate finer-scale temporal variations, study characteristics beyond physical distance, incorporate patient characteristics into clinical-wastewater comparisons, and implement wastewater monitoring into epidemiological studies.
5. Supporting Material

Additional details about the methods and results used in this study (Tables S1-S4 and Figures S1-S4).

6. Acknowledgements

We acknowledge the wastewater treatment plants that kindly provided samples for this study.

7. References


(accessed 2024-01-08).

